Springer Geophysics

For further volumes:
http://www.springer.com/series/10173

Erik W. Grafarend • Joseph L. Awange

Applications of Linear and Nonlinear Models

Fixed Effects, Random Effects, and Total Least Squares

Springer

Prof. Dr.-Ing. Erik W. Grafarend
University of Stuttgart
Geodetic Institute
Geschwister-Scholl-Strasse 24D
70174 Stuttgart
Germany
grafarend@gis.uni-stuttgart.de

Prof. Dr.-Ing. Joseph L. Awange
Professor of Environment and Earth Sciences
Maseno University
Maseno
Kenya

Curtin University
Perth
Australia

Karlsruhe Institute of Technology
Karlsruhe
Germany

Kyoto University
Kyoto
Japan

ISBN 978-3-642-22240-5 ISBN 978-3-642-22241-2 (eBook)
DOI 10.1007/978-3-642-22241-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011945776

© Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

This book is a source of knowledge and inspiration not only for geodesists and mathematicians, but also for engineers in general, as well as natural scientists and economists. Inference on effects which result in observations via linear and nonlinear functions is a general task in science. The authors provide a comprehensive in-depth treatise on the analysis and solution of such problems. They start with the treatment of underdetermined or/and overdetermined linear systems by methods of minimum norms, weighted least squares and regularization, respectively. Considering the observations as realizations of random variables leads to the question of how to balance bias and variance in linear stochastic regression estimation. In addition, the causal effects are allowed to be random (general linear Gauss–Markov model of estimation and prediction), even the model function; conditions are allowed, too (Gauss–Helmert model). Nonlinear functions are comprehended via Taylor expansion or by their special structures, e.g., within the context of directional observations.

The self-contained monograph is comprehensive in its inclusion of the more fundamental elements of the topic to the state of the art, complemented by an extensive list of references. All chapters have an introduction, often with an indication of fast track reading and with an explanation addressed to non-mathematicians which are of interest also to me as a mathematician. The book contains a great variety of case studies and numerical examples ranging from classical geodesy to dynamic systems theory. The background and much more is provided by appendices dealing with linear algebra, probability, statistics and stochastic processes as well as numerical methods. The presentation is lucid and elegant, with historical remarks and humorous interpretations, such that I took great pleasure in reading the individual chapters.

I wish all readers of this brilliant encyclopaedic book this pleasure and much benefit.

Prof. Dr. Harro Walk

Preface

"*Probability does not exist*", such a statement is at the beginning of the book by *B. de Finetti* dedicated to the foundation of probability, nowadays called "*probability theory*". We believe that such a statement is *overdone*. The target of the reference is made very clear, namely the analysis of the *random effects* or of a random experiment. The result of an experiment like throwing dice or a LOTTO outcome is *not* predictable. Such an *indeterminism* is characteristic analysis of an experiment. In general, we analyze here *nonlinear relations* between *random effects*, maybe stochastic or not. Obviously, the center of interest is on linear models, which are able to approximate nonlinear relations. There are typical nonlinear models, which *cannot* be linearized. Our experimental world is divided into linear or linearized *and* nonlinear models. Sometimes, nonlinear models can be transformed into *polynomial equations*, also called *algebraic*, which allow a rigorous solution set. A prominent example is the *resection problem of overdetermined type*. If we are able to express the characteristic rotation matrix between reference frames by *quaternions*, an algebraic equation of known high order arrives. But, in contrast to linear models, there exist a whole world of solutions, which have to be discussed. *P. Lohse* presented in his Ph.D thesis wonderful *examples* of which we refer.

A.N. Kolmogorov founded the *axiomatic (mathematical) theory* of probability. Typical for the axiomatic treatment, there are *special rules* relating to *measure theory*. Here, we only refer to excellent introductions of measure theory, for instance *H. Bauer (1992), J. Elstrodt (2005)* or *W. Rudin (1987)*. Of course, there are very good textbooks on the theory of probability, for instance *H. Bauer (2002), K.L. Chung (2001), O. Kallenberg (2002), M. Loeve (1977)*, and *A.N. Shiryaev (1996)*.

It is important to inform our readers about these many subjects, which are not treated here: robust estimation, Bayes statistics (only examples), stochastic integration, *ITÔ* stochastic processes, stochastic signal processing, martingales, Markov processes, limit theorems for associated random fields, modeling, measuring, and managing risks, genetic algorithms, chaos, turbulence, fractals, neural networks, wavelets, Gibbs field, Monte Carlo, and many, indeed *very important subjects*. This statement refers to an overcritical reviewer who complained of the subject, which we did *not* treat. *Unfortunately* we had to strongly restrict our subjects. Our *contribution* *is only an atom in the* 10^8 *atoms*.

What will be treated?

It is the task of the observer or the designer to decide: what is the best fit between a linear or nonlinear model or the "real world" of the observer. Both types of analysis have to make an a priori statement about a lot of facts.

- "Are the *observations* random or not, a stochastic effect *or* a fixed effect?"
- "What are the *parameters* in the model space or would you prefer a 'model free' concept?"
- "What is the nature of the *parameters, random effects* or *fixed effects*?"
- "Could you find the *conditions* between the observations or between the parameters? What is the nature of the conditions, *linear or nonlinear, random or not*?"
- "What is the bias between the model, and its *estimation or prediction*?"

There are various criteria, which characterize the observation space or the parameter space:

> linear or nonlinear, estimation theory, linear and nonlinear distributions, $l^r - optimality$, (LESS: l^2), restrictions, mixed models of fixed and random effects, zero-, first-, and second order design, nonlinear statistics on curved manifolds, minimal bias estimation, translational invariance, total least squares, generalized least squares, minimal distance estimation, optimal design, and optimal prediction.

Of key interest is to find *equivalence lemmas*, for instance: Assume direct observations. Then a least squares fit (*LESS*) is equivalent to a best linear uniformly unbiased estimation (*BLUUE*).

The Chap. 1 deals with the *first problem* of algebraic regression, the consistent system of linear observational equations or a system of *underdetermined linear equations*. From the first page, examples onwards, we treat MINOS and the "horizontal rank partitioning". In detail, we give the *eigenvalue decomposition of weighted minimum norm solution*. Examples are *FOURIER series, FOURIER-LEGENDRE series* including the *NYQUIST frequency* for special data. Special *nonlinear models* with *datum defects* in terms of *Taylor polynomials* and the *generalized Newton iterations* are introduced.

The Chap. 2 is based on the special *first problem*, the *bias problem* expressed by the Equivalence Theorem of the adjusted minimum norm solution (G_x-MINOS) and linear uniformly minimum biased estimator (S-LUMBE).

The *second problem* of algebraic regression, namely solving an *inconsistent system of linear observational equations* as an *overdetermined* system of linear equations is extensively treated in Chap. 3. We start with a front example for the *Least Squares Solution* (LESS). Indeed we discuss alternative solutions depending on the *metric of the observation space* like Second Order Design, the *Taylor-Karman* structure, an *optimal choice of the weight matrix* and the *Fuzzy sets*.

By an eiqenvalue decomposition of G_y-LESS, we introduce "*canonical LESS*". Our case studies range from partial redundancies, latent conditions, the theory of high leverage points versus breaking points, direct and inverse Grassman coordinates and PLÜCKER coordinates. We conclude with historical notes on *C.F. Gauss*, *A.M. Legendre* and generalizations.

In another extensive review, we concentrate on the second problem of probabilistic regression, namely the *special Gauss–Markov model without datum defect of* Chap. 5. We set up the *best, linear, uniformly unbiased estimator* (BLUUE) for moments of the *first order* and the *best, quadratically uniformly unbiased estimator for the central moments of the second order* (BIQUUE). We depart again from a detailed example BLUUE and BIQUUE. We set up Σ_y-BLUUE and note the *Equivalence Theorem* of G_y-LESS *and* Σ_y-BLUUE. For BIQUUE, we study the *block partitioning of the dispersion matrix*, the invariant quadratic estimation of types IQE, IQUUE, HIQUUE versus HIQE according to *F.R. Helmert*, variance-component estimation, simultaneous estimation of the first moments and the second central moments, the so called E-D correspondence, *inhomogeneous multilinear estimation, and BAYES design of moments estimation*.

The *third problem* of *algebraic regression* of Chap. 5 is defined as the problem to solve *inconsistent systems of linear observation equations with datum defect* as an overdetermined and indetermined system of *linear equations*. In the introduction, we begin with the front page example, subject to the minimum norm- least squares solution of the front page example, namely by the technique of *additive rank* partitioning as well as *multiplicative rank partitioning*. In more detail, we present MINOS in the second section including the weights of the MINOS as well as LESS. Especially, we interpret G_y, G_x-MINOS in terms of the *generalized inverse*. By *eigenvalue decomposition*, we find the solution to G_y, G_x-MINOS. A special section is devoted to *total least squares* in terms of α-*weighted hybrid approximate solution* (α-HAPS) within *Tykhonov-Phillips regularization*.

The *third problem of probabilistic regression* as a *special Gauss–Markov* model with *datum problem* is treated in Chap. 6. We set up *best linear minimum unbiased estimators* of type *BLUMBE* and best linear estimators of type *hom BLE, hom S-BLE and hom-α-BLE* depending on the weight assumption for the *bias*. For example, we introduce *continuous networks* of first and second derivatives. Finally, we discuss the limit process of discrete network into continuum.

The *overdetermined system of nonlinear equations on curved manifolds in* Chap. 7 is presented for inconsistent system of directional observation equations, namely by minimal *geodesic distance* (MINGEODISC). A special example is the directional observation from circular normal distribution of type *von MISES-FISHER* with a note of *angular metric*.

Chapter 8 is relating to the fourth probabilistic regression as a setup of type BLIP and VIP for the *central moments of first order*. We begin the definition of a *random effect model* according to our *magic triangle*. Three examples relate to

(i) *Nonlinear error propagation* with random effect components
(ii) *Nonlinear vector-valued error propagation* with random effect in *Geoinformatics*

(iii) *Nonlinear vector-valued error propagation* for *distance measurements* including *HESSIANS*.

The *fifth problem* of *algebraic regression* in Chap. 9 is defined as solving *conditional equations of type homogeneous and inhomogeneous*.

Alternatively, the *fifth problem* of *probabilistic regression in* Chap. 10 is treated as *inhomogeneous general linear GAUSS–MARKOV model including fixed and random effects*. In an explicit representation of errors in the *general GAUSS–MARKOV model with mixed effects*, we present an example, "*collocation*". Our comments relate to the *KOLMOGOROV–WIENER prediction* and more *up-to-date* literature.

The *sixth problem of probabilistic regression* in Chap. 11 is defined as a special random effect model called "*error-variables*". Here, we assume with respect to the *second order statistics* the *first order design matrix to be random*. Another name for this is the *Total Least Squares* as an algebraic analogue. At first, we relate the *Total Least Squares* to the nonlinear system "error-in-variables". Secondly, we introduce the models SIMEX and SYMEX from the approach of *Carroll–Cook–Stefanski–Polzehl–Zwanzig* depending on the variance-covariance matrix of the random effects.

As *special problem of nonlinearity*, we treat the popular *3d-datum transformation* and the *PROCRUSTES Algorithm in* Chap. 12.

As the *sixth problem* of *generalized algebraic regression*, we introduce finally the *GAUSS–HELMERT problem as a system of conditional equations with unknowns*. One part is the variance-covariance estimation $D\{y\}$, second the variance-covariance model of conditional equations, and finally the complete conditional equations with unknown in the third model $\mathbf{Ax} + \mathbf{By} = \mathbf{C}$. As a result of Chap. 13, we present the *block-structure of the dispersion matrix*.

As special problems of algebraic regression *as well as* stochastic estimation, we treat in Chap. 14 the *multivariate GAUSS–MARKOV model*, especially the *n-way classification model* and dynamical systems.

We conclude with *algebraic solution of systems of equations of type* linear, bilinear, quadratic and so on referring to *combinatoric subsets*, the *GROEBNER basis method*, and the *Multipolynomial resultants* in Chap. 15.

A very important part of our contribution is included in the appendices. Appendix A is an introduction to tensor algebra, linear algebra, matrix algebra *as well as* multilinear algebra. Topics are multilinear functions and their decomposition, matrix algebra and the Hodge star operator including self duality. But the major parts in *linear algebra* are "Ass", "Uni", and "Comm", Rings, spaces, division algebra, Lie algebra, *de Witt algebra*, composition algebra, generalized inverse, special matrix like *Helmert, Hankel, Vandemonte*, scalar measures of matrices, complex algebra, quaternion algebra, octonian algebra, and *Clifford algebra*.

A short introduction is given on *sampling distribution* and their uses in *confidence intervals and confidence region* in *Appendix B*. Sampling distribution for the *GAUSS–LAPLACE normal distribution* and its generalization to the *multidimensional GAUSS–LAPLACE normal distribution* as well as the distribution of the *sample mean* and *variance-covariance matrix* including the *correlation coefficients* are reviewed.

An introduction into statistical notions, random events and stochastic processes in *Appendix C* range from the *moment representation* of a probability, the *GAUSS–LAPLACE and other* normal distributions, to *error propagation*, scalar-, vector- and tensor valued stochastic processes of one- and multiparameter systems. Simple examples of one parameter include (i) cosine oscillation with random amplitudes and phases (ii) superposition of two uncorrelated random functions (iii) pulse modulation, (iv) random sequences with constant and moving averages of type ARMA (r,s), (v) *WIENER processes*, (vi) special analysis of the parameter *stationary* and *non-stationary stochastic processes* with discrete and continuous spectrum white and coloured noise with band limitation, (vii) *multiparameter systems homogeneous and isotropic* multipoint systems are given. Examples are (i) *two-dimensional EUCLIDEAN networks*, (ii) criterion matrices, (iii) *space gravity spectroscopy based on the TAYLOR-KARMAN structures*, (iv) *nonlinear prediction* and (v) nonlocal time series analysis.

Appendix D is devoted to GROEBNER basis algebra, the BUCHBERGER Algorithm, and C.F. GAUSS combinatorial formulation.

Acknowledgements

E.W. Grafarend wants to thank the colleagues B.G. Blaha (Bedford/Mass., USA), J. Cai (Stuttgart, Germany), A. Dermanis (Thessalonik, Greece), T. Krarup (Copenhagen, Denmark), K.R. Koch (Bonn/Germany), G. Kampmann (Uebach-Palenberg/Germany), L. Kubacek (Olomouc, Slovakia), P. Lohse (Paderborn, Germany), B. Palancz (Budapest, Hungary), F. Pukelsheim (Augsburg/Germany), R. Rummel (Munich, Germany), F. Sanso (Como, Italy), L.E. Sjoberg (Stockholm, Sweden), P. Teunissen (Deft, Netherlands), H. Walk (Stuttgart, Germany), P. Xu (Kyoto, Japan), P. Zaletnyik (Budapest, Hungary) and S. Zwanzig (Uppsala, Sweden).

J.L. Awange wishes to express his sincere thanks to B. Heck of the Department of Physical Geodesy (Karlsruhe Institute of Technology, Germany) for hosting him during the period of his Alexander von Humboldt Fellowship during the period 2008-2011 when part of the book was written. He is further grateful to B. Veenendaal (Head of Department, Spatial Sciences, Curtin University, Australia) and F. N. Onyango (Vice Chancellor, Maseno University, Kenya) for the support and motivation. Last, but not least, he acknowledges the support of *Curtin Research Fellowship*, while his stay at Karlsruhe Institute of Technology (KIT) was supported by a *Alexander von Humboldt's Ludwig Leichhardt's Memorial Fellowship*. He is very grateful for the financial support.

For the transfer of copy rights of our contributions, we thank the Springer Publishing House (Heidelberg-Berlin-New York) as well as Spectrum Verlag (Heidelberg), and European Geosciences Union. A proper reference is added to the

place we quote from. During the making of this edition of the book, more than 3000 references have been cited.

Finally, special thanks to I. Anjasmara and B. Arora of Curtin University who typeset parts of the book. This is an Institute for Geoscience Research (TIGeR) publication No. 260.

Stuttgart and Karlsruhe (Germany), *Erik W. Grafarend*
Perth (Australia) *Joseph L. Awange*

Contents

Chapter 1
The First Problem of Algebraic Regression

Consistent system of linear observational equations – underdetermined system of linear equations: $\mathbf{Ax} = \mathbf{y}|\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{y} \in \mathcal{R}(\mathbf{A}) \sim rk\mathbf{A} = n, n =$ dim $\in \mathbb{R}$.

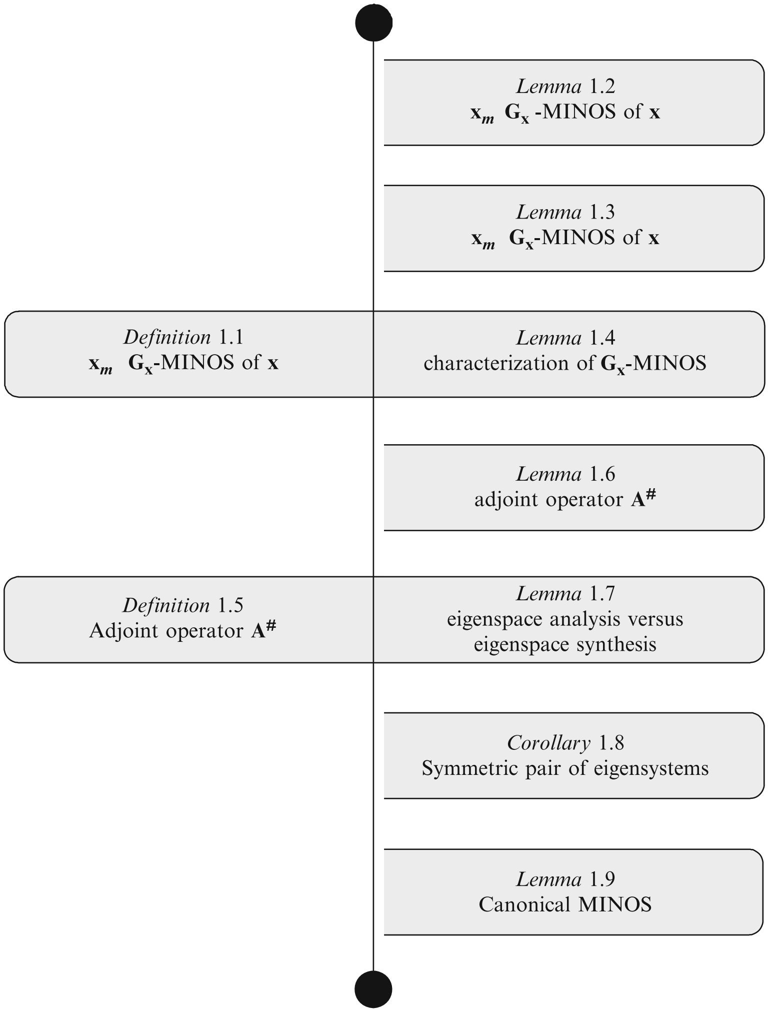
The optimisation problem which appears in treating underdetermined linear system equations and weakly nonlinear system equations is a standard topic in many textbooks on optimisation. Here we define a nearly nonlinear system of equations as a problem which allows Taylor expansion. We start in the first section with a front example, the transformation of a threedimensional parameter space into a twodimensional observation space. The Minimum Norm Solution, in short MINOS, leads to the solution (1.21) and (1.22). We determine as an example the range and the kernel of the mapping, namely by "horizontal rank partitioning". In particular,we refer to three types of analysis of underdetermined linear systems : (i) algebraic, (ii) geometric and (iii) set theoretical. The second section is based on a review of Minimum Norm Solution of an underdetermined linear system, for instance using the reflexive generalized inverse called RAO's Pandora Box. (C.R. Rao and S.K.Mitra 1971 : pages 44-47, "three basic g-inverses, section 3.1 : minimum norm solution). We apply the method of Lagrange multiplyer, or the Implicit Function Theorem, or the theory of extrema with side conditions. The numerical example is the eigenspace -eigenvector decomposition , namely the left and right eigenspace elements. While the theory of solving rank deficient linear equations is well known since the seventeenth of the 20st century, it is not so for many important applications.

Beside economical model theory we present applications in (i) Fourier series and (ii) Fourier-Legendre series enriched by NYQUIST Frequency Analysis of spherical data. For instance, for data on a grid 5 minutes by 5 minutes of arc resulting in n = 9,331,200 observations the NYQUIST Frequency N(f) = 2,160 as a limit when we develop the series by degree and order. Example (iii) introduces Taylor polynomials and generalized Newton iteration, namely in applying within twodimensial network analysis,in particular linearized models with Datum Defects. Such a problem arises when we have measured distances and want to derive absolute coordinates of network stations. Finally, as example (iv) we take reference of the General Datum Problem in R(n) including coordinate differences, distances, angles, angle ratios, cross-ratios and area elements presented in an extended reference list.

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_1,

© Springer-Verlag Berlin Heidelberg 2012

Fast track reading: read only *Lemma 1.3*. Compare with the guideline of Chap. 1.



The guideline of chapter one:definitions, lemmas and corollary

The *minimum norm solution* of a system of consistent linear equations $\mathbf{Ax} = \mathbf{y}$ subject to $\mathbf{A} \in \mathbb{R}^{n \times m}$, $rk\mathbf{A} = n, n < m$, is presented by *Definition 1.1, Lemma 1.2, and Lemma 1.3*. *Lemma 1.4* characterizes the solution of the quadratic optimization problem in terms of the (1, 2, 4)-*generalized inverse*, in particular the *right inverse*. The system of *consistent nonlinear equations* $\mathbf{Y} = F(\mathbf{X})$ is solved by means of two examples. Both examples are based on distance measurements in a planar network, namely a planar triangle. In the first example, $\mathbf{Y} = F(\mathbf{X})$ is linearized at the point $\mathbf{x}$, which is given by prior information and solved by means of *Newton iteration*. The minimum norm solution is applied to the consistent system of linear equations $\triangle\mathbf{y} = \mathbf{A}\triangle\mathbf{x}$ and interpreted by means of first and second moments of the nodal points. In contrast, the second example aims at solving the consistent system of nonlinear equations $\mathbf{Y} = F(\mathbf{X})$ in a closed form.

Since distance measurements as Euclidean distance functions are left equivariant under the action of the translation group as well as the rotation group – they are invariant under translation and rotation of the Cartesian coordinate system – at first a TR-basis (translation-rotation basis) is established. Namely the origin and the axes of the coordinate system are fixed. With respect to the TR-basis (a set of "*free parameters*" has been fixed), the "*bounded parameters*" are analytically fixed. Since no prior information is built in, we prove that two solutions of the consistent system of nonlinear equations $\mathbf{Y} = F(\mathbf{X})$ exist. The mapping $\mathbf{y} \longmapsto \mathbf{X})$ is $1 \longmapsto 2$. In the chosen TR-basis the solution vector $\mathbf{X}$ is not of minimum norm. Accordingly, we apply a datum transformation $\mathbf{X} \longmapsto \mathbf{X})$ of type "group of motion" (decomposed into the translation group and the rotation group). The parameters of the "*group of motion*" (2 for translation, 1 for rotation) are determined under the condition of minimum norm of the unknown vector $\mathbf{x}$, namely by means of a special *Procrustes algorithm*. As soon as the optimal datum parameters are determined, we are able to compute the unknown vector $\mathbf{x}$, which is minimum norm. Finally, the notes are an attempt to explain the origin of the injectivity rank deficiency, namely the dimension of the null space $\mathcal{N}(\mathbf{A}), m - rk\mathbf{A}$ of the consistent system of linear equations subject to $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $rk\mathbf{A} = n, n < m$ as well as of the consistent system of nonlinear equations subject to a Jacobi matrix $\mathbf{J} \in \mathbb{R}^{n \times m}$, $rk\mathbf{J} = n, n < m = dim\mathbb{X}$. The fundamental relation to the datum transformation, also called *transformation groups* (conformal group, dilatation group (scale), translation group, rotation group, and projective group), as well as to the "soft" *Implicit Function Theorem* is outlined.

By means of a certain algebraic objective function which geometrically are called *minimum distance functions*, we solve the first inverse problem of linear and nonlinear equations, in particular of algebraic type, which relate observations to parameters. The system of linear or nonlinear equations we are solving here is classified as underdetermined. The observations, also called *measurements*, are elements of a certain observation space $\mathbb{Y}$ of integer dimension $dim\mathbb{Y} = n$, which may be metrical, especially Euclidean, pseudo-Euclidean, in general a differentiable manifold. In contrast, the parameter space $\mathbb{X}$ of integer dimension $dim\mathbb{X} = m$ is metrical as well, especially Euclidean, pseudo-Euclidean, in general a differentiable manifold, but its metric is unknown. A typical feature of algebraic regression is the fact that the unknown metric of the parameter space $\mathbb{X}$ is induced by the functional relation between observations and parameters.

We shall outline three aspects of any discrete inverse problem: (i) set-theoretic (fibering), (ii) algebraic (rank partitioning, IPM, the Implicit Function Theorem) and (iii) geometrical (slicing). Here, we treat the first problem of algebraic regression. A consistent system of linear observational equations, which is also called *underdetermined system of linear equations* ("more unknowns than equations"), is

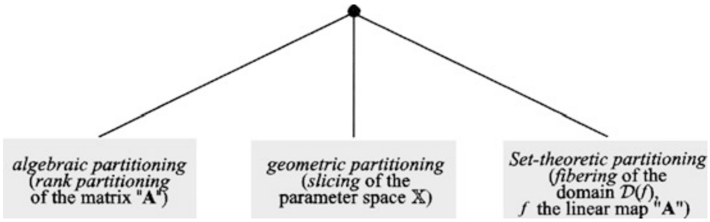


Fig. 1.1 The trinity

solved by means of an optimization problem. The following introduction presents us a front page example of two inhomogeneous linear equations with unknowns. In terms of five boxes and five figures, we review the minimum norm solution of such a consistent system of linear equations which is based upon the trinity that is shown in Fig. 1.1.

1-1 Introduction

With the introductory paragraph, we explain the fundamental concepts and basic notions of this section. For you, the analyst, who has the difficult task to deal with measurements, observational data, modeling, and modeling equations, we present numerical examples and graphical illustrations of all abstract notions. The elementary introduction is written not for a mathematician, but for you, the analyst, with limited remote control of the notions given hereafter. May we gain your interest.

Assume an n-dimensional observation space $\mathbb{Y}$, here a linear space parameterized by n observations (finite, discrete) as coordinates $\mathbf{y} = [y_1, \ldots, y_n]' \in \mathbb{R}^n$ in which an r-dimensional model manifold is embedded (immersed). The model manifold is described as the range of a linear operator $\hat{f}$ from an m-dimensional parameters space $\mathbb{X}$ into the observation space $\mathbb{Y}$. The mapping f is established by the mathematical equations which relate all observables to the unknown parameters. Here, the parameters space $\mathbb{X}$, the domain of the linear operator $\hat{f}$, will be restricted also to a linear space which is parameterized by coordinates $\mathbf{x} = [x_1, \ldots, x_n]' \in \mathbb{R}^n$.

In this manner, the linear operator $\hat{f}$ can be understood as a coordinate mapping of type $\mathbf{A} : \mathbf{x} \longmapsto \mathbf{y} = \mathbf{A}\mathbf{x}$. The linear mapping $f : \mathbb{X} \longrightarrow \mathbb{Y}$ is geometrically characterized by its range $\mathcal{R}(f)$, namely $\mathcal{R}(\mathbf{A})$, defined by $\mathcal{R}(f) := \{\mathbf{y} \in \mathbb{Y} \mid \mathbf{y} = f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{X}\}$, which in general is a linear subspace of $\mathbb{Y}$, and its null space defined by $\mathcal{N}(f) := \{\mathbf{y} \in \mathbb{Y} \mid f(x) = 0\}$. Here, we restrict the range $\mathcal{R}(f)$, namely $\mathcal{R}(\mathbf{A})$, to coincide with the $n = r$-dimensional observation space $\mathbb{Y}$ such that $\mathbf{y} \in \mathcal{R}(f)$, namely $\mathbf{y} \in \mathcal{R}(\mathbf{A})$. Note that Example 1.1 will therefore demonstrate the range space $\mathcal{R}(f)$, namely $\mathcal{R}(\mathbf{A})$, which coincides here with the observation space $\mathbb{Y}$ (f is surjective or "into") as well as the null space $\mathcal{N}(f)$, namely $\mathcal{N}(\mathbf{A})$, which is not empty (f is not injective or "one-to-one"). Furthermore, note that *Box 1.1* will introduce the special linear model of interest. By means of *Box 1.2*, it will be interpreted as a polynomial of degree two based upon two observations and three unknowns, namely as an underdetermined system of consistent linear equations. *Box 1.3* reviews the formal procedure in solving such a system of linear equations by means of "horizontal" rank partitioning and the postulate of the minimum norm solution of the unknown vector. In order to identify the range space $\mathcal{R}(\mathbf{A})$, the null space $\mathcal{N}(\mathbf{A})$, and its orthonormal compliment $\mathcal{N}^{\perp}(\mathbf{A})$, *Box 1.4* by means of algebraic partitioning ("horizontal" rank partitioning) outlines the general solution of a system of homogeneous linear equations approaching zero. With a background, *Box 1.5* presents the diagnostic algorithm for solving an underdetermined system of linear equations. In contrast, *Box 1.6* is a geometric interpretation of a special solution of a consistent system of inhomogeneous linear equations of type *MINOS* (minimum norm solution). The g-inverse $\mathbf{A}^{-}m$ of type *MINOS* is characterized by three conditions collected in *Box 1.7*. Moreover, note that Fig. 1.2 illustrates the range space $\mathcal{R}(\mathbf{A})$, while Figs. 1.3 and 1.4 demonstrate the null space $\mathcal{N}(\mathbf{A})$ as well as its orthogonal compliment $\mathcal{N}^{\perp}(\mathbf{A})$. Figure 1.5 illustrates the orthogonal projection of an element of the null space $\mathcal{N}(\mathbf{A})$ onto the range space $\mathcal{R}(\mathbf{A}^{-})$. In terms of fibering, the set of points of the parameter space as well as of the observation space, Fig. 1.6 introduces the related *Venn diagrams*.

1-11 The Front Page Example

Example 1.1. (Polynomial of degree two, consistent system of linear equations, i.e. $\{\mathbf{Ax} = \mathbf{y}, \mathbf{x} \in \mathbb{X} = \mathbb{R}^m, dim\mathbb{X} = m; \mathbf{y} \in \mathbb{Y} = \mathbb{R}^n, dim\mathbb{Y} = n, r = rk\mathbf{A} = dim\mathbb{Y}\}$.

As a front page example, consider the following front page consistent system of linear equations:

$$\begin{aligned} x_1 + x_2 + x_3 &= 2 \\ x_1 + 2x_2 + 4x_3 &= 3 \end{aligned} \tag{1.1}$$

Obviously, in general, it deals with the linear space $\mathbb{X} = \mathbb{R}^m \ni \mathbf{x}, dim\mathbb{X} = m$, here $m = 3$, called the *parameter space*, and the linear space $\mathbb{Y} = \mathbb{R}^n \ni \mathbf{y}, dim\mathbb{Y} = n$, here $n = 2$, called the *observation space*.

1-12 The Front Page Example: Matrix Algebra

By means of *Box 1.1* and according to A. Cayley's doctrine, let us specify (1.1) in terms of matrix algebra.

Box 1.1. (A special linear model: polynomial of degree two, two observations, and three unknowns).

$$
\begin{gathered}
y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
\Longleftrightarrow \\
\mathbf{x} = \mathbf{A}\mathbf{x} : \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
\Longleftrightarrow
\end{gathered}
\tag{1.2}
$$

$$
\begin{gathered}
x' = [x_1, x_2, x_3],\ y' = [y_1, y_2] = [2, 3] \\
x \in \mathbb{R}^{3\times 1},\ y \in \mathbb{Z}_+{}^{*2\times 1} \subset \mathbb{R}^{2\times 1}
\end{gathered}
\tag{1.3}
$$

$$
\begin{gathered}
A := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \in \mathbb{Z}_+{}^{*2\times 3} \subset \mathbb{R}^{2\times 3} \\
r = rk\mathbf{A} = dim\mathbb{Y} = n = 2
\end{gathered}
\tag{1.4}
$$

The matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is an element of $\mathbb{R}^{n\times m}$, the $n \times m$ array of real numbers. $dim\mathbb{X} = m$ defines the number of unknowns (here $m = 3$) and $dim\mathbb{Y} = n$ defines the number of observations (here $n = 2$). A mapping $\{$ is called linear if $f(x_1 + x_2) = f\ (x_1) + f(x_2)$ holds. Beside the range $\mathcal{R}(f)$, the range space $\mathcal{R}(A)$, the linear mapping is characterized by the kernel $\mathcal{N}(f) := \{x \in \mathbb{R}^n | f(x) = 0\}$, the null space $\mathcal{N}(A) := \{x \in \mathbb{R}^m | \mathcal{A}\S = 0\}$ to be specified later on.

Why is the front page consistent system of linear equations to be considered here called underdetermined? Just observe that we are left with only two linear equations for three unknowns $[x_1, x_2, x_3]$. Indeed, the system of inhomogeneous linear equations is underdetermined. Without any additional postulate, we shall be unable to inverse those equations for $[x_1, x_2, x_3]$. In particular, we shall outline how to find such an additional postulate.

Beforehand, we here have to introduce some special notions that are known from operator theory. Within matrix algebra, the index of the linear operator A is the rank $r = rk\mathbf{A}$, here $r = 2$, which coincides with the dimension of the observation space, here $n = dim\mathbb{Y} = 2$. A system of linear equations is called consistent if $rk\mathbf{A} = dim\mathbb{Y}$. Alternatively, we say that the mapping $f : x \longmapsto \mathbf{y} = f(x) \in \mathcal{R}(f)$ or $\mathbf{A} : \mathbf{x} \longmapsto \mathcal{A}\S = \dagger \in \mathcal{R}(A)$ takes an element $\mathbf{x} \in \mathbb{X}$ into the range $\mathcal{R}(f)$ or $\mathcal{R}(A)$, also called the column space of the matrix $\mathbf{A}$: compare with (1.5). Here, the column space is spanned by the first column c_1 and the second column c_2 of the matrix $\mathbf{A}$, the 2×3 array: compare with (1.6). The right complementary index of the linear operator $\mathbf{A} \in \mathbb{R}^{n\times m}$ accounts for the injectivity defect that is given by $d = m..rk\mathbf{A}$ (here $d = m..rk\mathbf{A} = 1$). *Injectivity* relates to the kernel $\mathcal{N}(f)$ or the null space we shall constructively introduce later on.

$$\begin{aligned} f &: \mathbf{x} \longmapsto \mathbf{y} = f(x), f(x) \in \mathcal{R}(f), \\ A &: \mathbf{x} \longmapsto \mathbf{Ax} = \mathbf{y}, \mathbf{y} \in \mathcal{R}(A), \end{aligned} \tag{1.5}$$

$$\mathcal{R}(A) = span\left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right]. \tag{1.6}$$

How can such a linear model of interest, namely a system of consistent linear equations, be generated? Let us assume that we have observed a dynamical system y(t) which is represented by a polynomial of degree two with respect to time $t \in \mathbb{R}$, namely $y(t) = x_1 + x_2 t + x_3 t^2$. Due to $y = 2x_3$, it is a dynamical system with constant acceleration or constant second derivative with respect to time t. The unknown polynomial coefficients are collected in the column array $x = [x_1, \ldots, x_n]', \mathbf{x} \in \mathbb{X} = \mathbb{R}^3, dim\mathbb{X} = 3$. They constitute the coordinates of the three-dimensional parameter space $\mathbb{X}$. If the dynamical system $y(t)$ is observed at two instants, say $y(t_1) = y_1 = 2$ and $y(t_2) = y_2 = 3$, say at $t_1 = 1$ and $t_2 = 2$, respectively, and if we collect the observations in the column array $y = [y_1, y_2]' = [2, 3]', \mathbf{y} \in \mathbb{Y} = \mathbb{R}^2, dim\mathbb{Y} = 2$, they constitute the coordinates of the two-dimensional observation space $\mathbb{Y}$. Thus, we are left with a special linear model interpreted in *Box 1.2*.

Box 1.2. (A special linear model: polynomial of degree two, two observations, and three unknowns).

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{1.7}$$

$$\Longleftrightarrow$$

$$\begin{gathered} t_1 = 1, y_1 = 2 \\ t_2 = 2, y_2 = 3 : \\ \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{gathered} \tag{1.8}$$

$$\sim$$

$$\begin{gathered} \mathbf{y} = \mathbf{Ax} \\ r = rk\mathbf{A} = dim\mathbb{Y} = n = 2 \end{gathered} \tag{1.9}$$

In the next section, let us begin with a more detailed analysis of the linear mapping $f : \mathbf{Ax} = \mathbf{y}$, namely of the linear operator $\mathbf{A} \in \mathbb{R}^{n \times m}, r = rk\mathbf{A} = dim\mathbb{Y} = n$. We shall pay special attention to the three fundamental partitioning, namely (i) algebraic partitioning, called *rank partitioning* of the matrix A, (ii) geometric partitioning, called slicing of the linear space $\mathbb{X}$, and (iii) set-theoretical partitioning, called *fibering* of the domain $\mathcal{D}(f)$.

1-13 The Front Page Example: MINOS, Horizontal Rank Partitioning

Let us go back to the front page consistent system of linear equations, namely the problem to determine three unknown polynomial coefficients from two sampling points which we classified as an underdetermined one. Nevertheless we are able to compute a unique solution of the underdetermined system of inhomogeneous linear equations $\mathbf{Ax} = \mathbf{y}, \mathbf{y} \in \mathcal{R}(A)$ or $rk\mathbf{A} = dim\mathbb{Y}$, here $\mathbf{A} \in \mathbb{R}^{2\times 3}$, $\mathbf{x} \in \mathbb{R}^{3\times 1}$, $\mathbf{y} \in \mathbb{R}^{2\times 1}$, if we determine the coordinates of the unknown vector $\mathbf{x}$ of minimum norm (minimal Euclidean length, $l_2 - norm$), here $\|\mathbf{x}\|_{\text{I}}^2 = \mathbf{xx} = x_1^2 + x_2^2 + x_3^2 = min$. *Box 1.3* outlines the solution of the related optimization problem.

Box 1.3. (MINOS (minimum norm solution) of the front page consistent system of inhomogeneous linear equations, horizontal rank partitioning).

The solution of the optimization problem $\|\mathbf{x}\|_I^2 = min_x \mathbf{Ax} = \mathbf{y}; rk\mathbf{A} = dim\mathbb{Y}g$ is directly based upon the horizontal rank partitioning of the linear mapping $f : x \longmapsto \mathbf{y} = \mathbf{Ax}, rk\mathbf{A} = dim\mathbb{Y}$, which we already have introduced. As soon as $\mathbf{x}_1 = -\mathbf{A}_1^1\mathbf{A}_2\mathbf{x}_2 + \mathbf{A}_1^1\mathbf{y}$ is implemented in the norm $\|\mathbf{x}\|_I^2$, we are prepared to compute the first derivatives of the unconstrained Lagrangean (1.10), which constitutes the necessary conditions. (The theory of vector derivatives is presented in Appendix B.) Following Appendix A, which is devoted to matrix algebra, namely applying $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$ and $(\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$ for appropriate dimensions of the involved matrices such that the identities (1.12) hold, we finally derive (1.13).

$$\begin{aligned}
L(x_1, x_2) &= \|\mathbf{x}\|_{\text{I}}^2 = x_1^2 + x_2^2 + x_3^2 \\
&= (\mathbf{y} - \mathbf{A}_2\mathbf{x}_2)'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}((\mathbf{y} - \mathbf{A}_2\mathbf{x}_2)) + \mathbf{x}_2{}'\mathbf{x}_2 \\
&= \mathbf{y}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y} - 2\mathbf{x}_2{}'\mathbf{A}_2(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y} + \mathbf{x}_2{}'\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2\mathbf{x}_2 + \mathbf{x}_2{}'\mathbf{x}_2 \\
&= min_x \\
\frac{\partial \mathcal{L}}{\partial \mathbf{x}_2}(\mathbf{x}_{2m}) &= 0 \qquad (1.10) \\
&\Longleftrightarrow
\end{aligned}$$

$$\begin{gathered}
-\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y} + [\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2 + \mathbf{I}]\mathbf{x}_{2m} = 0 \\
\Longleftrightarrow \\
\mathbf{x}_{2m} = [\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2 + \mathbf{I}]^{-1}\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y},
\end{gathered} \qquad (1.11)$$

$$\begin{aligned}\mathbf{x}_{2m} &= [\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2 + \mathbf{I}]^{-1}\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y} \\ &= \mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}[\mathbf{A}_2\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1} + \mathbf{I}]^{-1}\mathbf{y} \\ &= \mathbf{A}_2{}'[\mathbf{A}_2\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1} + \mathbf{I}]^{-1}(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{y},\end{aligned} \tag{1.12}$$

$$\mathbf{x}_{2m} = \mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}\mathbf{y} \tag{1.13}$$

The second derivatives that are given by (1.13), due to positive-definiteness of the matrix $\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2 + \mathbf{I}$, generate the sufficiency condition for obtaining the minimum of the unconstrained Lagrangean. Carrying out the backward transform $x_{2m} \longmapsto x_{1m} = -\mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{x}_2 + \mathbf{A}_1^{-1}\mathbf{y}$, we obtain (1.15).

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}_2 \partial \mathbf{x}_2{}'}(\mathbf{x}_{2m}) = 2[\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}')^{-1}\mathbf{A}_2 + \mathbf{I}] > 0, \tag{1.14}$$

$$\mathbf{x}_{2m} = \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}\mathbf{y} + \mathbf{A}_1^{-1}\mathbf{y}. \tag{1.15}$$

Let us additionally right multiply the identity $\mathbf{A}_1\mathbf{A}_1{}' = -\mathbf{A}_2\mathbf{A}_2{}' + (\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')$ by $(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}$, such that (1.16) holds, and left multiply by $\mathbf{A}_1{}'$, such that (1.17) holds. Obviously, we have generated the linear forms (1.18)–(1.20).

$$\mathbf{A}_1\mathbf{A}_1{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} = -\mathbf{A}_2\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} + \mathbf{I}, \tag{1.16}$$

$$\mathbf{A}_1{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} = \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} + \mathbf{A}_1^{-1}, \tag{1.17}$$

$$\begin{aligned}\mathbf{x}_{1m} &= \mathbf{A}_1{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}\mathbf{y}, \\ &= \mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}\mathbf{y}\end{aligned} \tag{1.18}$$

$$\begin{bmatrix}\mathbf{x}_{1m} \\ \mathbf{x}_{2m}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1{}' \\ \mathbf{A}_2{}'\end{bmatrix}(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1}\mathbf{y} \tag{1.19}$$

$$\mathbf{x}_m = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{y} \tag{1.20}$$

The following relations supply us with a numerical computation with respect to the introductory example.

$$\begin{aligned}&(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}') = \begin{bmatrix}3 & +7 \\ +7 & 21\end{bmatrix}, \ (\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} = \begin{bmatrix}21 & -7 \\ -7 & 3\end{bmatrix} \\ &\mathbf{A}_1{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} = \frac{1}{14}\begin{bmatrix}14 & -4 \\ 7 & -1\end{bmatrix}, \ \mathbf{A}_2{}'(\mathbf{A}_1\mathbf{A}_1{}' + \mathbf{A}_2\mathbf{A}_2{}')^{-1} = \frac{1}{14}[-7, 6],\end{aligned} \tag{1.21}$$

$$\mathbf{x}_{1m} = \begin{bmatrix} 3/7 \\ 11/14 \end{bmatrix}, \ \mathbf{x}_{2m} = 1/14, \ \|\mathbf{x}_m\|_I = 3\sqrt{42}/14,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}_2 \partial \mathbf{x}_2{}'}(\mathbf{x}_{2m}) = 28 > 0, \tag{1.22}$$

$$y(t) = \frac{8}{7} + \frac{11}{14}t + \frac{1}{14}t^2. \tag{1.23}$$

In the next section, let us go into the detailed analysis of $\mathcal{R}(f)$, $\mathcal{N}(f)$, and $\mathcal{N}^{\perp}(f)$ with respect to the front page example. Actually how can we identify the range space $\mathcal{R}(A)$, the null space $\mathcal{N}(\mathbf{A})$ or its orthogonal complement $\mathcal{N}^{\perp}(A)$.

1-14 The Range $\mathcal{R}(f)$ and the Kernel $\mathcal{N}(f)$

The range space $\mathcal{R}(A)$ is conveniently described by the first column $c_1 = [1, 1]'$ and the second column $c_2 = [1, 2]'$ of the matrix $\mathbf{A}$, namely 2-leg with respect to the orthogonal base vector $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively, attached to the origin O. Symbolically, we write the 2-leg in the form (1.24). Symbolically, we write $\mathcal{R}(A)$ in the form (1.25). As a linear space, $\mathcal{R}(A) \subset \mathbf{y}$ is illustrated by Fig. 1.2.

$$\{e_1 + e_2, e1 + 2e_2 | O\} \, or \, \{ec_1, ec_2 | O\}; \tag{1.24}$$

$$\mathcal{R}(A) = span\,[e_1 + e_2, e1 + 2e_2 | O]\,. \tag{1.25}$$

By means of *Box 1.4*, we identify $\mathcal{N}(f)$ or the null space $\mathcal{N}(A)$ and give its illustration by Fig. 1.2. Such a result has paved the way to the diagnostic algorithm for solving an underdetermined system of linear equations by means of rank partitioning presented in *Box 1.5*.

Box 1.4. (General solution of a system of homogeneous linear equations $\mathbf{Ax} = 0$, horizontal rank partitioning).

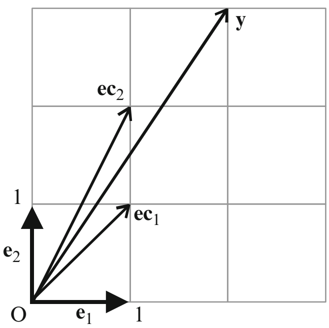


Fig. 1.2 Range $\mathcal{R}(f)$, range space $\mathcal{R}(A)$, $\mathbf{y} \in \mathcal{R}(A)$

The matrix A is called horizontally rank partitioned if (1.26) holds. (In the introductory example, $\mathbf{A} \in \mathbb{R}^{2\times 3}$, $\mathbf{A}_1 \in \mathbb{R}^{2\times 2}$, $\mathbf{A}_2 \in \mathbb{R}^{2\times 1}$, and $d(\mathbf{A}) = 1$ applies.) A consistent system of linear equations $\mathbf{Ax} = \mathbf{y}, rk\mathbf{A} = dim\mathbb{Y}$ is horizontally rank partitioned if (1.27) for a partitioned unknown vector (1.28) applies.

$$\mathbf{A} \in \mathbb{R}^{n\times m} \wedge \mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2] \begin{cases} r = rk\mathbf{A} = rk\mathbf{A}_1 = n, \\ \mathbf{A}_1 \in \mathbb{R}^{n\times r}, \mathbf{A}_2 \in \mathbb{R}^{n\times d}, \\ d = d(\mathbf{A}) = m - rk\mathbf{A}, \end{cases} \tag{1.26}$$

$$\mathbf{Ax} = \mathbf{y}, rk\mathbf{A} = dim\mathbb{Y} \Longleftrightarrow \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = \mathbf{y}, \tag{1.27}$$

$$\left\{ \mathbf{x} \in \mathbb{R}^{n\times m} \wedge \mathbf{A} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} | \mathbf{x}_1 \in \mathbb{R}^{r\times 1}, \mathbf{x}_2 \in \mathbb{R}^{d\times 1} \right\}. \tag{1.28}$$

The horizontal rank partitioning of the matrix $\mathbf{A}$ as well as the horizontally rank partitioned consistent system of linear equations $\mathbf{Ax} = \mathbf{y}, rk\mathbf{A} = dim\mathbb{Y}$ of the introductory example is

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, A_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \tag{1.29}$$

$$\begin{gathered} \mathbf{Ax} = \mathbf{y}, rk\mathbf{A} = dim\mathbb{Y} \Longleftrightarrow \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = \mathbf{y}, \\ \mathbf{x}_1 = [x_1, x_2] \in \mathbb{R}^{r\times 1}, \mathbf{x}_2 = [x_3] \in \mathbb{R}, \\ \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_3 = \mathbf{y}. \end{gathered} \tag{1.30}$$

By means of the horizontal rank partitioning of the system of homogenous linear equations, an identification of the null space (1.31) is provided by (1.32). In particular, in the introductory example, we are led to (1.33) and (1.34).

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{Ax} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = 0\}, \tag{1.31}$$

$$\mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = 0 \Longleftrightarrow \mathbf{x}_1 = -\mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{x}_2, \tag{1.32}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_3, \tag{1.33}$$

$$x_1 = 2x_3 = 2\tau, x_2 = -3x_3 = -3\tau, x_3 = \tau. \tag{1.34}$$

Here, the two equations $\mathbf{Ax} = 0$ for any $\mathbf{x} \in \mathbb{X} = \mathbb{R}^2$ constitute the linear space $\mathcal{N}(\mathbf{A}), dim\mathcal{N}(\mathbf{A}) = 1$, a one-dimensional subspace of $\mathbb{X} = \mathbb{R}^2$. For instance, if we introduce the parameter $x_3 = \tau$, the other coordinates of the parameter space $\mathbb{X} = \mathbb{R}^2$ amount to $x_1 = 2x_3 = 2\tau, x_2 = -3x_3 = -3$. In geometric language, the linear space $\mathcal{N}(\mathbf{A})$ is a parameterized straight line $\mathbb{L}_0^1$ through the origin O illustrated by Fig. 1.3. The parameter space $\mathbb{X} = \mathbb{R}^r$ (here r = 2) is sliced by the subspace, the

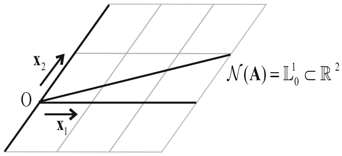


Fig. 1.3 The parameter space $\mathbb{X} = \mathbb{R}^3$ (x_3 is not displayed) sliced by the null space, the linear manifold $\mathcal{N}(\mathbf{A}) = \mathbb{L}_0^1 \subset \mathbb{R}^2$

linear space $\mathcal{N}(\mathbf{A})$, also called *linear manifold*, and $dim\mathcal{N}(\mathbf{A}) = d(A) = d$ (here a straight line $\mathbb{L}_0^1$ through the origin O).

In the following section, let us consider the interpretation of MINOS by the three fundamental partitionings: *algebraic* (*rank partitioning*), *geometric (slicing)*, and *set-theoretical (fibering)*.

1-15 The Interpretation of MINOS

The diagnostic algorithm for solving an underdetermined system of linear equations $\mathbf{Ax} = \mathbf{y}$, $rk\mathbf{A} = dim\,\mathbb{Y} = n$, $n < m = dim\mathbb{X}, \mathbf{y} \in \mathcal{R}(A)$ by means of rank partitioning is presented to you in *Box 1.5* While we have characterized the general solution of the system of homogenous linear equations $\mathbf{Ax} = 0$, we are left with the problem of solving the consistent system of inhomogeneous linear equations. Again, we take advantage of the rank partitioning of the matrix $\mathbf{A}$ summarized in *Box 1.6*.

Box 1.5. (A diagnostic algorithm for solving an underdetermined system of linear equations by means of rank partitioning).

A diagnostic algorithm for solving an underdetermined system of linear equations $\mathbf{Ax} = \mathbf{y}$, $rk\mathbf{A} = dim\mathbb{Y}, \mathbf{y} \in \mathcal{R}(A)$ by means of rank partitioning is immediately introduced by the following chain of statements.

$$\text{Determine the rank of the matrix } \mathbf{A}:$$
$$rk\mathbf{A} = dim\mathbb{Y} = n.$$
$$\Downarrow$$

$$\text{Compute the horizontal rank partitioning}:$$
$$\mathbf{A} = [A_1, A_2], A_1 \in \mathbb{R}^{r\times r} = \mathbb{R}^{n\times n}, A_2 \in \mathbb{R}^{n\times(m-r)} = \mathbb{R}^{n\times(m-n)}$$
$$m - r = m - n \text{ is called right complementary index.}$$
$$\mathbf{A} \text{ as a linear operator is not injective, but surjective.}$$
$$\Downarrow$$

Compute the null space $\mathcal{N}(\mathbf{A})$:
$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{A}\mathbf{x} = 0\} = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{x}_1 + \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{x}_2 = 0\}.$$
$$\Downarrow$$

Compute the unknown parameter vector of type MINOS:
$$\mathbf{x}_m = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{y}.$$

Box 1.6. (A special solution of a consistent system of inhomogeneous linear equations $\mathbf{A}\mathbf{x} = \mathbf{y}$, horizontal rank partitioning).

$$\mathbf{A}\mathbf{x} = \mathbf{y} \begin{cases} rk\mathbf{A} = dim\mathbb{Y} \\ \mathbf{y} \in \mathcal{R}(A) \end{cases} \Longleftrightarrow \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = \mathbf{y} \tag{1.35}$$

Since the matrix $\mathbf{A}_1$ is of full rank, it can be regularly inverted (Cayley inverse). In particular, we solve for $\mathbf{x}_1$:

$$\mathbf{x}_1 = -\mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{x}_2 + \mathbf{A}_1^{-1}\mathbf{y} \tag{1.36}$$

or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_3 + \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \tag{1.37}$$

$$x_1 = 2x_3 + 2y_1 - y_2, x_2 = -3x_3 - y_1 + y_2. \tag{1.38}$$

For instance, if we introduce the parameter $x_3 = \tau$, the other coordinates of the parameter space $\mathbb{X} = \mathbb{R}^2$ amount to (1.39). In geometric language, the admissible parameter space is a family of a one-dimensional linear space, a family of one-dimensional parallel straight lines dependent on $y = [y_1, y_2]'$, here $[2, 3]'$, in particular, the family (1.40), including the null space $\mathbb{L}^1_{0,0} = \mathcal{N}(\mathbf{A})$.

$$\begin{aligned} x_1 &= 2\tau + 2y_1 - y_2, \\ x_2 &= -3\tau - y_1 + y_2, \end{aligned} \tag{1.39}$$

$$\mathbb{L}^1_{y_1,y_2} = \{\mathbf{x} \in \mathbb{R}^3 | 2x_3 + 2y_1 - y_2, x_2 = -3x_3 - y_1 + y_2\}. \tag{1.40}$$

Figure 1.4 illustrates (i) the admissible parameter space $\mathbb{L}^1_{(y_1,y_2)}$, (ii) the line $\mathbb{L}^1_{\perp}$ which is orthogonal to the null space called $\mathcal{N}^{\perp}(A)$, and (iii) the intersection $\mathbb{L}^1_{(y_1,y_2)} \cap \mathcal{N}^{\perp}(A)$, generating the solution point $\mathbf{x}_m$ as will be proven now.

The geometric interpretation of the minimum norm solution is the following: With reference to Fig. 1.5, we decompose the vector $\mathbf{x} = \mathbf{x}_{\mathcal{N}(\mathbf{A})} + \mathbf{x}_{\mathcal{N}^{\perp}(\mathbf{A})}$, where $\mathbf{x}_{\mathcal{N}(\mathbf{A})}$ is an element of the null space $\mathcal{N}(\mathbf{A})$ (here the straight line $\mathbb{L}^1_{0,0}$), and $\mathbf{x}_{\mathcal{N}^{\perp}(\mathbf{A})} - \mathbf{x}_0$ is an element of the null space $\mathcal{N}(\mathbf{A})$ (here the straight line $\mathbb{L}^1_{0,0}$), while $\mathbf{i}_{\mathcal{N}(\mathbf{A})} = \mathbf{i}_m$ is an element of the rang space $\mathcal{R}(A^-)$ (here the above outlined

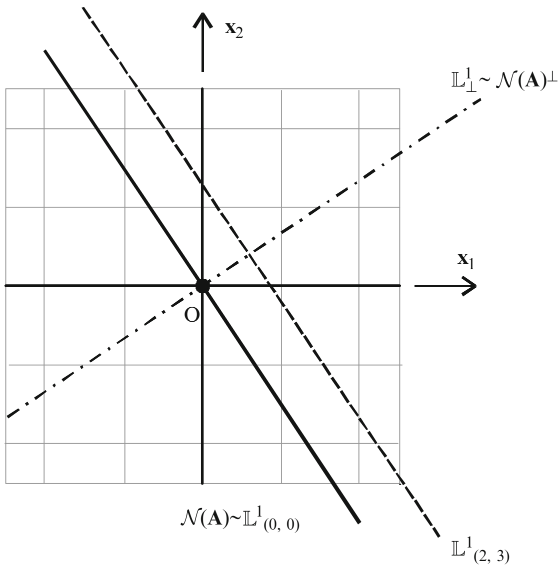


Fig. 1.4 The range space $\mathcal{R}(A^-)$ (the admissible parameter space), parallel straight lines $\mathbb{L}^1_{(y_1,y_2)}$, namely $\mathbb{L}^1_{(2,3)}$

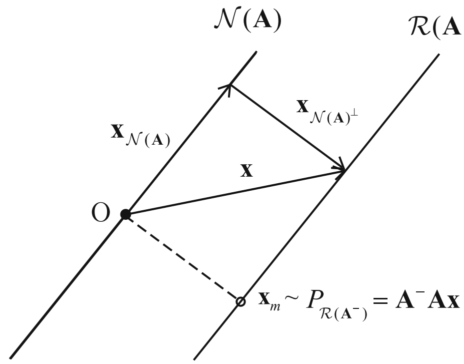


Fig. 1.5 Orthogonal projection of an element of $\mathcal{N}(\mathbf{A})$ onto the range space $\mathcal{R}(A^-)$

straight line $\mathbb{L}^1_{y_1,y_2}$, namely $\mathbb{L}^1_{2,3}$ of the generalized inverse matrix A^- of type *minimum norm solution*, which in this book mostly is abbreviated as "MINOS". $\| \mathbf{x} \|_I^2 = \| \mathbf{x}_{\mathcal{N}(\mathbf{A})} + \mathbf{x}_{\mathcal{N}^{\perp}(A)} \|^2 = \| \mathbf{x} \|_I^2 + 2 < \mathbf{x} \mid \mathbf{i} > + \| \mathbf{i} \|^2$ is minimal if and only if the inner product $< \mathbf{x} \mid \mathbf{i} >= 0$ and $\mathbf{x}_{\mathcal{N}(\mathbf{A})}$ and $\mathbf{i}_{\mathcal{N}^{\perp}(A)} = \mathbf{i}_m$ are orthogonal. The solution point $\mathbf{x}_m$ is the orthogonal projection of the null point onto $\mathcal{R}(A^-)$: $P_{\mathcal{R}(A^-)} = \mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{A}^-\mathbf{y}$ for all $\mathbf{x} \in D(\mathbf{A})$. Alternatively, if the vector $\mathbf{x}_m$ of minimal length is orthogonal to the null space $\mathcal{N}(\mathbf{A})$, being an element of $\mathcal{N}^{\perp}(\mathbf{A})$ (here the line $\mathbb{L}^1_{0,0}$), we may say that $\mathcal{N}^{\perp}(\mathbf{A})$ intersects $\mathcal{R}(A^-)$ in the solution point $\mathbf{x}_m$. Or we may say that the normal space $\mathbb{NL}^1_0$ with respect to the tangent space $\mathbb{TL}^1_0$ – which is in linear models identical to $\mathbb{L}^1_0$, the null space $\mathcal{N}(\mathbf{A})$ – intersects the tangent space $\mathbb{TL}^1_0$, the range space $\mathcal{R}(A^-)$ in the solution point $\mathbf{x}_m$. In summary, $\mathbf{x}_m \in \mathcal{N}^{\perp}(\mathbf{A}) \cap \mathcal{R}(A^-)$.

Let the algebraic partitioning and the geometric partitioning be merged to interpret the minimum norm solution of the consistent system of linear equations of type *underdetermined MINOS*. As a summary of such a merger, we take reference to *Box 1.7*.

Box 1.7. (General solution of a consistent system of inhomogeneous linear equations $\mathbf{Ax} = \mathbf{y}$).

Consistent system of equations $f : \mathbf{x} \longmapsto \mathbf{y} = \mathbf{Ax}, \mathbf{x} \in \mathbb{X} = \mathbb{R}^n$ (parameter space), $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n$(observation space), $r = rk\mathbf{A} = dim\mathbb{Y}$, $\mathcal{A}^-$ generalized inverse of type MINOS.

$$\begin{array}{cc} \textit{Condition}\#1: & \textit{Condition}\#1: \\ f(\mathbf{x}) = f(g(\mathbf{y})) & \mathbf{Ax} = \mathbf{AA}^-\mathbf{Ax} \\ \Longleftrightarrow & \Longleftrightarrow \\ f = f \circ g \circ f. & \mathbf{AA}^-\mathbf{A} = \mathbf{A}. \end{array} \tag{1.41}$$

$$\begin{array}{cc} & \textit{Condition}\#2: \\ \textit{Condition}\#2: & \text{(reflexive g-inverse)} \\ \text{(reflexive g-inverse mapping)} & \mathbf{x}_- = \mathbf{A}^-\mathbf{y} = \mathbf{A}^-\mathbf{AA}^-\mathbf{y} \\ \mathbf{x} = g(\mathbf{y}) = g(f(\mathbf{x})) & \Longleftrightarrow \\ & \mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-. \end{array} \tag{1.42}$$

$$\begin{array}{cc} \textit{Condition}\#3: & \textit{Condition}\#3: \\ g(f(\mathbf{x})) = \mathbf{x}_{\mathcal{R}(\mathbf{A}^-)} & \mathbf{A}^-\mathbf{A} = \mathbf{x}_{\mathcal{R}(\mathbf{A}^-)} \\ \Longleftrightarrow & \Longleftrightarrow \\ g \circ f = \mathrm{Proj}_{\mathcal{R}(\mathbf{A}^-)}. & \mathbf{A}^-\mathbf{A} = \mathrm{Proj}_{\mathcal{R}(\mathbf{A}^-)}. \end{array} \tag{1.43}$$

Regarding the above first condition $\mathbf{AA}^-\mathbf{A} = \mathbf{A}$ (let us here depart from MINOS of $\mathbf{y} = \mathbf{Ax}$, $\mathbf{x} \in \mathbb{X} = \mathbb{R}^n$, $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n$, $r = rk\mathbf{A} = n$, i.e., $\mathbf{x}_m = \mathbf{A}_m^-\mathbf{y} = \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{y}$: see (1.44). Concerning the above second condition $\mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-$ see (1.45). Concerning the above third condition $\mathbf{A}^-\mathbf{A} = \mathrm{Proj}_{\mathcal{R}(\mathbf{A}^-)}$: see (1.46).

$$\mathbf{Ax}_m = \mathbf{AA}_m^-\mathbf{y} = \mathbf{AA}_m^-\mathbf{Ax}_m \Longrightarrow \mathbf{Ax}_m = \mathbf{AA}_m^-\mathbf{Ax}_m \Longleftrightarrow \mathbf{AA}^-\mathbf{A} = \mathbf{A}, \tag{1.44}$$

$$\begin{gathered} \mathbf{x}_m = \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{y} = \mathbf{A}_m^-\mathbf{y} = \mathbf{AA}_m^-\mathbf{Ax} \Longrightarrow \mathbf{x}_m = \mathbf{A}_m^-\mathbf{y} = \mathbf{A}_m^-\mathbf{AA}_m^-\mathbf{y} \Longrightarrow \\ \Longrightarrow \mathbf{A}_m^-\mathbf{y} = \mathbf{A}_m^-\mathbf{AA}_m^-\mathbf{y} \Longleftrightarrow \mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-, \end{gathered} \tag{1.45}$$

$$\mathbf{x}_m = \mathbf{A}_m^-\mathbf{y} = \mathbf{A}_m^-\mathbf{Ax}_m \Longleftrightarrow \mathbf{A}^-\mathbf{A} = \mathrm{Proj}_{\mathcal{R}(\mathbf{A}^-)} := \mathrm{Proj}_{\mathcal{R}(\mathbf{A}^-)} \tag{1.46}$$

How are $rk\mathbf{A}_m^- = rk\mathbf{A}$ and $\mathbf{A}_m^-\mathbf{A}$ interpreted? On the interpretation of $rk\mathbf{A}_m^- = rk\mathbf{A}$: the g-inverse of type MINOS is the generalized inverse of maximal rank since in general $rk\mathbf{A}^- \leq rk\mathbf{A}$ holds. On the interpretation of $\mathbf{A}_m^-\mathbf{A}$: $\mathbf{A}_m^-\mathbf{A}$ is an orthogonal projection onto $\mathcal{R}(\mathbf{A}^-)$, but $\mathbf{i}_m = \mathbf{I} - \mathbf{A}_m^-\mathbf{A}$ onto its orthogonal complement $\mathcal{R}^\perp(\mathbf{A}^-)$.

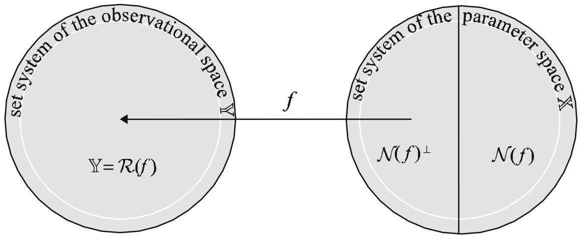


Fig. 1.6 Venn diagram, trivial fibering of the domain $\mathcal{D}(f)$, trivial fibers $\mathcal{N}(f)$ and $\mathcal{N}^{\perp}(f)$, $f : \mathbb{R}^n = \mathbb{X} \longrightarrow \mathbb{Y} = \mathbb{R}^n$, $\mathbb{Y} = \mathcal{R}(f)$, $\mathbb{X}$ is the set system of the parameter space, $\mathbb{Y}$ is the set system of the observation space

If the linear mapping $f : \mathbf{x} \longmapsto \mathbf{y} = f(\mathbf{x}), \mathbf{y} \in \mathcal{R}(f)$ is given, we are aiming at a generalized inverse (linear) mapping $\mathbf{y} \longmapsto \mathbf{x} = g(\mathbf{y})$ in such a manner that $\mathbf{y} = f(\mathbf{x}) = f(g(\mathbf{y})) = f(g(f(\mathbf{x})))$ or $f = f \circ g \circ f$ as a first condition is fulfilled. Alternatively, we are going to construct a generalized inverse $A\mathbf{A}^{-} : \mathbf{y} \longmapsto \mathbf{A}^{-}\mathbf{y} = \mathbf{x}_{-}$ in such a manner that the first condition $\mathbf{y} = \mathbf{A}\mathbf{x}_{-} = A\mathbf{A}^{-}\mathbf{A}\mathbf{y}$ or $\mathbf{A} = \mathbf{A}\mathbf{A}^{-}\mathbf{A}$ holds. Though the linear mapping $f : \mathbf{x} \longmapsto \mathbf{y} = f(\mathbf{x}), \mathbf{y} \in \mathcal{R}(f)$ or the system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{y}$, $rk\mathbf{A} = dim\mathbb{Y}$ is consistent, it suffers from the (injectivity) deficiency of the linear mapping $f(\mathbf{x})$ or of the matrix $\mathbf{A}$. Indeed, it recovers from the (injectivity) deficiency if we introduce the projection $\mathbf{x} \longmapsto g(f(\mathbf{x})) = q \in \mathcal{R}(g)$ or $\mathbf{x} \longmapsto \mathbf{A}^{-}\mathbf{A}\mathbf{x} = q \in \mathcal{R}(\mathbf{A}^{-})$ as the second condition. The projection matrix $\mathbf{A}^{-}\mathbf{A}$ is *idempotent*, which follows from $\mathbf{P}^2 = \mathbf{P}$ or $(\mathbf{A}^{-}\mathbf{A})(\mathbf{A}^{-}\mathbf{A}) = \mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}^{-}\mathbf{A}$.

Let us here finally outline the set-theoretical partitioning, the fibering of the set system of points which constitute the parameters space $\mathbb{X}$, the domain $\mathcal{D}(f)$. Since the set system $\mathbb{X}$ (the parameters space) is $\mathbb{R}^r$, the fibering is called *trivial*. Non-trivial fibering is reserved for nonlinear models in which case we are dealing with a parameters space $\mathbb{X}$ which is a *differentiable manifold*. Here, the fibering $\mathcal{D}(f) = \mathcal{N}(f) \cup \mathcal{N}^{\perp}(f)$ produces the trivial fibers $\mathcal{N}(f)$ and $\mathcal{N}^{\perp}(f)$, where the trivial fiber $\mathcal{N}^{\perp}(f)$ is the *quotient set* $\mathbb{R}^n/\mathcal{N}(f)$. By means of a Venn diagram (*John Venn* 1834–1928), also called Euler circles (*Leonhard Euler* 1707–1783), Fig. 1.6 illustrates the trivial fibers of the set system $\mathbb{X} = \mathbb{R}^m$ that is generated by $\mathcal{N}(f)$ and $\mathcal{N}^{\perp}(f)$. Of course, the set system of points which constitute the observation space $\mathbb{Y}$ is not subject to fibering since all points of the set system $\mathcal{D}(f)$ are mapped *into* the range $\mathcal{R}(f)$.

1-2 Minimum Norm Solution (MINOS)

The system of consistent linear equations $\mathbf{A}\mathbf{x} = \mathbf{y}(\mathbf{A} \in \mathbb{R}^{n \times m}, rk\mathbf{A} = n < m)$ allows certain solutions which we introduce by means of *Definition 1.1* as a solution of a certain optimization problem. *Lemma 1.2* contains the normal equations of the optimization problem. The solution of such a system of normal equations is presented in Lemma 1.3 as the minimum norm solution (MINOS) with respect to the G_x-seminorm. Finally, we discuss the metric of the parameter space and alternative choices of its metric before we identify by *Lemma 1.4* the solution of the quadratic optimization problem in terms of the $(1, 2, 4)$-generalized inverse.

By *Definition 1.1*, we characterize G_x-MINOS of the consistent system of linear equations $\mathbf{Ax} = \mathbf{y}$ subject to $\mathbf{A} \in \mathbb{R}^{n \times m}$, $rk\mathbf{A} = n < m$ (algebraic condition) or $\mathbf{y} \in \mathcal{R}(A)$ (geometric condition). Loosely speaking, we are confronted with a system of linear equations with more unknowns m than equations n, namely n < m. G_x-MINOS will enable us to find a solution of this underdetermined problem. By means of Lemma 1.2, we shall write down the "normal equations" of G_x-MINOS.

Definition 1.1. (The minimum norm solution (MINOS) with respect to the G_x-seminorm).

A vector $\mathbf{x}_m$ is called G_x-MINOS (minimum norm solution (MINOS) with respect to the G_x-seminorm) of the consistent system of linear equations (1.47) if both $\mathbf{Ax}_m = \mathbf{y}$ and – in comparison to all other vectors of solution $\mathbf{x} \in \mathbb{X} \equiv \mathbb{R}^n$ – the inequality $\| \mathbf{x}_m \|^2_{G_x} := \mathbf{x}'_m \mathbf{G}_x \mathbf{x}_m \leq \mathbf{x}'\mathbf{G}_x\mathbf{x} :=\| \mathbf{x} \|^2_{G_x}$ holds. The system of inverse linear equations $\mathbf{A}^{-}\mathbf{y} + \mathbf{i} = \mathbf{x}$, $rk\mathbf{A}^{-} \neq m$ or $\mathbf{x} \notin \mathcal{R}(\mathbf{A}^{-})$ is inconsistent.

$$\mathbf{Ax} = \mathbf{y}, \mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n \begin{cases} rk\mathbf{A} = rk[\mathbf{A}, \mathbf{y}] = n < m \\ \qquad \mathbf{A} \in \mathbb{R}^{n \times m} \\ \qquad \mathbf{y} \in \mathcal{R}(A). \end{cases} \tag{1.47}$$

Lemma 1.2. (The minimum norm solution (MINOS) with respect to the $\mathbf{G}_x$ seminorm).

A vector $\mathbf{x}_m \in \mathbb{X} \equiv \mathbb{R}^n$ is $\mathbf{G}_x$-MINOS of (1.47) if and only if the system of equations (1.48) with the vector $\lambda \in \mathbb{R}^n \times 1$ of *Lagrange multipliers* is fulfilled. x_m exists always and is, in particular, unique if $rk[\mathbf{G}_x, \mathbf{A}'] = m$ holds or, equivalently, if the matrix $\mathbf{G}_x + \mathbf{A}'\mathbf{A}$ is regular.

$$\begin{bmatrix} \mathbf{G}_x & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_m \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} \tag{1.48}$$

Proof. G_x-MINOS is based on the constrained Lagrangean (1.49) such that the first derivatives (1.50) constitute the necessary conditions. The second derivatives (1.51), according to the positive semi-definiteness of the matrix $\mathbf{G}_x$, generate the sufficiency condition for obtaining the *minimum* of the *constrained Lagrangean*. According to the assumption $rk\mathbf{A} = rk[\mathbf{A}, \mathbf{y}] = n$ or $\mathbf{y} \in \mathcal{R}(\mathbf{A})$ the existence of $\mathbf{G}_x$-MINOS $\mathbf{x}_m$ is guaranteed. In order to prove uniqueness of $\mathbf{G}_x$-MINOS $\mathbf{x}_m$, we have to consider case (i) ($\mathbf{G}_x$ positive-definite) and we have to consider case (ii) ($\mathbf{G}_x$ positive semi-definite).

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}'\mathbf{G}_x\mathbf{x} + 2\lambda'(\mathbf{Ax} - \mathbf{y}) = min_{\mathbf{x},\lambda} \tag{1.49}$$

$$\left.\begin{aligned}\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\mathbf{x}}(\mathbf{x}_m,\lambda_m) &= \mathbf{G}_x\mathbf{x}_m + \mathbf{A}'\lambda_m = 0\\ \frac{1}{2}\frac{\partial\mathcal{L}}{\partial\lambda}(\mathbf{x}_m,\lambda_m) &= \mathbf{A}\mathbf{x}_m - \mathbf{y} = 0\end{aligned}\right\} \Longleftrightarrow \begin{bmatrix}\mathbf{G}_x & \mathbf{A}'\\ \mathbf{A} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_m\\ \lambda\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{y}\end{bmatrix}, \tag{1.50}$$

$$\frac{1}{2}\frac{\partial^2\mathcal{L}}{\partial\mathbf{x}\partial\mathbf{x}'}(\mathbf{x}_m,\lambda_m) = \mathbf{G}_x \geq 0. \tag{1.51}$$

Case (i) ($\mathbf{G}_x$ positive-definite):

Due to $rk\mathbf{G}_x = m$ and $|\mathbf{G}_x| \neq 0$, the partitioned system of normal equations (ref) is uniquely solved. The theory of inverse partitioned matrices (IPM) is presented in Appendix A.

$$|\mathbf{G}_x| \neq 0,\ \begin{bmatrix}\mathbf{G}_x & \mathbf{A}'\\ \mathbf{A} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_m\\ \lambda\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{y}\end{bmatrix}. \tag{1.52}$$

Case (ii) ($\mathbf{G}_x$ positive semi-definite):

Follow these algorithmic steps. Multiply the second normal equation by $\mathbf{A}'$ in order to produce $\mathbf{A}'\mathbf{A}\mathbf{x} - \mathbf{A}'\mathbf{y} = 0$ or $\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{y}$ and add the result to the first normal equation, which generates $\mathbf{G}_x\mathbf{x}_m + \mathbf{A}'A\mathbf{x}_m + \mathbf{A}'\lambda m = \mathbf{A}'\mathbf{y}$ and therefore $(\mathbf{G}_x + \mathbf{A}'\mathbf{A})\mathbf{x}_m + \mathbf{A}'\lambda m = \mathbf{A}'\mathbf{y}$. The augmented first normal equation and the original second normal equation build up the equivalent system of normal equations (1.53), which is uniquely solved due to $rk[\mathbf{G}_x + \mathbf{A}'\mathbf{A}]= m, |\mathbf{G}_x + \mathbf{A}'\mathbf{A}| \neq 0$.

$$|\mathbf{G}_x + \mathbf{A}'\mathbf{A}| \neq 0,\ \begin{bmatrix}\mathbf{G}_x + \mathbf{A}'\mathbf{A} & \mathbf{A}'\\ \mathbf{A} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_m\\ \lambda\end{bmatrix} = \begin{bmatrix}\mathbf{A}'\\ \mathbf{I}_n\end{bmatrix}\mathbf{y}. \tag{1.53}$$

The solution of the system of normal equations directly leads to the linear form $\mathbf{x}_m = \mathbf{L}\mathbf{y}$, which is $\mathbf{G}_x$-MINOS of (1.48) and can be represented as follows.

Lemma 1.3. (The minimum norm solution (MINOS) with respect to the $\mathbf{G}_x$-seminorm).

$\mathbf{x}_m = \mathbf{L}\mathbf{y}$ is $\mathbf{G}_x$-MINOS of the consistent system of linear equations (1.48), i.e. $\mathbf{A}\mathbf{x} = \mathbf{y}$ and $rk\mathbf{A} = rk[\mathbf{A},\mathbf{y}] = n$ or $\mathbf{y} \in \mathcal{R}(A)$, if and only if $L \in \mathbb{R}^{m\times n}$ is represented by the following "relations".

Case (i) $(\mathbf{G}_x = \mathbf{I}_m)$:

$$\mathbf{L} = \mathbf{A}_g^- = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} \text{ (right inverse)}, \tag{1.54}$$

$$\mathbf{x}_m = \mathbf{A}_g^-\mathbf{y} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{y},\ \mathbf{x} = \mathbf{x}_m + \mathbf{i}_m. \tag{1.55}$$

$\mathbf{x} = \mathbf{x}_m + \mathbf{i}_m$ is an orthogonal decomposition of the unknown vector $\mathbf{x} \in \mathbb{X} \equiv \mathbb{R}^m$ into the I-MINOS vector $\mathbf{x}_m \in \mathbb{L}^n$ and the I-MINOS vector of inconsistency $\mathbf{i}_m \in \mathbb{L}^d$ subject to $\mathbf{x}_m = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}\mathbf{x}$ and $\mathbf{i}_m = \mathbf{x}-\mathbf{x}_m = [\mathbf{I}_m - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}]\mathbf{x}$. (According to $\mathbf{x}_m = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}\mathbf{x}$, I-MINOS has the reproducing property. As projection matrices, the matrix $\mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A}$ with $rk\mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A} = rk\mathbf{A} = n$ and the matrix $[\mathbf{I}_m - \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A}]$ with $rk[\mathbf{I}_m - \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A}] = n - rk\mathbf{A} = d$ are independent). Their corresponding norms are positive semi-definite, namely $\| \mathbf{x}_m \|^2_{I_m}$ and $\| \mathbf{i}_m \|^2_{I_m}$ yield $\| \mathbf{x}_m \|^2_{I_m} = \mathbf{y}'\mathbf{AA}')^{-1}\mathbf{y} = \mathbf{x}'\mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{Ax} = \mathbf{x}'\mathbf{G}_m\mathbf{x}$ and $\| \mathbf{i}_m \|^2_{I_m} = \mathbf{x}'[\mathbf{I}_m - \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A}]\mathbf{x}$. In particular, note that

$$\mathbf{x}_m = \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{Ax},\ \mathbf{i}_m = \mathbf{x} - \mathbf{x}_m = [\mathbf{I}_m - \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{A}]\mathbf{x}. \tag{1.56}$$

Case (ii) ($\mathbf{G}_x$ positive-definite) :

$$\mathbf{L} = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1} \text{ (weighted right inverse)}, \tag{1.57}$$

$$\mathbf{x}_m = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y},\ \mathbf{x} = \mathbf{x}_m + \mathbf{i}_m. \tag{1.58}$$

$\mathbf{x} = \mathbf{x}_m + \mathbf{i}_m$ is an orthogonal decomposition of the unknown vector $\mathbf{x} \in \mathbb{X} \equiv \mathbb{R}^m$ into the $\mathbf{G}_x$-MINOS vector $\mathbf{x}_m \in \mathbb{L}^n$ and the $\mathbf{G}_x$-MINOS vector of inconsistency $\mathbf{i}_m \in \mathbb{L}^d$ ($\mathbf{x}_m = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax},\ \mathbf{i}_m = [\mathbf{I}_m - \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax}$). (Due to $\mathbf{x}_m = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax}$, $\mathbf{G}_x$-MINOS has the reproducing property. As projection matrices, the matrix $\mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}$ with $rk\mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A} = rk\mathbf{A} = n$ and $[\mathbf{I}_m - \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}]$ with $rk[\mathbf{I}_m - \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}] = n - rk\mathbf{A} = d$ are independent). Their corresponding norms are positive semi-definite, namely these are given by $\| \mathbf{x}_m \|^2_{G_x} = \mathbf{y}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y} = \mathbf{x}'\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax} = \mathbf{x}'\mathbf{G}_m\mathbf{x}$ and $\| \mathbf{i}_m \|^2_{\mathbf{G}_x} = \mathbf{x}'[\mathbf{G}_x - \mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}]\mathbf{x}$. In particular, note that

$$\mathbf{x}_m = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax},\ \mathbf{i}_m = [\mathbf{I}_m - \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax}. \tag{1.59}$$

Case (iii) ($\mathbf{G}_x$ positive-definite):

$$\mathbf{L} = (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}')[\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}, \tag{1.60}$$

$$\mathbf{x}_m = (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}')[\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{y}, \tag{1.61}$$

$$\mathbf{x}_m = \mathbf{x}_m + \mathbf{i}_m. \tag{1.62}$$

$\mathbf{x} = \mathbf{x}_m + \mathbf{i}_m$ is an orthogonal decomposition of the unknown vector $\mathbf{x} \in \mathbb{X} \equiv \mathbb{R}^m$ into the $\mathbf{G}_x + \mathbf{AA}'$-MINOS vector $\mathbf{x}_m \in \mathbb{L}^n$ and the $\mathbf{G}_x + \mathbf{AA}'$-MINOS vector of inconsistency $\mathbf{i}_m \in \mathbb{L}^d$ subject to

$$\mathbf{x}_m = (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}')[\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{Ax}, \tag{1.63}$$

$$\mathbf{i}_m = (\mathbf{I}_m - (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A})\mathbf{x}. \tag{1.64}$$

Due to $\mathbf{x}_m = (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{Ax}$, $\mathbf{G}_x + \mathbf{AA}'$-MINOS has the reproducing property. As projection matrices, the two matrices (1.65) and (1.66)

$$\begin{gathered}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A} \\ rk[\mathbf{G}_x + \mathbf{AA}']^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A} = rk\mathbf{A} = n\end{gathered} \tag{1.65}$$

$$\begin{aligned} &(\mathbf{I}_m - (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A}) \\ rk&[\mathbf{I}_m - (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A} = n - rk\mathbf{A} = d. \end{aligned} \tag{1.66}$$

Their corresponding norms are positive semi-definite, namely the corresponding norms read

$$\begin{aligned} \| \mathbf{x}_m \|^2_{\mathbf{G}_x + \mathbf{AA}'} &= \mathbf{y}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{y} \\ &= \mathbf{x}'\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{Ax} = \mathbf{x}'\mathbf{G}_m\mathbf{x}, \end{aligned} \tag{1.67}$$

$$\| \mathbf{i}_m \|^2_{\mathbf{G}_x + \mathbf{AA}'} = \mathbf{x}'(\mathbf{I}_m - (\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A}']^{-1}\mathbf{A})\mathbf{x} \tag{1.68}$$

Proof. A basis of the proof could be C.R. Rao's Pandora Box, the theory of inverse partitioned matrices. (Appendix A. Fact: Inverse Partitioned Matrix (IPM) of a symmetric matrix). Due to the rank identity rk$\mathbf{A} = rk\mathbf{AG}_x^{-1}\mathbf{A}' = n < m$, the normal equations of the case (i) and the case (ii) ($\longrightarrow$(1.69)) can be directly solved by Gauss elimination. For the case (iii), we add to the first normal equation the term $\mathbf{AA}'\mathbf{x}_m = \mathbf{A}'\mathbf{y}$. Due to the rank identity rk$\mathbf{A}$ = rk$\mathbf{A}'(\mathbf{G}_x + \mathbf{AA}')^{-1}\mathbf{A} = n < m$, the modified normal equations of the case (i) and the case (ii) ($\longrightarrow$(1.70)) can be directly solved by Gauss elimination.

$$\begin{bmatrix} \mathbf{G}_x & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_m \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix}, \tag{1.69}$$

$$\begin{bmatrix} \mathbf{G}_x + \mathbf{A}'\mathbf{A} & \mathbf{A}' \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_m \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{A}' \\ \mathbf{I}_n \end{bmatrix} \mathbf{y}. \tag{1.70}$$

Multiply the first equation of (1.69) by $\mathbf{AG}_x^{-1}$ and subtract the second equation of (1.69) from the modified first one to obtain (1.71). The forward reduction step is followed by the backward reduction step. Implement λ_m into the first equation of (1.71) and solve for $\mathbf{x}_m$ to obtain (1.72). Thus, $\mathbf{G}_x$-MINOS $\mathbf{x}_m$ and λ_m are represented by (1.73).

$$\begin{aligned} \mathbf{Ax}_m + \mathbf{AG}_x^{-1}\mathbf{A}'\lambda_m &= \mathbf{0}, \\ \mathbf{Ax}_m &= \mathbf{y}, \\ \lambda_m &= -(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y}, \end{aligned} \tag{1.71}$$

$$\begin{aligned} \mathbf{G}_x\mathbf{x}_m - (\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y} &= \mathbf{0}, \\ \mathbf{x}_m &= \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y} \end{aligned} \tag{1.72}$$

$$\begin{aligned} \mathbf{x}_m &= \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y}, \\ \lambda_m &= -(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{y} \end{aligned} \tag{1.73}$$

Multiply the first equation of (1.70) by $\mathbf{AG}_x^{-1}$ and subtract the second equation of (1.70) from the modified first one to obtain (1.74). The forward reduction step is followed by the backward reduction step. Implement λ_m into the first equation of (1.74) and solve for $\mathbf{x}_m$ to obtain (1.75). Thus, $\mathbf{G}_x$-MINOS $\mathbf{x}_m$ and λ_m are represented by (1.76).

$$\begin{aligned} \mathbf{A}\mathbf{x}_m + \mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\lambda_m &= \mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}, \\ \mathbf{A}\mathbf{x}_m &= \mathbf{y} \end{aligned} \tag{1.74}$$

$$\begin{aligned} \lambda_m &= [\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' - \mathbf{I}_n]\mathbf{y} \\ &= -([\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1} - \mathbf{I}_n)\mathbf{y}. \\ (\mathbf{G}_x + \mathbf{A}'\mathbf{A})\mathbf{x}_m &- \mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}\mathbf{y} = \mathbf{0}, \\ \mathbf{x}_m &= (\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}\mathbf{y}, \end{aligned} \tag{1.75}$$

$$\begin{aligned} \mathbf{x}_m &= (\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}\mathbf{y}, \\ \lambda_m &= -([\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1} - \mathbf{I}_n)\mathbf{y}. \end{aligned} \tag{1.76}$$

1-21 A Discussion of the Metric of the Parameter Space $\mathbb{X}$

With the completion of the proof, we have to discuss in more detail the basic results of Lemma 1.3. At first, we have to observe that the matrix $\mathbf{G}_x$ of the metric of the parameter space $\mathbb{X}$ has to be given a priori. We classified MINOS according to (i) $\mathbf{G}_x = \mathbf{I}_m$, (ii) $\mathbf{G}_x$ positive-definite, and (iii) $\mathbf{G}_x$ positive semi-definite.

But how do we know the metric of the parameters space? Obviously, we need prior information about the geometry of the parameter space $\mathbb{X}$, namely from the empirical sciences like physics, chemistry, biology, geosciences, social sciences.

If the parameter space $\mathbb{X} \in \mathbb{R}^m$ is equipped with an inner product $< \mathbf{x}_1|\mathbf{x}_2 >$ of the type $< \mathbf{x}_1|\mathbf{x}_2 >= \mathbf{x}_1'\mathbf{G}_x\mathbf{x}_2$, with $x_1 \in \mathbb{X}$ and $x_2 \in \mathbb{X}$, where the matrix $\mathbf{G}_x$ of the $\|\mathbf{x}\|^2$ metric $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{G}_x\mathbf{x}$ is positive-definite, we refer to the metric space $\mathbb{X} \in \mathbb{R}^m$ as *Euclidean* $\mathbb{E}^m$. In contrast, if the parameter space $\mathbb{X} \in \mathbb{R}^m$ is restricted to a metric space with a matrix $\mathbf{G}_x$ of the metric which is positive semi-definite, we call the parameter space *semi-Euclidean* $\mathbb{E}^{m_1,m_2}$. m_1 is the number of positive eigenvalues, m_2 is the number of zero eigenvalues of the positive semi-definite matrix $\mathbf{G}_x$ of the metric ($m = m_1 + m_2$). In various applications, namely in the adjustment of observations which refer to *Special Relativity* or *General Relativity*, we have to generalize the metric structure of the parameter space $\mathbb{X}$: if the matrix $\mathbf{G}_x$ of the pseudometric $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{G}_x\mathbf{x}$ is built on m_1 positive eigenvalues (signature $+$), m_2 zero eigenvalues, and m_3 negative eigenvalues (signature $-$), we call the pseudometric parameter space *pseudo Euclidean* $\mathbb{E}^{m_1,m_2,m_3}$, $m = m_1+m_2+m_3$. For such a parameter space, MINOS has to be generalized to $\|\mathbf{x}\|^2$ = extr., for instance, *maximum norm solution*.

1-22 An Alternative Choice of the Metric of the Parameter Space $\mathbb{X}$

Another problem that is associated with the parameter space $\mathbb{X}$ is the so-called *norm choice problem*. Here, we have used the l_2-*norm*, for instance, the $l_2 vnorm$ (1.77). An alternative norm of choice is the l_p-*norm* (1.78) with $1 \leq p < \infty$. Beside the choice of the matrix $\mathbf{G}_x$ of the metric within the l_2-*norm* and of the l_p-*norm* itself, we like to discuss the result of the MINOS matrix $\mathbf{G}_m$ of the metric. Indeed, we have constructed MINOS from an a priori *choice* of the metric $\mathbf{G}$, called $\mathbf{G}_x$, and were led to the a posteriori *choice* of the metric $\mathbf{G}_m$ of the types (1.79). The matrices (1.79) are (i) *idempotent*, (ii) $\mathbf{G}_x$ *idempotent*, and (iii) $[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}$ *idempotent*, namely projection matrices. Similarly, the norms $\|\mathbf{i}_m\|^2$ of the types (1.80) measure the distance of the solution point $\mathbf{x}_m \in \mathbb{X}$ from the null space $\mathcal{N}(\mathbf{A})$. The matrices (1.80) are (i) *idempotent*, (ii) $\mathbf{G}_x^{-1}$ *idempotent*, and (iii) $(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}$ *idempotent*, namely projection matrices.

$$\begin{gathered} \|\mathbf{x}\|_2 := \\ := \sqrt{\mathbf{x}'\mathbf{x}} = \\ \sqrt{x_1^2 + x_2^2 + \ldots x_{m-1}^2 + x_m^2}, \end{gathered} \tag{1.77}$$

$$\begin{gathered} \|\mathbf{x}\|_p := \\ \sqrt[p]{x_1^p + x_2^p + \ldots x_{m-1}^p + x_m^p}, \end{gathered} \tag{1.78}$$

$$\begin{gathered} (i)\ \mathbf{G}_m = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}, \\ (ii)\ \mathbf{G}_m = \mathbf{A}'(\mathbf{A}\mathbf{G}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}, \\ (iii)\ \mathbf{G}_m = \mathbf{A}'[\mathbf{A}(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}\mathbf{A}, \end{gathered} \tag{1.79}$$

$$\begin{gathered} \mathbf{I}_m - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}, \\ \mathbf{G}_x - \mathbf{A}'(\mathbf{A}\mathbf{G}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}, \\ \mathbf{I}_m - (\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'[A(\mathbf{G}_x + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']^{-1}\mathbf{A} \end{gathered} \tag{1.80}$$

1-23 $\mathbf{G}_x$-MINOS and Its Generalized Inverse

By *Lemma 1.4*, a more formal version of the generalized inverse which is characteristic for $\mathbf{G}_x$-MINOS is presented.

Lemma 1.4. (Characterization of $\mathbf{G}_x$-MINOS).

$\mathbf{x}_m = \mathbf{L}\mathbf{y}$ is I-MINOS of the consistent system of linear equations (1.47) (i.e. $\mathbf{A}\mathbf{x} = \mathbf{y}$ and $rk\mathbf{A} = rk(\mathbf{A}, \mathbf{y})$ or $\mathbf{y} \in \mathcal{R}(\mathbf{A})$) if and only if the matrix $L \in \mathbb{R}^{m \times n}$ fulfils the two conditions (1.81). The matrix $\mathbf{L}$ is reflexive and is the $\mathbf{A}^{1,2,4}$ generalized inverse.

$$\mathbf{ALA} = \mathbf{A},\ \mathbf{LA} = \mathbf{LA}' \tag{1.81}$$

$\mathbf{x}_m = \mathbf{Ly}$ is $\mathbf{G}_x$-MINOS of the consistent system of linear equations (1.47) (i.e. $\mathbf{Ax} = \mathbf{y}$ and $rk\mathbf{A} = rk(\mathbf{A}, \mathbf{y})$ or $\mathbf{y} \in \mathcal{R}(\mathbf{A})$) if and only if the matrix $\mathbf{L} \in \mathbb{R}^{m\times n}$ fulfils the two conditions (1.82). The reflexive matrix $\mathbf{L}$ is the $\mathbf{G}_x$-weighted $\mathbf{A}^{1,2,4}$ generalized inverse.

$$\mathbf{ALA} = \mathbf{A},\ \mathbf{G}_x\mathbf{LA} = (\mathbf{G}_x\mathbf{LA})' \tag{1.82}$$

Proof.

According to the theory of the general solution of a system of linear equations which is presented in Appendix A, the conditions $\mathbf{ALA} = \mathbf{L}$ or $\mathbf{L} = \mathbf{A}^-$ guarantee the solution $\mathbf{x} = \mathbf{Ly}$ of (1.47) ($rk\mathbf{A} = rk(\mathbf{A}, \mathbf{y})$). The general solution (1.83) with an arbitrary vector $\mathbf{x} \in \mathbb{R}^{m\times 1}$ leads to the appropriate representation of the $\mathbf{G}_x$-seminorm by means of (1.84), where the arbitrary vector $\mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n$ holds, if and only if $\mathbf{y}'\mathbf{L}'\mathbf{G}_x(\mathbf{I} - \mathbf{LA})\mathbf{z} = \mathbf{0}$ for all $\mathbf{z} \in \mathbb{R}^m \times 1$ or $\mathbf{A}'\mathbf{L}'\mathbf{G}_x(\mathbf{I} - \mathbf{LA}) = \mathbf{0}$ or $\mathbf{A}'\mathbf{L}'\mathbf{G}_\mathbf{x} = \mathbf{A}'\mathbf{L}'\mathbf{G}_x\mathbf{LA}$. The right hand side is a symmetric matrix. Accordingly, the left hand side must have this property, too, namely $\mathbf{G}_x\mathbf{LA} = (\mathbf{G}_x\mathbf{LA})'$, what had to be shown.

$$\mathbf{x} = \mathbf{x}_m + (\mathbf{I} - \mathbf{LA})\mathbf{z} \tag{1.83}$$

$$\begin{aligned}
\|\mathbf{x}_m\|^2_{\mathbf{G}_x} &= \|\mathbf{Ly}\|^2_{\mathbf{G}_x} \le \|\mathbf{x}\|^2_{\mathbf{G}_x} = \|\mathbf{x}_m + (\mathbf{I} - \mathbf{LA})\mathbf{z}\|^2_{\mathbf{G}_x} \\
&= \|\mathbf{x}_m\|^2_{\mathbf{G}_x} + 2\mathbf{x}'{}_m\mathbf{G}_x(\mathbf{I} - \mathbf{LA})\mathbf{z} + \|(\mathbf{I} - \mathbf{LA})\mathbf{z}\|^2_{\mathbf{G}_x} \\
&= \|\mathbf{Ly}\|^2_{\mathbf{G}_x} + 2\mathbf{y}'\mathbf{L}'\mathbf{G}_x(\mathbf{I} - \mathbf{LA})\mathbf{z} + \|(\mathbf{I} - \mathbf{LA})\mathbf{z}\|^2_{\mathbf{G}_x} \\
&= \mathbf{y}'\mathbf{L}'\mathbf{G}_x\mathbf{Ly} + 2\mathbf{y}'\mathbf{L}'\mathbf{G}_x(\mathbf{I} - \mathbf{LA})\mathbf{z} + \mathbf{z}'(\mathbf{I} - \mathbf{A}'\mathbf{L}')\mathbf{G}_x(\mathbf{I} - \mathbf{LA})\mathbf{z}
\end{aligned} \tag{1.84}$$

Reflexivity of the matrix $\mathbf{L}$ originates from the consistency condition, namely $(\mathbf{I} - \mathbf{AL})\mathbf{y} = \mathbf{0}$ for all $\mathbf{y} \in \mathbb{R}^{m\times 1}$ or $\mathbf{AL} = \mathbf{I}$. The reflexive condition of the $\mathbf{G}_x$ weighted,minimum norm generalized inverse $\mathbf{i}_m[\mathbf{I}_m - \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{A}]\mathbf{x}$, $\mathbf{G}_x\mathbf{LAL} = \mathbf{G}_x\mathbf{L}$, is a direct consequence. *Consistency* of the normal equations (1.48) or, equivalently, the *uniqueness* of $\mathbf{G}_x\mathbf{x}_m$ follows from (1.85) for arbitrary matrices $\mathbf{L}_1 \in \mathbb{R}^{m\times n}$ and $\mathbf{L}_2 \in \mathbb{R}^{m\times n}$ which satisfy $\mathbf{x}_m = \mathbf{G}_x^{-1}\mathbf{A}'(\mathbf{AG}_x^{-1}\mathbf{A}')^{-1}\mathbf{Ax}$.

$$\begin{aligned}
\mathbf{G}_x\mathbf{L}_1\mathbf{y} &= \mathbf{A}'\mathbf{L}'{}_1\mathbf{G}_x\mathbf{L}_1\mathbf{y} = \mathbf{G}_x\mathbf{L}_1\mathbf{AL}_1\mathbf{y} = \mathbf{G}_x\mathbf{L}_1\mathbf{AL}_2\mathbf{y} = \mathbf{A}'\mathbf{L}'{}_1\mathbf{AL}'2\mathbf{G}_x\mathbf{L}_2\mathbf{y} \\
&= \mathbf{A}'\mathbf{L}_2\mathbf{G}_x\mathbf{L}_2\mathbf{y} = \mathbf{G}_x\mathbf{L}_2\mathbf{y}
\end{aligned} \tag{1.85}$$

1-24 *Eigenvalue Decomposition of* $\mathbf{G}_x$*-MINOS: Canonical MINOS*

In the empirical sciences, we meet quite often the inverse problem to determine the infinite set of coefficients of a series expansion of a function or a functional (*Taylor polynomials*) from a finite set of observations.

Mind the following examples.

(i) Determine the *Fourier coefficients* (discrete Fourier transform, trigonometric polynomials) of a harmonic function with circular support from observations in a one-dimensional lattice. (ii) Determine the *spherical harmonic coefficients* (discrete Fourier–Legendre transform) of a harmonic function with spherical support from observations in a two-dimensional lattice.

End of examples.

Typically, such expansions generate an infinite dimensional linear model based upon orthogonal (orthonormal) functions. Naturally, such a linear model is under-determined since a finite set of observations is confronted with an infinite set of unknown parameters. In order to make such an infinite dimensional linear model accessible to the computer, the expansion into orthogonal (orthonormal) functions is truncated or band-limited. Note that the observables $\mathbf{y} \in \mathbb{Y}$ and $\dim \mathbb{Y} = n$ are related to parameters $\mathbf{x} \in \mathbb{X}$ and $dim\mathbb{X} = m \gg n = dim\mathbb{Y}$ (namely the unknown coefficients) by a linear operator $\mathbf{A} \in \mathbb{R}^{m \times n}$, which is given in the form of an *eigenvalue decomposition*. Furthermore, note that we are confronted with the problem to construct *canonical* MINOS, also called *eigenvalue decomposition* of $\mathbf{G}_x$-MINOS.

First, we intend to canonically represent the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$. Here, we shall assume that both spaces are Euclidean, equipped with a symmetric, positive-definite matrix of the metric $\mathbf{G}_x$ and $\mathbf{G}_y$, respectively. Enjoy the *diagonalization procedure* of both matrices that is reviewed in Box 1.8. The inner products $\mathbf{aa}'$ and $\mathbf{bb}'$, respectively, constitute the matrix of the metric $\mathbf{G}_x$ and $\mathbf{G}_y$, respectively. The base vectors $\{\mathbf{a}_1, \ldots, \mathbf{a}_m \mid 0\}$ span the parameter space $\mathbb{X}$, $\dim \mathbb{X} = m$, and the base vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_m \mid 0\}$ span the observation space $\mathbb{Y}$, $dim\mathbb{Y} = n$. Note the rank identities $\text{rk}\mathbf{G}_\mathbf{x} = m$ and $\text{rk}\mathbf{G}_y = n$, respectively. The left norm $\|\mathbf{x}\|^2_{\mathbf{G}_x} = \mathbf{x}'\mathbf{G}_x\mathbf{x}$ is taken with respect to the left matrix of the metric $\mathbf{G}_x$. In contrast, the right norm $\|\mathbf{y}\|^2_{\mathbf{G}_y} = \mathbf{y}'\mathbf{G}_y\mathbf{y}$ refers to the right matrix of the metric $\mathbf{G}_y$. In order to diagonalize the left quadratic form as well as the right quadratic form, we transform $\mathbf{G}_x \longmapsto \mathbf{G}^*_x = diag[\lambda^x_1, \ldots, \lambda^x_m] = \mathbf{V}'\mathbf{G}_x\mathbf{V}$ as well as $\mathbf{G}_y \longmapsto \mathbf{G}^*_y = diag[\lambda^y_1, \ldots, \lambda^y_n] = \mathbf{U}'\mathbf{G}_y\mathbf{U}$ into the canonical form by means of the left orthonormal matrix $\mathbf{V}$ and by means of the right orthonormal matrix $\mathbf{U}$. Such a procedure is called *eigenspace analysis* of the matrix $\mathbf{G}_x$ as well as *eigenspace analysis* of the matrix $\mathbf{G}_y$. $\boldsymbol{\Lambda}_x$ constitutes the diagonal matrix of the *left positive eigenvalues* $(\lambda^x_1, \ldots, \lambda^x_m)$, the *right positive eigenvalues* $(\lambda^y_1, \ldots, \lambda^y_n)$ the n-dimensional right spectrum. The inverse transformation $\mathbf{G}^*\mathbf{x} = \boldsymbol{\Lambda}_x \longmapsto \mathbf{G}_x$

as well as $\mathbf{G}_y^* = \mathbf{\Lambda}_y \longmapsto \mathbf{G}_y$ is denoted as left eigenspace synthesis as well as right eigenspace synthesis.

Box 1.8. (The canonical representation of the matrix of the metric, parameter space versus observation space).

$$\begin{array}{cc} \text{Parameter Space :} & \text{Observation Space:} \\ \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_m] = \mathbb{X} & \text{span}[\mathbf{b}_1, \ldots, \mathbf{b}_n] = \mathbb{Y} \\ <\mathbf{a}_{j_1}|\mathbf{a}_{j_2}>= g_{j_1 j_2} & <\mathbf{b}_{k_1}|\mathbf{b}_{k_2}>= g_{k_1 k_2} \\ \forall j_1 j_2 \in \{1, \ldots, m\} & \forall k_1 k_2 \in \{1, \ldots, n\} \end{array} \tag{1.86}$$

$$\begin{array}{cc} \mathbf{aa}' = \mathbf{G}_x, & \mathbf{bb}' = \mathbf{G}_y, \\ rk\mathbf{G}_x = m. & rk\mathbf{G}_y = n. \end{array}$$

$$\begin{array}{cc} \text{Left norms :} & \text{Right norms :} \\ \|\mathbf{x}\|^2_{\mathbf{G}_x} = \mathbf{x}'\mathbf{G}_x\mathbf{x} = \mathbf{x}^{*\prime}\mathbf{x}^* & \|\mathbf{y}\|^2_{\mathbf{G}_y} = \mathbf{y}'\mathbf{G}_y\mathbf{y} = \mathbf{y}^{*\prime}\mathbf{y}^*. \end{array} \tag{1.87}$$

$$\begin{array}{cc} \text{Eigenspace analysis}(\mathbf{G}_x) : & \text{Eigenspace analysis}(\mathbf{G}_y) : \\ \mathbf{G}_x^* = \mathbf{V}'\mathbf{G}_x\mathbf{V} = & \mathbf{G}_y^* = \mathbf{U}'\mathbf{G}_y\mathbf{U} = \\ = \text{diag}[\lambda_1^x, \ldots, \lambda_m^x] = \mathbf{\Lambda}_x, & = \text{diag}[\lambda_1^y, \ldots, \lambda_n^y] = \mathbf{\Lambda}_y, \\ \text{subject to.} & \text{subject to.} \end{array} \tag{1.88}$$

$$\begin{array}{cc} \mathbf{VV}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_m, & \mathbf{UU}' = \mathbf{U}'\mathbf{U} = \mathbf{I}_n, \\ (\mathbf{G}_x - \lambda_j^x\mathbf{I}_m)\mathbf{v}_j = \mathbf{0}. & (\mathbf{G}_y - \lambda_k^y\mathbf{I}_n)\mathbf{u}_j = \mathbf{0}. \end{array} \tag{1.89}$$

$$\begin{array}{cc} \text{Eigenspace synthesis}(\mathbf{G}_x) : & \text{Eigenspace synthesis}(\mathbf{G}_y) : \\ \mathbf{G}_x = \mathbf{VG}_x^*\mathbf{V}' = \mathbf{V\Lambda}_x\mathbf{V}'. & \mathbf{G}_y = \mathbf{UG}_y^*\mathbf{U}' = \mathbf{U\Lambda}_y\mathbf{U}'. \end{array} \tag{1.90}$$

Second, let us here study the impact of the left diagonalization of the metric $\mathbf{G}_x$ as well as right diagonalization of the matrix of the metric $\mathbf{G}_y$ on the coordinates $\mathbf{x} \in \mathbb{X}$ and $\mathbf{y} \in \mathbb{Y}$, the parameter systems of the left Euclidean space $\mathbb{X}$, dim $\mathbb{X} = m$, and of the right Euclidean space $\mathbb{Y}$. Enjoy the way how we have established in Box 1.9 the canonical coordinates $\mathbf{x}^* := [\mathbf{x}_1^*, \ldots, \mathbf{x}_m^*]'$ of $\mathbb{X}$ as well as the canonical coordinates $\mathbf{y}^* := [\mathbf{y}_1^*, \ldots, \mathbf{y}_n^*]$, which are called the left and right star coordinates of $\mathbb{X}$ and $\mathbb{Y}$, respectively. In terms of the star coordinates (1.92), the left norm $\|\mathbf{x}^*\|^2$ of the type (1.90, left) as well as the right norm $\|\mathbf{y}^*\|^2$ of type (1.90, right) take the canonical left and right quadratic form. The transformations $\mathbf{x} \longmapsto \mathbf{x}^*$ as well as $\mathbf{y} \longmapsto \mathbf{y}^*$ of type (1.92, left) and (1.92, right) are special versions of the left and right polar decomposition: a rotation constituted by the matrices $\mathbf{U}, \mathbf{V}$ is followed by a stretch constituted by the matrices $\{\mathbf{\Lambda}_x^{1/2}, \mathbf{\Lambda}_y^{1/2}\}$ as diagonal matrices. The forward transformations $\mathbf{x} \longmapsto \mathbf{x}^*$ and $\mathbf{y} \longmapsto \mathbf{y}^*$ are computed by the backward transformations $\mathbf{x}^* \longmapsto \mathbf{x}$ and $\mathbf{y}^* \longmapsto \mathbf{y}$. $\{\mathbf{\Lambda}_x^{1/2}$ and $\mathbf{\Lambda}_y^{1/2}\}$, respectively, denote those diagonal matrices which are generated by the positive roots of the left and right eigenvalues, respectively. Equation (1.94) defines corresponding direct and inverse matrix identities. We conclude with the proof that the ansatz (1.92) indeed leads to the canonical representation (1.91) of the left and right norms.

Box 1.9. (The canonical coordinates $\mathbf{x}^* \in \mathbb{X}$ and $\mathbf{y}^* \in \mathbb{Y}$, parameter space versus observation space).

$$\begin{array}{cc} \text{Canonical coordinates} & \text{Canonical coordinates} \\ \text{(parameter space) :} & \text{(observation space) :} \\ \|\mathbf{x}^*\|^2 = \mathbf{x}^{*\prime}\mathbf{x}^* = & \|\mathbf{y}^*\|^2 = \mathbf{y}^{*\prime}\mathbf{y}^* = \end{array} \tag{1.91}$$

$$\begin{array}{cc} \text{Ansatz :} & \text{Ansatz) :} \\ \mathbf{x}^* = \mathbf{V}'\boldsymbol{\Lambda}_x^{1/2}\mathbf{x} & \mathbf{y}^* = \mathbf{U}'\boldsymbol{\Lambda}_y^{1/2}\mathbf{y} \end{array} \tag{1.92}$$

$$\begin{array}{cc} \text{versus} & \text{versus)} \\ \mathbf{x} = \boldsymbol{\Lambda}_x^{-1/2}\mathbf{V}\mathbf{x}^* & \mathbf{y} = \boldsymbol{\Lambda}_y^{-1/2}\mathbf{U}\mathbf{y}^* \end{array} \tag{1.93}$$

$$\begin{array}{cc} \boldsymbol{\Lambda}_x^{+1/2} := \operatorname{diag}\left[\sqrt{\lambda_1^x}, \ldots, \sqrt{\lambda_m^x}\right], & \boldsymbol{\Lambda}_y^{+1/2} := \operatorname{diag}\left[\sqrt{\lambda_1^y}, \ldots, \sqrt{\lambda_m^y}\right], \\ \boldsymbol{\Lambda}_x^{-1/2} =:= \operatorname{diag}\left[\frac{1}{\sqrt{\lambda_1^x}}, \ldots, \frac{1}{\sqrt{\lambda_m^x}}\right]. & \boldsymbol{\Lambda}_y^{-1/2} = \left[\frac{1}{\sqrt{\lambda_1^y}}, \ldots, \frac{1}{\sqrt{\lambda_m^y}}\right]. \end{array} \tag{1.94}$$

$$\begin{array}{cc} \text{Ansatz proof :} & \text{Ansatz proof :} \\ \mathbf{G}_x = \mathbf{V}\boldsymbol{\Lambda}_x\mathbf{V}', & \mathbf{G}_y = \mathbf{U}\boldsymbol{\Lambda}_y\mathbf{U}', \\ \|\mathbf{x}\|^2_{\mathbf{G}_x} = & \|\mathbf{y}\|^2_{\mathbf{G}_y} = \\ = \mathbf{x}'\mathbf{G}_x\mathbf{x} = & = \mathbf{y}'\mathbf{G}_y\mathbf{y} = \\ = \mathbf{x}'\mathbf{V}\boldsymbol{\Lambda}_x^{1/2}\boldsymbol{\Lambda}_x^{1/2}\mathbf{V}'\mathbf{x} = & = \mathbf{y}'\mathbf{U}\boldsymbol{\Lambda}_y^{1/2}\boldsymbol{\Lambda}_y^{1/2}\mathbf{U}'\mathbf{y} = \\ \mathbf{x}^{*\prime}\boldsymbol{\Lambda}_x^{-1/2}\mathbf{V}'\mathbf{V}\boldsymbol{\Lambda}_x\boldsymbol{\Lambda}_x^{1/2}\times & \mathbf{y}^{*\prime}\boldsymbol{\Lambda}_y^{-1/2}\mathbf{U}'\mathbf{U}\boldsymbol{\Lambda}_y\boldsymbol{\Lambda}_y^{1/2}\times \\ \times\mathbf{V}'\mathbf{V}\boldsymbol{\Lambda}_x^{-1/2}\mathbf{x}^* = & \times\mathbf{U}'\mathbf{U}\boldsymbol{\Lambda}_y^{-1/2}\mathbf{y}^* = \\ = \mathbf{x}^{*\prime}\mathbf{x}^* = & = \mathbf{y}^{*\prime}\mathbf{y}^* = \\ = \|\mathbf{x}^*\|^2. & = \|\mathbf{y}^*\|^2. \end{array} \tag{1.95}$$

Third, let us here study the dual operations of coordinate transformations $\mathbf{x} \longmapsto \mathbf{x}^*$ and $\mathbf{y} \longmapsto \mathbf{y}^*$, namely the behavior of canonical bases, which are also called orthonormal bases $\mathbf{e}_x$ and $\mathbf{e}_y$, or Cartan frames of reference $\{\mathbf{e}_1^x, \ldots, \mathbf{e}_m^x|\mathbf{0}\}$ spanning the parameter space $\mathbb{X}$ and $\{\mathbf{e}_1^y, \ldots, \mathbf{e}_n^y|\mathbf{0}\}$ spanning the observation space $\mathbb{Y}$, here $\mathbf{a} \longmapsto \mathbf{e}_x$ and $\mathbf{b} \longmapsto \mathbf{e}_y$. In terms of orthonormal bases $\mathbf{e}_x$ and $\mathbf{e}_y$, as outlined in Box 1.10, the matrix of the metric $\mathbf{e}_x\mathbf{e}'_x = \mathbf{I}_m$ and $\mathbf{e}'_y\mathbf{e}_y = \mathbf{I}_n$ takes the canonical form ("modular"). Compare the relations (1.97, left) and (1.99, left). Equations (1.97, right) and (1.99, right) are achieved by the changes of bases ("CBS") of type left $\mathbf{e}_x \longmapsto \mathbf{a}$, $\mathbf{a} \longmapsto \mathbf{e}_x$, compare with (1.100, left) and (1.101, left), and of type right $\mathbf{e}_y \longmapsto \mathbf{b}$, $\mathbf{b} \longmapsto \mathbf{e}_y$, compare with (1.100, right) and (1.101, right). Indeed, the transformations $\mathbf{x} \longmapsto \mathbf{x}^*$, $\mathbf{a} \longmapsto \mathbf{e}_x$, and $\mathbf{y} \longmapsto \mathbf{y}^*$, $\mathbf{b} \longmapsto \mathbf{e}_y$ are dual or inverse. Fourth, let us begin the eigenspace analysis versus eigenspace synthesis of the rectangular matrix, $\mathbf{A} \in \mathbb{R}^{n\times m}$, $r := rk\mathbf{A} = n, n < m$. Indeed, the eigenspace of the rectangular matrix looks differently when compared to the eigenspace of the quadratic, symmetric, positive-definite matrix $\mathbf{G}_x \in \mathbb{R}^{m\times m}$, $rk\mathbf{G}_x = m$ and $\mathbf{G}_y \in \mathbb{R}^{n\times n}$, $rk\mathbf{G}_y = n$ of the left and the right metric. At first, we have to generalize the transpose of a rectangular matrix by introducing the adjoint operator $\mathbf{A}^{\#}$, which takes into account the matrices $\{\mathbf{G}_x, \mathbf{G}_y\}$ of the left and the right metric. The definition of the adjoint operator $\mathbf{A}^{\#}$, compare with Definition 1.5, is followed by its representation, compare with Lemma 1.6.

Box 1.10. (The general bases versus the orthonormal bases spanning the parameter space as well as the observation space).

$$\begin{array}{cc} \text{Left} & \text{Right} \\ \text{(parameter space}\mathbb{X}, & \text{(observation space}\mathbb{Y}, \\ \text{general left base) :} & \text{general right base) :} \\ \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_m] = \mathbb{X}. & \text{span}[\mathbf{b}_1, \ldots, \mathbf{b}_n] = \mathbb{Y}. \end{array} \tag{1.96}$$

$$\begin{array}{cc} \text{Matrix of the metric} & \text{Matrix of the metric} \\ \mathbf{a}\mathbf{a}' = \mathbf{G}_x. & \mathbf{b}\mathbf{b}' = \mathbf{G}_y. \end{array} \tag{1.97}$$

$$\begin{array}{cc} \text{Orthonormal left base} & \text{Orthonormal right base} \\ \text{span}[\mathbf{e}_1^x, \ldots, \mathbf{e}_m^x] = \mathbb{X}. & \text{span}[\mathbf{e}_1^y, \ldots, \mathbf{e}_n^y] = \mathbb{Y}. \end{array} \tag{1.98}$$

$$\begin{array}{cc} \text{Matrix of the metric} & \text{Matrix of the metric} \\ \mathbf{e}_x \mathbf{e}'_x = \mathbf{I}_m. & \mathbf{e}'_y \mathbf{e}_y = \mathbf{I}_n. \end{array} \tag{1.99}$$

$$\begin{array}{cc} \text{Base transformation} & \text{Base transformation} \\ \mathbf{a} = \boldsymbol{\Lambda}_x^{1/2} \mathbf{V} \mathbf{e}_x. & \mathbf{b} = \boldsymbol{\Lambda}_y^{1/2} \mathbf{U} \mathbf{e}_y. \end{array} \tag{1.100}$$

$$\begin{array}{cc} \text{versus} & \text{versus} \\ \mathbf{e}_x = \mathbf{V}' \boldsymbol{\Lambda}_x^{-1/2} \mathbf{a}, & \mathbf{e}_y = \mathbf{U}' \boldsymbol{\Lambda}_y^{-1/2} \mathbf{b}, \\ \text{span}[\mathbf{e}_1^x, \ldots, \mathbf{e}_m^x] = \mathbb{X}. & \text{span}[\mathbf{e}_1^y, \ldots, \mathbf{e}_n^y] = \mathbb{Y}. \end{array} \tag{1.101}$$

Definition 1.5. (Adjoint operators $\mathbf{A}^{\#}$).

The adjoint operator $\mathbf{A}^{\#} \in \mathbb{R}^{m\times n}$ of the matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is defined by the inner product identity

$$< \mathbf{y}|\mathbf{A}\mathbf{x} >_{G_y} = < \mathbf{x}|\mathbf{A}^{\#}\mathbf{y} >_{G_x}, \tag{1.102}$$

where the left inner product operates on the symmetric, full rank matrix $\mathbf{G}_y$ of the observation space $\mathbb{Y}$, while the right inner product is taken with respect to the symmetric full rank matrix $\mathbf{G}_x$ of the parameter space $\mathbb{X}$.

Lemma 1.6. (Adjoint operators $\mathbf{A}^{\#}$).

A representation of the adjoint operator $\mathbf{A}^{\#} \in \mathbb{R}^{m \times n}$ of the matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ is provided by

$$\mathbf{A}^{\#} = \mathbf{G}_x^{-1}\mathbf{A}'\mathbf{G}_y. \tag{1.103}$$

Proof.

For the proof, we take advantage of the symmetry of the left inner product, namely

$$< \mathbf{y}|\mathbf{Ax} >_{G_y} = \mathbf{y}'\mathbf{G}_y\mathbf{Ax} \quad \text{versus} \quad < \mathbf{x}|\mathbf{A}^{\#}\mathbf{y} >_{G_x} = \mathbf{x}'\mathbf{G}_x\mathbf{A}^{\#}\mathbf{y}, \tag{1.104}$$

$$\begin{gathered} \mathbf{y}'\mathbf{G}_y\mathbf{Ax} = \mathbf{x}'\mathbf{A}'\mathbf{G}_y\mathbf{y} = \mathbf{x}'\mathbf{G}_x\mathbf{A}^{\#}\mathbf{y} \\ \Longleftrightarrow \\ \mathbf{A}'\mathbf{G}_y = \mathbf{G}_x\mathbf{A}^{\#} \\ \Longleftrightarrow \\ \mathbf{G}_x^{-1}\mathbf{A}'\mathbf{G}_y = \mathbf{A}^{\#} \end{gathered} \tag{1.105}$$

Fifth, we additionally solve the underdetermined system of linear equations $\{\mathbf{Ax} = \mathbf{y}|\mathbf{A} \in \mathbb{R}^{n \times m}, rk\mathbf{A} = n, n < m\}$ by introducing the eigenspace of the rectangular matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ of rank $r := rk\mathbf{A} = n, n < m(\mathbf{A} \longmapsto \mathbf{A}^*)$ and the left and right canonical coordinates $\mathbf{x} \longmapsto \mathbf{x}^*$ and $\mathbf{y} \longmapsto \mathbf{y}^*$. Compare with the relations that are supported by *Box 1.11*.

Box 1.11. (Canonical representation, underdetermined system of linear equations).

$$\begin{array}{cc} \text{Left} & \text{Right} \\ (\text{parameter space } \mathbb{X}: & (\text{observation space } \mathbb{Y}: \\ \mathbf{x}^* = \mathbf{V}'\mathbf{G}_x^{1/2}\mathbf{x}, & \mathbf{y}^* = \mathbf{U}'\mathbf{G}_y^{1/2}\mathbf{y}, \\ \mathbf{x} = \mathbf{G}_x^{-1/2}\mathbf{V}\mathbf{x}^*. & \mathbf{y} = \mathbf{G}_y^{-1/2}\mathbf{U}\mathbf{y}^*. \end{array} \tag{1.106}$$

$$\boxed{\begin{array}{c} \textit{Underdetermined system of linear equations} \\ (\{\mathbf{Ax} = \mathbf{y}|\mathbf{A} \in \mathbb{R}^{n \times m}, rk\mathbf{A} = n, n < m\}): \end{array}}$$

$$\begin{array}{cc} \mathbf{y} = \mathbf{Ax}, & \mathbf{y}^* = \mathbf{A}^*\mathbf{x}^*, \\ \mathbf{G}_y^{-1/2}\mathbf{U}\mathbf{y}^* = \mathbf{A}\mathbf{G}_x^{-1/2}\mathbf{V}\mathbf{x}^*, & \mathbf{U}'\mathbf{G}_y^{1/2}\mathbf{y} = \mathbf{A}^*\mathbf{V}'\mathbf{G}_x^{1/2}\mathbf{x}, \\ \mathbf{y}^* = (\mathbf{U}'\mathbf{G}_y^{1/2}\mathbf{A}\mathbf{G}_x^{-1/2}\mathbf{V})\mathbf{x}^* & \mathbf{y} = (\mathbf{G}_y^{-1/2}\mathbf{U}\mathbf{A}^*\mathbf{V}'\mathbf{G}_x^{1/2})\mathbf{x} \end{array} \tag{1.107}$$

$$\begin{array}{cc} \text{subject to} & \text{subject to} \\ \mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_n. & \mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_m. \end{array} \tag{1.108}$$

Left and right eigenspace.

$\boxed{\texttt{Left}}$
eigenspace analysis :
$\mathbf{A}^* = \mathbf{U}'\mathbf{G}_y^{1/2}\mathbf{A}\mathbf{G}_x^{-1/2}[\mathbf{V}_1, \mathbf{V}_2] = [\mathbf{V}, \mathbf{0}].$

$\boxed{\texttt{Right}}$
eigenspace synthesis :
$\mathbf{A} = \mathbf{G}_y^{-1/2}\mathbf{U}[\mathbf{V}, \mathbf{0}]\begin{bmatrix}\mathbf{V}'_1 \\ \mathbf{V}'_2\end{bmatrix}\mathbf{G}_x^{1/2}.$

(1.109)

$\boxed{\text{Dimension identities}}$

$$\mathbf{\Lambda} \in \mathbb{R}^{r\times r},\ \mathbf{0} \in \mathbb{R}^{r\times(m-r)},\ r := \mathrm{rk}\mathbf{A} = n,\ n < m$$

$$\mathbf{V_1} \in \mathbb{R}^{m\times r}, \mathbf{V_2} \in \mathbb{R}^{m\times(m-r)}, \mathbf{U} \in R^{r\times r}$$

$\boxed{\texttt{Left eigenspace:}}$

$$\mathbf{L} := \mathbf{G_y}^{-\frac{1}{2}}\mathbf{U} \Rightarrow \mathbf{L}^{-1} = \mathbf{U}'\mathbf{G_y}^{\frac{1}{2}}$$

$\boxed{\texttt{Right eigenspace:}}$

$$\mathbf{R} := \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V} \Rightarrow \mathbf{R}^{-1} = \mathbf{V}'\mathbf{G_x}^{\frac{1}{2}}$$
$$\mathbf{R_1} := \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V}_1,\ \ \mathbf{R_2} := \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V}_2 \tag{1.110}$$
$$\mathbf{R_1^-} := \mathbf{V_1'}\mathbf{G_x}^{\frac{1}{2}},\ \ \mathbf{R_2^-} := \mathbf{V_2'}\mathbf{G_x}^{\frac{1}{2}}$$

$$\mathbf{LL}' = \mathbf{G_y}^{-1} \Rightarrow (\mathbf{L}^{-1})'\mathbf{L}^{-1} = \mathbf{G_y} \qquad \mathbf{RR}' = \mathbf{G_x}^{-1} \Rightarrow (\mathbf{R}^{-1})'\mathbf{R}^{-1} = \mathbf{G_x} \tag{1.111}$$

$$\mathbf{A} = \mathbf{L}\mathbf{A}^*\mathbf{R}^{-1}\ \text{ versus }\ \mathbf{A}^* = \mathbf{L}^{-1}\mathbf{A}\mathbf{R} \tag{1.112}$$

$$\mathbf{A} = \mathbf{L}\,[\mathbf{\Lambda}, \mathbf{0}]\begin{bmatrix}\mathbf{R_1^-} \\ \mathbf{R_2^-}\end{bmatrix} \ \text{ versus }\ \begin{array}{c}\mathbf{A}^* = [\mathbf{\Lambda}, \mathbf{0}] = \\ = \mathbf{L}^{-1}\mathbf{A}\,[\mathbf{R_1}, \mathbf{R_2}]\end{array} \tag{1.113}$$

$$\mathbf{AA}^{\#}\mathbf{L} = \mathbf{L}\mathbf{\Lambda}^2\ \text{ versus }\ \left[\begin{array}{l}\mathbf{A}^{\#}\mathbf{AR_1} = \mathbf{R_1}\mathbf{\Lambda}^2 \\ \mathbf{A}^{\#}\mathbf{AR_2} = \mathbf{0}\end{array}\right. \tag{1.114}$$

$\boxed{\text{Underdetermined system of linear equations solved in canonical coordinates:}}$

$$\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* = [\mathbf{\Lambda}, \mathbf{0}]\begin{bmatrix}\mathbf{x_1^*} \\ \mathbf{x_2^*}\end{bmatrix} = \mathbf{\Lambda}\mathbf{x_1^*}, \mathbf{x_1^*} \in \mathbb{R}^{r\times 1}, \mathbf{x_2^*} \in \mathbb{R}^{(m-r)\times 1}$$
$$\Longrightarrow \tag{1.115}$$
$$\begin{bmatrix}\mathbf{x_1^*} \\ \mathbf{x_2^*}\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{y}^* \\ \mathbf{x_2^*}\end{bmatrix}$$

if $\mathbf{x}^*$ is MINOS, then $\mathbf{x_2^*} = 0 : \left(\mathbf{x_1^*}\right)_m = \mathbf{\Lambda}^{-1}\mathbf{y}^*$.

The transformations $\mathbf{x} \mapsto \mathbf{x}^*$, $\mathbf{y} \mapsto \mathbf{y}^*$ from the original coordinates $[x_1, \ldots, x_m]$), the parameters of the parameter space $\mathbb{X}$ $\mathbb{X}$, to the canonical coordinates $[x_1^*, \ldots, x_m^*]$, the left star coordinates, as well as from the original coordinates

$[y_1, \ldots, y_n]$, the parameters of the observation space $\mathbb{Y}$ $\mathbb{Y}$, to the canonical coordinates $[y_1^*, \ldots, y_n^*]$, the right star coordinates are polar decompositions: a rotation $\{\mathbf{U}, \mathbf{V}\}$ is followed by a general stretch $\{\mathbf{G_y^{1/2}}, \mathbf{G_x^{1/2}}\}$. The matrices $\mathbf{G_y^{1/2}}$ as well as $\mathbf{G_x^{1/2}}$ are product decompositions of type $\mathbf{G_y} = \mathbf{S_y S_y'}$ and $\mathbf{G_x} = \mathbf{S_x' S_x}$. If we substitute $\mathbf{S_y} = \mathbf{G_y^{1/2}}$ or $\mathbf{S_x} = \mathbf{G_x^{1/2}}$ symbolically, we are led to the methods of general stretches $\mathbf{G_y^{1/2}}$ and $\mathbf{G_x^{1/2}}$ respectively. Let us substitute the inverse transformations $\mathbf{x}^* \mapsto \mathbf{x} = \mathbf{G_x^{-1/2} V x}^*$ and $\mathbf{y}^* \mapsto \mathbf{y} = \mathbf{G_y^{-1/2} U y}^*$ into our system of linear equations $\mathbf{y} = \mathbf{Ax}$ or its dual $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^*$. Such an operation leads us to $\mathbf{y}^* = f(\mathbf{x}^*)$ as well as $\mathbf{y} = f(\mathbf{x})$. Subject to the orthonormality conditions $\mathbf{U'U} = \mathbf{I_n}$ and $\mathbf{V'V} = \mathbf{I_m}$ we have generated the matrix $\mathbf{A}^*$ of left-right eigenspace analysis (1.109, left), namely (1.116), subject to the horizontal rank partitioning of the matrix $\mathbf{V} = [\mathbf{V_1}, \mathbf{V_2}]$. Alternatively, the left-right eigenspace synthesis (1.109, right), namely (1.117), is based upon the left matrix $\mathbf{L} := \mathbf{G_y^{-1/2} U}$ and the right matrix $\mathbf{R} := \mathbf{G_x^{-1/2} V}$.

$$\mathbf{A}^* = [\mathbf{\Lambda}, \mathbf{0}]\,, \tag{1.116}$$

$$\mathbf{A} = \mathbf{G_y^{-\frac{1}{2}}} \mathbf{U}\,[\mathbf{\Lambda}, \mathbf{0}] \begin{bmatrix} \mathbf{V_1'} \\ \mathbf{V_2'} \end{bmatrix} \mathbf{G_x^{1/2}}. \tag{1.117}$$

Indeed the left matrix $\mathbf{L}$ by means of $\mathbf{LL'} = \mathbf{G_y^{-1}}$ reconstructs the inverse matrix of the metric of the observation space $\mathbb{Y}$. Similarly, the right matrix $\mathbf{R}$ by means of $\mathbf{RR'} = \mathbf{G_x^{-1}}$ generates the inverse matrix of the metric of the parameter space $\mathbb{X}$. In terms of $\mathbf{L}$ and $\mathbf{R}$ we have summarized the eigenvalue decompositions (1.112)–(1.114). Such an eigenvalue decomposition helps us to canonically invert $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^*$ by means of (1.115), namely the rank partitioning of the canonical unknown vector $\mathbf{x}^*$ into $\mathbf{x}_1^* \in \mathbb{R}^r$ and $\mathbf{x_2^*} \in \mathbb{R}^{\mathbf{m-r}}$ to determine $\mathbf{x_1^*} = \mathbf{\Lambda}^{-1}\mathbf{y}^*$, but leaving $\mathbf{x_2^*}$ underdetermined. Consult the commutative diagram that is shown in Fig. 1.7 for a short hand summary of the introduced transformations of coordinates of both the parameter space $\mathbb{X}$ and the observation space $\mathbb{Y}$. Next we shall proof that $\mathbf{x_2^*} = 0$ if $\mathbf{x}^*$ is MINOS.

Six, we prepare ourselves for MINOS of the underdetermined system of linear equations $\{\mathbf{y} = \mathbf{Ax} | \mathbf{A} \in \mathbb{R}^{n \times m}, \text{rk}\mathbf{A} = n, n < m,\ ||\mathbf{x}||^2_{\mathbf{G_x}} = \min\}$ by introducing *Lemma 1.7*, namely the eigenvalue–eigencolumn equations of the matrices $\mathbf{A}^\#\mathbf{A}$ and $\mathbf{AA}^\#$, respectively, as well as *Lemma 1.8*, our basic result on canonical MINOS, subsequently completed by proofs.

Lemma 1.7. (eigenspace analysis versus eigenspace synthesis of the matrix $\{\mathbf{A} \in \mathbb{R}^{n \times m}, r := rk\mathbf{A} = n < m\}$)

The pair of matrices $\{\mathbf{L}, \mathbf{R}\}$ for the eigenspace analysis and the eigenspace synthesis of the rectangular matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ of rank $r := rk\mathbf{A} = n < m$, namely

Fig. 1.7 Commutative diagram of coordinate transformations

$$
\begin{tikzcd}
\mathbb{X} \ni \mathbf{x} \arrow[r, "\mathbf{A}"] \arrow[d, "\mathbf{V}'\mathbf{G}_{\mathbf{x}}^{\frac{1}{2}}"'] & \mathbf{y} \in \mathbb{Y} \arrow[d, "\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{\frac{1}{2}}"] \\
\mathbb{X} \ni \mathbf{x}^* \arrow[r, "\mathbf{A}^*"'] & \mathbf{y}^* \in \mathbb{Y}
\end{tikzcd}
$$

$$\mathbf{A}^* = \mathbf{L}^{-1}\mathbf{A}\mathbf{R} \ \text{ versus } \ \mathbf{A} = \mathbf{L}\mathbf{A}^*\mathbf{R}^{-1} \tag{1.118}$$

or

$$\mathbf{A}^* = [\boldsymbol{\Lambda}, \mathbf{0}] = \mathbf{L}^{-1}\mathbf{A}\,[\mathbf{R}_1, \mathbf{R}_2] \ \text{ versus } \ \mathbf{A} = \mathbf{L}\,[\boldsymbol{\Lambda}, \mathbf{0}]\begin{bmatrix} \mathbf{R}_1^{-1} \\ \mathbf{R}_2^{-1} \end{bmatrix}, \tag{1.119}$$

is determined by the eigenvalue–eigencolumn equations (eigenspace equations) of the matrices $\mathbf{A}^{\#}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{\#}$, respectively, namely

$$\mathbf{A}^{\#}\mathbf{A}\mathbf{R}_1 = \mathbf{R}_1\boldsymbol{\Lambda}^2 \ \text{ versus } \ \mathbf{A}\mathbf{A}^{\#}\mathbf{L} = \boldsymbol{\Lambda}^2 \tag{1.120}$$

subject to

$$\boldsymbol{\Lambda}^2 = \begin{bmatrix} \lambda_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r^2 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \operatorname{Diag}\left(+\sqrt{\lambda_1^2}, \dots, +\sqrt{\lambda_r^2}\right). \tag{1.121}$$

Proof.

Let us prove first $\mathbf{A}\mathbf{A}^{\#}\mathbf{L} = \mathbf{L}\boldsymbol{\Lambda}^2$, second $\mathbf{A}^{\#}\mathbf{A}\mathbf{R}_1 = \mathbf{R}_1\boldsymbol{\Lambda}^2$.

$$\begin{gathered}\mathbf{A}\mathbf{A}^{\#}\mathbf{L} = \mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{L} \\ = \mathbf{L}\,[\boldsymbol{\Lambda}, \mathbf{0}]\begin{bmatrix} \mathbf{V}_1' \\ \mathbf{V}_2' \end{bmatrix} \mathbf{G}_{\mathbf{x}}^{1/2}\mathbf{G}_{\mathbf{x}}^{-1}(\mathbf{G}_{\mathbf{x}}^{-\frac{1}{2}})'\,[\mathbf{V}_1, \mathbf{V}_2]\begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0}' \end{bmatrix} \mathbf{U}'(\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}})'\mathbf{G}_{\mathbf{y}}\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{U},\end{gathered} \tag{1.122}$$

$$\mathbf{A}\mathbf{A}^{\#}\mathbf{L} = \mathbf{L}\,[\boldsymbol{\Lambda}, \mathbf{0}]\begin{bmatrix} \mathbf{V}_1'\mathbf{V}_1 & \mathbf{V}_1'\mathbf{V}_2 \\ \mathbf{V}_2'\mathbf{V}_1 & \mathbf{V}_2'\mathbf{V}_2 \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0}' \end{bmatrix} = \mathbf{L}\,[\boldsymbol{\Lambda}, \mathbf{0}]\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m-r} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0}' \end{bmatrix}. \tag{1.123}$$

$$\begin{aligned}\mathbf{A}^{\#}\mathbf{A}\mathbf{R} &= \mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{A}\mathbf{R}\\ &= \mathbf{G}_{\mathbf{x}}^{-1}\mathbf{G}_{\mathbf{x}}^{1/2}\mathbf{V}\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}'\end{bmatrix}\mathbf{U}'(\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}})'\mathbf{G}_{\mathbf{y}}\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{U}\,[\mathbf{\Lambda},\mathbf{0}]\,\mathbf{V}'\mathbf{G}_{\mathbf{x}}^{\frac{1}{2}}\mathbf{G}_{\mathbf{x}}^{-\frac{1}{2}}\mathbf{V},\end{aligned} \tag{1.124}$$

$$\mathbf{A}^{\#}\mathbf{A}\mathbf{R} = \mathbf{G}_{\mathbf{x}}^{-\frac{1}{2}}\mathbf{V}\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}'\end{bmatrix}[\mathbf{\Lambda},\mathbf{0}] = \mathbf{G}_{\mathbf{x}}^{-\frac{1}{2}}[\mathbf{V}_1,\mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},$$

$$\mathbf{A}^{\#}\mathbf{A}\,[\mathbf{R}_1,\mathbf{R}_2] = \mathbf{G}_{\mathbf{x}}^{-\frac{1}{2}}\left[\mathbf{V}_1\mathbf{\Lambda}^2,\mathbf{0}\right],\ \mathbf{A}^{\#}\mathbf{A}\mathbf{R}_1 = \mathbf{R}_1\mathbf{\Lambda}^2. \tag{1.125}$$

Lemma 1.8. (The canonical MINOS).

Let $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^*$ be a canonical representation of the underdetermined system of linear equations $\{\mathbf{y} = \mathbf{A}\mathbf{x}|\mathbf{A} \in \mathbb{R}^{\mathbf{n}\times\mathbf{m}}, \mathbf{r} := \mathbf{rkA} = \mathbf{n}, \mathbf{n} < \mathbf{m}\}$. Then the rank partitioning of $\mathbf{x}_{\mathbf{m}}^*$ is $\mathbf{G}_{\mathbf{x}}$-MINOS. In terms if the original coordinates $[x_1,\ldots,x_m]'$ of the parameter space $\mathbb{X}$ a canonical representation of $\mathbf{G}_{\mathbf{x}}$-MINOS is provided by (1.127). The $\mathbf{G}_{\mathbf{x}}$-MINOS solution $\mathbf{x}_{\mathbf{m}}$ is built on the canonical $\{\mathbf{G}_{\mathbf{x}},\mathbf{G}_{\mathbf{y}}\}$ weighted reflexive inverse of $\mathbf{A}$ provided by (1.128).

$$\mathbf{x}_{\mathbf{m}}^* = \begin{bmatrix}\mathbf{x}_1^*\\ \mathbf{x}_2^*\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{y}^*\\ \mathbf{0}\end{bmatrix}, \mathbf{x}_1^* \in \mathbb{R}^{r\times 1}, \mathbf{x}_2^* \in \mathbb{R}^{(m-r)\times 1}, \tag{1.126}$$

$$\mathbf{x}_{\mathbf{m}} = \mathbf{G}_{\mathbf{x}}^{-1/2}[\mathbf{V}_1,\mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1}\\ \mathbf{0}'\end{bmatrix}\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{1/2}\mathbf{y} = \mathbf{G}_{\mathbf{x}}^{-1/2}\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{1/2} \tag{1.127}$$

$$\begin{aligned}&= \mathbf{R}_1\mathbf{\Lambda}^{-1}\mathbf{L}^{-1}\mathbf{y},\\ \mathbf{A}_{\mathbf{m}}^- &= \mathbf{G}_{\mathbf{x}}^{-1/2}\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{1/2}.\end{aligned} \tag{1.128}$$

Proof.

For the proof we depart from $\mathbf{G}_{\mathbf{x}}$-MINOS (1.58) and replace the matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ by its canonical representation, namely eigenspace synthesis.

$$\mathbf{x}_{\mathbf{m}} = \mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\left(\mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\right)^{-1}\mathbf{y},\ \mathbf{A} = \mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{U}\,[\mathbf{\Lambda},\mathbf{0}]\begin{bmatrix}\mathbf{V}_1'\\ \mathbf{V}_2'\end{bmatrix}\mathbf{G}_{\mathbf{x}}^{1/2}, \tag{1.129}$$

$$\mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}' = \mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{U}\,[\mathbf{\Lambda},\mathbf{0}]\begin{bmatrix}\mathbf{V}_1'\\ \mathbf{V}_2'\end{bmatrix}\mathbf{G}_{\mathbf{x}}^{1/2}\mathbf{G}_{\mathbf{x}}^{-1}(\mathbf{G}_{\mathbf{x}}^{1/2})'\,[\mathbf{V}_1,\mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}\mathbf{U}'(\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}})' \tag{1.130}$$

$$= \mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}'(\mathbf{G}_{\mathbf{y}}^{-\frac{1}{2}})',\ \left(\mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\right)^{-1} = \left(\mathbf{G}_{\mathbf{y}}^{1/2}\right)'\mathbf{U}\mathbf{\Lambda}^{-2}\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{\frac{1}{2}},$$

$$\begin{aligned}\mathbf{x_m} &= \mathbf{G_x^{-1}}\left(\mathbf{G_x^{1/2}}\right)'[\mathbf{V_1},\mathbf{V_2}]\begin{bmatrix}\boldsymbol{\Lambda}\\ \mathbf{0}\end{bmatrix}\mathbf{U}'\left(\mathbf{G_y^{-\frac{1}{2}}}\right)'\left(\mathbf{G_y^{1/2}}\right)'\mathbf{U}\boldsymbol{\Lambda}^{-2}\mathbf{U}'\mathbf{G_y^{\frac{1}{2}}}\mathbf{y}\\ &= \mathbf{G_x^{-\frac{1}{2}}}[\mathbf{V_1},\mathbf{V_2}]\begin{bmatrix}\boldsymbol{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix}\mathbf{U}'\mathbf{G_y^{\frac{1}{2}}}\mathbf{y} = \mathbf{G_x^{-\frac{1}{2}}}\mathbf{V_1}\boldsymbol{\Lambda}^{-1}\mathbf{U}'\mathbf{G_y^{\frac{1}{2}}}\mathbf{y} = \mathbf{A_m^-}\mathbf{y},\end{aligned} \tag{1.131}$$

$$\mathbf{A_m^-} = \mathbf{G_x^{-\frac{1}{2}}}\mathbf{V_1}\boldsymbol{\Lambda}^{-1}\mathbf{U}'\mathbf{G_y^{\frac{1}{2}}} \in \mathbf{A^{1,2,4}_{G_x}}(\mathbf{G_x} \text{ weighted reflexive inverse of } \mathbf{A}), \tag{1.132}$$

$$\mathbf{x_m^*} = \begin{bmatrix}\mathbf{x_1^*}\\ \mathbf{x_2^*}\end{bmatrix} = \mathbf{V}'\mathbf{G_x^{\frac{1}{2}}}\mathbf{x_m} = \begin{bmatrix}\boldsymbol{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix}\mathbf{U}'\mathbf{G_y^{\frac{1}{2}}}\mathbf{y} = \begin{bmatrix}\boldsymbol{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix}\mathbf{y}^* = \begin{bmatrix}\boldsymbol{\Lambda}^{-1}\mathbf{y}^*\\ \mathbf{0}\end{bmatrix}. \tag{1.133}$$

The pair of eigensystems $\mathbf{AA^\#L} = \mathbf{L}\boldsymbol{\Lambda}^2, \mathbf{A^\#AR_1} = \mathbf{R_1}\boldsymbol{\Lambda}^2$ is unfortunately based upon non-symmetric matrices $\mathbf{AA^\#} = \mathbf{AG_x^{-1}A'G_y}$ and $\mathbf{A^\#A} = \mathbf{G_x^{-1}A'G_yA}$ which make the left and right eigenspace analysis numerically more complex. It appears that we are forced to use the *Arnoldi method* rather than the more efficient *Lanczos method* used for symmetric matrices. In this situation we look out for an alternative. Indeed when we substitute $\{\mathbf{L},\mathbf{R}\}$ by $\{\mathbf{G_y^{-1/2}U},\mathbf{G_x^{-1/2}V}\}$ into the pair of eigensystems and consequently left multiply $\mathbf{AA^\#L}$ by $\mathbf{G_x^{-1/2}}$, we achieve a pair of eigensystems identified in Corollary 1.9 relying on symmetric matrices. In addition, such a symmetric pair of eigensystems produces the canonical base, namely orthonormal eigencolumns. Such a procedure requires two factorizations, namely $\mathbf{G_x} = (\mathbf{G_x^{1/2}})'\mathbf{G_x^{1/2}}, \mathbf{G_x^{-1}} = \mathbf{G_x^{-1/2}}(\mathbf{G_x^{-1/2}})'$ and $\mathbf{G_y} = (\mathbf{G_y^{1/2}})'\mathbf{G_y}, \mathbf{G_y^{-1}} = \mathbf{G_y^{-1/2}}(\mathbf{G_y^{-1/2}})'$ via Cholesky factorization or eigenvalue decomposition of the matrices $\mathbf{G_x}$ and $\mathbf{G_y}$.

Corollary 1.9. *(symmetric pair of eigensystems)*

The pair of eigensystems (1.134) is based upon symmetric matrices. The left and right eigencolumns are orthogonal.

$$\mathbf{G_y^{1/2}AG_x^{-1}A'(G_y^{1/2})'U} = \boldsymbol{\Lambda}^2\mathbf{U} \text{ versus } (\mathbf{G_x^{-1/2}})'\mathbf{A'G_yAG_x^{-1/2}V_1} = \mathbf{V_1}\boldsymbol{\Lambda}^2, \tag{1.134}$$

$$\begin{aligned}\left|\mathbf{G_y^{1/2}AG_x^{-1}A'(G_y^{-1/2})'} - \boldsymbol{\lambda}_\mathbf{i}^2\mathbf{I_r}\right| &= 0,\\ \left|(\mathbf{G_x^{-1/2}})'\mathbf{A'G_yAG_x^{-1/2}} - \boldsymbol{\lambda}_\mathbf{j}^2\mathbf{I_m}\right| &= 0.\end{aligned} \tag{1.135}$$

The important result of $\mathbf{x_m^*}$ based on the canonical $\mathbf{G_x}$-MINOS of (1.136) needs a short comment. The rank partitioning of the canonical unknown vector $\mathbf{x}^*$, namely $\mathbf{x_1^*} \in \mathbb{R}^r$, $\mathbf{x_2^*} \in \mathbb{R}^{m-r}$ again paved the way for an interpretation. First, we acknowledge the "direct inversion" (1.137), for instance (1.138). Second, $\mathbf{x_2^*} = \mathbf{0}$, for instance (1.139) introduces a fixed datum for the canonical coordinates that are given by $[x_{r+1},\ldots,x_m]$.

$$\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^*|\mathbf{A}^* \in \mathbb{R}^{n\times m}, rk\mathbf{A}^* = rk\mathbf{A} = n, n < m, \tag{1.136}$$

$$\begin{tikzcd}
\mathbb{Y} \ni \mathrm{y} \arrow[r, "\mathrm{A}_\mathrm{m}^-"] \arrow[d, "\mathrm{U}'\mathrm{G}_\mathrm{y}^{\frac{1}{2}}"'] & \mathrm{x}_\mathrm{m} \in \mathbb{X} \arrow[d, "\mathrm{V}'\mathrm{G}_\mathrm{x}^{\frac{1}{2}}"] \\
\mathbb{Y} \ni \mathrm{y}^* \arrow[r, "(\mathrm{A}^*)_\mathrm{m}^-"'] & \mathrm{x}_\mathrm{m}^* \in \mathbb{X}
\end{tikzcd}$$

Fig. 1.8 Commutative diagram of inverse coordinate transformations

$$\mathbf{x}_1^* = \boldsymbol{\Lambda}^{-1}\mathbf{y}^*, \ \ \boldsymbol{\Lambda} = \mathrm{Diag}\left[+\sqrt{\lambda_1^2}, \ldots, +\sqrt{\lambda_r^2}\right], \tag{1.137}$$

$$[x_1^*, \ldots, x_r^*]' = [\lambda_1^{-1} y_1, \ldots, \lambda_r^{-1} y_r]' \tag{1.138}$$

$$[x_{r+1}^*, \ldots, x_m^*]' = [0, \ldots, 0]' \tag{1.139}$$

Beyond it, enjoy the commutative diagram of Fig. 1.8 illustrating our previously introduced transformations of type MINOS and canonical MINOS, by means of $\mathbf{A}_\mathbf{m}^-$ and $(\mathbf{A}^*)_\mathbf{m}^-$. Finally in *Box 1.12*, let us compute canonical MINOS for the Front Page Example, specialized by $\mathbf{G_x} = \mathbf{I_3}$ and $\mathbf{G_y} = \mathbf{I_2}$. Note that the eigenspace analysis, in summary, leads to the result

$$\boldsymbol{\Lambda} = \mathrm{Diag}\left[\sqrt{12 + \sqrt{130}}, \sqrt{12 - \sqrt{130}}\right], \tag{1.140}$$

$$\mathbf{U} = \begin{bmatrix} \dfrac{7}{\sqrt{260 + 18\sqrt{130}}} & \dfrac{7}{\sqrt{260 - 18\sqrt{130}}} \\ \dfrac{\sqrt{211 + 18\sqrt{130}}}{\sqrt{260 + 18\sqrt{130}}} & \dfrac{\sqrt{211 + 18\sqrt{130}}}{\sqrt{260 - 18\sqrt{130}}} \end{bmatrix}, \tag{1.141}$$

$$\mathbf{V} = \left[\begin{array}{cc|c} \dfrac{62 + 5\sqrt{130}}{\sqrt{102700 + 9004\sqrt{130}}} & \dfrac{62 - 5\sqrt{130}}{\sqrt{102700 - 9004\sqrt{130}}} & \dfrac{2}{\sqrt{14}} \\ \dfrac{105 + 9\sqrt{130}}{\sqrt{102700 + 9004\sqrt{130}}} & \dfrac{105 - 9\sqrt{130}}{\sqrt{102700 - 9004\sqrt{130}}} & \dfrac{3}{\sqrt{14}} \\ \dfrac{191 + 17\sqrt{130}}{\sqrt{102700 + 9004\sqrt{130}}} & \dfrac{-191 + 17\sqrt{130}}{\sqrt{102700 - 9004\sqrt{130}}} & \dfrac{1}{\sqrt{14}} \end{array}\right] = [\mathbf{V_1}, \mathbf{V_2}]. \tag{1.142}$$

Box 1.12. (The computation of canonical MINOS for the Front Page Example):

$$\mathbf{y} = \mathbf{A}\mathbf{x} : \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, r := rk\mathbf{A} = 2 \tag{1.143}$$

Left eigenspace — Right eigenspace

$$\mathbf{A}\mathbf{A}^{\#}\mathbf{U} = \mathbf{A}\mathbf{A}'\mathbf{U} = \mathbf{U}\mathbf{\Lambda}^2, \qquad \mathbf{A}^{\#}\mathbf{A}\mathbf{V}_1 = \mathbf{A}'\mathbf{A}\mathbf{V}_1 = \mathbf{V}_1\mathbf{\Lambda}^2,$$

$$\mathbf{A}^{\#}\mathbf{A}\mathbf{V}_2 = \mathbf{A}'\mathbf{A}\mathbf{V}_2 = 0,$$

$$\mathbf{A}\mathbf{A}' = \begin{bmatrix} 3 & 7 \\ 7 & 21 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 3 & 5 \\ 3 & 5 & 9 \\ 5 & 9 & 17 \end{bmatrix} = \mathbf{A}'\mathbf{A}. \tag{1.144}$$

Eigenvalues:

$$\left|\mathbf{A}\mathbf{A}' - \lambda_i^2\mathbf{I_2}\right| = 0 \quad \Leftrightarrow \quad \left|\mathbf{A}'\mathbf{A} - \lambda_j^2\mathbf{I_3}\right| = 0$$

$$\lambda_1^2 = 12 + \sqrt{130} \qquad \lambda_2^2 = 12 - \sqrt{130}, \lambda_3^2 = 0 \tag{1.145}$$

Left eigencolumns (1st): — Right eigencolumns (1st):

$$\begin{bmatrix} 3-\lambda_1^2 & 7 \\ 7 & 21-\lambda_1^2 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} = 0 \qquad \begin{bmatrix} 2-\lambda_1^2 & 3 & 5 \\ 3 & 5-\lambda_1^2 & 9 \\ 5 & 9 & 17-\lambda_1^2 \end{bmatrix} \begin{bmatrix} \mathrm{v}_{11} \\ \mathrm{v}_{21} \\ \mathrm{v}_{31} \end{bmatrix} = 0$$

$$(\text{subject to } u_{11}^2 + u_{21}^2 = 1) \qquad (\text{subject to } \mathrm{v}_{11}^2 + \mathrm{v}_{21}^2 + \mathrm{v}_{31}^2 = 1) \tag{1.146}$$

$$(3-\lambda_1^2)u_{11} + 7u_{21} = 0, \qquad \begin{aligned} (2-\lambda_1^2)\mathrm{v}_{11} + 3\mathrm{v}_{21} + 5\mathrm{v}_{31} &= 0 \\ 3\mathrm{v}_{11} + (5-\lambda_1^2)\mathrm{v}_{21} + 9\mathrm{v}_{31} &= 0 \end{aligned} \tag{1.147}$$

$$\begin{aligned} u_{11}^2 &= \frac{49}{49 + (3-\lambda_1^2)^2} = \frac{49}{260 + 18\sqrt{130}}, \\ u_{21}^2 &= \frac{(3-\lambda_1^2)^2}{49 + (3-\lambda_1^2)^2} = \frac{211 + 18\sqrt{130}}{260 + 18\sqrt{130}}, \end{aligned} \tag{1.148}$$

$$\begin{bmatrix} \mathrm{v}_{11}^2 \\ \mathrm{v}_{21}^2 \\ \mathrm{v}_{31}^2 \end{bmatrix} = \frac{1}{(2+5\lambda_1^2)^2 + (3-9\lambda_1^2)^2 + (-1+7\lambda_1^2-\lambda_1^4)^2} \begin{bmatrix} (2+5\lambda_1^2)^2 \\ (3-9\lambda_1^2)^2 \\ (1-7\lambda_1^2+\lambda_1^4)^2 \end{bmatrix}, \tag{1.149}$$

$$\begin{bmatrix} v_{11}^2 \\ v_{21}^2 \\ v_{31}^2 \end{bmatrix} = \frac{1}{102700 + 9004\sqrt{130}} \begin{bmatrix} \left(62 + 5\sqrt{130}\right)^2 \\ \left(-105 - 9\sqrt{130}\right)^2 \\ \left(191 + 17\sqrt{130}\right)^2 \end{bmatrix}.$$

Left eigencolumns (2nd):	Right eigencolumns (2nd):

$$\begin{bmatrix} 3-\lambda_2^2 & 7 \\ 7 & 21-\lambda_2^2 \end{bmatrix} \begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} = 0 \qquad \begin{bmatrix} 2-\lambda_2^2 & 3 & 5 \\ 3 & 5-\lambda_2^2 & 9 \\ 5 & 9 & 17-\lambda_2^2 \end{bmatrix} \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \end{bmatrix} = 0$$

$$\text{(subject to } u_{12}^2 + u_{22}^2 = 1\text{)}, \qquad \text{(subject to } v_{12}^2 + v_{22}^2 + v_{32}^2 = 1\text{)}, \tag{1.150}$$

$$(3-\lambda_2^2)u_{12} + 7u_{22} = 0, \qquad \begin{aligned} (2-\lambda_2^2)v_{12} + 3v_{22} + 5v_{32} = 0, \\ 3v_{12} + (5-\lambda_2^2)v_{22} + 9v_{32} = 0, \end{aligned} \tag{1.151}$$

$$\begin{aligned} u_{12}^2 &= \frac{49}{49 + (3-\lambda_2^2)^2} = \frac{49}{260 - 18\sqrt{130}}, \\ u_{22}^2 &= \frac{(3-\lambda_2^2)^2}{49 + (3-\lambda_2^2)^2} = \frac{211 - 18\sqrt{130}}{260 - 18\sqrt{130}}, \end{aligned} \tag{1.152}$$

$$\begin{bmatrix} v_{12}^2 \\ v_{22}^2 \\ v_{32}^2 \end{bmatrix} = \frac{1}{(2+5\lambda_2^2)^2 + (3-9\lambda_2^2)^2 + (-1+7\lambda_2^2-\lambda_2^4)^2} \begin{bmatrix} (2+5\lambda_2^2)^2 \\ (3-9\lambda_2^2)^2 \\ (1-7\lambda_2^2+\lambda_2^4)^2 \end{bmatrix},$$

$$\begin{bmatrix} v_{12}^2 \\ v_{22}^2 \\ v_{32}^2 \end{bmatrix} = \frac{1}{102700 - 9004\sqrt{130}} \begin{bmatrix} \left(62 - 5\sqrt{130}\right)^2 \\ \left(-105 + 9\sqrt{130}\right)^2 \\ \left(191 - 17\sqrt{130}\right)^2 \end{bmatrix}. \tag{1.153}$$

Right eigencolumns (3rd):

$$\begin{bmatrix} 2 & 3 & 5 \\ 3 & 5 & 9 \\ 5 & 9 & 17 \end{bmatrix} \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \end{bmatrix} = 0 \tag{1.154}$$

$$\text{(subject to } v_{13}^2 + v_{23}^2 + v_{33}^2 = 1\text{)},$$

$$\begin{aligned} 2v_{13} + 3v_{23} + 5v_{33} = 0 \\ 3v_{13} + 5v_{23} + 9v_{33} = 0 \end{aligned} \tag{1.155}$$

$$\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} v_{13} \\ v_{23} \end{bmatrix} = \begin{bmatrix} -5 \\ -9 \end{bmatrix} v_{33} \Leftrightarrow \begin{bmatrix} v_{13} \\ v_{23} \end{bmatrix} = -\begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 9 \end{bmatrix} v_{33} \tag{1.156}$$

$$v_{13} = 2v_{33}, v_{23} = -3v_{33}, v_{13}^2 = \frac{2}{7}, v_{23}^2 = \frac{9}{14}, v_{33}^2 = \frac{1}{14}. \tag{1.157}$$

There are four combinatorial solutions to generate square roots.

$$\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{u_{11}^2} & \pm\sqrt{u_{12}^2} \\ \pm\sqrt{u_{21}^2} & \pm\sqrt{u_{22}^2} \end{bmatrix}, \tag{1.158}$$

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{v_{11}^2} & \pm\sqrt{v_{12}^2} & \pm\sqrt{v_{13}^2} \\ \pm\sqrt{v_{21}^2} & \pm\sqrt{v_{22}^2} & \pm\sqrt{v_{23}^2} \\ \pm\sqrt{v_{31}^2} & \pm\sqrt{v_{32}^2} & \pm\sqrt{v_{33}^2} \end{bmatrix}. \tag{1.159}$$

Here we have chosen the one with the positive sign exclusively.

$$\mathbf{\Lambda} = \mathrm{Diag}\left[\sqrt{12+\sqrt{130}}, \sqrt{12-\sqrt{130}}\right], \tag{1.160}$$

$$\mathbf{U} = \begin{bmatrix} \dfrac{7}{\sqrt{260+18\sqrt{130}}} & \dfrac{7}{\sqrt{260-18\sqrt{130}}} \\ \dfrac{\sqrt{211+18\sqrt{130}}}{\sqrt{260+18\sqrt{130}}} & \dfrac{\sqrt{211+18\sqrt{130}}}{\sqrt{260-18\sqrt{130}}} \end{bmatrix}, \tag{1.161}$$

$$\mathbf{V} = \left[\begin{array}{cc|c} \dfrac{62+5\sqrt{130}}{\sqrt{102700+9004\sqrt{130}}} & \dfrac{62-5\sqrt{130}}{\sqrt{102700-9004\sqrt{130}}} & \dfrac{2}{\sqrt{14}} \\ \dfrac{105+9\sqrt{130}}{\sqrt{102700+9004\sqrt{130}}} & \dfrac{105-9\sqrt{130}}{\sqrt{102700-9004\sqrt{130}}} & \dfrac{3}{\sqrt{14}} \\ \dfrac{191+17\sqrt{130}}{\sqrt{102700+9004\sqrt{130}}} & \dfrac{-191+17\sqrt{130}}{\sqrt{102700-9004\sqrt{130}}} & \dfrac{1}{\sqrt{14}} \end{array}\right] = [\mathbf{V_1}, \mathbf{V_2}]. \tag{1.162}$$

1-3 Case Study

Let us continue this chapter by considering some important case studies. Let us consider orthogonal functions and let us study Fourier series versus *Fourier–Legendre* series and circular harmonic versus spherical harmonic regression. In empirical sciences, we continuously meet the problems of underdetermined linear equations. Typically we develop a characteristic field variable into orthogonal series, for instance into circular harmonic functions (discrete Fourier transform) or into spherical harmonics (discrete *Fourier–Legendre* transform) with respect to a reference sphere. We are left with the problem of algebraic regression to determine the values of the function at sample points, an infinite set of coefficients of the series expansion from a finite set of observations. An infinite set of coefficients, the coordinates in an infinite-dimensional Hilbert space, cannot be determined by finite computer manipulations. Instead, band-limited functions are introduced. Only a finite set of coefficients of a circular harmonic expansion or of a spherical harmonic expansion can be determined. It is the art of the analyst to fix the degree/order of the expansion properly. In a peculiar way the choice of the highest degree/order of the expansion is related to the Uncertainty Principle, namely to the width of lattice of the sampling points.

Another aspect of any series expansion is the choice of the function space. For instance, if we develop scalar-valued, vector-valued or tensor-valued functions into scalar-valued, vector-valued or tensor-valued circular or spherical harmonics, we generate orthogonal functions with respect to a special inner product, also called *scalar product* on the circle or spherical harmonics are eigenfunctions of the circular or spherical *Laplace–Beltrami operator*. Under the postulate of the *Sturm–Liouville boundary conditions*, the spectrum (namely the eigenvalues) of the Laplace–Beltrami operator are *positive* and *integer*. We here note that the eigenvalues of the circular Laplace–Beltrami operator are l^2 for integer values $l \in \{0, 1, \ldots, \infty\}$, of the spherical Laplace–Beltrami operator $\left\{k(k+1), l^2\right\}$ for integer values $k \in \{0, 1, \ldots, \infty\}$, $l \in \{-k, -k+1, \ldots, -1, 0, 1, \ldots, k-1, k\}$. Thanks to such a structure of the infinite-dimensional eigenspace of the Laplace–Beltrami operator we discuss the solutions of the underdetermined regression problem (linear algebraic regression) in the context of "canonical MINOS". We solve the system of linear equations $\{\mathbf{Ax} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m}, rk\mathbf{A} = n, n \sim m\}$ by singular value decomposition as shortly outlined in Appendix A.

1-31 Fourier Series

Fourier series represent the periodic behavior of a function $\mathbf{x}(\lambda)$ on a circle $\mathbb{S}^1$. Fourier series are also called *trigonometric series* since trigonometric functions $\{1, \sin\lambda, \cos\lambda, \sin 2\lambda, \cos 2\lambda, \ldots, \sin\ell\lambda, \cos\ell\lambda\}$ represent such a periodic signal. Here we have chosen the parameter "longitude λ" to locate a point on $\mathbb{S}^1$. Instead we could exchange the parameter λ by time t, if clock readings would substitute longitude, a conventional technique in classical navigation. In such a setting,

$$\lambda = \omega t = \frac{2\pi}{T} t = 2\pi\nu t, \ell\lambda = \ell\omega t = 2\pi\ell\frac{t}{T} = 2\pi\ell\nu t. \tag{1.163}$$

(In such a setting, the longitude λ would be exchanged by 2π, the product of ground period T and time t or by 2π, the product of ground frequency ν. In contrast, $\ell\lambda$ for all $\ell \in \{0, 1, \ldots, L\}$ would be substituted by 2π the product of overtones ℓ/T or $\ell\nu$ and time t, where ℓ is integer, i.e., $l \in \mathbb{Z}$. According to classical navigation, ω would represent the rotational speed of the Earth.) Let us refer to Box 1.13 as a summary of the Fourier representation of a function $\mathbf{x}(\lambda)$, $\lambda \in \mathbb{S}^1$. Note that a Fourier series (1.164) or (1.165) can be understood as an infinite-dimensional vector space (linear space, Hilbert space) since the base functions $\mathbf{e}_\ell(\lambda)$ generate a complete orthogonal (orthonormal) system based on trigonometric functions. The countable base, namely the base functions $\mathbf{e}_\ell(\lambda)$ or $\{1, \sin\lambda, \cos\lambda, \sin 2\lambda, \cos 2\lambda, \ldots, \sin\ell\lambda, \cos\ell\lambda\}$ span the Fourier space $L^2[0, 2\pi]$. According to the ordering by means of positive and negative indices $\{-L, -L+1, \ldots, -1, 0, +1, \ldots, L-1, L\}$ (0.4) $\mathbf{x}^\wedge(\lambda)$ is an approximation of the function $\mathbf{x}(\lambda)$ up to order three, also denoted by $\mathbf{x}_L$.

Box 1.13. (Fourier series):

$$\begin{aligned}\mathbf{x}(\lambda) = x_1 + (\sin\lambda)x_2 + (\cos\lambda)x_3 + (\sin 2\lambda)x_4 + (\cos 2\lambda)x_5 + \\ + O_3(\sin\ell\lambda, \cos\ell\lambda)\end{aligned} \tag{1.164}$$

$$\mathbf{x}(\lambda) = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda)x_\ell, \ \mathbf{e}_\ell(\lambda) := \left[\begin{array}{cc} \cos\ell\lambda & \forall\ell > 0 \\ 1 & \forall\ell = 0 \\ \sin|\ell|\,\lambda & \forall\ell < 0. \end{array}\right. \tag{1.165}$$

Example (approximation of order three):

$$\mathbf{x}^\wedge(\lambda) = \mathbf{e}_0 x_1 + \mathbf{e}_{-1}x_2 + \mathbf{e}_{+1}x_3 + \mathbf{e}_{-2}x_4 + \mathbf{e}_{+2}x_5 + O_3. \tag{1.166}$$

What is an infinite dimensional vector space? What is a Hilbert space? What makes up the Fourier space? An infinite dimensional vector space (linear space) is similar to a finite dimensional vector space: As in an Euclidean space an inner product and a norm is defined. While the inner product and the norm in a finite dimensional vector space required summation of their components, the inner product (1.167), (1.168) and the norm (1.170) in an infinite-dimensional vector space force us to integration. Indeed the inner products (scalar products) (1.167) and (1.168) are integrals over the line element of $\mathbb{S}^1_r$ applied to the vectors $\mathbf{x}(\lambda)$, $\mathbf{y}(\lambda)$ or $\mathbf{e}_{\ell_1}$, $\mathbf{e}_{\ell_2}$, respectively. Those integrals are divided by the length s of a total arc of $\mathbb{S}^1_r$. Alternative representations of $< \mathbf{x}|\mathbf{y} >$ and $< \mathbf{e}_{\ell_1}|\mathbf{e}_{\ell_2} >$ (Dirac's notation of brackets, decomposed into "bra" and "ket") based upon $ds = rd\lambda$, $s = 2\pi r$, lead us directly to the integration over $\mathbb{S}^1$, the unit circle. Let us refer to *Box 1.14* as a summary of the most essential features of the Fourier space.

Box 1.14. (Fourier space):

The base functions $\mathbf{e}_\ell(\lambda)$, $\ell \in \{-L, -L+1, \ldots, -1, 0, +1, \ldots, L-1, L\}$, span the Fourier space $\mathbf{L}^2\,[0, 2\pi]$: they generate a complete orthogonal (orthonormal) system of trigonometric functions.

$$\boxed{\textit{Inner product}\ (\mathbf{x} \in \text{FOURIER},\ \mathbf{y} \in \text{FOURIER}) :}$$

$$\langle \mathbf{x} \mid \mathbf{y} \rangle := \frac{1}{s}\int_0^\infty ds\,x(s)y(s) = \frac{1}{2\pi}\int_0^{2\pi} d\lambda\,x(\lambda)y(\lambda) \tag{1.167}$$

Normalization:

$$< \mathbf{e}_{\ell_1}(\lambda)|\mathbf{e}_{\ell_2}(\lambda) >:= \frac{1}{2\pi}\int_0^{2\pi} d\lambda\,\mathbf{e}_{\ell_1}(\lambda)\,\mathbf{e}_{\ell_2}(\lambda) = \lambda_{\ell_1}\delta_{\ell_1\ell_2} \tag{1.168}$$

subject to

$$\begin{bmatrix} \lambda_{\ell_1} = 1 \forall \ell_1 = 0, \\ \lambda_{\ell_1} = \frac{1}{2} \forall \ell_1 \neq 0 \end{bmatrix} \tag{1.169}$$

Norms, convergence

$$\begin{gathered} ||\mathbf{x}||^2 = \tfrac{1}{2\pi} \int_0^{2\pi} d\lambda\, \mathbf{x}^2(\lambda) = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \lambda_\ell \mathbf{x}_\ell^2 < \infty, \\ \lim_{L\to\infty} ||\mathbf{x} - \mathbf{x}_L^\wedge||^2 = 0 \text{ (convergence in the mean).} \end{gathered} \tag{1.170}$$

Synthesis versus analysis:

$$\mathbf{x} = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \mathbf{e}_\ell x_\ell \text{ versus } x_\ell = \frac{1}{\lambda_\ell} < \mathbf{e}_\ell | \mathbf{x} > := \frac{1}{2\pi\lambda_\ell} \int_0^{2\pi} d\lambda\, \mathbf{e}_\ell(\lambda)\, \mathbf{x}(\lambda), \tag{1.171}$$

$$\mathbf{x} = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \frac{1}{\lambda_\ell} \mathbf{e}_\ell < \mathbf{x} | \mathbf{e}_\ell > . \tag{1.172}$$

Canonical basis of the Hilbert space “FOURIER” (orthonormal basis)

$$\mathbf{e}_\ell^* := \begin{bmatrix} \sqrt{2} \sin \ell\lambda & \forall \ell > 0, \\ 1 & \forall \ell = 0, \\ \sqrt{2} \cos \ell\lambda & \forall \ell < 0. \end{bmatrix} \tag{1.173}$$

$$\begin{gathered} \mathbf{e}_\ell^* = \tfrac{1}{\sqrt{\lambda_\ell}} \mathbf{e}_\ell \ \ versus\ \ \mathbf{e}_\ell = \sqrt{\lambda_\ell} \mathbf{e}_\ell^*, \\ x_\ell^* = \sqrt{\lambda_\ell} x_\ell \ \ versus\ \ x_\ell = \tfrac{1}{\sqrt{\lambda_\ell}} x_\ell^*, \end{gathered} \tag{1.174}$$

$$\mathbf{x} = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \mathbf{e}_\ell^* < \mathbf{x} | \mathbf{e}_\ell^* > \tag{1.175}$$

Orthonormality:

$$< \mathbf{e}_{\ell_1}^*(\mathbf{x}) | \mathbf{e}_{\ell_2}^*(\mathbf{x}) >= \delta_{\ell_1 \ell_2} \tag{1.176}$$

Fourier space $L \to \infty$

$$\begin{gathered} FOURIER = \operatorname{span}\{\mathbf{e}_{-L}, \mathbf{e}_{-L+1}, \dots, \mathbf{e}_{-1}, \mathbf{e}_0, \mathbf{e}_1, \dots, \mathbf{e}_{L-1}, \mathbf{e}_L\}, \\ \dim FOURIER = \lim_{L\to\infty} (2L+1) = \infty, \end{gathered} \tag{1.177}$$

$$\text{“FOURIER”} = \text{“HARM”}_{L^2(\mathbb{S}^1)}. \tag{1.178}$$

A further comment has to be made to the normalization (1.168). Thanks to $< \mathbf{e}_{\ell_1}(\lambda) | \mathbf{e}_{\ell_2}(\lambda) >= 0$ for all $\ell_1 \neq \ell_2$.

for instance $< \mathbf{e}_1(\lambda) | \mathbf{e}_1(\lambda) > = 0$, the base functions $\mathbf{e}_\ell(\lambda)$ are called orthogonal. But according to $< \mathbf{e}_\ell(\lambda) | \mathbf{e}_\ell(\lambda) >= \frac{1}{2}$, for instance $< \mathbf{e}_1(\lambda) | \mathbf{e}_1(\lambda) >= ||\mathbf{e}_1(\lambda)||^2 = \frac{1}{2}$, $< \mathbf{e}_2(\lambda) | \mathbf{e}_2(\lambda) >= ||\mathbf{e}_2(\lambda)||^2 = \frac{1}{2}$, they are not normalized to

1. Note that a canonical basis of the Hilbert space "FOURIER" has been introduced by $\mathbf{e}_\ell^*$. Indeed the base functions $\mathbf{e}_\ell^*(\lambda)$ fulfil the condition of orthonormality.

The crucial point of an infinite dimensional vector space is convergency. When we write $\mathbf{x}(\lambda)$ ($\longrightarrow$ (1.164)) as an identity of infinite series we must be sure that the series converge. In infinite dimensional vector space no pointwise convergency is required. In contrast, *convergence in the mean* ($\longrightarrow$ (1.170)) is postulated. The norm $||\mathbf{x}||^2$ ($\longrightarrow$ (1.170)) equals the limes of the infinite sum of the λ_ℓ weighted, squared coordinate x_ℓ, the coefficient in the trigonometric function (1.164), $||\mathbf{x}||^2 = \lim_{L\to\infty} \sum_{\ell=-L}^{+L} \lambda_\ell x_\ell^2 < \infty$, which must be finite. As soon as *convergence in the mean* is guaranteed, we move from a pre-Fourier space of trigonometric functions to a Fourier space we shall define more precisely later on. (i) Fourier analysis as well as Fourier synthesis, represented by (1.171, left) versus (1.171, right), is meanwhile well prepared. First, given the Fourier coefficients $\mathbf{x}_\ell$ we are able to systematically represent the vector $\mathbf{x} \in$ "*FOURIER*" in the orthogonal base $\mathbf{e}_\ell(\lambda)$. Second, the projection of the vector $\mathbf{x} \in$ "*FOURIER*" onto the base vectors $\mathbf{e}_\ell(\lambda)$ agrees analytically to the Fourier coefficients as soon as we take into account the proper matrix of the metric of the Fourier space. Note the reproducing representation "from $\mathbf{x}$ to $\mathbf{x}$" ($\longrightarrow$ (1.172)). (ii) The transformation from the orthogonal base $\mathbf{e}_\ell(\lambda)$ to the orthonormal base $\mathbf{e}_\ell^*$, also called *canonical* or *modular* as well as its inverse is summarized by (1.174). The dual transformations from Fourier coefficients $\mathbf{x}_\ell$ to canonical Fourier coefficients $\mathbf{x}_\ell^*$ as well as its inverse is also summarized by (1.174). Note the canonical reproducing representation "from $\mathbf{x}$ to $\mathbf{x}$" ($\longrightarrow$ (1.175)). (iii) The space (1.177) has the dimension of hyperreal number ∞. As already mentioned in the introduction, (1.178) is identical with the Hilbert space $L^2(\mathbb{S}^1)$ of harmonic functions on the circle $\mathbb{S}^1$.

What is a harmonic function which has the unit circle $\mathbb{S}^1$ as a support? A harmonic function "on the unit circle" $\mathbb{S}^1$ is a function $x(\lambda), \lambda \in \mathbb{S}^1$, which fulfils on one hand fulfills the one-dimensional Laplace equation (the differential equation) of a harmonic oscillator) and on the other hand fulfills a special Sturm–Liouville boundary condition. The special Sturm–Liouville equation force the frequency to be integer, which is proven subsequently.

(1st):

$$\Delta_1 x(\lambda) = 0 \Leftrightarrow \left(\frac{d^2}{d\lambda^2} + \omega^2\right)x(\lambda) = 0 \tag{1.179}$$

(2nd):

$$\left[\begin{array}{c} x(0) = x(2\pi) \\ \left[\frac{d}{d\lambda}x(\lambda)\right](0) = \left[\frac{d}{d\lambda}x(\lambda)\right](2\pi). \end{array}\right. \tag{1.180}$$

Proof.

"Ansatz":

$$x(\lambda) = c_\omega \cos \omega\lambda + s_\omega \sin \omega\lambda \tag{1.181}$$

"Initial conditions":

$$x(0) = x(2\pi) \Leftrightarrow \tag{1.182}$$

Consequences:

$$c_\omega = c_\omega \cos 2\pi\omega + s_\omega \sin 2\pi\omega,$$

$$\left[\frac{d}{d\lambda}x(\lambda)\right](0) = \left[\frac{d}{d\lambda}x(2\pi)\right](2\pi) \tag{1.183}$$

$$\Longleftrightarrow$$

$$s_\omega\omega = -c_\omega\omega \sin 2\pi\omega + s_\omega\omega \cos 2\pi\omega$$

$$\Longleftrightarrow$$

$$\Leftrightarrow \left.\begin{array}{r} \cos 2\pi\omega = 0 \\ \sin 2\pi\omega = 0 \end{array}\right] \Rightarrow \omega = \ell \ \forall \ell \in \{0, 1, \ldots, L-1, L\}. \tag{1.184}$$

$\omega = \ell, \ell \in \{0, 1, \ldots, L-1, L\}$ concludes the proof.

How can we setup a linear model for Fourier analysis? The linear model of Fourier analysis which relates the elements $\mathbf{x} \in \mathbb{X}$ of the parameter space $\mathbb{X}$ to the elements $\mathbf{y} \in \mathbb{Y}$ of the observation space $\mathbb{Y}$ is setup in Box 1.15. Here we shall assume that the observed data have been made available on an equidistant angular grid, in short *equidistant lattice* of the unit circle parameterized by $(\lambda_1, \ldots, \lambda_n)$. For the optimal design of the Fourier linear model it has been proven that the equidistant lattice $\lambda_i = (i-1)\frac{2\pi}{I} \ \forall i \in \{1, \ldots, I\}$ is "D-optimal". *Box 1.15* contains four examples for such a lattice. Figure 1.9 shows such a lattice. In summary, the finite dimensional observation space $\mathbb{Y}$, $dim\mathbb{Y} = n, n = I$, has integer dimension I. In contrast, the parameter space $\mathbb{X}$, $dim\mathbb{X} = \infty$, is infinite dimensional. The unknown Fourier coefficients, conventionally collected in a Pascal triangular graph is shown in Fig. 1.10, are vectorized by (1.188) in a peculiar order (1.190). Indeed, the linear model (1.189) contains $m = 2L+1$, $L \to \infty$, $m \to \infty$, unknowns, a hyperreal number. The linear operator $\mathbf{A} : \mathbb{X} \longmapsto \mathbb{X}$ is generated by the base functions of lattice points. Equation (1.191) is a representation of the linear observational equations (1.189) in Ricci calculus which is characteristic for Fourier analysis.

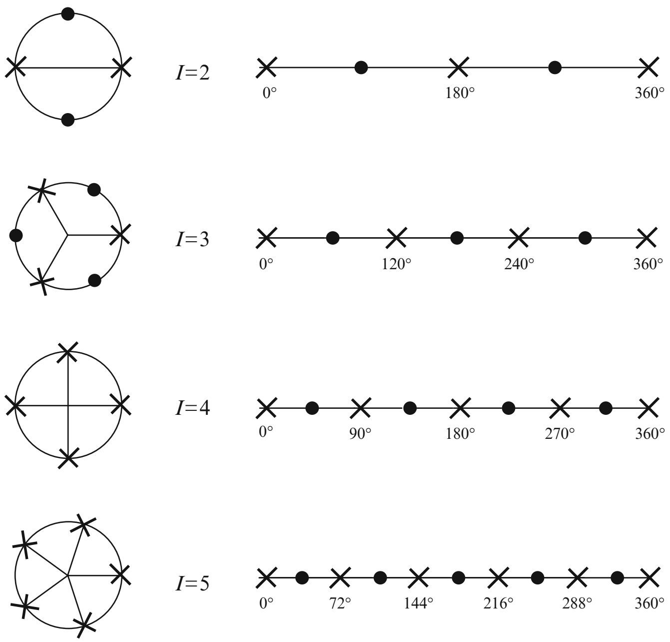


Fig. 1.9 Equidistance lattice on $\mathbb{S}^1$. I = 2 or 3 or 4 or 5

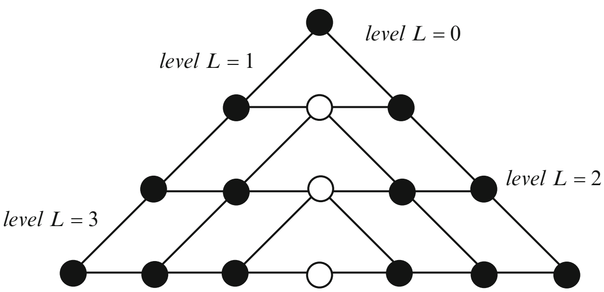


Fig. 1.10 Fourier series. Pascal triangular graph, weights of the graph: unknown coefficients of Fourier series

Box 1.15. (Fourier analysis as an underdetermined linear model)

The observation space $\mathbb{Y}$:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix} := \begin{bmatrix} x(\lambda_1) \\ x(\lambda_2) \\ \vdots \\ x(\lambda_{n-1}) \\ x(\lambda_n) \end{bmatrix} = [x(\lambda_i)] \;\; \forall i \in \{1, .., I\}, \; \lambda \in [0, 2\pi] \tag{1.185}$$

Equidistant lattice on $\mathbb{S}^1$:

$$\lambda_i = (i-1)\frac{2\pi}{I} \;\; \forall i \in \{1,\dots,I\}. \tag{1.186}$$

Examples:

$$\begin{aligned} I=2:\;& \lambda_1 = 0,\ \lambda_2 = \pi \sim 180^\circ,\\ I=3:\;& \lambda_1 = 0,\ \lambda_2 = \frac{2\pi}{3} \sim 120^\circ,\ \lambda_3 = \frac{4\pi}{3} \sim 240^\circ,\\ I=4:\;& \lambda_1 = 0,\ \lambda_2 = \frac{2\pi}{4} \sim 90^\circ,\ \lambda_3 = \pi \sim 180^\circ,\\ & \lambda_4 = \frac{3\pi}{2} \sim 270^\circ,\\ I=5:\;& \lambda_1 = 0,\ \lambda_2 = \frac{2\pi}{5} \sim 72^\circ,\ \lambda_3 = \frac{4\pi}{5} \sim 144^\circ,\\ & \lambda_4 = \tfrac{6\pi}{5} \sim 216^\circ,\ \lambda_5 = \frac{8\pi}{5} \sim 288^\circ \end{aligned} \quad \mathit{Example} \tag{1.187}$$

$\boxed{\text{The parameter space } \mathbb{X}}$

$$\begin{gathered} x_1 = x_0,\ x_2 = x_{-1},\ x_3 = x_{+1},\ x_4 = x_{-2},\\ x_5 = x_{+2},\ \dots, x_{m-1} = x_{-L},\ x_m = x_L,\ ,\\ dim\mathbb{X} = m \sim 2L+1. \end{gathered} \tag{1.188}$$

$\boxed{\text{The underdetermined linear model } (n < m:\ I < 2L+1:}$

$$\mathbf{y} := \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{n-1} \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & \sin\lambda_1 & \cos\lambda_1 & \dots & \sin L\lambda_1 & \cos L\lambda_1 \\ 1 & \sin\lambda_2 & \cos\lambda_2 & \dots & \sin L\lambda_2 & \cos L\lambda_2 \\ & & & & & \\ 1 & \sin\lambda_{n-1} & \cos\lambda_{n-1} & \dots & \sin L\lambda_{n-1} & \cos L\lambda_{n-1} \\ 1 & \sin\lambda_n & \cos\lambda_n & \dots & \sin L\lambda_n & \cos L\lambda_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_{m-1} \\ x_m \end{bmatrix}. \tag{1.189}$$

$$\mathbb{X} = \underset{L\to\infty}{\text{span}} \{x_0, x_{-1}, x_{+1}, \dots, x_{-L}, x_{+L}\},\ \dim X = m = \infty \tag{1.190}$$

$$y_i = y(\lambda_i) = \lim_{L\to\infty} \sum_{\ell=-L}^{L} \mathbf{e}_\ell(\lambda_i) x_\ell \;\; \forall i \in \{1,\dots,n\} \tag{1.191}$$

The portray of Fourier analysis which for the readers' convenience is outlined in *Box 1.16* summarizes its peculiarities effectively. A finite number of observations is confronted with an infinite number of observations. Such a linear model of type "underdetermined of power 2" cannot be solved in finite computer time. Instead one has to truncate the Fourier series, a technique or approximation to make up Fourier series "finite" or "bandlimited". We have to consider three cases that are shown in *Box 1.16*.

Box 1.16. The portray of Fourier analysis

Number of observed data at versus number of unknown

lattice points: Fourier coefficients:

$n = I$ $m = 2L + 1 \to \infty$

(finite) (infinite)

The three cases:

$n > m$

(Overdetermined case)

$n = m$

(regular case)

$n < m$

(underdetermined case).

First, we can truncate the infinite Fourier series such that $n > m$ holds. In this case of an overdetermined problem, we have more observations than equations. Second, we alternatively balance the number of unknown Fourier coefficients such that $n = m$ holds. Such a model choice assures a regular linear system. Both linear Fourier models which are tuned to the number of observations suffer from a typical uncertainty. What is the effect of the forgotten unknown Fourier coefficients $m > n$? Indeed a significance test has to decide upon any truncation to be admissible. We are in need of an objective criterion to decide upon the degree m of bandlimit. Third, in order to be as objective as possible we follow the third case of "less observations than unknowns" such that $n < m$ holds. Such a Fourier linear model which generates an underdetermined system of linear equations is explicitly considered in *Box 1.17* and *Box 1.18*. Note that the first example that is presented in Box 1.17 ($n - m = 1$) as well as the second example that is present in *Box 1.18* ($n - m = 2$) demonstrate MINOS of the Fourier linear model.

Box 1.17. (The first example).

Fourier analysis as an underdetermined linear model:

$$\begin{gathered} n - \operatorname{rk}\mathbf{A} = n - m = 1,\ L = 1 \\ dim\mathbb{Y} = n = 2,\ dim\mathbb{X} = m = 3, \end{gathered} \tag{1.192}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & \sin\lambda_1 & \cos\lambda_1 \\ 1 & \sin\lambda_2 & \cos\lambda_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \sim \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{1.193}$$

$$\text{Example } (I = 2): \ \lambda_1 = 0°,\ \lambda_2 = 180°:$$

$$\begin{gathered}\sin\lambda_1 = 0,\ \cos\lambda_1 = 1,\ \sin\lambda_2 = 0,\ \cos\lambda_2 = -1 \\ A := \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & \sin\lambda_1 & \cos\lambda_1 \\ 1 & \sin\lambda_2 & \cos\lambda_2 \end{bmatrix} \in \mathbb{R}^{2\times 3}\end{gathered} \tag{1.194}$$

$$\mathbf{AA}' = 2\mathbf{I_2} \Leftrightarrow (\mathbf{AA}')^{-1} = \frac{1}{2}\mathbf{I_2} \tag{1.195}$$

$$\mathbf{AA}' = \begin{bmatrix} 2 & 1 + \sin\lambda_1\sin\lambda_2 + \cos\lambda_1\cos\lambda_2 \\ 1 + \sin\lambda_2\sin\lambda_1 + \cos\lambda_2\cos\lambda_1 & 2 \end{bmatrix}. \tag{1.196}$$

$$\text{if } \lambda_i = (i - \lambda)\tfrac{2\pi}{I}, \text{ then}$$

$$1 + 2\sin\lambda_1\sin\lambda_2 + 2\cos\lambda_1\cos\lambda_2 = 0 \tag{1.197}$$

or

$$\begin{gathered} L = 1: \ \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda_{i_1})\mathbf{e}_\ell(\lambda_{i_2}) = 0 \ \forall i_1 \neq i_2, \\ L = 1: \ \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda_{i_1})\mathbf{e}_\ell(\lambda_{i_2}) = L + 1 \ \forall i_1 = i_2, \end{gathered} \tag{1.198}$$

$$\begin{gathered}\mathbf{x}_\ell = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_\ell = \mathbf{A}'(\mathbf{AA}')^{-1}\mathbf{y} = \tfrac{1}{2}\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}\mathbf{y} = \frac{1}{2}\begin{bmatrix} y_1 + y_2 \\ 0 \\ y_1 - y_2 \end{bmatrix}, \\ ||\mathbf{x}_\ell||^2 = \frac{1}{2}\mathbf{y}'\mathbf{y}\end{gathered} \tag{1.199}$$

Box 1.18. (The second example)

Fourier analysis as an underdetermined linear model:

$$\begin{gathered} n - \operatorname{rk}\mathbf{A} = n - m = 2,\ L = 2 \\ dim\mathbb{Y} = n = 3,\ dim\mathbb{X} = m = 5, \end{gathered} \tag{1.200}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & \sin\lambda_1 & \cos\lambda_1 & \sin 2\lambda_1 & \cos 2\lambda_1 \\ 1 & \sin\lambda_2 & \cos\lambda_2 & \sin 2\lambda_2 & \cos 2\lambda_2 \\ 1 & \sin\lambda_3 & \cos\lambda_3 & \sin 2\lambda_3 & \cos 2\lambda_3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}. \tag{1.201}$$

$$\text{Example } (I = 3): \ \lambda_1 = 0°,\ \lambda_2 = 120°,\ \lambda_3 = 240°$$

$$
\begin{gathered}
\sin\lambda_1 = 0,\ \sin\lambda_2 = \frac{1}{2}\sqrt{3},\ \sin\lambda_3 = -\frac{1}{2}\sqrt{3},\\
\cos\lambda_1 = 1,\ \cos\lambda_2 = -\frac{1}{2},\ \cos\lambda_3 = -\frac{1}{2},\\
\sin 2\lambda_1 = 0,\ \sin 2\lambda_2 = -\frac{1}{2}\sqrt{3},\ \sin 2\lambda_3 = \frac{1}{2}\sqrt{3},\\
\cos 2\lambda_1 = 1,\ \cos 2\lambda_2 = -\frac{1}{2},\ \cos 2\lambda_3 = -\frac{1}{2},
\end{gathered} \tag{1.202}
$$

$$
\mathbf{A} := \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & \frac{1}{2}\sqrt{3} & -\frac{1}{2} & -\frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ 1 & -\frac{1}{2}\sqrt{3} & -\frac{1}{2} & \frac{1}{2}\sqrt{3} & -\frac{1}{2} \end{bmatrix}, \tag{1.203}
$$

$$
\mathbf{AA}' = 3\mathbf{I_3} \Leftrightarrow (\mathbf{AA}')^{-1} = \tfrac{1}{3}\mathbf{I_3} \tag{1.204}
$$

$$
\mathbf{AA}' = \begin{bmatrix} 3 & (\mathbf{AA}')_{12} & (\mathbf{AA}')_{13} \\ (\mathbf{AA}')_{21} & 3 & (\mathbf{AA}')_{23} \\ (\mathbf{AA}')_{31} & (\mathbf{AA}')_{32} & 3 \end{bmatrix}, \tag{1.205}
$$

$$
\begin{gathered}
(\mathbf{AA}')_{12} =\\
1 + \sin\lambda_1 \sin\lambda_2 + \cos\lambda_1 \cos\lambda_2 + + \sin 2\lambda_1 \sin 2\lambda_2 + \cos 2\lambda_1 \cos 2\lambda_2\\
(\mathbf{AA}')_{13} =\\
1 + \sin\lambda_1 \sin\lambda_3 + \cos\lambda_1 \cos\lambda_3 + + \sin 2\lambda_1 \sin 2\lambda_3 + \cos 2\lambda_1 \cos 2\lambda_3\\
(\mathbf{AA}')_{21} =\\
1 + \sin\lambda_2 \sin\lambda_1 + \cos\lambda_2 \cos\lambda_1 + + \sin 2\lambda_2 \sin 2\lambda_1 + \cos 2\lambda_2 \cos 2\lambda_1\\
(\mathbf{AA}')_{23} =\\
1 + \sin\lambda_2 \sin\lambda_3 + \cos\lambda_2 \cos\lambda_3 + + \sin 2\lambda_2 \sin 2\lambda_3 + \cos 2\lambda_2 \cos 2\lambda_3\\
(\mathbf{AA}')_{31} =\\
1 + \sin\lambda_3 \sin\lambda_1 + \cos\lambda_3 \cos\lambda_1 + + \sin 2\lambda_3 \sin 2\lambda_1 + \cos 2\lambda_3 \cos 2\lambda_1\\
(\mathbf{AA}')_{32} =\\
1 + \sin\lambda_3 \sin\lambda_2 + \cos\lambda_3 \cos\lambda_2 + + \sin 2\lambda_3 \sin 2\lambda_2 + \cos 2\lambda_3 \cos 2\lambda_2
\end{gathered} \tag{1.206}
$$

$$
\text{if } \lambda_i = (i-1)\frac{2\pi}{I}, \text{ then}
$$

$$
\begin{gathered}
1 + \sin\lambda_1 \sin\lambda_2 + \cos\lambda_1 \cos\lambda_2 + \sin 2\lambda_1 \sin 2\lambda_2 + \cos 2\lambda_1 \cos 2\lambda_2 =\\
= 1 - \frac{1}{2} - \frac{1}{2} = 0\\
1 + \sin\lambda_1 \sin\lambda_3 + \cos\lambda_1 \cos\lambda_3 + \sin 2\lambda_1 \sin 2\lambda_3 + \cos 2\lambda_1 \cos 2\lambda_3 =\\
= 1 - \frac{1}{2} - \frac{1}{2} = 0\\
1 + \sin\lambda_2 \sin\lambda_3 + \cos\lambda_2 \cos\lambda_3 + \sin 2\lambda_2 \sin 2\lambda_3 + \cos 2\lambda_2 \cos 2\lambda_3 =\\
= 1 - \frac{3}{4} - \frac{1}{4} - \frac{1}{4} + \frac{1}{4} = 0
\end{gathered} \tag{1.207}
$$

$$
\begin{aligned}
L = 2: \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda_{i_1})\mathbf{e}_\ell(\lambda_{i_2}) &= 0 \forall i_1 \neq i_2, \\
L = 2: \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda_{i_1})\mathbf{e}_\ell(\lambda_{i_2}) &= L + 1 \forall i_1 = i_2,
\end{aligned} \tag{1.208}
$$

$$
\mathbf{x}_\ell = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & \frac{1}{2}\sqrt{3} & -\frac{1}{2}\sqrt{3} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2}\sqrt{3} & \frac{1}{2}\sqrt{3} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}
$$

$$
= \mathbf{x}_\ell = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}_\ell = \frac{1}{3}\begin{bmatrix} y_1 + y_2 + y_3 \\ \frac{1}{2}\sqrt{3}y_2 - \frac{1}{2}\sqrt{3}y_3 \\ y_1 - \frac{1}{2}y_2 - \frac{1}{2}y_3 \\ -\frac{1}{2}\sqrt{3}y_2 + \frac{1}{2}\sqrt{3}y_3 \\ y_1 - \frac{1}{2}y_2 - \frac{1}{2}y_3 \end{bmatrix},
$$

$$
||\mathbf{x}||^2 = \frac{1}{3}\mathbf{y}'\mathbf{y}. \tag{1.209}
$$

Lemma 1.9. (Fourier analysis).

If finite Fourier series (1.210) or (1.211) are sampled at observations points $\lambda_l \in \mathbb{S}^1$ on an equidistance lattice (equiangular lattice) (1.212), then discrete orthogonality (1.213) applies.

$$
\begin{aligned}
y_i &= y(\lambda_i) \\
&= [1, \sin\lambda_i, \cos\lambda_i, \ldots, \cos(L-1)\lambda_i, \sin L\lambda_i, \cos L\lambda_i]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{m-2} \\ x_{m-1} \\ x_m \end{bmatrix},
\end{aligned} \tag{1.210}
$$

$$
\begin{aligned}
\mathbf{y} = \mathbf{A}\mathbf{x},\ \mathbf{A} \in \mathbb{R}^{n\times m}, \mathrm{rk}\mathbf{A} = n, I = n < m = 2L+1 \\
\mathbf{A} \in \mathbb{O}(n) := \{\mathbf{A} \in \mathbb{R}^{n\times m} | \mathbf{A}\mathbf{A}' = (L+1)\mathbf{I_n}\}
\end{aligned} \tag{1.211}
$$

$$\lambda_i = (i-1)\frac{2\pi}{I}\forall i, i_1, i_2 \in \{1, \ldots, I\}, \tag{1.212}$$

$$\mathbf{A}\mathbf{A}' = (L+1)\mathbf{I_n} \sim \sum_{\ell=-L}^{+L} \mathbf{e}_\ell(\lambda_{i_1})\mathbf{e}_\ell(\lambda_{i_2}) = \left[\begin{array}{ll} 0 & \forall i_1 \neq i_2 \\ L+1 & \forall i_1 = i_2 \end{array}\right. \tag{1.213}$$

$\mathbf{A}$ is an element of the orthogonal group $O(n)$. MINOS of the underdetermined system of linear equations is

$$\mathbf{x}_m = \frac{1}{L+1}\mathbf{A}'\mathbf{y},\ \|\mathbf{x}_m\|^2 = \frac{1}{L+1}\mathbf{y}'\mathbf{y}. \tag{1.214}$$

1-32 Fourier–Legendre Series

Fourier–Legendre series represent the periodic behavior of a function $\mathbf{x}(\lambda, \phi)$ on a sphere $\mathbb{S}^2$. They are called *spherical harmonic functions* since $\{1, P_{11}(\sin\phi)\sin\lambda,$ $P_{10}(\sin\phi),\ P_{11}(\sin\phi)\cos\phi, \ldots, P_{kk}(\sin\phi)\cos k\lambda\}$ represent such a periodic signal. Indeed they are a pelicular combination of Fourier's trigonometric polynomials $\{\sin\ell\lambda, \cos\ell\lambda\}$ and Ferrer's associated Legendre polynomials $P_{k\ell}(\sin\phi)$. Here we have chosen the parameters longitude λ and latitude ϕ to locate a point on $\mathbb{S}^2$. Instead we could exchange the parameter λ by time t, if clock readings would submit longitude, a conventional technique in classical navigation. In such a setting,

$$\lambda = \omega t = \frac{2\pi}{T}t = 2\pi\nu t, \ell\lambda = \ell\omega t = 2\pi\ell\frac{t}{T} = 2\pi\ell\nu t. \tag{1.215}$$

In such a setting, longitude λ would be exchanged by 2π multiplied with the time t and divided by the ground period T or by the product of 2π, the ground frequency ν, and the time t. In contrast, $\ell\lambda$ for all $\ell \in \{-k, -k+1, \ldots, -1, 0, 1, \ldots, k-1, k\}$ would be substituted by 2π the product of overtones ℓ/T or $\ell\nu$ and time t, where k, ℓ are integer, i.e., $k, \ell \in \mathbb{Z}$. According to classical navigation, ω would represent the rotational speed of the Earth. Let us refer to *Box 1.19* as a summary of the Fourier–Legendre representation of a function $\mathbf{x}(\lambda, \phi)$, where $\lambda \in [0, 2\pi]$ and $\phi \in [-\pi/2, +\pi/2]$.

Box 1.19. (Fourier–Legendre series).

$$\begin{aligned}\mathbf{x}(\lambda, \phi) = {} & P_{00}(\sin\phi)x_1 \\ & + P_{1-1}(\sin\phi)\sin\lambda x_2 + P_{10}(\sin\phi)x_3 + P_{1+1}(\sin\phi)\cos\lambda x_4 \\ & + P_{2-2}(\sin\phi)\sin 2\lambda x_5 + P_{2-1}(\sin\phi)\sin\lambda x_6 \\ & + P_{20}(\sin\phi)x_7 + P_{21}(\sin\phi)\cos\lambda x_8 + P_{22}(\sin\phi)\cos 2\lambda x_9 \\ & + \mathcal{O}_3[P_{k\ell}(\sin\phi)\{\cos\ell\lambda, \sin\ell\lambda\}]\end{aligned} \tag{1.216}$$

$$\mathbf{x}(\lambda,\phi) = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{\ell=-k}^{+k} \mathbf{e}_{k\ell}(\lambda,\phi) x_{k\ell} \tag{1.217}$$

$$e_{k\ell}(\lambda,\phi) := \begin{cases} P_{k\ell}(\sin\phi)\cos\ell\lambda & \forall \ell > 0 \\ P_{k0}(\sin\phi) & \forall \ell = 0 \\ P_{k|\ell|}(\sin\phi)\sin|\ell|\lambda & \forall \ell < 0 \end{cases} \tag{1.218}$$

$$\mathbf{x}(\lambda,\phi) = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{\ell=-k}^{k} P_{k|\ell|}(\sin\phi)(c_{k\ell}\cos\ell\lambda + s_{k\ell}\sin\ell\lambda) \tag{1.219}$$

“Legendre polynomials of the first kind” (recurrence relation):

$$\left.\begin{array}{c} kP_k(t) = (2k-1)tP_{k-1}(t) - (k-1)P_{k-2}(t) \\ \textit{initial data}: P_0(t) = 1,\, P_1(t) = t \end{array}\right] \Rightarrow \tag{1.220}$$

Example:

$$2P_2(t) = 3tP_1(t) - P_0(t) = 3t^2 - 1 \Rightarrow P_2(t) = \frac{3}{2}t^2 - \frac{1}{2} \tag{1.221}$$

$$\text{if } t = \sin\phi, \text{ then } P_2(\sin\phi) = \frac{3}{2}\sin^2\phi - \frac{1}{2}$$

Ferrer’s associates Legendre polynomials of the first kind:

$$P_{k\ell}(t) := (1-t^2)^{l/2} \frac{d^\ell P_k(t)}{dt^\ell} \tag{1.222}$$

Example:

$$P_{11}(t) = \sqrt{1-t^2}\frac{d}{dt}P_1(t),\, P_{11}(t) = \sqrt{1-t^2} \tag{1.223}$$

$$\text{if } t = \sin\phi, \text{ then } P_{11}(\sin\phi) = \cos\phi.$$

Example:

$$\begin{array}{c} P_{21}(t) = \sqrt{1-t^2}\dfrac{d}{dt}P_2(t),\, P_{21}(t) = \sqrt{1-t^2}\dfrac{d}{dt}\left(\dfrac{3}{2}t^2 - \dfrac{1}{2}\right) \\ P_{21}(t) = 3t\sqrt{1-t^2} \end{array} \tag{1.224}$$

$$\text{if } t = \sin\phi, \text{ then } P_{21}(\sin\phi) = 3\sin\phi\cos\phi.$$

Example:

$$P_{22}(t) = (1-t^2)\frac{d^2}{dt^2}P_2(t),\, P_{22}(t) = 3(1-t^2) \tag{1.225}$$

$$\text{if } t = \sin\phi, \text{ then } P_{22}(\sin\phi) = 3\cos^2\phi$$

Example (approximation of order three):

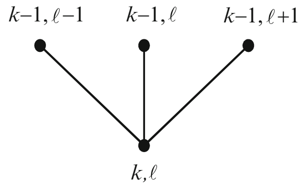


Fig. 1.11 A formal representation of the vertical recurrence relation

$$\begin{aligned}&\mathbf{x}\hat{\ }(\lambda,\phi)\\ &= \mathbf{e}_{00}x_1 + \mathbf{e}_{1-1}x_2 + \mathbf{e}_{10}x_3 + \mathbf{e}_{11}x_4\\ &+\mathbf{e}_{2-2}x_5 + \mathbf{e}_{2-1}x_6 + \mathbf{e}_{20}x_7 + \mathbf{e}_{21}x_8 + \mathbf{e}_{22}x_9 + O_3\end{aligned} \tag{1.226}$$

$$P_{k\ell}(\sin\phi) = \sin\phi P_{k-1,\ell}(\sin\phi) - \cos\phi\,[P_{k-1,\ell+1}(\sin\phi) - P_{k-1,\ell-1}(\sin\phi)]$$
(vertical recurrence relation)
$$P_{00}(\sin\phi) = 1,\, P_{k|\ell|} = P_{k-\ell}\forall\ell < 0 \tag{1.227}$$
(initial data).
Compare with Fig. 1.11.

We note that the Fourier–Legendre series (1.216) can be understood as an infinite-dimensional vector space (linear space, Hilbert space) since the base functions $\mathbf{e}_{k\ell}(\lambda,\phi)$ generate a complete orthogonal (orthonormal) system based on surface spherical functions. The countable base, i.e., the base functions $\mathbf{e}_{k\ell}(\lambda,\phi)$ or $\{1, \cos\phi\sin\lambda,\ \sin\phi,\ \cos\phi\cos\lambda, \ldots, P_{k|\ell|}(\sin\phi)\sin\ell\lambda\}$, span the *Fourier–Legendre space* $L^2\{[0, 2\pi[\times] - \pi/2, +\pi/2\}$. According to our order $\hat{\mathbf{x}}(\lambda,\phi)$ is an approximation of the function $\mathbf{x}(\lambda,\phi)$ up to order $P_{k\ell}(\sin\phi)\,\{\cos\ell\lambda, \sin|l|\,\lambda\}$ for all $\ell > 0, \ell = 0$ and $\ell < 0$, respectively. Let us refer to Box 1.20 as a summary of the essential features of the Fourier–Legendre space.

Box 1.20. (The Fourier–Legendre space)

The base functions $\mathbf{e}_{k\ell}(\lambda,\phi)$, $k \in \{1,\ldots,K\}$, $\ell \in \{-K, -K+1, \ldots, -1, 0, 1, \ldots, K-1, K\}$ span the Fourier–Legendre space $L^2\{[0, 2\pi]\times] - \pi/2, +\pi/2\}$. The base function generate a complete orthogonal (orthonormal) system of surface spherical functions.

Inner product
($\mathbf{x} \in$ *FOURIER–LEGENDRE* and $\mathbf{y} \in$ *FOURIER–LEGENDRE*)

$$\begin{aligned}<\mathbf{x}|\mathbf{y}> &= \frac{1}{S}\int dS\mathbf{x}(\lambda,\phi)_T\mathbf{y}(\lambda,\phi)\\ &= \frac{1}{4\pi}\int_0^{2\pi} d\lambda \int_{-\pi/2}^{+\pi/2} d\phi\,\cos\phi\mathbf{x}(\lambda,\phi)\mathbf{y}(\lambda,\phi).\end{aligned} \tag{1.228}$$

$\boxed{\text{Normalization:}}$

$$
\begin{aligned}
&< \mathbf{e}_{k_1\ell_1}(\lambda,\phi)|\mathbf{e}_{k_2\ell_2} > (\lambda,\phi) > \\
&= \tfrac{1}{4\pi}\int_0^{2\pi} d\lambda \int\limits_{-\pi/2}^{+\pi/2} d\phi \cos\phi\, \mathbf{e}_{k_1\ell_1}(\lambda,\phi)\mathbf{e}_{k_2\ell_2}(\lambda,\phi) \\
&= \lambda_{k_1\ell_1}\delta_{k_1k_2}\delta_{\ell_1\ell_2} \forall k_1,k_2 \in \{0,\ldots,K\}, \ell_1,\ell_2 \in \{-k,\ldots,+k\},
\end{aligned}
\tag{1.229}
$$

$$
\lambda_{k_1\ell_1} = \frac{1}{2k_1+1}\frac{(k_1-|\ell_1|)!}{(k_1+\ell_1)!}.
$$

$\boxed{\text{Norms, convergence:}}$

$$
\begin{aligned}
||\mathbf{x}||^2 &= \frac{1}{4\pi}\int\limits_0^{2\pi} d\lambda \int\limits_{-\pi/2}^{+\pi/2} d\phi \cos\phi\, \mathbf{x}^2(\lambda,\phi) \\
&= \lim_{K\to\infty}\sum_{k=0}^{K}\sum_{\ell=-k}^{+k} \lambda_{k\ell}\mathbf{x}_{k\ell}^2 < \infty,
\end{aligned}
\tag{1.230}
$$

$$
\lim_{K\to\infty} ||\mathbf{x}-\mathbf{x}_K^{\wedge}||^2 = 0 \text{ (convergence in the mean).} \tag{1.231}
$$

$\boxed{\text{Synthesis versus analysis:}}$

$$
\begin{aligned}
\mathbf{x} &= \lim_{K\to\infty}\sum_{k=0}^{K}\sum_{\ell=-k}^{+k} \mathbf{e}_{k\ell}x_{k\ell} \text{ versus } x_{k\ell} = \frac{1}{\lambda} < \mathbf{e}_{k\ell}|\mathbf{x} > \\
&:= \frac{1}{4\pi\lambda_{k\ell}}\int\limits_0^{2\pi} d\lambda \int\limits_{-\pi/2}^{+\pi/2} d\phi \cos\phi\, \mathbf{e}_{k\ell}(\lambda,\phi)\mathbf{x}(\lambda,\phi),
\end{aligned}
\tag{1.232}
$$

$$
\mathbf{x} = \lim_{K\to\infty}\sum_{k=0}^{K}\sum_{\ell=-k}^{+k} \frac{1}{\lambda_{k\ell}}\mathbf{e}_{k\ell} < \mathbf{x}|\mathbf{e}_{k\ell} > \tag{1.233}
$$

$\boxed{\text{Canonical basis of the Hilbert space : Fourier–Legendre series}}$

$$
\mathbf{e}^*_{k\ell}(\lambda,\phi) := \sqrt{2k+1}\sqrt{\frac{(k+|\ell|)!}{(k-|\ell|)!}}P_{k\ell}(\sin\phi)\left[\begin{array}{ll} \sqrt{2}\cos\ell\lambda & \forall\ell>0 \\ 1 & \forall\ell=0 \\ \sqrt{2}\sin|\ell|\,\lambda & \forall\ell<0 \end{array}\right. \tag{1.234}
$$

(orthonormal basis),

$$
\begin{aligned}
\mathbf{e}^*_{k\ell} &= \frac{1}{\sqrt{\lambda_{k\ell}}}\mathbf{e}_{k\ell} \;\; \textit{versus} \;\; \mathbf{e}_{k\ell} = \sqrt{\lambda_{k\ell}}\mathbf{e}^*_{k\ell}, \\
x^*_{k\ell} &= \sqrt{\lambda_{k\ell}}x_{k\ell} \;\; \textit{versus} \;\; x_{k\ell} = \frac{1}{\sqrt{\lambda_{k\ell}}}x^*_{k\ell},
\end{aligned}
\tag{1.235}
$$

$$\mathbf{x} = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{\ell=-k}^{+k} \mathbf{e}^*_{k\ell} \; < \mathbf{x}|.\, \mathbf{e}^*_{k\ell} > \tag{1.236}$$

$$\boxed{\text{Orthonormality:}}$$

$$< \mathbf{e}^*_{k\ell}(\lambda,\phi)|\mathbf{e}^*_{k\ell}(\lambda,\phi) > = \delta_{k_1k_2}\delta_{\ell_1\ell_2}. \tag{1.237}$$

Fourier–Legendre space
$K \to \infty$:

"*FOURIER – LEGENDRE*"

$$= span[\mathbf{e}_{K,-L}, \mathbf{e}_{K,-L+1}, \ldots, \mathbf{e}_{K,-1}, \mathbf{e}_{K,0}, \mathbf{e}_{K,1}, \ldots, \mathbf{e}_{K,L-1}, \mathbf{e}_{K,L}],$$

$$\dim \text{“}\textit{FOURIER v LEGENDRE}\text{”} \tag{1.238}$$

$$= \lim_{K\to\infty} (K+1)^2$$
$$= \infty,$$

$$\boxed{\text{“FOURIER – LEGENDRE” = “HARM”}_{L^2(\mathbb{S}^2)}.}$$

We note that an infinite-dimensional vector space (linear space) is similar to a finite-dimensional vector space: As in an Euclidean space an inner product and a norm is defined. While the inner product and the norm in a finite-dimensional vector space required summation of their components, the inner product (1.228) and (1.229) and the norm (1.230) in an infinite-dimensional vector space force us to integration. Indeed the inner products (scalar products) (1.228) and (1.229) are integrals over the surface element of $\mathbb{S}^2_r$ applied to the vectors $\mathbf{x}(\lambda,\phi),$, $\mathbf{y}(\lambda,\phi)$ or $\mathbf{e}_{k_1\ell_1}$, $\mathbf{e}_{k_2\ell_2}$ respectively. Those integrals are divided by the size of the surface element 4π of $\mathbb{S}^2_r$. Alternative representations of $< \mathbf{x}, \mathbf{y} >$ and $<\mathbf{e}_{k_1\ell_1}, \mathbf{e}_{k_2\ell_2}>$ (Dirac's notation of a bracket decomposed into "bra" and "txt") based upon $dS = rd\lambda d\phi \cos\phi$, and $S = 4\pi r^2$, lead us directly to the integration over $\mathbb{S}^2_r$, the unit sphere.

Let us adopt the definitions of Fourier–Legendre analysis as well as Fourier–Legendre synthesis following (1.228)–(1.237). Here we concentrate on the key problem: What is a harmonic function which has the sphere as a support? A harmonic function "on the unit sphere $\mathbb{S}^2_r$"is a function $\mathbf{x}(\lambda,\phi)$, $(\lambda,\phi) \in [0, 2\pi[\, \times\,] - \pi/2, +\pi/2$ which fulfils (i) the two-dimensional Laplace equation (the differential equation of a two-dimensional harmonic osculator) and (ii) a special Sturm–Liouville boundary condition (1.239) plus the harmonicity condition (1.240) for Δ_k. Note that the special Strum–Liouville equation force the frequency to be integer!

(1st)

$$\Delta_{k|\ell|}\mathbf{x}(\lambda,\phi) = 0 \Leftrightarrow \left(\frac{d^2}{d\lambda^2} + \omega\right)\mathbf{x}(\lambda) = 0. \tag{1.239}$$

(2nd)

$$\mathbf{x}(0) = \mathbf{x}(2\pi), \quad \left[\frac{d}{d\lambda}\mathbf{x}(\lambda)\right](0) = \left[\frac{d}{d\lambda}\mathbf{x}(\lambda)\right](2\pi). \tag{1.240}$$

How can we setup a linear model for Fourier–Legendre analysis? The linear model of Fourier–Legendre analysis which relates the elements $\mathbf{x} \in \mathbb{X}$ of the parameter space $\mathbb{X}$ to the elements $\mathbf{y} \in \mathbb{Y}$ of the observations space $\mathbb{Y}$ is again setup in *Box 1.21*. Here we shall assume that the observed data have been made available on a special grid which extents to $\lambda \in [0, 2\pi], \phi \in [-\pi/2, +\pi/2]$ and $I = 2J$, $\phi_j \forall j \in \{1, \ldots, I\}$, $\lambda_i = (i-1)\frac{2\pi I}{\forall} i \in \{1, \ldots, I\}$. Longitudinal interval: $\Delta\lambda =: \lambda_{i+1} - \lambda_i = \frac{2\pi}{I}$. Lateral interval: $J(even) : \Delta\phi =: \phi_{j+i} - \phi_j = \frac{\pi}{J}$, $J(odd) : \Delta\phi =: \phi_{j+1} - \phi_j = \frac{\pi}{J+1}$. In addition, we shall review the data sets fix J even as well as for J odd. Examples are given for (i) $J = 1, I = 2$ and (ii) $J = 2, I = 4$. The number of observations which correspond to these data sets have been (i) $n = 18$ and (ii) $n = 32$.

Box 1.21. The Fourier–Legendre analysis as an underdetermined linear model. The observation space $\mathbb{Y}$

Equidistant lattice on $\mathbb{S}^2$ (equiangular):

$$\lambda \in [0, 2\pi], \phi \in \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right], \tag{1.241}$$

$$I = 2J \begin{cases} \lambda_i = (i-1)\frac{2\pi}{I} & \forall i \in \{1, \ldots, I\}, \\ \phi_j & \forall j \in \{1, \ldots, I\}, \end{cases} \tag{1.242}$$

$$\begin{aligned} &J(even) \begin{cases} \phi_k = \dfrac{\pi}{J} + (k-1)\dfrac{\pi}{J} & \forall k \in \left\{1, \ldots, \dfrac{J}{2}\right\}, \\ \phi_k = -\dfrac{\pi}{J} - (k-1)\dfrac{\pi}{J} & \forall k \in \left\{\dfrac{J+2}{2}, \ldots, J\right\}, \end{cases} \\ &J(odd) \begin{cases} \phi_k = (k-1)\dfrac{\pi}{J+1} & \forall k \in \left\{1, \ldots, \dfrac{J+1}{2}\right\} \\ \phi_k = -(k-1)\dfrac{\pi}{J+1} & \forall k \in \{\dfrac{J+3}{2}, \ldots, J\} \end{cases} \end{aligned} \tag{1.243}$$

$\boxed{\text{longitudinal interval:}}$

$$\Delta\lambda := \lambda_{i+1} - \lambda_i = \frac{2\pi}{I} \tag{1.244}$$

$\boxed{\text{lateral interval:}}$

$$\begin{cases} J\,even{:}\Delta\phi := \phi_{j+1} - \phi_j = \dfrac{\pi}{J} \\ J\,odd{:}\Delta\phi := \phi_{j+1} - \phi_j = \dfrac{\pi}{J+1} \end{cases} \tag{1.245}$$

Initiation: choose J, derive $I = 2J$

$$J(even)\left\{\left[\begin{array}{l}\phi_k = \dfrac{\Delta\phi}{2} + (k-1)\Delta\phi \quad \forall k \in \left\{1,\ldots,\dfrac{J}{2}\right\}\\[2ex] \phi_k = -\dfrac{\Delta\phi}{2} - (k-1)\Delta\phi \;\forall k \in \left\{\dfrac{J+2}{2},\ldots,J\right\}\end{array}\right.\right.$$
$$J(odd)\left\{\begin{array}{l}\phi_k = (k-1)\Delta\phi \quad \forall k \in \left\{1,\ldots,\dfrac{J+1}{2}\right\}\\[2ex] \phi_k = -(k-1)\Delta\phi \;\forall k \in \left\{\dfrac{J+3}{2},\ldots,J\right\}\end{array}\right. \tag{1.246}$$

$$\lambda_i = (i-1)\Delta\lambda \forall i \in \{1,\ldots,I\}, I = 2J. \tag{1.247}$$

$$\boxed{\text{Multivariate setup of the observation space:}}$$

$$y_{ij} = x(\lambda_i, \phi_j) \tag{1.248}$$

$$\boxed{\text{Vectorizations of the matrix of observations:}}$$

Example ($J = 1, I = 2$)

Sample points	Observation vector $\mathbf{y}$
(λ_1,ϕ_1) (λ_2,ϕ_1)	$\begin{bmatrix}(\lambda_1,\phi_1)\\(\lambda_2,\phi_1)\end{bmatrix} == \mathbf{y} \in \mathbb{R}^{8x1}$

Example ($J = 2, I = 4$)

Sample points	Observation vector $\mathbf{y}$
(λ_1,ϕ_1) (λ_2,ϕ_1) (λ_3,ϕ_1) (λ_4,ϕ_1) (λ_1,ϕ_2) (λ_2,ϕ_2) (λ_3,ϕ_2) (λ_4,ϕ_2)	$\begin{bmatrix}(\lambda_1,\phi_1)\\(\lambda_2,\phi_1)\\(\lambda_3,\phi_1)\\(\lambda_4,\phi_1)\\(\lambda_1,\phi_2)\\(\lambda_2,\phi_2)\\(\lambda_3,\phi_2)\\(\lambda_4,\phi_2)\end{bmatrix} == \mathbf{y} \in \mathbb{R}^{8x1}$

Number of observations: $n = IJ = 2J^2$

$$\boxed{\text{Example:}}$$

$$\begin{aligned} J = 1 &\Rightarrow n = 2,\\ J = 3 &\Rightarrow n = 18,\\ J = 2 &\Rightarrow n = 8,\\ J = 4 &\Rightarrow n = 32. \end{aligned} \tag{1.249}$$

Table 1.1. Equidistant lattice on $\mathbb{S}^2$. The lateral lattice

J	$\Delta\phi$	lateral grid									
		1	2	3	4	5	6	7	8	9	10
1	-	0°									
2	90°	45°	-45°								
3	45°	0°	45°	-45°							
4	45°	22.5°	67.5°	-22.5°	-67.5°						
5	30°	0°	30°	60°	-30°	-60°					
6	30°	15°	45°	75°	-15°	-45°	-75°				
7	22.5°	0°	22.5°	45°	67.5°	-22.5°	-45°	-67.5°			
8	22.5°	11.25°	33.75°	56.25°	78.75°	-11.25°	-33.75°	-56.25°	-78.75°		
9	18°	0°	18°	36°	54°	72°	-18°	-36°	-54°	-72°	
10	18°	90°	27°	45°	63°	81°	-9°	-27°	-45°	-63°	-81°

Table 1.2. Equidistant lattice on $\mathbb{S}^2$. The longitudinal lattice

J	$I = 2J$	$\Delta\lambda$	Longitudinal grid									
			1	2	3	4	5	6	7	8	9	10
1	2	180°	0°	180°								
2	4	90°	0°	90°	180°	270°						
3	6	60°	0°	60°	120°	180°	240°	300°				
4	8	45°	0°	45°	90°	135°	180°	225°	270°	315°		
5	10	36°	0°	36°	72°	108°	144°	180°	216°	252°	288°	324°

For the optimal design of the Fourier–Legendre linear model it has been shown that the equidistant lattice (1.242) and (1.243) is "D-optimal". Tables 1.1 and 1.2 contain samples of an equidistant lattice on $\mathbb{S}$, especially in a lateral and a longitudinal lattice. As samples, via Fig. 1.12, we have computed various horizontal and vertical sections of spherical lattices, for instants, $\{J = 1, I = 2\}$, $\{J = 2, I = 4\}$, $\{J = 3, I = 6\}$, $\{J = 4, I = 8\}$, and $\{J = 5, I = 10\}$. By means of Fig. 1.13, we have added the corresponding Platt–Carré Maps. The unknown Fourier–Legendre coefficients, collected in a Pascal triangular graph of Fig. 1.12 are vectorized by (1.250). Indeed, the linear model (1.175) contains $m = IJ = 2J^2$ observations and $m \to \infty$ unknowns, a hyperreal number. The linear operator $A : \mathbb{X} \to \mathbb{Y}$ is generated by the base functions of lattice points (1.251). Equation 1.251 is a representation of the linear observational equations (1.175) in Ricci calculus which is characteristic for Fourier–Legendre analysis.

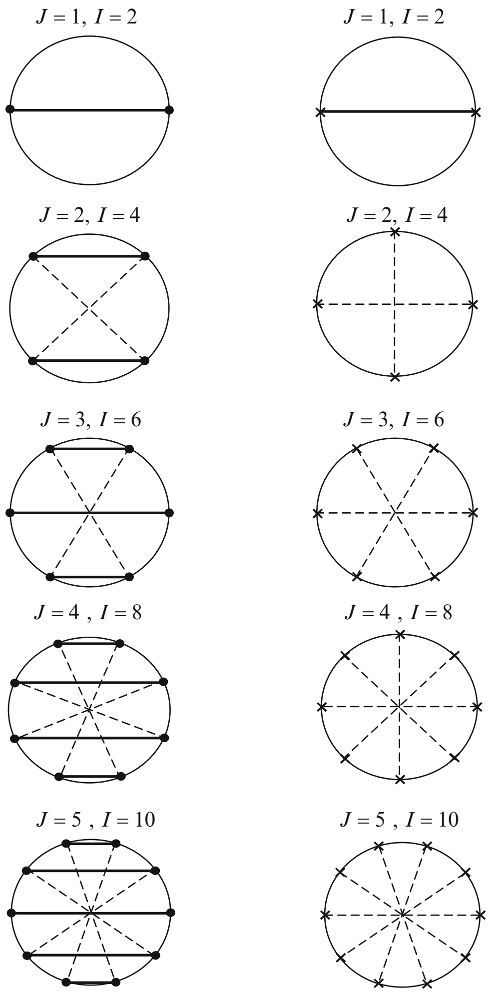
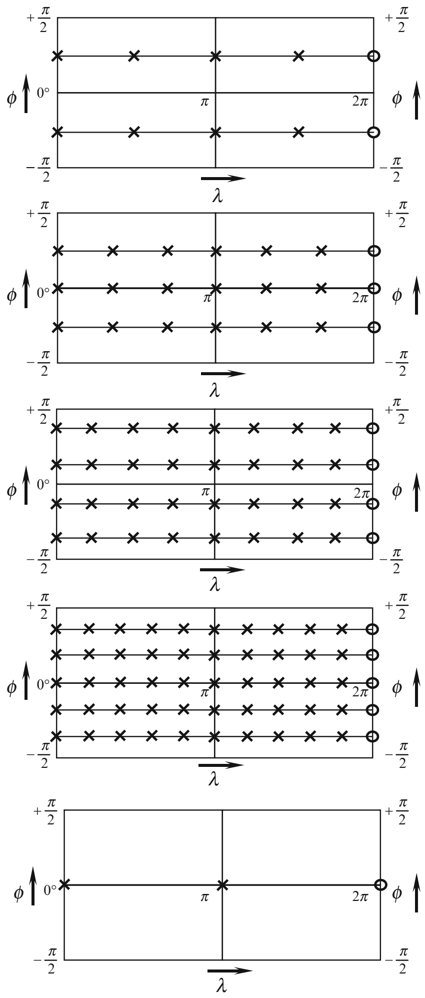


Fig. 1.12 The spherical lattice. *Left*: vertical section, trace of parallel circles. *Right row*: horizontal section, trace of meridians

$$
\begin{aligned}
&X = \operatorname{span}\{x_{00}, x_{1-1}, x_{10}, x_{11}, \ldots, x_{k|\ell|}\} \\
&k \to \infty, k = 0 \ldots k, |\ell| = 0 \ldots k \qquad (1.250) \\
&dim\mathbb{X} = m = \infty,
\end{aligned}
$$

$$
j_{ij} = y(x_{ij}) = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{\ell=-k}^{+k} \mathbf{e}_{k\ell}(\lambda_i, \phi_j)\mathbf{x}_{k\ell} \forall i, j \in \{1, \ldots, n\}. \qquad (1.251)
$$

Fig. 1.13 Platt–Carré Map of longitude-latitude lattice



Note that the finite dimensional observation space $\mathbb{Y}$ has integer dimension, i.e. $\dim \mathbb{Y} = n = IJ$, $I = 2J$ and the parameter space $\mathbb{X}$ is infinite dimensional, i.e. $\dim \mathbb{X} = m = 1$. The portray of Fourier–Legendre analysis which for the readers' convenience is outlined in *Box 1.22* summarizes effectively its peculiarities. A finite number of observations is confronted with an infinite number of observations. Such a linear model of type "underdetermined of power 2" cannot be solved in finite computer time. Instead, one has to truncate the Fourier–Legendre series, leaving the series "bandlimited". We consider the three cases shown in *Box 1.22*.

Box 1.22. (The portray of Fourier–Legendre analysis)

Number of observed data at lattice points	versus	Number of unknown Fourier–Legendre coefficients:
$n = IJ = 2J^2$		$m = \lim_{K\to\infty} \sum_{k=0}^{K} \sum_{\ell=-k}^{+k} \mathbf{e}_{k\ell}$
(finite).		(infinite).
	The three cases:	
$n > m$	$n = m$	$n < m$
(overdetermined case),	(regular cases),	(underdetermined case).

First, we have to truncate the. infinite Fourier–Legendre series that $n > m$ hold. In this case of an overdetermined problem, we have more observations than equations. Second, we alternatively balance the number of unknown Fourier–Legendre coefficients such that $n = m$ holds. Such a model choice assures a regular linear system. Both linear Fourier–Legendre models which are tuned to the number of observations suffer from a typical uncertainty. What is the effect of the forgotten unknown Fourier–Legendre coefficients $m > n$? Indeed a significance test has to decide upon any truncation to be admissible. We need an objective criterion to decide upon the degree m of bandlimit. Third, in order to be as objective as possible we again follow the third case of "less observations than unknowns" such that $n < m$ holds. Such a Fourier–Legendre linear model generating an underdetermined system of linear equations will consequently be considered in *Boxes 1.23* and *1.24*. Note that the first example that is presented in *Box 1.23* demonstrates MINOS of the *Fourier–Legendre linear model* for $n = IJ = 2J^2 = 2$ and $k = 1, m = (k+1)^2 = 4$ as observations and unknowns. Furthermore, note that the second example presented in *Box 1.24* refers to MINOS of the Fourier–Legendre linear model for $n = 8$ and $m = 9$. In the framework of the second example, we have computed the design matrix $\mathbf{A} \in \mathbb{R}^{8\times 9}$.

Box 1.23. (Example)

Fourier–Legendre analysis as an underdetermined linear model:

$$m - \mathrm{rk}\mathbf{A} = m - n = 2,$$

$$\mathtt{dim}\ \mathbb{Y} = n = 2 \ \textit{versus}\ \mathtt{dim}\ \mathbb{X} = m = 4,$$

$$J = 1, I = 2J = 2 \Rightarrow n = IJ = 2J^2 = 2 \ \textit{versus}\ K = 1 \Rightarrow m = (k+1)^2 = 4$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & P_{11}(\sin\phi_1)\sin\lambda_1 & P_{10}(\sin\phi_1) & P_{11}(\sin\phi_1)\cos\lambda_1 \\ 1 & P_{11}(\sin\phi_2)\sin\lambda_2 & P_{10}(\sin\phi_2) & P_{11}(\sin\phi_2)\cos\lambda_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (1.252)$$

subject to $(\lambda_1, \phi_1) = (0°, 0°)$ and $(\lambda_2, \phi_2) = (180°, 0°)$,

$$\{\mathbf{y} = \mathbf{A}\mathbf{x} \mid \mathbf{A} \in \mathbb{R}^{2\times 4}, rk\mathbf{A} = n = 2, m = 4, n = m = 2\}, \tag{1.253}$$

$$\mathbf{A} := \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & P_{11}(\sin\phi_1)\sin\lambda_1 & P_{10}(\sin\phi_1) & P_{11}(\sin\phi_1)\cos\lambda_1 \\ 1 & P_{11}(\sin\phi_2)\sin\lambda_2 & P_{10}(\sin\phi_2) & P_{11}(\sin\phi_2)\cos\lambda_2 \end{bmatrix}, \tag{1.254}$$

$$\begin{aligned} P_{11}(\sin\phi) &= \cos\phi, \\ P_{10}(\sin\phi) &= \sin\phi, \\ P_{11}(\sin\phi_1) &= P_{11}(\sin\phi_2) = 1, \\ P_{10}(\sin\phi_1) &= P_{10}(\sin\phi_2) = 0, \end{aligned} \tag{1.255}$$

$$\sin\lambda_1 = \sin\lambda_2 = 0,\ \cos\lambda_1 = 1,\ \cos\lambda_2 = -1, \tag{1.256}$$

$$\mathbf{A}\mathbf{A}' = \begin{bmatrix} (AA')_{11} & (AA')_{12} \\ (AA')_{21} & (AA')_{22} \end{bmatrix} \tag{1.257}$$

$$\begin{aligned} (AA')_{11} &= 1 + P_{11}^2(\sin\phi_1) + P_{10}^2(\sin\phi_1), \\ (AA')_{12} &= 1 + P_{11}(\sin\phi_1)\, P_{11}(\sin\phi_2)\sin\lambda_1\sin\lambda_2 \\ &\quad + P_{10}(\sin\phi_1)\, P_{10}(\sin\phi_2) \\ &\quad + P_{11}(\sin\phi_1)\, P_{11}(\sin\phi_2)\cos\lambda_1\cos\lambda_2 \\ (AA')_{21} &= 1 + P_{11}(\sin\phi_2)\, P_{11}(\sin\phi_1)\sin\lambda_2\sin\lambda_1 \\ &\quad + P_{10}(\sin\phi_2)\, P_{10}(\sin\phi_2) \\ &\quad + P_{11}(\sin\phi_2)\, P_{11}(\sin\phi_1)\cos\lambda_2\cos\lambda_1 \\ (AA')_{22} &= 1 + P_{11}^2(\sin\phi_2) + P_{10}^2(\sin\phi_2), \end{aligned} \tag{1.258}$$

$$\boxed{\mathbf{A}\mathbf{A}' = 2\mathbf{I}_2 \Leftrightarrow (\mathbf{A}\mathbf{A}')^{-1} = \frac{1}{2}\mathbf{I}_2,} \tag{1.259}$$

$$\begin{aligned} &\sum_{k_1,k_2=0}^{K=1}\ \sum_{\ell_1=-k_1,\ell_2=-k_2}^{+k_1,+k_2} \mathbf{e}_{k_1\ell_1}(\lambda_{i_1}, \phi_{i_1})\mathbf{e}_{k_2\ell_2}(\lambda_{i_2}, \phi_{i_2}) = 0, \forall i_1 \neq i_2, \\ &\sum_{k_1,k_2=0}^{K=1}\ \sum_{\ell_1=-k_1,\ell_2=-k_2}^{+k_1,+k_2} \mathbf{e}_{k_1\ell_1}(\lambda_{i_1}, \phi_{i_1})\mathbf{e}_{k_2\ell_2}(\lambda_{i_2}, \phi_{i_2}) = 2, \forall i_1 = i_2, \end{aligned} \tag{1.260}$$

$$x_\ell = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} (MINOS) = \begin{bmatrix} c_{00} \\ s_{11} \\ c_{10} \\ c_{11} \end{bmatrix} (MINOS) = \frac{1}{2}\mathbf{A}'\mathbf{y}$$
$$= \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} y_1 + y_2 \\ 0 \\ 0 \\ y_1 - y_2 \end{bmatrix}. \tag{1.261}$$

Fourier–Legendre analysis as an underdetermined linear model:

$$m - \mathrm{rk}\mathbf{A} = m - n = 1,$$

$$\mathtt{dim}\ \mathbb{Y} = n = 8 \textit{ versus } \mathtt{dim}\ \mathbb{X} = m = 9,$$

$$J = 2, I = 2J = 4 \Rightarrow n = IJ = 2J^2 = 8\, versus\, k = 2 \Rightarrow m = (k+1)^2 = 9,$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & P_{11}(\sin\phi_1)\sin\lambda_1 & P_{10}(\sin\phi_1) & P_{11}(\sin\phi_1)\cos\lambda_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & P_{11}(\sin\phi_8)\sin\lambda_8 & P_{10}(\sin\phi_8) & P_{11}(\sin\phi_8)\cos\lambda_8 \end{bmatrix} \cdots$$
$$\cdots \begin{matrix} P_{22}(\sin\phi_1)\sin 2\lambda_1 & P_{21}(\sin\phi_1)\sin\lambda_1 & P_{20}(\sin\phi_1) \\ \vdots & \vdots & \vdots \\ P_{22}(\sin\phi_8)\sin 2\lambda_8 & P_{21}(\sin\phi_8)\sin\lambda_8 & P_{20}(\sin\phi_8) \end{matrix} \cdots \tag{1.262}$$
$$\cdots \begin{matrix} P_{21}(\sin\phi_1)\cos\lambda_1 & P_{22}(\sin\phi_1)\cos 2\lambda_1 \\ \vdots & \vdots \\ P_{21}(\sin\phi_8)\cos\lambda_8 & P_{22}(\sin\phi_8)\cos 2\lambda_8 \end{matrix} \Bigg] \begin{bmatrix} x_1 \\ \vdots \\ x_9 \end{bmatrix}.$$

Equidistant lattice (longitudinal width $\Delta\lambda$, lateral width $\Delta\phi$ "
$\Delta\lambda = 90°, \Delta\phi = 90°$):

$$\begin{aligned} (\lambda_1, \phi_1) &= (0°, +45°), \\ (\lambda_2, \phi_2) &= (90°, +45°), \\ (\lambda_3, \phi_3) &= (180°, +45°), \\ (\lambda_4, \phi_4) &= (270°, +45°), \\ (\lambda_5, \phi_5) &= (0°, -45°), \\ (\lambda_6, \phi_6) &= (90°, -45°), \\ (\lambda_7, \phi_7) &= (180°, -45°), \\ (\lambda_8, \phi_8) &= (270°, -45°), \end{aligned} \tag{1.263}$$

$$\begin{gathered} P_{11}(\sin\phi) = \cos\phi, P_{10}(\sin\phi) = \sin\phi, \\ P_{22}(\sin\phi) = 3\cos^2\phi, P_{21}(\sin\phi) = 3\sin\phi\cos\phi, \\ P_{20}(\sin\phi) = (3/2)\sin^2\phi - (1/2), \end{gathered} \tag{1.264}$$

Table 1.3 (Number of observed data at lattice points versus the number of Fourier–Legendre Coefficients)

Finite	*Infinite*
Number of observed data at lattice	Number of unknown Fourier–Legendre Coefficients
$n = IJ = 2J^2$	$m = \lim_{K\to\infty} \sum_{K=0}^{K} (2K+1) =$
	$= \lim_{K\to\infty} (K+1)^2$
	finite
$J = 1 \Longrightarrow n = 2$	MINOS:$K = 1, (K+1)^2 = 4, m - n = 4 - 2 = 2$
	LESS:$K = 0, (K+1)^2 = 1, m - n = 2 - 1 = 1$
$J = 2 \Longrightarrow n = 8$	MINOS:$K = 2, (K+1)^2 = 9, m - n = 1$
	LESS:$K = 1, (K+1)^2 = 4, m - n = 4$
$J = 3 \Longrightarrow n = 18$	MINOS:$K = 4, (K+1)^2 = 25, m - n = 17$
	LESS:$K = 2, (K+1)^2 = 9, m - n = 9$
	$K = 3, (K+1)^2 = 16$
$J = 4 \Longrightarrow n = 32$	MINOS:$K = 5, (K+1)^2 = 36, m - n = 4$
	$K = 3, (K+1)^2 = 16, n - m = 16$
	$K = 4, (K+1)^2 = 25$

1-33 Nyquist Frequency for Spherical Data

We repeat our introduction of a regular grid analysis for data on the unit sphere with identical intervals in longitude and latitude. See our *Box 1.21* which is enlarged by Table 1.3 generalizing to MINOS and LESS (**LE**ast **S**quares **S**olution) Table 1.4 lists the various Nyguist frequencies for data on the unit sphere depending on

(i) The number of observations
(ii) The number of unknown parameters
(iii) The Nyguist frequencies $N_f = \pi/\Delta\phi = \pi/\Delta\lambda$.

For instance the number of estimable parameters for a $5' \times 5'$ *regular grid* and 9,326,916 unknowns account for $N_f = 2{,}160$ quantities. But for a $5'' \times 5''$ *regular grid* and $3,356,224 \times 10^{10}$ unknowns, the number of the estimable quantities sum up to $N_f = 129{,}600$ as the *Nyguist frequencies*.

Table 1.4 (Nyquist frequency, data on the reference sphere)

Number of observations $n = IJ = 2J^2$	Number of parameters $m = (K+1)^2$	Nyquist frequency $N_f = \pi/\Delta\phi = \pi/\Delta\lambda$ $\sqrt{n} - 1 \geq K$
$5° \times 5°$ regular grid $n = 2{,}529$	$m = 2{,}500$ $K = 49$	$N_f = 36$
$1° \times 1°$ regular grid $n = 64{,}800$	$m = 64{,}516$ $K = 253$	$N_f = 180$
$5' \times 5'$ regular grid $n = 9{,}331{,}200$	$m = 9{,}326{,}916$ $K = 3{,}053$	$N_f = 2{,}160$
$5'' \times 5''$ regular grid $n = 3{,}359{,}232 \times 10^{10}$	$m = 3{,}356{,}224 \times 10^{10}$ $K = 1{,}831$	$N_f = 129{,}600$
$1'' \times 1''$ regular grid $n = 8{,}398{,}08 \times 10^{11}$	$m = 8{,}398{,}072{,}88 \times 10^{11}$ $K = 916{,}409$	$N_f = 648{,}000$

1-4 Special Nonlinear Models

Let us finish this chapter by considering an example of a consistent system of linearized observational equations $\mathbf{Ax} = \mathbf{y}$ and $\mathrm{rk}\mathbf{A} = \mathrm{rk}(\mathbf{A}, \mathbf{y})$ where the matrix $\mathbf{Ax} = \mathbb{R}^{\mathbf{n}\times\mathbf{m}}$ is the Jacobi matrix (Jacobi map) of the nonlinear model. Let us present a planar triangle whose nodal points have to be coordinated from three distance measurements and the minimum norm solution of type $\mathbf{I}$-MINOS.

1-41 Taylor Polynomials, Generalized Newton Iteration

In addition we review the invariance properties of the observational equations with respect to a particular transformation group which makes the a priori indeterminism of the consistent system of linearized observational equations plausible. The observation vector is an element of the column space $\mathbf{Y} \in \mathcal{R}(\mathbf{A})$. The geometry of the planar triangle is illustrated in Fig. 1.14. The point of departure for the linearization process of nonlinear observational equations is the nonlinear mapping $\mathbf{X} \mapsto \mathbf{F}(\mathbf{X}) = \mathbf{Y}$. The Taylor expansion is truncated to the order $\mathcal{O}\left[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})\right]$,

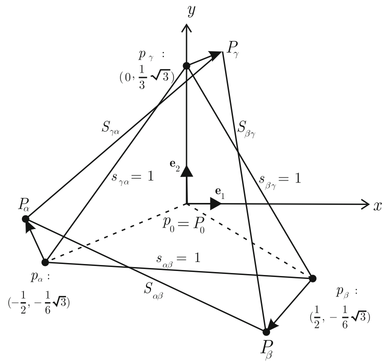


Fig. 1.14 Barycentric rectangular coordinates of the nodal points, namely of the equilateral triangle $p_\alpha p_\beta p_\gamma$.

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}) = \mathbf{F}(\mathbf{x}) + \mathbf{J}(\mathbf{x})(\mathbf{X}-\mathbf{x}) + \mathbf{H}(\mathbf{x})(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})$$
$$+\mathcal{O}\,[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})]. \tag{1.265}$$

$\mathbf{J}(\mathbf{x})$ represents the Jacobi matrix of the first partial derivatives, while $\mathbf{H}(\mathbf{x})$ is the Hesse matrix of second derivatives of the vector-valued function $\mathbf{F}(\mathbf{X})$ with respect to the coordinates of the vector $\mathbf{X}$, both taken at the evaluation point $\mathbf{x}$. A linearized nonlinear model is generated by truncating the vector-valued function $\mathbf{F}(\mathbf{x})$ to the order $\mathcal{O}\,[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})]$, namely

$$\Delta\mathbf{y} := \mathbf{F}(\mathbf{X}) - \mathbf{F}(\mathbf{x}) = \mathbf{J}(\mathbf{x})(\mathbf{X}-\mathbf{x}) + \mathcal{O}\,[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})]. \tag{1.266}$$

A generalized Newton iteration process for solving the nonlinear observational equations by solving a sequence of linear equations of (injectivity) defect by means of the right inverse of type $\mathbf{G_x}$-MINOS is the algorithm that is shown in *Box 1.24*.

Box 1.24. Newton iteration, $\mathbf{I}$-MINOS, $\mathrm{rk}\mathbf{A} = \mathrm{rk}(\mathbf{A}, \mathbf{y})$

$$\begin{gathered}\texttt{Level 0}:\\ \mathbf{x}_0 = \mathbf{x}_0,\ \Delta\mathbf{y}_0 = \mathbf{F}(\mathbf{X}) - \mathbf{F}(\mathbf{x}_0),\ \Delta\mathbf{x}_1 = [\mathbf{J}(\mathbf{x}_0)]^-_{\mathbf{R}}\,\Delta\mathbf{y}_0.\end{gathered} \tag{1.267}$$

$$\begin{gathered}\texttt{Level 1}:\\ \mathbf{x}_1 = \mathbf{x}_0 + \Delta\mathbf{x}_1\,,\ \Delta\mathbf{y}_1 = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_1),\ \Delta\mathbf{x}_2 = [\mathbf{J}(\mathbf{x}_1)]^-_{\mathbf{R}}\,\Delta\mathbf{y}_1.\end{gathered} \tag{1.268}$$

$$\begin{gathered}\texttt{Level } i:\\ \mathbf{x}_i = \mathbf{x}_{i-1} + \Delta\mathbf{x}_i\,,\ \Delta\mathbf{y}_i = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_i)\ \ \Delta\mathbf{x}_{i+1} = [\mathbf{J}(\mathbf{x}_i)]^-_{\mathbf{R}}\,\Delta\mathbf{y}_i.\end{gathered} \tag{1.269}$$

Level n (i.e. the reproducing point in the computer arithmetric): $\Delta\mathbf{x}_{n+1} = \Delta\mathbf{x}_n$.

The planar triangle $P_\alpha P_\beta P_\gamma$ is approximately an equilateral triangle $p_\alpha p_\beta p_\gamma$ whose nodal points are a priori coordinated by *Box 1.25*. Obviously, the approximate coordinates of the three nodal points are barycentric, namely characterized by *Boxes 1.27* and *1.28*. Their sum as well as their product sum vanish.

Box 1.25. (Barycentric rectangular coordinates of the equilateral triangle $p_\alpha p_\beta p_\gamma$)

$$p_\alpha = \begin{bmatrix} x_\alpha = -\dfrac{1}{2} \\ y_\alpha = -\dfrac{1}{6}\sqrt{3} \end{bmatrix} \qquad p_\beta = \begin{bmatrix} x_\beta = \dfrac{1}{2} \\ y_\beta = -\dfrac{1}{6}\sqrt{3} \end{bmatrix} \qquad p_\gamma = \begin{bmatrix} x_\gamma = 0 \\ y_\gamma = \dfrac{1}{3}\sqrt{3}. \end{bmatrix} \tag{1.270}$$

Box 1.26. (First and second moments of nodal points, approximate coordinates).

$$x_\alpha + x_\beta + x_\gamma = 0, \ \ y_\alpha + y_\beta + y_\gamma = 0, \tag{1.271}$$

$$\begin{aligned} J_{xy} &= x_\alpha y_\alpha + x_\beta y_\beta + x_\gamma y_\gamma = 0, \\ J_{xx} &= -(y_\alpha^2 + y_\beta^2 + y_\gamma^2) = -\frac{1}{2}, \ \ J_{yy} = -(x_\alpha^2 + x_\beta^2 + x_\gamma^2) = -\frac{1}{2}, \end{aligned} \tag{1.272}$$

$$[I_i] = \begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} x_\alpha + x_\beta + x_\gamma \\ y_\alpha + y_\beta + y_\gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \ \ \forall i \in \{1, 2\}, \tag{1.273}$$

$$\begin{aligned} [I_{ij}] &= \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix} \\ &= \begin{bmatrix} -(y_\alpha^2 + y_\beta^2 + y_\gamma^2) & x_\alpha y_\alpha + y_\beta x_\beta + x_\gamma y_\gamma \\ x_\alpha y_\alpha + y_\beta x_\beta + x_\gamma y_\gamma & -(x_\alpha^2 + x_\beta^2 + x_\gamma^2) \end{bmatrix} \\ &= \begin{bmatrix} -\dfrac{1}{2} & 0 \\ 0 & -\dfrac{1}{2} \end{bmatrix} = -\frac{1}{2}\mathbf{I}_2 \ \ \forall i, j \in \{1, 2\}. \end{aligned} \tag{1.274}$$

Box 1.27. (First and second moments of nodal points, inertia tensors).

$$\begin{aligned} \mathbf{I}_1 &= \sum_{i=1}^{2} \mathbf{e}^i I_i = \mathbf{e}^1 I_1 + \mathbf{e}^2 I_2, \\ I_i &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \rho(x, y) x_i \ \ \forall i \in \{1, 2\}, \end{aligned} \tag{1.275}$$

$$\begin{aligned}\mathbf{I}_2 &= \sum_{i,j=1}^{2} \mathbf{e}^i \otimes \mathbf{e}^j I_{ij} \\ &= \mathbf{e}^1 \otimes \mathbf{e}^1 I_{11} + \mathbf{e}^1 \otimes \mathbf{e}^2 I_{12} + \mathbf{e}^2 \otimes \mathbf{e}^1 I_{21} + \mathbf{e}^2 \otimes \mathbf{e}^2 I_{22}, \\ I_{ij} &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \rho(x,y)(x_i x_j - r^2 \delta_{ij}) \ \forall i,j \in \{1,2\}.\end{aligned} \tag{1.276}$$

subject to

$$r^2 = x^2 + y^2, \tag{1.277}$$

$$\rho(x,y) = \delta(x,y,x_\alpha y_\alpha) + \delta(x,y,x_\beta y_\beta) + \delta(x,y,x_\gamma y_\gamma). \tag{1.278}$$

The product sum of the approximate coordinates of the nodal points constitute the rectangular coordinates of the inertia tensor (1.279) and (1.280). The mass density distribution $\rho(x,y)$ generates directly the coordinates I_{xy}, I_{xx}, I_{yy} of the inertia tensor. ($\delta(...)$ denotes the Dirac generalized function. The nonlinear observational equations of distance measurements are generated by the Pythagoras representation presented in *Box 1.28*.

$$\mathbf{I} = \sum_{i,j=1}^{2} \mathbf{e}^i \otimes \mathbf{e}^j I_{ij} \tag{1.279}$$

$$I_{ij} = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \rho(x,y)(x_i x_j - r^2 \delta_{ij}) \ \forall i,j \in \{1,2\}, \tag{1.280}$$

$$\begin{aligned} r^2 &= x^2 + y^2, \\ \rho(x,y) &= \delta(x,y,x_\alpha y_\alpha) + \delta(x,y,x_\beta y_\beta) + \delta(x,y,x_\gamma y_\gamma). \end{aligned} \tag{1.281}$$

Box 1.28. (Nonlinear observational equations of distance measurements in the plane, geometric notation versus algebraic notation).

Taylor expansion of the nonlinear distance observational equations

$$\begin{aligned} Y_1 &= \mathrm{F}_1(\mathbf{X}) = S^2_{\alpha\beta} = (X_\beta - X_\alpha)^2 + (Y_\beta - Y_\alpha)^2 \\ Y_2 &= \mathrm{F}_2(\mathbf{X}) = S^2_{\beta\gamma} = (X_\gamma - X_\beta)^2 + (Y_\gamma - Y_\beta)^2 \\ Y_3 &= \mathrm{F}_3(\mathbf{X}) = S^2_{\gamma\alpha} = (X_\alpha - X_\gamma)^2 + (Y_\alpha - Y_\gamma)^2. \end{aligned} \tag{1.282}$$

$$\begin{aligned} \mathbf{Y}' &:= \left[S^2_{\alpha\beta}, S^2_{\beta\gamma}, S^2_{\gamma\alpha}\right], \quad \mathbf{X}' := \left[X_\alpha, Y_\alpha, X_\beta, Y_\beta, X_\gamma, Y_\gamma\right] \\ \mathbf{x}' &= \left[x_\alpha, y_\alpha, x_\beta, y_\beta, x_\gamma, y_\gamma\right] = \left[-\frac{1}{2}, -\frac{1}{6}\sqrt{3}, \frac{1}{2}, -\frac{1}{6}\sqrt{3}, 0, \frac{1}{3}\sqrt{3}\right]. \end{aligned} \tag{1.283}$$

Jacobi map:

$$
\mathbf{J}(\mathbf{x}) := \begin{bmatrix} \frac{\partial F_1}{\partial X_\alpha} & \frac{\partial F_1}{\partial Y_\alpha} & \frac{\partial F_1}{\partial X_\beta} & \frac{\partial F_1}{\partial Y_\beta} & \frac{\partial F_1}{\partial X_\gamma} & \frac{\partial F_1}{\partial Y_\gamma} \\ \frac{\partial F_2}{\partial X_\alpha} & \frac{\partial F_2}{\partial Y_\alpha} & \frac{\partial F_2}{\partial X_\beta} & \frac{\partial F_2}{\partial Y_\beta} & \frac{\partial F_2}{\partial X_\gamma} & \frac{\partial F_2}{\partial Y_\gamma} \\ \frac{\partial F_3}{\partial X_\alpha} & \frac{\partial F_3}{\partial Y_\alpha} & \frac{\partial F_3}{\partial X_\beta} & \frac{\partial F_3}{\partial Y_\beta} & \frac{\partial F_3}{\partial X_\gamma} & \frac{\partial F_3}{\partial Y_\gamma} \end{bmatrix} (\mathbf{x})
$$

$$
= \begin{bmatrix} -2(x_\beta - x_\alpha) & -2(y_\beta - y_\alpha) & 2(x_\beta - x_\alpha) & 2(y_\beta - y_\alpha) \\ 0 & 0 & -2(x_\gamma - x_\beta) & -2(y_\gamma - y_\beta) & \ldots \\ 2(x_\alpha - x_\gamma) & 2(y_\alpha - y_\gamma) & 0 & 0 \end{bmatrix} \tag{1.284}
$$

$$
\begin{bmatrix} & 0 & 0 \\ \ldots & 2(x_\gamma - x_\beta) & 2(y_\gamma - y_\beta) \\ & -2(x_\alpha - x_\gamma) & -2(y_\alpha - y_\gamma) \end{bmatrix} = \begin{bmatrix} -2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\sqrt{3} & -1 & \sqrt{3} \\ -1 & -\sqrt{3} & 0 & 0 & 1 & \sqrt{3} \end{bmatrix}.
$$

Let us analyze "observed minus computed", namely the expression (1.285), here specialized to *Box 1.29*. The sum of the final coordinates is zero. However, due to the non-symmetric displacement field $[\Delta x_\alpha,\ \Delta y_\alpha,\ \Delta x_\beta,\ \Delta y_\beta,\ \Delta x_\gamma,\ \Delta y_\gamma]'$ the coordinate J_{xy} of the inertia tensor does not vanish. These results are collected in *Box 1.30*.

$$
\begin{aligned} \Delta\mathbf{y} &:= \mathbf{F}(\mathbf{X}) - \mathbf{F}(\mathbf{x}) = \mathbf{J}(\mathbf{x})(\mathbf{X}-\mathbf{x}) + \mathcal{O}\left[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})\right] \\ &= \mathbf{J}\Delta\mathbf{x} + \mathcal{O}\left[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})\right], \end{aligned} \tag{1.285}
$$

Box 1.29. (Linearized observational equations of distance measurements in the plane, **I**-MINOS, rk**A**= dim $\mathbb{Y}$).

"Observed minus computed":

$$
\begin{aligned} \Delta\mathbf{y} &:= \mathbf{F}(\mathbf{X}) - \mathbf{F}(\mathbf{x}) = \mathbf{J}(\mathbf{x})(\mathbf{X}-\mathbf{x}) + \mathcal{O}\left[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})\right] \\ &= \mathbf{J}\Delta\mathbf{x} + \mathcal{O}\left[(\mathbf{X}-\mathbf{x}) \otimes (\mathbf{X}-\mathbf{x})\right], \end{aligned} \tag{1.286}
$$

$$
\begin{bmatrix} \Delta s^2_{\alpha\beta} \\ \Delta s^2_{\beta\gamma} \\ \Delta s^2_{\gamma\alpha} \end{bmatrix} = \begin{bmatrix} S^2_{\alpha\beta} - s^2_{\alpha\beta} \\ S^2_{\beta\gamma} - s^2_{\beta\gamma} \\ S^2_{\gamma\alpha} - s^2_{\gamma\alpha} \end{bmatrix} = \begin{bmatrix} 1.1 - 1 \\ 0.9 - 1 \\ 1.2 - 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} \\ -\frac{1}{10} \\ \frac{1}{5} \end{bmatrix}, \tag{1.287}
$$

$$
\begin{bmatrix} \Delta s^2_{\alpha\beta} \\ \Delta s^2_{\beta\gamma} \\ \Delta s^2_{\gamma\alpha} \end{bmatrix} = \begin{bmatrix} a_{\alpha\beta} & b_{\alpha\beta} & -a_{\alpha\beta} & -b_{\alpha\beta} & 0 & 0 \\ 0 & 0 & a_{\beta\gamma} & b_{\beta\gamma} & -a_{\beta\gamma} & -b_{\beta\gamma} \\ -a_{\gamma\alpha} & -b_{\gamma\alpha} & 0 & 0 & a_{\gamma\alpha} & b_{\gamma\alpha} \end{bmatrix} \begin{bmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x_\beta \\ \Delta y_\beta \\ \Delta x_\gamma \\ \Delta y_\gamma \end{bmatrix}. \tag{1.288}
$$

linearized observational equations:

$$\mathbf{A} = \begin{bmatrix} -2 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\sqrt{3} & -1 & \sqrt{3} \\ -1 & -\sqrt{3} & 0 & 0 & 1 & \sqrt{3} \end{bmatrix}, \tag{1.289}$$

$$\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} = \frac{1}{36}\begin{bmatrix} -9 & 3 & -3 \\ \sqrt{3} & \sqrt{3} & -5\sqrt{3} \\ 9 & 3 & -3 \\ \sqrt{3} & -5\sqrt{3} & \sqrt{3} \\ 0 & -6 & 6 \\ -2\sqrt{3} & 4\sqrt{3} & 4\sqrt{3} \end{bmatrix}. \tag{1.290}$$

Minimum norm solution:

$$\mathbf{x}_m = \begin{bmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x_\beta \\ \Delta y_\beta \\ \Delta x_\gamma \\ \Delta y_\gamma \end{bmatrix} = \frac{1}{36}\begin{bmatrix} -9y_1 + 3y_2 - 3y_3 \\ \sqrt{3}y_1 + \sqrt{3}y_2 - 5\sqrt{3}y_3 \\ 9y_1 + 3y_2 - 3y_3 \\ \sqrt{3}y_1 - 5\sqrt{3}y_2 + \sqrt{3}y_3 \\ -6y_2 + 6y_3 \\ -2\sqrt{3}y_1 + 4\sqrt{3}y_2 + 4\sqrt{3}y_3 \end{bmatrix}, \tag{1.291}$$

$$\mathbf{x}'_m = \frac{1}{180}\left[-9, -5\sqrt{3}, 0, 4\sqrt{3}, +9, \sqrt{3}\right], \tag{1.292}$$

$$\begin{aligned} &\mathbf{x} + \Delta\mathbf{x})' \\ &= \left[x_\alpha + \Delta x_\alpha, y_\alpha + \Delta y_\alpha, x_\beta + \Delta x_\beta, y_\beta + \Delta y_\beta, x_\gamma + \Delta x_\gamma, y_\gamma + \Delta y_\gamma\right] \\ &= \frac{1}{180}\left[-99, -35\sqrt{3}, +90, -26\sqrt{3}, +9, +61\sqrt{3}\right]. \end{aligned} \tag{1.293}$$

Box 1.30. (First and second moments of nodal points, final coordinates):

$$\begin{aligned} &y_\alpha + \Delta y_\alpha + y_\beta + \Delta y_\beta + y_\gamma + \Delta y_\gamma \\ &= y_\alpha + y_\beta + y_\gamma + \Delta y_\alpha + \Delta y_\beta + \Delta y_\gamma = 0, \end{aligned} \tag{1.294}$$

$$\begin{aligned} J_{xy} &= I_{xy} + \Delta I_{xy} \\ &= (x_\alpha + \Delta x_\alpha)(y_\alpha + \Delta y_\alpha) + (x_\beta + \Delta x_\beta)(y_\beta + \Delta y_\beta) + (x_\gamma + \Delta x_\gamma)(y_\gamma + \Delta y_\gamma) \\ &= x_\alpha y_\alpha + x_\beta y_\beta + x_\gamma y_\gamma + x_\alpha \Delta y_\alpha + y_\alpha \Delta x_\alpha + x_\beta \Delta y_\beta + y_\beta \Delta x_\beta + x_\gamma \Delta y_\gamma + y_\gamma \Delta x_\gamma \\ &\quad + \mathcal{O}(\Delta x_\alpha \Delta y_\alpha, \Delta x_\beta \Delta y_\beta, \Delta x_\gamma \Delta y_\gamma) \\ &= \sqrt{3}/15, \end{aligned} \tag{1.295}$$

$$
\begin{aligned}
J_{xx} &= I_{xx} + \Delta I_{xx} \\
&= -(y_\alpha + \Delta y_\alpha)^2 - (y_\beta + \Delta y_\beta)^2 - (\Delta y_\gamma - y_\gamma)^2 \\
&= -(y_\alpha^2 + y_\beta^2 + y_\gamma^2) - 2y_\alpha \Delta y_\alpha - 2y_\beta \Delta y_\beta - 2y_\gamma \Delta y_\gamma - O(\Delta y_\alpha^2, \Delta y_\beta^2, \Delta y_\gamma^2) \\
&= -7/12,
\end{aligned} \tag{1.296}
$$

$$
\begin{aligned}
J_{yy} &= I_{yy} + \Delta I_{yy} \\
&= -(x_\alpha + \Delta x_\alpha)^2 - (x_\beta + \Delta x_\beta)^2 - (\Delta x_\gamma - x_\gamma)^2 \\
&= -(x_\alpha^2 + x_\beta^2 + x_\gamma^2) - 2x_\alpha \Delta x_\alpha - 2x_\beta \Delta x_\beta - 2x_\gamma \Delta x_\gamma - O(\Delta x_\alpha^2, \Delta x_\beta^2, \Delta x_\gamma^2) \\
&= -11/20.
\end{aligned} \tag{1.297}
$$

1-42 Linearized Models with Datum Defect

More insight into the structure of a consistent system of observational equations with *datum defect* is gained in the case of a nonlinear model. Such a nonlinear model may be written $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ subject to $\mathbf{Y} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{R}^m$, or $\{Y_i = F_i(X_j) \mid i \in \{1, \ldots, n\},\ j \in \{1, \ldots, m\}\}$. A classification of such a nonlinear function can be based upon the "soft" Implicit Function Theorem, which is a substitute for the theory of *algebraic partitioning*, i.e., rank partitioning, and which is reviewed in Appendix C. Let us compute the matrix of first derivatives $[\partial F_i/\partial X_j \in \mathbb{R}^{n\times m}]$, which is a rectangular matrix of dimension $n \times m$. The set of n independent columns builds up the *Jacobi matrix* (1.298), which is the rectangular matrix of first derivatives (1.299) subject to (1.300). The term $m - rk\mathbf{A}$ is called the *datum defect* of the consistent system of nonlinear equations $\mathbf{y} = \mathbf{F}(\mathbf{X})$, which is a priori known. By means of such a rank partitioning, we decompose the vector of unknowns (1.301) into "bounded parameters" $\mathbf{X}_1$ and "free parameters" $\mathbf{X}_2$ subject to (1.302). Let us apply the "soft" Implicit Function Theorem to the nonlinear observational equations of distance measurements in the plane which we already introduced in the previous example. *Box 1.31* outlines the nonlinear observational equations for $Y_1 = S^2_{\alpha\beta},\ Y_2 = S^2_{\beta\gamma},\ Y_3 = S^2_{\gamma\alpha}$.. The columns $\mathbf{c}_1,\ \mathbf{c}_2,\ \mathbf{c}_3$ of the matrix $[\partial F_i/\partial X_j] \in \mathbb{R}^{n\times m}$ are linearly independent and accordingly build up the Jacobi matrix $\mathbf{J}$ of full rank. Let us partition the unknown vector $\mathbf{X}' = [\mathbf{X}'_1,\ \mathbf{X}'_2]$, namely into the "free parameters" $[X_\alpha,\ Y_\alpha,\ Y_\beta]$ and the "bounded parameters" $[X_\beta,\ X_\gamma,\ Y_\gamma]'$.

$$\mathbf{A} := \begin{bmatrix} \frac{\partial F_1}{\partial X_1} & \frac{\partial F_1}{\partial X_2} & \cdots & \frac{\partial F_1}{\partial X_n} \\ \frac{\partial F_2}{\partial X_1} & \frac{\partial F_2}{\partial X_2} & \cdots & \frac{\partial F_2}{\partial X_n} \\ \cdots & & & \\ \frac{\partial F_n}{\partial X_1} & \frac{\partial F_n}{\partial X_2} & \cdots & \frac{\partial F_n}{\partial X_n} \end{bmatrix}, \tag{1.298}$$

$$\begin{gathered} \mathbf{A} := [\mathbf{A}_1,\ \mathbf{A}_2] = [\mathbf{J},\ \mathbf{K}], \\ r = \mathrm{rk}\mathbf{A} = n, \end{gathered} \tag{1.299}$$

$$\begin{gathered} \mathbf{A} \in \mathbb{R}^{n\times m},\ \mathbf{A}_1 = \mathbf{J} \in \mathbb{R}^{n\times n} = \mathbb{R}^{r\times r}, \\ \mathbf{A}_2 = \mathbf{K} \in \mathbb{R}^{n\times(m-n)} = \mathbb{R}^{r\times(n-r)}, \end{gathered} \tag{1.300}$$

$$\mathbf{X}' = [\mathbf{X}'_1,\ \mathbf{X}'_2, \tag{1.301}$$

$$\mathbf{X}_1 \in \mathbb{R}^n = \mathbb{R}^r,\ \mathbf{X}_2 \in \mathbb{R}^{m-n} = \mathbb{R}^{m-r}. \tag{1.302}$$

Box 1.31. (Nonlinear observational equations of distance measurements in the plane, geometric notation versus algebraic notation).

Geometric notation:

$$\begin{aligned} S^2_{\alpha\beta} &= (X_\beta - X_\alpha)^2 + (Y_\beta - Y_\alpha)^2, \\ S^2_{\beta\gamma} &= (X_\gamma - X_\beta)^2 + (Y_\gamma - Y_\beta)^2, \\ S^2_{\gamma\alpha} &= (X_\alpha - X_\gamma)^2 + (Y_\alpha - Y_\gamma)^2. \end{aligned} \tag{1.303}$$

Algebraic notation:

$$\begin{aligned} Y_1 &= F_1(\mathbf{X}) = S^2_{\alpha\beta} = (X_\beta - X_\alpha)^2 + (Y_\beta - Y_\alpha)^2, \\ Y_2 &= F_2(\mathbf{X}) = S^2_{\beta\gamma} = (X_\gamma - X_\beta)^2 + (Y_\gamma - Y_\beta)^2, \\ Y_3 &= F_3(\mathbf{X}) = S^2_{\gamma\alpha} = (X_\alpha - X_\gamma)^2 + (Y_\alpha - Y_\gamma)^2. \end{aligned} \tag{1.304}$$

$$\begin{gathered} \mathbf{Y}' := [Y_1, Y_2, Y_3] = [S^2_{\alpha\beta}, S^2_{\beta\gamma}, S^2_{\gamma\alpha}] \\ \mathbf{X}' := [X_1, X_2, X_3, X_4, X_5, X_6] = [X_\alpha, Y_\alpha, X_\beta, Y_\beta, X_\gamma, Y_\gamma] \end{gathered} \tag{1.305}$$

Jacobi matrix:

$$\begin{gathered} \left[\frac{\partial F_i}{\partial X_j}\right] = 2 \left[\begin{matrix} -(X_3 - X_1) & -(X_4 - X_2) & (X_3 - X_1) \\ 0 & 0 & -(X_5 - X_3) \\ (X_1 - X_5) & (X_2 - X_6) & 0 \end{matrix}\right. \cdots \\ \cdots \left.\begin{matrix} (X_4 - X_2) & 0 & 0 \\ -(X_6 - X_4) & (X_5 - X_3) & (X_6 - X_4) \\ 0 & -(X_1 - X_5) & -(X_2 - X_6) \end{matrix}\right], \end{gathered} \tag{1.306}$$

$$\mathrm{rk}\left[\frac{\partial F_i}{\partial X_j}\right] = 3, \tag{1.307}$$

$$\left[\frac{\partial F_i}{\partial X_j}\right] = 3 \times 6, \tag{1.308}$$

$$\mathbf{J} = \begin{bmatrix} -(X_3 - X_1) & -(X_4 - X_2) & (X_3 - X_1) \\ 0 & 0 & -(X_5 - X_3) \\ (X_1 - X_5) & (X_2 - X_6) & 0 \end{bmatrix}, \ \text{rk}\mathbf{J} = 3, \tag{1.309}$$

$$\mathbf{K} = \begin{bmatrix} (X_4 - X_2) & 0 & 0 \\ -(X_6 - X_4) & (X_5 - X_3) & (X_6 - X_4) \\ 0 & -(X_1 - X_5) & -(X_2 - X_6) \end{bmatrix} \tag{1.310}$$

$$\begin{array}{ll} \text{“}\texttt{Free parameters}\text{”:} & \text{“}\texttt{Bounded parameters}\text{”:} \\ X_1 = X_\alpha = 0, & X_3 = X_\beta = \underset{(-)}{+} S_{\alpha\beta}, \\ X_2 = Y_\alpha = 0, & X_5 = X_\gamma = \underset{(-)}{+} \sqrt{S^2_{\alpha\beta} - S^2_{\beta\gamma} + S^2_{\gamma\alpha}} \\ & \quad = \underset{(-)}{+} \sqrt{Y_3^2 - Y_2^2 + Y_1^2} \\ X_4 = Y_\beta = 0. & X_6 = Y_\gamma = \underset{(-)}{+} \sqrt{S^2_{\beta\gamma} - S^2_{\alpha\beta}} \\ & \quad \underset{(-)}{+} \sqrt{Y_2^2 - Y_1^2}. \end{array} \tag{1.311}$$

Here, we have made the following choice for the “free parameters”. The origin of the coordinate system, we have fixed by $\{X_\alpha = 0,\ Y_\alpha = 0\}$. Obviously, the point P_α is this origin. The orientation of the X axis is given by $Y_\beta = 0$. In consequence, the “bounded parameters” are now derived by solving a quadratic equation, indeed a very simple one: due to the datum choice, we find (1.312). Indeed, we meet the characteristic problem of nonlinear observational equations. There are two solutions which we indicate by $\pm$. Only prior information can tell us what is the realized one in our experiment. Such a prior information has been built into by “approximate coordinates” in the previous example, a prior information, we lack now. For special reasons, we have chosen here the $+$ solution, which is in agreement with *Box 1.25*.

$$\begin{aligned} X_\beta &= \pm\sqrt{S^2_{\alpha\beta}} = \pm\sqrt{Y_1} \\ X_\gamma &= \pm(S^2_{\alpha\beta} - S^2_{\beta\gamma} + S^2_{\gamma\alpha})/(2S_{\alpha\beta}) = \pm(Y_1 - Y_2 + Y_3)/(2\sqrt{Y_1}) \\ Y_\gamma &= \pm\sqrt{S^2_{\gamma\alpha} - (S^2_{\alpha\beta} - S^2_{\beta\gamma} + S^2_{\gamma\alpha})^2/(4S^2_{\alpha\beta})} \\ &= \pm\sqrt{Y_3 - (Y_1 - Y_2 + Y_3)^2/(4Y_1)}. \end{aligned} \tag{1.312}$$

An intermediate summary of our first solution of a set of nonlinear observational equations is the following. By the choice of the datum parameters (here, the choice of origin and orientation of the coordinate system) as “free parameters”, we were able to compute the “bounded parameters” by solving a quadratic equation.

The solution space, which could be constructed in a closed form, was nonunique. Uniqueness was only achieved by prior information. Note that the closed form solution $\mathbf{X} = [X_1, X_2, X_3, X_4, X_5, X_6]' = [X_\alpha, Y_\alpha, X_\beta, Y_\beta, X_\gamma, Y_\gamma]'$ has another deficiency, namely $\mathbf{X}$ is not MINOS: it is for this reason that we apply the datum transformation $(X, Y) \mapsto (x, y)$ that is outlined in *Box 1.32* subject to $\| \mathbf{x} \|^2 = \min$, namely I-MINOS. Since we have assumed distance observations, the datum transformation is described as a rotation (rotation group SO(2)) and a translation (translation group T(2)) with three parameters (1 rotation parameter called ϕ and two translational parameters called $t_x,\ t_y$). A pointwise transformation $(X_\alpha, Y_\alpha) \mapsto (x_\alpha, y_\alpha),\ (X_\beta, Y_\beta) \mapsto (x_\beta, y_\beta)$ and $(X_\gamma, Y_\gamma) \mapsto (x_\gamma, y_\gamma)$ is presented in *Box 1.32*. The datum parameters ϕ, t_x and t_y are determined by I-MINOS, in particular, by a special *Procrustes algorithm*. There are various representations of the Lagrangean of type MINOS. For instance, as it is outlined in *Box 1.33*, we could use the representation of $\| \mathbf{x} \|^2$ in terms of observations $Y_1 = S^2_{\alpha\beta}$, $Y_2 = S^2_{\beta\gamma}$, and $Y_3 = S^2_{\gamma\alpha}$, which transforms $\| \mathbf{x} \|^2$ of (1.318) into $\| \mathbf{x} \|^2$ (Y_1, Y_2, Y_3) of (1.319). $\| \mathbf{x} \|^2$ of (1.329) is equivalent to minimizing the product sums of Cartesian coordinates. As soon as we substitute the datum transformation of *Box 1.32*, which we illustrate by Figs. 1.15 and 1.16, into the Lagrangean $\mathcal{L}(t_x, t_y, \phi)$ of type MINOS ($\| \mathbf{x} \|^2 = \min$), we arrive at the quadratic objective function of *Box 1.34*. In the first forward step of the special Procrustes algorithm, we obtain the minimal solution for the translation parameters $\hat{t}_x$ and $\hat{t}_y$. The second forward step of the special Procrustes algorithm is built on (i) the substitution of $\hat{t}_x$ and $\hat{t}_y$ in the original *Lagrangean*, which leads to the reduced Lagrangean of *Box 1.35*, and (ii) the minimization of the reduced Lagrangean $\mathrm{L}(\phi)$ with respect to the rotation parameter ϕ. In an intermediate phase, we introduce "centralized coordinates" $(\Delta X, \Delta Y)$, namely coordinate differences with respect to the centre $P_o = (X_o, Y_o)$ of the polyhedron, namely the triangle $P_\alpha,\ P_\beta,\ P_\gamma$. In this way, we are able to generate the simple (standard form) $\tan 2\phi^\wedge$ of the solution $\phi^\wedge$, the argument of $\mathcal{L}_1 = \mathcal{L}_1(\phi) = \min$ or $\mathcal{L}_2 = \mathcal{L}_2(\phi)$. The special Procrustes algorithm is completed by the backforward steps outlined in *Box 1.36*. Firstly, we convert $\tan 2\phi^\wedge$ to $(\cos\phi^\wedge, \sin\phi^\wedge)$. Secondly, we substitute $(\cos\phi^\wedge, \sin\phi^\wedge)$ into the translation formula $(t_x^\wedge, t_y^\wedge)$. Thirdly, we substitute $(t_x^\wedge, t_y^\wedge, \cos\phi^\wedge, \sin\phi^\wedge)$ into the Lagrangean $\mathcal{L}(t_x, t_x, \phi)$, thus generating the optimal objective function $\| \mathbf{x} \|^2$

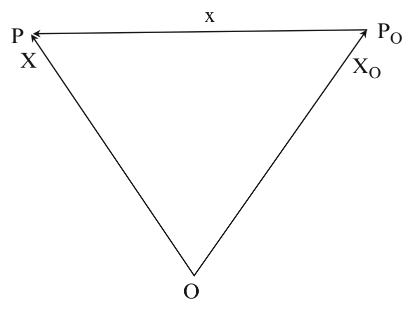


Fig. 1.15 Commutative diagram (P diagram). P_O = centre of polyhedron (triangle $p_\alpha p_\beta p_\gamma$). Action of the translation group

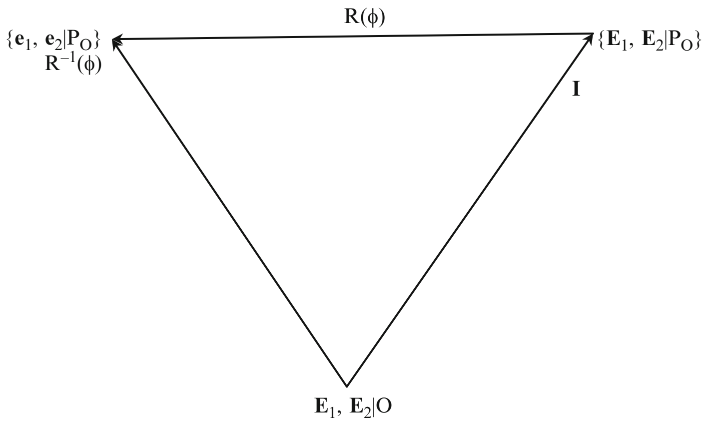


Fig. 1.16 Commutative diagram (E diagram). Orthonormal 2-legs $\{\mathbf{E}_1, \mathbf{E}_2|P_O$ at P_O and $\{\mathbf{e}_1, \mathbf{e}_2|P_O\}$ at P_O at P_O. Action of the translation group

at $(t_x^\wedge, t_y^\wedge, \phi^\wedge)$. Finally, as step four, we succeed to compute the centric coordinates (1.313, left) with respect to the orthonormal 2-leg $\{\mathbf{e}_1, \mathbf{e}_2|P_O$ at P_O from the given coordinates (1.313, right) with respect to the orthonormal 2-leg $\{\mathbf{e}_1, \mathbf{e}_2|O$ at O and the optimal datum parameters $(t_x^\wedge, t_y^\wedge, \cos\phi^\wedge, \sin\phi^\wedge)$.

$$\begin{bmatrix} x_\alpha & x_\beta & x_\gamma \\ y_\alpha & y_\beta & y_\gamma \end{bmatrix}, \quad \begin{bmatrix} X_\alpha & X_\beta & X_\gamma \\ Y_\alpha & Y_\beta & Y_\gamma \end{bmatrix} \tag{1.313}$$

Box 1.32. (Datum transformation of Cartesian coordinates):

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{R} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{1.314}$$

$$\mathbf{R} \in SO(2) := \{\mathbf{R} \in \mathbb{R}^{2\times 2} | \mathbf{R}'\mathbf{R} = \mathbf{I}_2, \mathtt{det}\mathbf{R} = +1\}. \tag{1.315}$$

(representation of a 2×2 orthonormal matrix):

$$\mathbf{R} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{1.316}$$

$$\begin{aligned} x_\alpha &= X_\alpha \cos\phi + Y_\alpha \sin\phi - t_x \\ y_\alpha &= -X_\alpha \sin\phi + Y_\alpha \cos\phi - t_y \\ \\ x_\beta &= X_\beta \cos\phi + Y_\beta \sin\phi - t_x \\ y_\beta &= -X_\beta \sin\phi + Y_\beta \cos\phi - t_y \\ \\ x_\gamma &= X_\gamma \cos\phi + Y_\gamma \sin\phi - t_x \\ y_\gamma &= -X_\gamma \sin\phi + Y_\gamma \cos\phi - t_y. \end{aligned} \tag{1.317}$$

Box 1.33. (Various forms of MINOS).

(i)

$$\| \mathbf{x} \|^2 = x_\alpha^2 + y_\alpha^2 + x_\beta^2 + y_\beta^2 + x_\gamma^2 + y_\gamma^2 = \min_{\phi, t_x, t_y}; \tag{1.318}$$

(ii)

$$\begin{gathered} \| \mathbf{x} \|^2 = \tfrac{1}{2}(S_{\alpha\beta}^2 + S_{\beta\gamma}^2 + S_{\gamma\alpha}^2) + \\ + x_\alpha x_\beta + x_\beta x_\gamma + x_\gamma x_\alpha + y_\alpha y_\beta + y_\beta y_\gamma + y_\gamma y_\alpha = \min_{\phi, t_x, t_y}; \end{gathered} \tag{1.319}$$

(iii)

$$\| \mathbf{x} \|^2 = \min \quad \Leftrightarrow \quad \left[\begin{array}{l} x_\alpha x_\beta + x_\beta x_\gamma + x_\gamma x_\alpha = \min \\ y_\alpha y_\beta + y_\beta y_\gamma + y_\gamma y_\alpha = \min . \end{array} \right. \tag{1.320}$$

Proof.

The representation of the objective function of type MINOS in terms of the observations $Y_1 = S_{\alpha\beta}^2$, $Y_2 = S_{\beta\gamma}^2$, and $Y_3 = S_{\gamma\alpha}^2$ can be proven as follows:

$$\begin{aligned} S_{\alpha\beta}^2 &= (x_\beta - x_\alpha)^2 + (y_\beta - y_\alpha)^2 = x_\alpha^2 + y_\alpha^2 + x_\beta^2 + y_\beta^2 - 2(x_\alpha x_\beta + y_\alpha y_\beta) \\ &= \Rightarrow \frac{1}{2} S_{\alpha\beta}^2 + x_\alpha x_\beta + y_\alpha y_\beta = \frac{1}{2}(x_\alpha^2 + y_\alpha^2 + x_\beta^2 + y_\beta^2), \end{aligned} \tag{1.321}$$

$$\begin{aligned} \| \mathbf{x} \|^2 &= x_\alpha^2 + y_\alpha^2 + x_\beta^2 + y_\beta^2 + x_\gamma^2 + y_\gamma^2 \\ &= \frac{1}{2}(S_{\alpha\beta}^2 + S_{\beta\gamma}^2 + S_{\gamma\alpha}^2) + \\ &\quad + x_\alpha x_\beta + x_\beta x_\gamma + x_\gamma x_\alpha + y_\alpha y_\beta + y_\beta y_\gamma + y_\gamma y_\alpha \\ &= \frac{1}{2}(Y_1 + Y_2 + Y_3) \\ &\quad + x_\alpha x_\beta + x_\beta x_\gamma + x_\gamma x_\alpha + y_\alpha y_\beta + y_\beta y_\gamma + y_\gamma y_\alpha . \end{aligned} \tag{1.322}$$

Box 1.34. (Minimum norm solution, special Procrustes algorithm, 1st forward step).

$$\begin{gathered} \| \mathbf{x} \|^2 \\ := x_\alpha^2 + y_\alpha^2 + x_\beta^2 + y_\beta^2 + x_\gamma^2 + y_\gamma^2 \\ = \min_{t_x, t_y, \phi} \end{gathered} \tag{1.323}$$

$\boxed{\text{Lagrangean}}$

$$
\begin{aligned}
&\mathcal{L}(t_x, t_y, \phi) \\
:= \;& (X_\alpha \cos\phi + Y_\alpha \sin\phi - t_x)^2 \\
&+(-X_\alpha \sin\phi + Y_\alpha \cos\phi - t_y)^2 \\
&+(X_\beta \cos\phi + Y_\beta \sin\phi - t_x)^2 \\
&+(-X_\beta \sin\phi + Y_\beta \cos\phi - t_y)^2 \\
&+(X_\gamma \cos\phi + Y_\gamma \sin\phi - t_x)^2 \\
&+(-X_\gamma \sin\phi + Y_\gamma \cos\phi - t_y)^2.
\end{aligned}
\tag{1.324}
$$

1st forward step

$$
\begin{aligned}
\frac{1}{2}\frac{\partial\mathcal{L}}{\partial t_x}(t_x^\wedge) &= (X_\alpha + X_\beta + X_\gamma)\cos\phi + (Y_\alpha + Y_\beta + Y_\gamma)\sin\phi - 3t_x^\wedge = 0, \\
\frac{1}{2}\frac{\partial\mathcal{L}}{\partial t_y}(t_y^\wedge) &= -(X_\alpha + X_\beta + X_\gamma)\sin\phi + (Y_\alpha + Y_\beta + Y_\gamma)\cos\phi - 3t_y^\wedge = 0,
\end{aligned}
\tag{1.325}
$$

$$
\begin{aligned}
t_x^\wedge &= +\frac{1}{3}\{(X_\alpha + X_\beta + X_\gamma)\cos\phi + (Y_\alpha + Y_\beta + Y_\gamma)\sin\phi\}, \\
t_y^\wedge &= +\frac{1}{3}\{-(X_\alpha + X_\beta + X_\gamma)\sin\phi + (Y_\alpha + Y_\beta + Y_\gamma)\cos\phi\},
\end{aligned}
\tag{1.326}
$$

$$
\{t_x^\wedge, t_y^\wedge\} = \arg[\mathcal{L}(t_x, t_y, \phi) = \min]. \tag{1.327}
$$

Box 1.35. (Minimum norm solution, special Procrustes algorithm, second forward step).

Solution $\{t_x^\wedge, t_y^\wedge\}$ in Lagrangean (reduced Lagrangean):

$$
\begin{aligned}
&\mathcal{L}(\phi) := \\
&:= \left\{X_\alpha \cos\phi + Y_\alpha \sin\phi - \tfrac{1}{3}[(X_\alpha + X_\beta + X_\gamma)\cos\phi + (Y_\alpha + Y_\beta + Y_\gamma)\sin\phi]\right\}^2 \\
&\quad + \left\{X_\beta \cos\phi + Y_\beta \sin\phi - \tfrac{1}{3}[(X_\alpha + X_\beta + X_\gamma)\cos\phi + (Y_\alpha + Y_\beta + Y_\gamma)\sin\phi]\right\}^2 \\
&\quad + \left\{X_\gamma \cos\phi + Y_\gamma \sin\phi - \tfrac{1}{3}[(X_\alpha + X_\beta + X_\gamma)\cos\phi + (Y_\alpha + Y_\beta + Y_\gamma)\sin\phi]\right\}^2 \\
&\quad + \left\{-X_\alpha \sin\phi + Y_\alpha \cos\phi - \tfrac{1}{3}[-(X_\alpha + X_\beta + X_\gamma)\sin\phi + (Y_\alpha + Y_\beta + Y_\gamma)\cos\phi]\right\}^2 \\
&\quad + \left\{-X_\beta \sin\phi + Y_\beta \cos\phi - \tfrac{1}{3}[-(X_\alpha + X_\beta + X_\gamma)\sin\phi + (Y_\alpha + Y_\beta + Y_\gamma)\cos\phi]\right\}^2 \\
&\quad + \{-X_\gamma \sin\phi + Y_\gamma \cos\phi - \tfrac{1}{3}[-(X_\alpha + X_\beta + X_\gamma)\sin\phi + (Y_\alpha + Y_\beta + Y_\gamma)\cos\phi]\}^2 \\
&= \min_\phi
\end{aligned}
\tag{1.328}
$$

$$
\begin{aligned}
\mathcal{L}(\phi) &= \\
&= \left\{\left[X_\alpha - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)\right]\cos\phi + \left[Y_\alpha - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)\right]\sin\phi\right\}^2 \\
&\quad + \left\{\left[X_\beta - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)\right]\cos\phi + \left[Y_\beta - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)\right]\sin\phi\right\}^2 \\
&\quad + \left\{\left[X_\gamma - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)\right]\cos\phi + \left[Y_\gamma - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)\right]\sin\phi\right\}^2 \\
&\quad + \left\{-\left[X_\alpha - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)\right]\sin\phi + \left[Y_\alpha - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)\right]\cos\phi\right\}^2 \\
&\quad + \{-[X_\beta - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)]\sin\phi + [Y_\beta - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)]\cos\phi\}^2 \\
&\quad + \left\{-\left[X_\gamma - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma)]\sin\phi + [Y_\gamma - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma)\right]\cos\phi\right\}^2 .
\end{aligned}
\tag{1.329}
$$

Centralized coordinate:

$$
\begin{aligned}
\Delta X &:= X_\alpha - \frac{1}{3}(X_\alpha + X_\beta + X_\gamma) = \frac{1}{3}(2X_\alpha - X_\beta - X_\gamma) \\
\Delta Y &:= Y_\alpha - \frac{1}{3}(Y_\alpha + Y_\beta + Y_\gamma) = \frac{1}{3}(2Y_\alpha - Y_\beta - Y_\gamma)
\end{aligned}
\tag{1.330}
$$

Reduced Lagrangean:

$$
\begin{aligned}
\mathcal{L}_1(\phi) &= (\Delta X_\alpha \cos\phi + \Delta Y_\alpha \sin\phi)^2 + (\Delta X_\beta \cos\phi + \Delta Y_\beta \sin\phi)^2 \\
&\quad + (\Delta X_\gamma \cos\phi + \Delta Y_\gamma \sin\phi)^2 \\
\mathcal{L}_2(\phi) &= (-\Delta X_\alpha \sin\phi + \Delta Y_\alpha \cos\phi)^2 + (-\Delta X_\beta \sin\phi + \Delta Y_\beta \cos\phi)^2 \\
&\quad + (-\Delta X_\gamma \sin\phi + \Delta Y_\gamma \cos\phi)^2
\end{aligned}
\tag{1.331}
$$

$$
\begin{gathered}
\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\phi}(\phi^\wedge) = 0 \\
\Leftrightarrow \\
+(\Delta X_\alpha \cos\phi^\wedge + \Delta Y_\alpha \sin\phi^\wedge)(-\Delta X_\alpha \sin\phi^\wedge + \Delta Y_\alpha \cos\phi^\wedge) \\
+(\Delta X_\beta \cos\phi^\wedge + \Delta Y_\beta \sin\phi^\wedge)^2(-\Delta X_\beta \sin\phi^\wedge + \Delta Y_\beta \cos\phi^\wedge) \\
+(\Delta X_\gamma \cos\phi^\wedge + \Delta Y_\gamma \sin\phi^\wedge)^2(-\Delta X_\gamma \sin\phi^\wedge + \Delta Y_\gamma \cos\phi^\wedge) = 0
\end{gathered}
\tag{1.332}
$$

$$\Leftrightarrow$$

$$\begin{aligned}
&-(\Delta X_\alpha^2 + \Delta X_\beta^2 + \Delta X_\gamma^2)\sin\phi^\wedge\cos\phi^\wedge \\
&+(\Delta X_\alpha \Delta Y_\alpha + \Delta X_\beta \Delta Y_\beta + \Delta X_\gamma \Delta Y_\gamma)\cos^2\phi^\wedge \\
&-(\Delta X_\alpha \Delta Y_\alpha + \Delta X_\beta \Delta Y_\beta + \Delta X_\gamma \Delta Y_\gamma)\sin^2\phi^\wedge \\
&+(\Delta Y_\alpha^2 + \Delta Y_\beta^2 + \Delta Y_\gamma^2)\sin\phi^\wedge\cos\phi^\wedge = 0
\end{aligned} \tag{1.333}$$

$$\Leftrightarrow$$

$$\begin{aligned}
&[(\Delta X_\alpha^2 + \Delta X_\beta^2 + \Delta X_\gamma^2) - (\Delta Y_\alpha^2 + \Delta Y_\beta^2 + \Delta Y_\gamma^2)]\sin 2\phi^\wedge \\
&= 2[\Delta X_\alpha \Delta Y_\alpha + \Delta X_\beta \Delta Y_\beta + \Delta X_\gamma \Delta Y]\cos 2\phi^\wedge,
\end{aligned}$$

$$\tan 2\phi^\wedge = 2\frac{\Delta X_\alpha \Delta Y_\alpha + \Delta X_\beta \Delta Y_\beta + \Delta X_\gamma \Delta Y}{(\Delta X_\alpha^2 + \Delta X_\beta^2 + \Delta X_\gamma^2) - (\Delta Y_\alpha^2 + \Delta Y_\beta^2 + \Delta Y_\gamma^2)}. \tag{1.334}$$

Orientation parameter in terms of Gauss brackets:

$$\begin{aligned}
\tan 2\phi^\wedge &= \frac{2[\Delta \mathbf{X} \Delta \mathbf{Y}]}{[\Delta \mathbf{X}^2] - [\Delta \mathbf{Y}^2]} \\
\phi^\wedge &= \arg\{\mathcal{L}_1(\phi) = \min\} = \arg[\mathcal{L}_2(\phi) = \min].
\end{aligned} \tag{1.335}$$

Box 1.36. (Special Procrustes algorithm, backward steps).

Step one:

$$\tan 2\phi^\wedge = \frac{2[\Delta \mathbf{X} \Delta \mathbf{Y}]}{[\Delta \mathbf{X}^2] - [\Delta \mathbf{Y}^2]} \quad \Rightarrow \quad \left[\begin{array}{l} \cos\phi^\wedge, \\ \sin\phi^\wedge. \end{array}\right. \tag{1.336}$$

Step two:

$$\begin{aligned}
t_x^\wedge &= \frac{1}{3}([\mathbf{X}]\cos\phi^\wedge + [\mathbf{Y}]\sin\phi^\wedge), \\
t_y^\wedge &= \frac{1}{3}(-[\mathbf{X}]\sin\phi^\wedge + [\mathbf{Y}]\cos\phi^\wedge).
\end{aligned} \tag{1.337}$$

Step three:

$$\| \mathbf{x}^\wedge \|^2 = \mathcal{L}(t_x^\wedge, t_y^\wedge, \phi^\wedge). \tag{1.338}$$

Step four:

$$\begin{bmatrix} x_\alpha & x_\beta & x_\gamma \\ y_\alpha & y_\beta & y_\gamma \end{bmatrix} = \begin{bmatrix} \cos\phi^\wedge & \sin\phi^\wedge \\ -\sin\phi^\wedge & \cos\phi^\wedge \end{bmatrix} \begin{bmatrix} X_\alpha & X_\beta & X_\gamma \\ Y_\alpha & Y_\beta & Y_\gamma \end{bmatrix} - \begin{bmatrix} t_x^\wedge \\ t_y^\wedge \end{bmatrix} \mathbf{1}'. \tag{1.339}$$

We leave the proof of the following relations as an exercise to the reader: $[\mathbf{x}] = x_\alpha + x_\beta + x_\gamma = 0$, $[\mathbf{y}] = y_\alpha + y_\beta + y_\gamma = 0$, and $[\mathbf{xy}] = x_\alpha y_\alpha + x_\beta y_\beta + x_\gamma y_\gamma \neq 0$. A numerical example is provided by *Example 1.2*.

Example 1.2. (A numerical example).

$$S_{\alpha\beta} = 1.1, \quad S_{\beta\gamma} = 0.9, \quad S_{\gamma\alpha} = 1.2,$$
$$Y_1 = 1.21\ , \quad Y_2 = 0.81, \quad Y_3 = 1.44, \tag{1.340}$$

$$X_\alpha = 0, \quad X_\beta = 1.10, \quad X_\gamma = 0.84,$$
$$Y_\alpha = 0\ , \quad Y_\beta = 0 \quad , \quad Y_\gamma = 0.86, \tag{1.341}$$

$$\Delta X_\alpha = -0.647, \quad \Delta X_\beta = +0.453, \quad \Delta X_\gamma = +0.193,$$
$$\Delta Y_\alpha = -0.287\ , \quad \Delta Y_\beta = -0.287, \quad \Delta Y_\gamma = +0.573. \tag{1.342}$$

$$\boxed{\text{Test:}}$$

$$[\Delta\mathbf{X}] = 0, \quad [\Delta\mathbf{Y}] = 0, \tag{1.343}$$

$$[\Delta\mathbf{X}\Delta\mathbf{Y}] = 0.166, \quad [\Delta\mathbf{X}^2] = 0.661, \quad [\Delta\mathbf{Y}^2] = 0.493, \tag{1.344}$$

$$\begin{gathered} \tan 2\phi^\wedge = 1.979, \quad \phi^\wedge = 31^\circ.598, 828, 457 \\ \phi^\wedge = 31^\circ\, 35'\ 55''.782, \\ \cos\phi^\wedge = 0.851, 738, \quad \sin\phi^\wedge = 0.523, 968, \\ t_x^\wedge = 0.701, \quad t_y^\wedge = -0.095, \end{gathered} \tag{1.345}$$

$$\left[\begin{array}{l} x_\alpha = -0.701, \quad x_\beta = +0.236, \quad x_\gamma = +0.465, \\ y_\alpha = +0.095, \quad y_\beta = -0.481, \quad y_\gamma = +0.387, \end{array}\right. \tag{1.346}$$

$$\boxed{\text{Test:}}$$

$$[\mathbf{x}] = x_\alpha + x_\beta + x_\gamma = 0, [\mathbf{y}] = y_\alpha + y_\beta + y_\gamma = 0, \tag{1.347}$$

$$[\mathbf{xy}] = +0.019 \neq 0 \tag{1.348}$$

1-5 Notes

Table 1.5 contains a list of observables in $\mathbb{R}^n$, equipped with a metric and their corresponding transformation groups. The number of the datum parameters coincides with the injectivity rank deficiency in a consistent system of linear (linearized) observational equations $\mathbf{Ax} = \mathbf{y}$ subject to $d(\mathbf{A}) = m - \text{rk}\mathbf{A}$.

What is the origin of the rank deficiency three of the linearized observational equations, namely the three distance functions observed in a planar triangular network, we present in paragraph three. In geometric terms, the a priori indeterminancy of relating observed distances to absolute coordinates placing points in the plane can be interpreted easily: the observational equation of distances in the plane is invariant with respect to a translation and a rotation of the coordinate system.

Table 1.5. Observables and transformation groups

Observed quantities	Transformation group	Datum parameters
Coordinate differences in $\mathbb{R}^2$	Translation group T(2)	2
Coordinate differences in $\mathbb{R}^3$	Translation group T(3)	3
Coordinate differences in $\mathbb{R}^n$	Translation group T(n)	n
Distances in $\mathbb{R}^2$	Group of motion T(2),SO(2)	3
Distances in $\mathbb{R}^3$	Group of motion T(3),SO(3)	3+3 = 6
Distances in $\mathbb{R}^n$	Group of motion T(n),SO(n)	n + (n + 1) = 2
Angles, distance ratios in $\mathbb{R}^2$	Conformal group C4(2)	4
Angles, distance ratios in $\mathbb{R}^3$	Conformal group C7(3)	7
Angles, distance ratios in $\mathbb{R}^n$	Conformal group $\mathrm{C}_{(n+1)(n+2)/2}(n)$	(n + 1)(n + 2)/2
Cross-ratios of area elements in the projective plane	Projective group	8

Mind the following note. The structure group of the two-dimensional Euclidean space is the group of motion decomposed into the translation group (two parameters) and the rotation group (one parameter). Under the action of the group of motion (three parameters) Euclidean distance functions are left equivariant. The three parameters of the group of motion cannot be determined from distance measurements: they produce a rank deficiency of three in the linearized observational equations. A detailed analysis of the relation between the transformation groups and the observational equations has been presented by Grafarend (1974, 1976).

Mind also the following notes. (i) More generally, the structure group of a three-dimensional *Weitzenboeck space is* the conformal group which is decomposed into the translation group (3 parameters), the special orthogonal group SO(3) (3 parameters) and the dilatation group ("scale", 1 parameter). Under the action of the conformal group – in total 7 parameters – distance ratios and angles are left equivariant. The conformal group generates a transformation of Cartesian coordinates covering which is called *similarity transformation* or *datum transformation*. Any choice of an origin of the coordinate system, of the axes orientation, and of the scale constitutes an S-base following *W. Baarda* (1962, 1967, 1973, 1979, 1995), *J. Bossler* (1973), *M. Berger* (1994), *A. Dermanis* (1998), *A. Dermanis, and E. Grafarend* (1993), *A. Fotiou and D. Rossikopoulis* (1993), *E. Grafarend* (1973, 1979, 1983), *E. Grafarend, E. H. Knickmeyer, and B. Schaffrin* (1982), *E. Grafarend and G. Kampmann* (1996), *G. Heindl* (1986), *M. Molenaar* (1981), *H. Quee* (1983), *P.J.G. Teunissen* (1960, 1985), and *H. Wolf* (1990). (ii) In projective networks (image processing, photogrammetry, robot vision), the projective group is active. The projective group generates a perspective transformation which is outlined in *E. Grafarend and J. Shan* (1997). Under the action of the projective group, cross-ratios of areal elements in the projective plane are left equivariant. For more details, let us refer to *M. Berger* (1994), *M.H. Brill, and E.B. Barrett* (1983), *R.O. Duda and P.E. Heart* (1973), *E. Grafarend and J. Shan* (1997), F. Gronwald and *F.W. Hehl* (1996), *M.R. Haralick* (1980), *R.J. Holt, and A.N. Netrawalli* (1995), *R.L. Mohr, L. Morin, and E. Grosso* (1992), J.*L. Mundy and A. Zisserman* (1992a, b), *R.F. Riesenfeldt* (1981), *J.A. Schouten* (1954). (iii) In electromagnetism (*Maxwell equations*), the conformal group is active. The conformal group generates a transformation of "space-time" by means of 16 parameters (6 rotational parameters three for rotation, three for "hyperbolic rotation", 4 translational parameters, 5 "involutary" parameters, 1 dilatation (scale) parameter) which leaves the *Maxwell equations* in vacuum as well as pseudo-distance ratios and angles equivariant. Sample references are *A.O. Barut (1972), H. Bateman (1910), F. Bayen (1976), J. Beckers, J. Harnard, M. Perroud, and M. Winternitz (1976), D.G. Boulware, L.S. Brown, R.D. Peccei (1970), P. Carruthers (1971), E. Cunningham (1910), T. tom Dieck (1967), N. Euler, and W.H. Steeb (1992), P.G.O. Freund (1974), T. Fulton, F. Rohrlich, and L. Witten (1962), J. Haantjes (1937), H.A. Kastrup (1962, 1966), R. Kotecky, and J. Niederle (1975), K.H. Marivalla (1971), D.H. Mayer (1975), J.A. Schouten, and J. Haantjes (1936), D.E. Soper (1976) and J. Wess (1990).*

Chapter 2
The First Problem of Probabilistic Regression: The Bias Problem

Minimum Bias solution of problems with datum defects. LUMBE of fixed effects.

The *bias problem in probabilistic* regression has been the subject of Sect. 4-37 for simultaneous determination of first moments as well as second central moments by inhomogeneous multilinear, namely bilinear, estimation. Based on the review of the first author "Variance-covariance component estimation: theoretical results and geodetic application" (Statistical and Decision, Supplement Issue No. 2, pages 401–441, 105 references, Oldenbourg Verlag, München 1989), we collected 5 postulates for simultaneous determination of first and second central moments. A first reference is *J. Kleffe* (1978). It forms the basis of Sect. 4-37:

ansatz

$$E\{y\} = \boldsymbol{A}\mu = \boldsymbol{A}\boldsymbol{x}_1, \quad vech\ D\{y\} = \boldsymbol{B}\sigma = \boldsymbol{B}\boldsymbol{x}_2$$

estimation:
ansatz: bilinearity

$$\begin{bmatrix} \widehat{\boldsymbol{x}_1} = \boldsymbol{K}_1 + \boldsymbol{L}_1\boldsymbol{y} + \boldsymbol{M}_1(\boldsymbol{y}\otimes\boldsymbol{y}) \\ \widehat{\boldsymbol{x}_2} = \boldsymbol{K}_2 + \boldsymbol{L}_2\boldsymbol{y} + \boldsymbol{M}_2(\boldsymbol{y}\otimes\boldsymbol{y}) \end{bmatrix} \widehat{\boldsymbol{x}} :\leq \begin{bmatrix} \widehat{\boldsymbol{x}_1} \\ \widehat{\boldsymbol{x}_2} \end{bmatrix}$$

1st postulate

"inhomogeneous, bilinear estimation"

$$\boldsymbol{X} := \begin{bmatrix} \boldsymbol{K}_1 & \boldsymbol{L}_1 & \boldsymbol{M}_1 \\ \boldsymbol{K}_2 & \boldsymbol{L}_2 & \boldsymbol{M}_2 \end{bmatrix}, \quad \boldsymbol{Y} := \begin{bmatrix} \mathbf{1} \\ \mathbf{y} \\ \mathbf{y}\otimes\mathbf{y} \end{bmatrix}$$

2nd postulate

"invariance"
$\tilde{\sigma}$ is invariant in the sense $y \mapsto y - \boldsymbol{A}\mu$

3rd postulate

"unbiased in the mean" versus "minimum bias"
bias vector, bias matrix

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2_2,
© Springer-Verlag Berlin Heidelberg 2012

$$\boldsymbol{b} := E\{\hat{x}\} - x = \boldsymbol{S}x, \boldsymbol{S} := \boldsymbol{X}\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{B}\end{bmatrix} - \boldsymbol{I}$$

4th postulate

"$D\{\hat{x}\}$ contains second, third and fourth order moments which are based on a *quasi-Gauss normal* structure."

5th postulate

"best estimation"

$$\|D\{\hat{x}\}\| = min$$

Here we need only two postulates, namely *(i) linear*: $(x_1) = \boldsymbol{L_1 y}$ and *(ii) minimum bias in the mean* when we consider the simple model $E\{\boldsymbol{y}\} = \boldsymbol{A\mu} = \boldsymbol{A}x_1(\mathbf{A} \in \mathcal{R}^{\mathrm{n}\times\mathrm{m}},\ \mathrm{n} < m)$ for the moment of first order and $D\{\boldsymbol{y}\} = I\ \sigma^2 = I\ \boldsymbol{x}_2$ for the central moment of the second order ($\sigma^2 > 0$), one parameter of type *variance*. You find the estimate of type *linear* $\widehat{\boldsymbol{x_1}} = \boldsymbol{L_1 y}$ and type *minimum bias* in the mean and of type *FROBENIUS matrix norm* $\|\boldsymbol{B}\|^2$ for the case $m > n$, here $\|\boldsymbol{B}\|^2 = d = m - n$.

Real World Problems are nonlinear. We have to divide the solution space into those classes: (a) strong nonlinear, (b) weak nonlinear, and (c) linear. Weak nonlinearity is defined by nonlinear system which allows a *Taylor series approximation*. Linear system equations, for instance the minimum bias problem in the *FROBENIUS matrix norm* with respect to *linear estimation theory*, have the advantage that its *norm of equations* has only one minimum solution. We call that solution "*global*" and "*uniform*". In contrast to linear estimation, weak nonlinear equations produce "*local*" solutions. They experience a great disadvantage: there exist *many local solutions*. A typical example is a geodetic network analysis by *P. LOHSE (1994)*. Another example is also *C.R. RAO and J. KLEFFE (1999, pages 161–180)*.

Minimum bias solutions of rank deficient linear systems have been discussed in detail by *C.R. RAO (1974)*. He introduces the notation of the *substitute matrix* $\boldsymbol{S}$ referring the matrix $\boldsymbol{\xi\xi'}$ by the matrix $\mathbf{S}$ of arbitrary rank, for instance $rk\boldsymbol{S} = m$. Note that the rank of the matrix $\boldsymbol{\xi\xi'}$ is one by the technique of *LAGRANGE multiplyer* the minimization of the S weighted *FROBENIUS norm* $\|(\boldsymbol{I}_m - LA)'\|_L^2$ within *Theorem 2.3*. Obviously, **S**-weighted homogeneous linear uniform by minimum bias estimator ("hom S-LUMBE") is equivalent to the weighted minimum norm solutions ("**G**-MINOS") subject of the RAO-Theorem. Our example is the special model for $\boldsymbol{S} = \boldsymbol{I}_m$ called "*unity substitute matrix*", obviously $\|\boldsymbol{B}\|^2 = d = m - n = m - rk\boldsymbol{A}$.

In practice, it is difficult to determine the rank of matrix A. For instance, in problems of Configuration Defects in Geodetic Sciences E. Grafarend and E. Livieratos (1979), E. Grafarend and K. Heinz (1978), E. Grafarend and V.

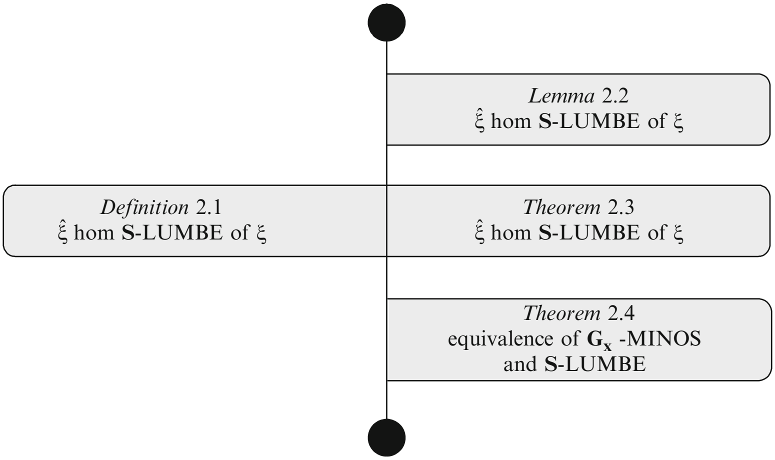


Fig. 2.1 The guideline of chapter two: definition, lemma and theorems

Miller (1985) and E. Grafarend and A. Mader (1989). Critical comments of the RAO bias problem solution have been raised by Arne Bjerhammar and Jujit Kumar Mitra (private communication to the first author). They argue that bias has nothing to do with rank definition of the type m-n. They use the argument that you need PRIOR INFORMATION, for instance of BAYES type: Bayes Estimation x_0. Here we leave this question open.

In summary, the minimum of the trace of the matrix valued bias leads to the rank deficiency d subject to $\mathrm{d} = \mathrm{m} - \mathrm{n}, \mathrm{m} > \mathrm{n}$.

Please pay attention to the guideline of Chap. 2 shown in Fig. 2.1. In particular, mind the structure of definitions, lemma, and theorems. Fast track reading: consult Box 2.1 and read only Theorem 2.3.

In the first chapter, we have solved a special algebraic regression problem, namely the inversion of a system of consistent linear equations which is classified as underdetermined. By means of the postulate of a minimum norm solution, namely $||\mathbf{x}||^2 = \min$, we were able to determine m unknowns ($m > n$, say $m = 10^6$) from n observations (more unknowns m than equations n, say $n = 10$). Indeed, such a mathematical solution may surprise the analyst: in the example, MINOS produces one million unknowns from 10 observations. Though MINOS generates a rigorous solution, we are left with some doubts.

How can we conveniently interpret such an "unbelievable solution"? The key for an evaluation of MINOS is handed over to us by treating the special algebraic regression problem with datum defect. The bias generated by any solution of an underdetermined or ill-posed problem will be introduced as a decisive criterion for evaluating MINOS, now in the context of a probabilistic regression problem. In particular, a special form of LUMBE, the *linear uniformly minimum bias estimator* $||\mathbf{LA}-\mathbf{I}||^2 =$ min leads us to a solution which is equivalent to "MINOS". Alternatively we may say that in the various classes of solving an underdetermined problem "LUMBE" generates a solution of minimal bias.

What is a *underdetermined regression problem* of type MINOS? By means of a certain objective function, here of type "*minimum bias*", we solve the inverse problem of linear and nonlinear equations with fixed effects. In particular, in order to minimize the bias vector $\boldsymbol{E}\{\hat{\boldsymbol{\xi}}\}-\boldsymbol{\xi}$ we meet the problem that its minimum norm $\|\boldsymbol{E}\{\hat{\boldsymbol{\xi}}\}-\boldsymbol{\xi}\|^2 = tr[(\boldsymbol{E}\{\hat{\boldsymbol{\xi}}\}-\boldsymbol{\xi})(\boldsymbol{E}\{\hat{\boldsymbol{\xi}}\}-\boldsymbol{\xi})']$ subject to $\hat{\boldsymbol{\xi}} = \mathbf{Ly}$ *depends on the product* $\boldsymbol{\xi}\boldsymbol{\xi}'$ where $\boldsymbol{\xi}$ denotes the *unknown parameter vector*. In this section *C.R. Rao* proposed the use instead of the matrix $\boldsymbol{\xi}\boldsymbol{\xi}'$, $rk\boldsymbol{\xi}\boldsymbol{\xi}' = 1$, the *substitute matrix* $\mathbf{S}$ with full rank, for instance.

2-1 Linear Uniformly Minimum Bias Estimator (LUMBE)

Let us introduce the special consistent linear Gauss–Markov model specified in Box 2.1, which is given form of a *consistent system* of linear equations relating non-stochastic, real-valued vector $\boldsymbol{\xi}$ of unknowns to the observation vector $\mathbf{y}$. Here, the rank rk$\mathbf{A}$ of the matrix $\mathbf{A}$ connecting the observations and unknown parameters equals the number n of observations $\mathbf{y} \in \mathbb{R}^n$.

Box 2.1. (Special consistent linear Gauss–Markov model of fixed effects. Bias vector, bias matrix, vector and matrix bias norms).

$$\{\mathbf{y} = \mathbf{A}\boldsymbol{\xi} | \mathbf{A} \in R^{n\times m}, \mathrm{rk}\mathbf{A} = n, n < m\}$$

ξ unknown

"*ansatz*"

$$\hat{\boldsymbol{\xi}} = \mathbf{Ly} \tag{2.1}$$

"*bias vector*"

$$\boldsymbol{\beta} := \boldsymbol{E}\{\hat{\boldsymbol{\xi}}\} - \boldsymbol{\xi} = -[\mathbf{I_m} - \mathbf{LA}]\,\boldsymbol{\xi},\ \ \forall \boldsymbol{\xi} \in \mathbb{R}^m \tag{2.2}$$

"*bias matrix*"

$$\mathbf{B}' = \mathbf{I_m} - \mathbf{LA} \tag{2.3}$$

"*bias norms*"

$$||\boldsymbol{\beta}||^2 = \boldsymbol{\beta}'\boldsymbol{\beta} = \boldsymbol{\xi}'[\mathbf{I_m} - \mathbf{LA}]'[\mathbf{I_m} - \mathbf{LA}]\boldsymbol{\xi} \tag{2.4}$$

$$||\boldsymbol{\beta}||^2 = \mathrm{tr}\boldsymbol{\beta}\boldsymbol{\beta}' = \mathrm{tr}[\mathbf{I_m} - \mathbf{LA}]\boldsymbol{\xi}\boldsymbol{\xi}'[\mathbf{I_m} - \mathbf{LA}]' = \ ||\mathbf{B}||^2_{\xi\xi'} \tag{2.5}$$

$$||\boldsymbol{\beta}||^2_{\mathbf{S}} : = \mathrm{tr}[\mathbf{I_m} - \mathbf{LA}]\mathbf{S}[\mathbf{I_m} - \mathbf{LA}]' =: \ ||\mathbf{B}||^2_{\mathbf{S}} \tag{2.6}$$

Since we deal with a linear model, it is "a natural choice" to setup a homogeneous linear form to estimate the parameters $\boldsymbol{\xi}$ of fixed effects, at first, namely (2.1) where is a matrix-valued fixed unknown. In order to determine the real-valued $m \times n$ matrix $\mathbf{L}$, the homogeneous linear estimation $\hat{\boldsymbol{\xi}}$ of the vector $\boldsymbol{\xi}$ of fixed effects has to fulfil a certain optimality condition, which we will outline. Let us try to analyze the bias in solving an underdetermined system of linear equations. Take reference to Box 2.1, where we systematically introduce (i) the bias vector $\boldsymbol{\beta}$, (ii) the bias matrix, (iii) the $\mathbf{S}$-modified bias matrix norm as a weighted Frobenius norm. In detail, let us discuss the bias terminology: For a homogeneous linear estimation (2.1), the vector-valued bias $\boldsymbol{\beta} := [\boldsymbol{E}\{\hat{\boldsymbol{\xi}}\} - \boldsymbol{\xi}]$ takes over the special form (2.2), which led us to the definition of the bias matrix $(\mathbf{I} - \mathbf{LA})'$. The norm of the bias vector $\boldsymbol{\beta}$, namely $||\boldsymbol{\beta}||^2 := \boldsymbol{\beta}'\boldsymbol{\beta}$, coincides with the $\boldsymbol{\xi}\boldsymbol{\xi}'$ weighted Frobenius norm of the bias matrix $\mathbf{B}$, namely $||\mathbf{B}||^2_{\xi\xi'}$. Here, we meet the central problem that the weight matrix $\boldsymbol{\xi}\boldsymbol{\xi}'$, $\mathrm{rk}\boldsymbol{\xi}\boldsymbol{\xi}' = 1$, has rank one. In addition, $\boldsymbol{\xi}\boldsymbol{\xi}'$ is not accessible since $\boldsymbol{\xi}$ is unknown. In this problematic case we replace the matrix $\boldsymbol{\xi}\boldsymbol{\xi}'$ by a fixed positive-definite $m \times m$ matrix $\mathbf{S}$, $\mathrm{rk}\mathbf{S} = m$, C.R. Rao's substitute matrix and define the $\mathbf{S}$-weighted Frobenius matrix norm (2.6) Indeed, the substitute matrix $\mathbf{S}$ constitutes the matrix of the metric of the bias space.

Being prepared for optimality criteria we give a precise definition of $\hat{\boldsymbol{\xi}}$ of type hom $\mathbf{S}$-LUMBE in Definition 2.1. The estimations $\hat{\boldsymbol{\xi}}$ of type hom $\mathbf{S}$-LUMBE can be characterized by Lemma 2.2.

Definition 2.1. ($\hat{\boldsymbol{\xi}}$ hom **S**-LUMBE of $\boldsymbol{\xi}$)

An $m \times 1$ vector $\hat{\boldsymbol{\xi}}$ is called hom S-LUMBE (homogeneous Linear Uniformly Minimum Bias Estimation) of $\boldsymbol{\xi}$ in the special consistent linear Gauss–Markov model of fixed effects of Box 2.1, if (1st) $\hat{\boldsymbol{\xi}}$ is a homogeneous linear form (2.7) (2nd) – in comparison to all other linear estimation $\hat{\boldsymbol{\xi}}$ has the minimum bias in the sense of (2.8).

$$\hat{\boldsymbol{\xi}} = \mathbf{Ly}, \tag{2.7}$$

$$||\mathbf{B}||^2_{\mathbf{S}} := \ ||(\mathbf{I_m} - \mathbf{LA})'||^2_{\mathbf{S}}. \tag{2.8}$$

Lemma 2.2. ($\hat{\boldsymbol{\xi}}$ hom **S**-LUMBE of $\boldsymbol{\xi}$).

An $m \times 1$ vector $\hat{\boldsymbol{\xi}}$ is hom S-LUMBE of $\boldsymbol{\xi}$ in the special consistent linear Gauss–Markov model with fixed effects of Box 2.1, if and only if the matrix $\hat{\boldsymbol{L}}$ fulfils the normal equations

$$\mathbf{ASA}'\hat{\mathbf{L}}' = \mathbf{AS}. \tag{2.9}$$

Proof.

The $\mathbf{S}$-weighted Frobenius norm $||\,(\mathbf{I_m} - \mathbf{LA})'\,||^2_{\mathbf{S}}$ establishes the Lagrangean (2.10) for $\mathbf{S}$-LUMBE. The necessary conditions for the minimum of the quadratic Lagrangean $\mathcal{L}(\mathbf{L})$ are provided by (2.11).

$$\mathcal{L}(\mathbf{L}) := \operatorname{tr}(\mathbf{I_m} - \mathbf{LA})\,\mathbf{S}\,(\mathbf{I_m} - \mathbf{LA})' = \min \mathbf{L}, \tag{2.10}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}\left(\hat{L}\right) := 2\left[\mathbf{ASA}'\hat{\mathbf{L}}' - \mathbf{AS}\right]' = 0. \tag{2.11}$$

The second derivatives (2.12) at the "point" $\hat{\mathbf{L}}$ constitute the sufficiency conditions. In order to compute such a $mn \times mn$ matrix of second derivatives we have to vectorize the matrix normal equation by (2.13)

$$\frac{\partial^2 \mathcal{L}}{\partial(\mathrm{vec}\mathbf{L})\partial(\mathrm{vec}\mathbf{L})'}\left(\hat{\mathbf{L}}\right) > 0, \tag{2.12}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}\left(\hat{\mathbf{L}}\right) = 2\left[\hat{\mathbf{L}}\mathbf{ASA}' - \mathbf{SA}'\right], \tag{2.13}$$

$$\frac{\partial \mathcal{L}}{\partial(\mathrm{vec}\mathbf{L})}\left(\hat{\mathbf{L}}\right) = \mathrm{vec}2\left[\hat{\mathbf{L}}\mathbf{ASA}' - \mathbf{SA}'\right], \tag{2.14}$$

$$\frac{\partial \mathcal{L}}{\partial(\mathrm{vec}\mathbf{L})}\left(\hat{\mathbf{L}}\right) = 2\left[\mathbf{ASA}' \otimes \mathbf{I_m}\right]\mathrm{vec}\hat{\mathbf{L}} - 2\mathrm{vec}\left(\mathbf{SA}'\right). \tag{2.15}$$

The Kronecker–Zehfuss product $\mathbf{A} \otimes \mathbf{B}$ of two arbitrary matrices as well as Kronecker-Zehfuss product $(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C}+\mathbf{B}\otimes\mathbf{C}$ of three arbitrary matrices subject to the dimension condition $\dim \mathbf{A} = \dim \mathbf{B}$ is introduced in Appendix A. The vec operation (vectorization of an array) is reviewed in Appendix A as well as $\mathrm{vec}\mathbf{AB} = (\mathbf{B}' \otimes \mathbf{I}'_\ell)\mathrm{vec}\mathbf{A}$ for suitable matrices $\mathbf{A}$ and $\mathbf{B}$. Now, we are prepared to compute (2.16) as a positive-definite matrix.

$$\frac{\partial^2 \mathcal{L}}{\partial(\mathrm{vec}\mathbf{L})\partial(\mathrm{vec}\mathbf{L})'}\left(\mathbf{L}'\right) = 2(\mathbf{ASA}') \otimes \mathbf{I_m} > 0 \tag{2.16}$$

For an explicit representation of $\hat{\boldsymbol{\xi}}$ of type hom LUMBE in the special *consistent* linear Gauss–Markov model of fixed effects of Box 2.1, we solve the normal equations for (2.17). Beside the explicit representation of $\hat{\boldsymbol{\xi}}$ of type hom LUMBE in (2.15) we attempt to calculate the bias by (2.19). Of course, the bias computation depends on C.R. Rao's substitute matrix $\mathbf{S}$, $\mathrm{rk}\mathbf{S} = m$. Indeed we can associate any element of the solution vector, the dispersion matrix as well as the bias vector with a particular weight which can be "designed" by the analyst.

$$\hat{\mathbf{L}} = \arg\{\mathcal{L}(\mathbf{L}) = \min_{\mathbf{L}}\} \tag{2.17}$$

Theorem 2.3. ($\hat{\boldsymbol{\xi}}$ hom LUMBE of $\boldsymbol{\xi}$)

Let $\hat{\boldsymbol{\xi}} = \mathbf{Ly}$ be hom LUMBE in the special consistent linear Gauss–Markov model of fixed effects of Box 2.1. Then the solution of the normal equation is completed by bias vector

$$\hat{\boldsymbol{\xi}} = \mathbf{SA}'(\mathbf{ASA}')^{-1}\mathbf{y} \tag{2.18}$$

$$\boldsymbol{\beta} := \boldsymbol{E}\{\hat{\boldsymbol{\xi}}\} - \boldsymbol{\xi} = -\left[\mathbf{I_m} - \mathbf{SA}'(\mathbf{ASA}')^{-1}\mathbf{A}\right]\boldsymbol{\xi},\ \forall \boldsymbol{\xi} \in \mathbb{R}^m. \tag{2.19}$$

Indeed, the bias computation is problematic since we have no *prior information* of the parameter vector $\boldsymbol{\xi}$. We could indeed use an a posteriori *i*nformation, for instance $\boldsymbol{\xi}_0 = \hat{\boldsymbol{\xi}}$.

2-2 The Equivalence Theorem of $\mathbf{G_x}$-MINOS and S-LUMBE

Of course, we have included the second chapter on hom $\mathbf{S}$-LUMBE in order to interpret $\mathbf{G_x}$-MINOS of the first chapter. When are hom $\mathbf{S}$-LUMBE and $\mathbf{G_x}$-MINOS equivalent is answered will by Theorem 2.4.

Theorem 2.4. (equivalence of $\mathbf{G}_x$-MINOS and $\mathbf{S}$-LUMBE).

With respect to the special consistent linear Gauss–Markov model (2.1), (2.2) $\hat{\boldsymbol{\xi}} = \mathbf{Ly}$ is hom $\mathbf{S}$-LUMBE for a positive-definite matrix $\mathbf{S}$ if $\boldsymbol{\xi}_\mathbf{m} = \mathbf{Ly}$ is $\mathbf{G_x}$-MINOS of the underdetermined system of linear equations (1.47) for

$$\mathbf{G_x} = \mathbf{S^{-1}} \sim \mathbf{G_x^{-1}} = \mathbf{S} \tag{2.20}$$

The proof is straight forward if we compare directly the solution (1.55) of $\mathbf{G_x}$-MINOS and (2.20) of hom $\mathbf{S}$-LUMBE. Obviously the inverse matrix of the metric of the parameter space $\mathbb{X}$ is equivalent to the matrix of the metric of the bias space $\mathbb{B}$. Or conversely, the inverse matrix of the metric of the bias space $\mathbb{B}$ determines the matrix of the metric of the parameter space $\mathbb{X}$. In particular, the bias vector $\boldsymbol{\beta}$ of type (2.19) depends on the vector $\boldsymbol{\xi}$ which is *inaccessible*. The situation is similar to the one in hypothesis testing. We can produce only an estimation $\hat{\boldsymbol{\beta}}$ of the bias vector $\boldsymbol{\beta}$ if we identify $\boldsymbol{\xi}$ by the hypothesis $\boldsymbol{\xi}_\mathbf{0} = \hat{\boldsymbol{\xi}}$.

2-3 Example

Due to the *Equivalence Theorem* $\mathbf{G_x}$-MINOS is equivalent to **S**-LUMBE the only new item is the *bias matrix* $\mathbf{B}\{\hat{\boldsymbol{\xi}}\,|\,\text{hom}\,LUMBE\}$. For our example, let us use the simple assumption $\mathbf{S} = \mathbf{I_m}$. Such an assumption is called "u.s." or "unity substituted" or unity substitute matrix. For our case the matrix $\mathbf{I_m} - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$ is idempotent. Note the fact that tr**A** = rk**A** is idempotent. Indeed the Frobenius norm of the u.s. bias matrix **B** (hom *LUMBE*) equalizes the square root $\sqrt{m-n} = \sqrt{d}$ of the right complementary index of the matrix **A**

$$||\mathbf{B}||^2 = \text{tr}[\mathbf{I_m} - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}] \tag{2.21}$$

$$||\mathbf{B}||^2 = d = m - n = m - \text{rk}\mathbf{A} \tag{2.22}$$

Box 2.2 summarizes those data outputs of the front examples of the first chapter relating to $||\mathbf{B}(\text{hom}\,BLUMBE)||$.

Box 2.2. Simple matrix of type u.s., Frobenius norm of the simple bias matrix, front page example.

$$\mathbf{A} \in \mathbb{R}^{2\times 3},$$

$$n = 2,\ m = 3,$$

$$\mathbf{A} := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \mathbf{A}\mathbf{A}' = \begin{bmatrix} 3 & 7 \\ 7 & 21 \end{bmatrix}, (\mathbf{A}\mathbf{A}')^{-1} = \frac{1}{14}\begin{bmatrix} 21 & -7 \\ -7 & 3 \end{bmatrix}$$

$$\text{rk}\mathbf{A} = 2$$

$$\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} = \frac{1}{14}\begin{bmatrix} 14 & -4 \\ 7 & -1 \\ -7 & 5 \end{bmatrix}, \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A} = \frac{1}{14}\begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix}$$

$$(\mathbf{A}\mathbf{A}')^{-2} = \frac{1}{98}\begin{bmatrix} 245 & -84 \\ -84 & 29 \end{bmatrix}, \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-2}\mathbf{A} = \frac{1}{98}\begin{bmatrix} 106 & 51 & -59 \\ 51 & 25 & -27 \\ -59 & -27 & 37 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\xi}}} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-2}\mathbf{A}\sigma^2_{\mathbf{y}} = \frac{1}{98}\begin{bmatrix} 106 & 51 & -59 \\ 51 & 25 & -27 \\ -59 & -27 & 37 \end{bmatrix}\sigma^2_{\mathbf{y}}$$

$$||\mathbf{B}||^2 = \text{tr}\left[\mathbf{I_m} - \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}\right] = \text{tr}\mathbf{I_3} - \text{tr}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$$

$$||\mathbf{B}||^2 = 3 - \tfrac{1}{14}(10 + 5 + 13) = 3 - 2 = 1 = d$$

$$||\mathbf{B}|| = 1 = \sqrt{d}.$$

Chapter 3
The Second Problem of Algebraic Regression

Overdetermined system of linear equations: $\{\mathbf{AX} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m}$ $\mathbf{y} \in \mathcal{R}(A) \sim rk\mathbf{A} = m,\ m = dim\,\mathbb{X}\}$

The optimization problem which appears in treating *overdetermined linear system* equations is a standard topic in any textbook on optimization. Here we consider again a weak nonlinear system as a problem which allows *a Taylor expansion*. We start from the *first section* with a front page example, an inconsistent linear system of a threedimensional observation space with a parameter space of two dimensions. Its *Least Squares Solution*, in short "LESS", for such an overdetermined linear system leads to the solution $\mathbf{x}_l = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$ subject to the condition of a positive-definite second derivative. Again – as an example for non-mathematicians – we derive the range and the kernel of the mapping, namely by "vertical rank partitioning." Finally, we give a geometric interpretation of the *Least Squares Solution*. In the *second section* we are turned to *weighted Least Squares* with weight matrix $\mathbf{G}_y$ in its semi-norm. As a generalized inverse (*C.R. Rao and S.K. Mitra* 1971, page 48–50, "three basic g-inverses, Sect. 3-2: g-inverse for a least squares solution") it produces the *Left Inverse*. We apply alternatively by the *method of Lagrange* multiplier, or the theorem of *extrema with side conditions* in (3.16)–(3.18), a technique which is based on the theory of inverse partitioned matrices of type "*C.R. Rao's Pandora Box*". We extend the theory by means of a canonical representation of a rank deficient matrix within the overdetermined observation space, namely the *eigenspace analysis* and the *eigenspace synthesis*, as an example for *non-mathematicians*. The discussion of the *metric of* the observation space leads us to the notion of the *weighted Least Squares Solution* with a remarkable extention to a *non-Euclidean* observation space. An *example* is the metric of type *von Mises-Fisher*, measuring the *spherical distance of a circle, a sphere* or *a hyper-sphere in Chap.* 7.

A *geodetic example* is the problem of an *optimal choice of the weight matrix* called *Second Order Design*. We call the *First Order Design* as the problem of *Datum Determination*. We take special reference to the *excellent textbook* by *Friedrich Pukelsheim* (*Optimal Design of Experiments*, J. Wiley Publication, New York, 1990). Here we base our analysis on the *Taylor-Karman structure* of a homogeneous and isotropic tensor-valued, two-point function of a twodimensional planar network.

For the example of *Gravimetric Leveling* we analyze the *projection matrices* by *horizontal rank partitioning*. The importance of the "*Hat Matrix*" for *robust estimates* is highlighted as well as the $\mathbf{G}_y$ *orthogonality between the matrices* **A** *and* **B**, namely *parameter fitting* as well as *applying conditions following G. Kampmann* et al. (1992, 1994, 1997), for example. The *Grassmann-Plücker coordinates* which

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_3,
© Springer-Verlag Berlin Heidelberg 2012

span the normal space R(A$^+$) will be discussed in *Chap. 10* when we *discuss condition equations*.

Of key importance is the application of the concept of *FUZZI Sets* which we define carefully. An example is given for *Deformation Analysis of Surface* referring to *E. Grafarend (2006), E. Grafarend and F. Krumm (2006), B. Voosoghi (2006)*, as well as *B. Waelder (2008)*. We give more than 20 references to application of *FUZZI SETS*.

A *special subsection* is devoted to the genesis of Gy-LESS and its related *Generalized Inverse* characterized by (3.38)–(3.39). Associated to this topic of our review of the special *Eigenvalue Decomposition* of $\mathbf{G}_y$-LESS called *canonical* and *canonical LESS*, illustrated by an *extensive example*.

A special section gives a number of *important case studies*:

1. Partial redundancies, latent conditions, high leverage points versus break points, direct and inverse GRASSMANN coordinates, Plcker coordinates: overview
2. *Multilinear Algebra*, "Join" and "Meet" operators, the HODGE Star operator
3. From A to B: latent restrictions, *Grassmann and Plücker coordinates*
4. From B to A: Latent parameter equations, *dual GRASSMANN coordinates, dual PLÜCKER coordinates*
5. *Break points, Gauss-Jacobi combinatorial Algorithms*

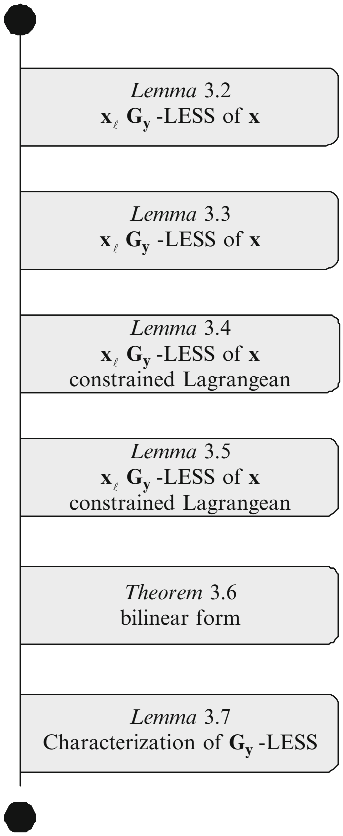
We conclude with a review of a family of means by direct observations as well as a *historical note of C.F. GAUSS and A.M. Legendre*.

Fast track reading: Read only *Lemma 3.7*.

By means of a certain algebraic objective function which geometrically is called a minimum distance function, we solve the *second inverse problem* of linear and nonlinear equations, in particular of algebraic type, which relate *observations to parameters*. The system of linear or nonlinear equations we are solving here is classified as *overdetermined*. The *observations*, also called *measurements*, are elements of a certain observation space $\mathbb{Y}$ of integer dimension, dim $\mathbb{Y} = n$, which may be metrical, especially Euclidean, pseudo-Euclidean, in general a differentiable manifold. In contrast, the *parameter space* $\mathbb{X}$ of integer dimension, dim $\mathbb{X} = m$, is *metrical* as well, especially *Euclidean*, *pseudo-Euclidean*, in general a differentiable manifold, but its metric is unknown. A typical feature of *algebraic regression* is the fact that the unknown metric of the *parameter space* $\mathbb{X}$ is induced by the functional relation between observations and parameters.

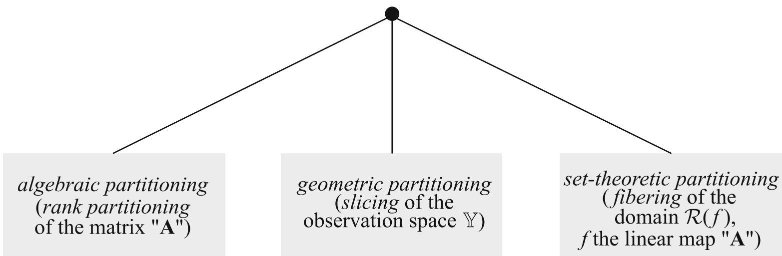
We shall outline three aspects of any *discrete inverse problem*: (i) set-theoretic (*fibering*), (ii) algebraic (*rank partitioning*, "IPM", the *Implicit Function Theorem*) and (iii) geometrical (*slicing*).

Here we treat the second problem of algebraic regression: An inconsistent system of linear observational equations: $\mathbf{AX} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m}$, $rk\mathbf{A} = m$, $n > m$, also called "overdetermined system of linear equations", in short



"*more observations than unknowns*"

is solved by means of an optimization problem. The *introduction* presents us a *front page example* of three inhomogeneous equations with two unknowns. In terms of 31 *boxes* and 12 *figures* we review the *least-squares solution* of such a inconsistent system of linear equations which is based upon the trinity.



3-1 Introduction

With the introductory paragraph we explain the *fundamental concepts* and *basic notions* of section. For you, *the analyst*, who has the difficult task to deal with measurements, observational data, modeling and modeling equations we present *numerical examples* and *graphical illustrations* of all *abstract notions*. The elementary introduction is written *not for a mathematician*, but for you, *the analyst*, with limited remote control of the notions given hereafter. *May we gain your interest*

Assume an n-dimensional *observation space* $\mathbb{Y}$, here a *linear space* parameterized by n observations (finite, discrete) as coordinates $\mathbf{y} = [y_1, \ldots, y_n]' \in \mathbb{R}^n$ in which an m-dimensional *model manifold* is embedded (immersed). The model manifold is described as the *range* of a *linear operator* f from an m-dimensional *parameter space* $\mathbb{X}$ into the *observation space* $\mathbb{Y}$. The mapping f is established by the mathematical equations which relate all observables to the unknown parameters. Here the *parameter space* $\mathbb{X}$, the domain of the linear operator f, will also be restricted to a *linear space* which is parameterized by coordinates $\mathbf{X} = [x_1, \ldots, x_n]' \in \mathbb{R}^m$. In this way the linear operator f can be understood as a coordinate mapping $\mathbf{A} : \mathbf{x} \mapsto \mathbf{y} = \mathbf{Ax}$. The linear mapping $f : \mathbb{X} \longrightarrow \mathbb{Y}$ is geometrically characterized by its *range* $\mathcal{R}(f)$, namely $\mathcal{R}(\mathbf{A})$, defined by $\mathcal{R}(f) := \{\mathbf{y} \in \mathbb{Y} | \mathbf{y} = f(\mathbf{x}\})\forall \mathbf{x} \in \mathbb{X}$ which in general is a linear subspace of $\mathbb{Y}$ and its *kernel* $\mathcal{N}(f)$, namely $\mathcal{N}(f)$, defined by $\mathcal{N}(f) := \mathbf{x} \in \mathbb{X} | f(\mathbf{x}) = 0\}$. Here the range $\mathcal{R}(f)$, namely $\mathcal{R}(\mathbf{A})$, *does not coincide* with the n-dimensional *observation space* $\mathbb{Y}$ such that $\mathbf{y} \notin \mathcal{R}(f)$, namely $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$. In contrast, we shall assume that the *null space* element $\mathcal{N}(f) = 0$ "*is empty*": it contains only the element $x = 0$.

Example 3.1 will therefore demonstrate the *range space* $\mathcal{R}(f)$, namely the *range space* $\mathcal{R}(\mathbf{A})$, which does not coincide with the *observation space* $\mathbb{Y}$, (f is not *surjective* or "onto") as well as the *null space* $\mathcal{N}(f)$, namely $\mathcal{N}(\mathbf{A})$, which is empty. f is *not* surjective, but *injective*. *Box 3.21* will introduce the special linear model of interest. By means of *Box 3.22* it will be interpreted.

3-11 The Front Page Example

Example 3.1. (polynomial of degree two, inconsistent system of linear equations $\{\mathbf{AX} + \mathbf{i} = \mathbf{y}, \mathbf{x} \in \mathbb{X} = \mathbb{R}^m,$ *dim*$\mathbb{X} = m, \mathbf{y} \in \mathbb{Y} = \mathbb{R}^n,$ $rk\mathbf{A} =$ *dim*$\mathbb{X} = m,$ $\mathbf{y} \in \mathcal{N}(A)$:

First, the *introductory example* solves the front page *inconsistent system of linear equations*,

$$\begin{aligned} x_1 + x_2 &\doteq 1 \\ x_1 + 2x_2 &\doteq 3 \\ x_1 + 3x_2 &\doteq 4 \end{aligned} \quad \texttt{or} \quad \begin{aligned} x_1 + x_2 + i_1 &= 1, \\ x_1 + 2x_2 + i_2 &= 3 \\ x_1 + 3x_2 + i_3 &= 4, \end{aligned}$$

obviously in general dealing with the linear space $\mathbf{X} = \mathbb{R}^m \ni \mathbf{x}$, dim $\mathbb{X} = m$, *here* m=2, called the *parameter space*, and the linear space dim $\mathbb{Y} = \mathbb{R}^n \ni \mathbf{y}$, *here* $n = 3$, called the *observation space*.

3-12 The Front Page Example in Matrix Algebra

Second, by means of *Box 3.1* and according to *A. Cayley's* doctrine let us specify the inconsistent system of linear equations in terms of *matrix algebra*.

Box 3.1. (Special linear model: polynomial of degree one, three observations, two unknowns).

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\Leftrightarrow \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} : \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$\mathbf{x}' = [x_1,\ x_2],\ \mathbf{y}' = [y_1,\ y_2,\ y_3] = [1,\ 2,\ 3],\ \mathbf{i}' = [i_1,\ i_2,\ i_3,]$$

$$\mathbf{A} \in \mathbb{Z}_+^{*3\times 2} \subset \mathbb{R}^{3\times 2},\ \mathbf{x} \in \mathbb{R}^{2\times 1},\ \mathbf{y} \in \mathbb{Z}_+^{*3\times 1} \subset \mathbb{R}^{3\times 1}$$

$$r = rk\mathbf{A} = \mathtt{dim}\,\mathbb{X} = m = 2.$$

As a linear mapping $f : \mathbf{x} \mapsto \mathbf{y} \doteq \mathbf{A}\mathbf{x}$ can be classified as following: f is *injective*, but *not surjective*. (A mapping f is called *linear* if $f(x_1+x_2) = f(x_1)+f(x_2)$ holds.) Denote the set of all $\mathbf{x} \in \mathbb{X}$ by the *domain* $\mathfrak{D}(f)$ *or* the domain space $\mathfrak{D}(\mathbf{A})$. Under the mapping f we generate a particular set called the *range* $\mathcal{R}(f)$ or the *range space* $\mathcal{R}(\mathbf{A})$. Since the set of all $\mathbf{y} \in \mathbb{Y}$ is *not* in the *range* $\mathcal{R}(f)$ *or* the *range space* $\mathcal{R}(\mathbf{A})$, namely $\mathbf{y} \notin \mathcal{R}(f)$ *or* $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$, the mapping f is *not surjective*. Beside the range $\mathcal{R}(f)$, the range space $\mathcal{R}(\mathbf{A})$, the linear mapping is characterized by the *kernel* $\mathcal{N}(\mathbf{f}) := \{\mathbf{x} \in \mathbb{R}^m | \mathbf{A}\mathbf{x} = \mathbf{0}\}$ or the null space $\mathcal{N}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{R}^m | \mathbf{A}\mathbf{x} = \mathbf{0}\}$. Since the inverse mapping $g : \mathcal{R}(f) \ni \mathbf{y}^\wedge \longmapsto \mathbf{x} \in \mathfrak{D}(f)$ is one-to-one, the mapping f is *injective*. Alternatively we may identify the *kernel* $\mathcal{N}(f)$, or the *null space* $\mathcal{N}(\mathbf{A})$ with $\{\mathbf{0}\}$.

? Why is the front page system of linear equations called inconsistent?

For instance, let us solve the first two equations, namely $x_1 = 0, x_2 = 1$. As soon as we substitute this solution in the third one, the inconsistency $3 \neq 4$ is met. Obviously such a system of linear equations needs general inconsistency parameters $(i_1,\ i_2,\ i_3)$ in order to avoid contradiction. Since the right-hand side of the equations, namely the inhomogeneity of the system of linear equations, has been measured as well as the model (the model equations) have been fixed, we have no alternative but inconsistency.

Within matrix algebra the index of the *linear operator* $\mathbf{A}$ is the *rank* $r = rk\mathbf{A}$, here $r = 2$, which coincides with the dimension of the *parameter space* $\mathbb{X}$, dim $\mathbb{X} = m$, namely $r = rk\mathbf{A} = dim\,\mathbb{X} = m$, *here* $r = m = 2$. In the terminology of the linear mapping f, f is *not* "*onto*" (surjective), but "one-to-one" (injective). The *left complementary index* of the linear operator $\mathbf{A} \in \mathbb{R}^{n\times m}$, which account for the *surjectivity defect* is given by $d_s = n - \mathrm{rk}\mathbf{A}$, also called "*degree of freedom*" (*here* $d_s = n - \mathrm{rk}\mathbf{A} = 1$). While "*surjectivity*" related to the range $\mathcal{R}(f)$ or "the *range space* $\mathcal{R}(\mathbf{A})$" and "*injectivity*" to the *kernel* $\mathcal{N}(f)$ or "the *null space* $\mathcal{N}(\mathbf{A})$" we shall constructively introduce the notion of

range $\mathcal{R}(f)$	*versus*	kernel $\mathcal{N}(f)$
range space $\mathcal{R}(\mathbf{A})$		*null space* $\mathcal{N}(f)$

by consequently solving the inconsistent system of linear equations. But beforehand let us ask:

How can such a linear model of interest, namely a system of inconsistent linear equations, be generated ? With reference to Box 3.2 let us assume that we have observed a dynamical system y(t) which is represented by a *polynomial of degree one* with respect to time namely

$$\mathrm{y}(t) = x_1 + x_2 t.$$

(Due to $\mathrm{y}^{\bullet}(t) = x_2$ it is a dynamical system with constant velocity or constant first derivative with respect to time t.) The *unknown* polynomial coefficients are collected in the *column array* $\mathbf{x} = [x_1, x_2]'$, $\mathbf{x} \in \mathbb{X} = \mathbb{R}^2$, $\mathtt{dim}\,\mathbb{X} = 2$, and constitute the coordinates of the two-dimensional *parameter space* $\mathbb{X}$. If the dynamical system y(t) is *observed* at *three instants*, say $y(t_1) = y_1 = 1, y(t_2) = y_2 = 2, y(t_3) = y_3 = 4$, and if we collect the *observations* in the column array $\mathbf{y} = [y_1, y_2, y_3]' = [1, 2, 4]'$, $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^3$, $\mathtt{dim}\,\mathbb{Y} = 3$, they constitute the coordinates of the three-dimensional *observation space* $\mathbb{Y}$. Thus we are left with the problem to compute *two unknown polynomial coefficients* from *three measurements*.

Box 3.2. (Special linear model: polynomial of degree one, three observations, two unknowns).

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} t_1 = 1, & y_1 = 1 \\ t_2 = 2, & y_2 = 2 \\ t_3 = 3, & y_3 = 4 \end{bmatrix} : \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}.$$

Thirdly, let us begin with a more detailed analysis of the linear mapping f : $\mathbf{Ax} = \mathbf{y}$ or $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, namely of the *linear operator*, $\mathbf{A} \in \mathbb{R}^{n \times m}$, $r = rk\mathbf{A} = dim\,\mathbb{X} = m$. We shall pay special attention to the *three fundamental partitionings*, namely

(i) Algebraic partitioning called *rank partitioning* of the matrix $\mathbf{A}$
(ii) Geometric partitioning called *slicing* of the linear space $\mathbb{Y}$ (observation space)
(iii) Set-theoretical partitioning called *fibering* of the set $\mathbb{Y}$ of observations.

3-13 Least Squares Solution of the Front Page Example by Means of Vertical Rank Partitioning

Let us go back to the front page inconsistent system of linear equations, namely *the problem* to determine two unknown polynomial coefficients from three sampling points which we classified as an *overdetermined one*. Nevertheless we are able to compute a unique solution of the overdetermined system of inhomogeneous linear equations $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$ or $r = rk\mathbf{A} = dim\,\mathbb{X}$, here $\mathbf{A} \in \mathbb{R}^{3\times 2}, \mathbf{x} \in \mathbb{R}^{2\times 1}, \mathbf{y} \in \mathbb{R}^{3\times 1}$ if we determine the coordinates of the unknown vector x as well as the vector i of the inconsistency by *least squares* (minimal Euclidean length, l_2-norm), here $\| \mathbf{i} \|_{\mathrm{I}}^2 = i_1^2 + i_2^2 + i_3^2 = \min$.

Box 3.3 outlines the solution of the related optimization problem.

Box 3.3. (Least squares solution of the inconsistent system of inhomogeneous linear equations, *vertical rank partitioning*).

The solution of the optimization problem

$$\{ \| \mathbf{i} \|_{\mathrm{I}}^2 = \min_{\mathbf{x}} | \mathbf{Ax} + \mathbf{i} = \mathbf{y},\ rk\mathbf{A} = \mathtt{dim}\,\mathbb{X}\}$$

is based upon the *vertical rank partitioning* of the linear mapping $f : \mathbf{x} \mapsto \mathbf{y} = \mathbf{Ax} + \mathbf{i}$, $\mathrm{rk}\mathbf{A} = \mathtt{dim}\,\mathbb{X}$ which we already introduced. As soon as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} \quad \text{subjectto}\mathbf{A}_1 \in \mathbb{R}^{r\times r}$$

$$\mathbf{x} = -\mathbf{A}_1^{-1}\mathbf{i}_1 + \mathbf{A}_1^{-1}\mathbf{y}_1$$

$$\mathbf{y}_2 = -\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{i}_1 + \mathbf{i}_2 + \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1$$

$$\Rightarrow \qquad \mathbf{i}_2 = \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{i}_1 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 + \mathbf{y}_2$$

is implemented in the norm $\| \mathbf{i} \|_{\mathbf{I}}^2$ we are prepared to compute the *first derivatives* of the *unconstrained Lagrangean*

$$\begin{aligned}
\mathcal{L}(\mathbf{i}_1, \mathbf{i}_2) :=\ & \| \mathbf{i} \|_{\mathbf{I}}^2 = \mathbf{i}'_1\mathbf{i}_1 + \mathbf{i}'_2\mathbf{i}_2 \\
&= \mathbf{i}'_1\mathbf{i}_1 + \mathbf{i}'_1\mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{i}_1 - 2\mathbf{i}'_1\mathbf{A}'^{-1}_1\mathbf{A}'_2(\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 - \mathbf{y}_2) \\
&\quad +(\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 - \mathbf{y}_2)'(\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 - \mathbf{y}_2) \\
&= \min_{\mathbf{i}_1} \\
\frac{\partial\mathcal{L}}{\partial\mathbf{i}_1}(\mathbf{i}_{1l}) &= 0 \\
&\Leftrightarrow -\mathbf{A}'^{-1}_1\mathbf{A}'_2(\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 - \mathbf{y}_2) + [\mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1} + \mathbf{I}]\,\mathbf{i}_{1l} = 0 \\
&\Leftrightarrow \quad \mathbf{i}_{1l} = [\mathbf{I} + \mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1}]^{-1}\mathbf{A}'^{-1}_1\mathbf{A}'_2(\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 - \mathbf{y}_2)
\end{aligned}$$

which constitute the *necessary conditions*.

1st term

$$\begin{aligned}
&(\mathbf{I} + \mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1})^{-1}\mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = (\mathbf{A}'_1 + \mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1})^{-1}\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 \\
&= \mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = -\mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_1\mathbf{y}_1 \\
&\quad + \mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}(\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1} + \mathbf{A}'_1)\mathbf{y}_1 = -\mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_1\mathbf{y}_1 \\
&\quad + (\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}(\mathbf{A}'_2\mathbf{A}_2 + \mathbf{A}'_1\mathbf{A}_1)\mathbf{y}_1 = -\mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_1\mathbf{y}_1 + \mathbf{y}_1
\end{aligned}$$

2nd term

$$\begin{aligned}
&-(\mathbf{I} + \mathbf{A}'^{-1}_1\mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1})^{-1}\mathbf{A}'^{-1}_1\mathbf{A}_2\mathbf{y}_2 = -(\mathbf{A}'_1 + \mathbf{A}'_2\mathbf{A}_2\mathbf{A}_1^{-1})^{-1}\mathbf{A}_2\mathbf{y}_2 \\
&= -\mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}\mathbf{A}'_2\mathbf{y}_2 \quad \Rightarrow \\
&\Leftrightarrow \quad \mathbf{i}_{1l} = -\mathbf{A}_1(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}(\mathbf{A}'_1\mathbf{y}_1 + \mathbf{A}'_2\mathbf{y}_2) + \mathbf{y}_1.
\end{aligned}$$

The *second derivatives*

$$\frac{\partial^2\mathcal{L}}{\partial\mathbf{i}_1\partial\mathbf{i}'_1}(\mathbf{i}_{1l}) = 2[(\mathbf{A}_2\mathbf{A}_1^{-1})'(\mathbf{A}_2\mathbf{A}_1^{-1}) + \mathbf{I}] > 0$$

due to positive-definiteness of the matrix $(\mathbf{A}_2\mathbf{A}_1^{-1})'(\mathbf{A}_2\mathbf{A}_1^{-1}) + \mathbf{I}$ generate the *sufficiency condition* for obtaining the minimum of the *unconstrained Lagrangean*. Finally let us backward transform $\mathbf{i}_{1l} \mapsto \mathbf{i}_{2l} = \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{i}_{1l} - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 + \mathbf{y}_2$.

$$\mathbf{i}_{2l} = -\mathbf{A}_2(\mathbf{A}'_1\mathbf{A}_1 + \mathbf{A}'_2\mathbf{A}_2)^{-1}(\mathbf{A}'_1\mathbf{y}_1 + \mathbf{A}'_2\mathbf{y}_2) + \mathbf{y}_2.$$

Obviously we have generated the linear form

$$\mathbf{i}_{1l} = -\mathbf{A}_1(\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1}(\mathbf{A'}_1\mathbf{y}_1 + \mathbf{A'}_2\mathbf{y}_2) + \mathbf{y}_1$$
$$\mathbf{i}_{2l} = -\mathbf{A}_2(\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1}(\mathbf{A'}_1\mathbf{y}_1 + \mathbf{A'}_2\mathbf{y}_2) + \mathbf{y}_2$$

or

$$\begin{bmatrix}\mathbf{i}_{1l}\\ \mathbf{i}_{2l}\end{bmatrix} = -\begin{bmatrix}\mathbf{A}_1\\ \mathbf{A}_2\end{bmatrix}(\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1}[\mathbf{A'}_1,\ \mathbf{A'}]\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} + \begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}$$

or

$$\mathbf{i}_l = -\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{y} + \mathbf{y}.$$

Finally we are left with the backward step to compute the unknown vector of parameters $\mathbf{x} \in \mathbb{X}$:

$$\mathbf{x}_l = -\mathbf{A}_1^{-1}\mathbf{i}_{1l} + \mathbf{A}_1^{-1}\mathbf{y}_1$$

$$\mathbf{x}_l = (\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1}(\mathbf{A'}_1\mathbf{y}_1 + \mathbf{A'}_2\mathbf{y}_2)$$

or

$$\mathbf{x}_l = (\mathbf{A'A})^{-1}\mathbf{A'y}.$$

A numerical computation with respect to the introductory example is

$$\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2 = \begin{bmatrix}3 & 6\\ 6 & 14\end{bmatrix},\quad (\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1} = \frac{1}{6}\begin{bmatrix}14 & -6\\ -6 & 3\end{bmatrix},$$

$$\mathbf{A}_1(\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1} = \frac{1}{6}\begin{bmatrix}8 & -3\\ 2 & 0\end{bmatrix},$$

$$\mathbf{A}_2(\mathbf{A'}_1\mathbf{A}_1 + \mathbf{A'}_2\mathbf{A}_2)^{-1} = \frac{1}{6}[-4,\ 3],$$

$$\mathbf{A'}_1\mathbf{y}_1 + \mathbf{A'}_2\mathbf{y}_2 = \begin{bmatrix}8\\ 19\end{bmatrix},\quad \mathbf{y}_1 = \begin{bmatrix}1\\ 3\end{bmatrix},\quad \mathbf{y}_2 = 4$$

$$\mathbf{i}_{1l} = \frac{1}{6}\begin{bmatrix}-1\\ 2\end{bmatrix},\quad \mathbf{i}_{2l} = -\frac{1}{6},\quad \|\ \mathbf{i}_l\ \|_{\mathbf{I}} = \frac{1}{6}\sqrt{6},$$

$$\mathbf{x}_l = \frac{1}{6}\begin{bmatrix}-2\\ 9\end{bmatrix},\quad \|\ \mathbf{x}_l\ \| = \frac{1}{6}\sqrt{85},$$

$$\mathrm{y}(t) = -\frac{1}{3} + \frac{3}{2}t$$

$$\frac{1}{2}\frac{\partial^2\mathcal{L}}{\partial\mathbf{x_2}\partial\mathbf{x'_2}}(\mathbf{x}_{2m}) = [(\mathbf{A}_2\mathbf{A}_1^{-1})'(\mathbf{A}_2\mathbf{A}_1^{-1}) + \mathbf{I}] = \begin{bmatrix}2 & -2\\ -2 & 5\end{bmatrix} > 0,$$

"First eigenvalue $\lambda_1 \left(\begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix} \right) = 6$", "Second eigenvalue λ_2 $\left(\begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix} \right) = 1$."

The *diagnostic algorithm* for solving an overdetermined system of linear equations $\mathbf{y} = \mathbf{Ax}$, $\text{rk}\mathbf{A} = \dim \mathbb{X} = m$, $m < n = \dim \mathbb{Y}$, $\mathbf{y} \neq \mathbb{Y}$ by means of rank partitioning is presented to you by *Box 3.4*.

3-14 *The Range $\mathcal{R}(f)$ and the Kernel $\mathcal{N}(f)$, Interpretation of "LESS" by Three Partitionings*

(i) algebraic (rank partitioning)
(ii) geometric (slicing)
(iii) set-theoretical (fibering)

Fourthly, let us go into the detailed analysis of $\mathcal{R}(f)$, $\mathcal{R}(f)^{\perp}$, $\mathcal{N}(f)$, with respect to the front page example. Beforehand we begin with a comment.

We want to emphasize the two step procedure of the *least squares solution* (LESS) once more: The *first step* of LESS maps the observation vector y onto the range space $\mathcal{R}(f)$ while in the *second step* the LESS point $\mathbf{y} \in \mathcal{R}(\mathbf{A})$ is uniquely mapped to the point $\mathbf{x}_l \in \mathbb{X}$, an element of the parameter space. Of course, we directly produce $\mathbf{x}_l = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$ just by substituting the inconsistency vector $\mathbf{i} = \mathbf{y} - \mathbf{Ax}$ into the $l_2 -$ norm $\| \mathbf{i} \|_{\mathbf{I}}^2 = (\mathbf{y} - \mathbf{Ax})'(\mathbf{y} - \mathbf{Ax}) = \min$. Such a direct procedure which is *common practice* in LESS does not give any insight into the geometric structure of LESS.

But how to identify the range $\mathcal{R}(f)$, namely the *range space* $\mathcal{R}(\mathbf{A})$, or the kernel $\mathcal{N}(f)$, namely the null space $\mathcal{N}(\mathbf{A})$ in the front page example?

By means of *Box 3.4* we identify $\mathcal{R}(f)$ or "*the null space* $\mathcal{R}(\mathbf{A})$" and give its illustration by *Fig. 3.1*. Such a result has paved the way to the diagnostic algorithm for solving an overdetermined system of linear equations by means of rank partitioning presented in *Box 3.5*.

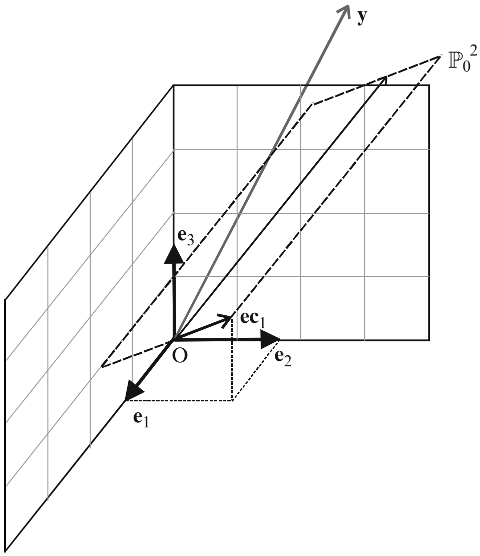
The *kernel* $\mathcal{N}(f)$ or "the null space" is immediately identified as $\{0\} = \mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{Ax} = 0\} = \{\mathbf{x} \in \mathbb{R}^m | \mathbf{A}_1\mathbf{x} = 0\}$ by means of rank partitioning ($\mathbf{A}_1\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} = \mathbf{0}$).

Box 3.4. (The range space of the system of inconsistent linear equations $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, "vertical" rank partitioning)

The matrix $\mathbf{A}$ is called "*vertically rank partitioned*", if

$$\left\{ \mathbf{A} \in \mathbb{R}^{n \times m} \wedge \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \middle| \begin{array}{l} r = \text{rk}\mathbf{A} = \text{rk}\mathbf{A}_1 = m, \\ \mathbf{A}_1 \in \mathbb{R}^{r \times r},\ \mathbf{A}_2 \in \mathbb{R}^{d \times r} \\ d = d(\mathbf{A}) = m - rk\mathbf{A} \end{array} \right\}$$

Fig. 3.1 Range $\mathcal{R}(f)$, range space $\mathcal{R}(\mathbf{A})$, $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$, observation space $\mathbb{Y} = \mathbb{R}^3$, slice by $\mathbf{y} = \mathbf{e}_1 u + \mathbf{e}_2 \mathrm{v} + \mathbf{e}_3(-u + 2\mathrm{v}) \in \mathcal{R}(\mathbf{A})$



holds. (In the introductory example $\mathbf{A} \in \mathbb{R}^{3\times 2}$, $\mathbf{A}_1 \in \mathbb{R}^{2\times 2}$, $\mathbf{A}_2 \in \mathbb{R}^{1\times 2}$, $rk\mathbf{A} = 2$, $d(\mathbf{A}) = 1$ applies.) An inconsistent system of linear equations is "vertically rank partitioned" if

$$\mathbf{Ax} = \mathbf{y},\ \mathrm{rk}\mathbf{A} = \mathtt{dim}\mathbb{X} \Leftrightarrow \ \mathbf{y} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix}$$
$$\Leftrightarrow \quad \begin{bmatrix} \mathbf{y}_1 = \mathbf{A}_1\mathbf{x} + \mathbf{i}_1 \\ \mathbf{y}_2 = \mathbf{A}_2\mathbf{x} + \mathbf{i}_2 \end{bmatrix}$$

for a partitioned observation vector

$$\{\mathbf{y} \in \mathbb{R}^n \wedge \ \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \ \mid \ \mathbf{y}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{y}_2 \in \mathbb{R}^{d\times 1}\}$$

and a partitioned inconsistency vector

$$\{\mathbf{i} \in \mathbb{R}^n \wedge \ \mathbf{i} = \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} \ \mid \ \mathbf{i}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{i}_2 \in \mathbb{R}^{d\times 1}\},$$

respectively, applies. (The "vertical" *rank partitioning* of the matrix $\mathbf{A}$ as well as the "vertically rank partitioned" inconsistent system of linear equations $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, $rk\mathbf{A} = \mathtt{dim}\mathbb{X} = m$, of the introductory example is

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \mathbf{A} = \left[\frac{\mathbf{A}_1}{\mathbf{A}_2}\right] = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix},$$

$$\left[\frac{\mathbf{y}_1}{\mathbf{y}_2}\right] = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{y}_1 \in \mathbb{R}^{2\times 1}, \quad \mathbf{y}_2 \in \mathbb{R}.$$

By means of the vertical rank partitioning of the inconsistent system of *inhomogeneous linear equations* an identification of the range space $\mathcal{R}(\mathbf{A})$, namely

$$\mathcal{R}(A) = \{y \in \mathbb{R}^n | \mathbf{y}_2 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = \mathbf{0}\}$$

is based upon

$$\begin{aligned} \mathbf{y}_1 = \mathbf{A}_1\mathbf{x} + \mathbf{i}_1 \quad &\Rightarrow \quad \mathbf{x}_1 = -\mathbf{A}_1^{-1}(\mathbf{y}_1 - \mathbf{i}_1) \\ \mathbf{y}_2 = \mathbf{A}_2\mathbf{x} + \mathbf{i}_2 \quad &\Rightarrow \quad \mathbf{x}_2 = -\mathbf{A}_2\mathbf{A}_1^{-1}(\mathbf{y}_1 - \mathbf{i}_1) + \mathbf{i}_2 \quad \Rightarrow \end{aligned}$$

$$\mathbf{y}_2 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = \mathbf{i}_2 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{i}_1$$

which leads to the *range space* $\mathcal{R}(\mathbf{A})$ for inconsistency zero, particularly in the *introductory example*

$$y_3 - [1, \; 3]\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0.$$

For instance, if we introduce the coordinates $y_1 = u, \; y_2 = v$, the other coordinate y_3 of the range space $\mathcal{R}(\mathbf{A}) \subset \mathbb{Y} = \mathbb{R}^3$ amounts to

$$y_3 = [1, \; 3]\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = [-1, \; 2]\begin{bmatrix} u \\ v \end{bmatrix}$$
$$\Rightarrow y_3 = -u + 2v.$$

In geometric language the linear space $\mathcal{R}(\mathbf{A})$ is a parameterized plane through the origin illustrated by *Fig. 3.1*. The *observation space* $\mathbb{Y} = \mathbb{R}^n$ (here $n = 3$) is *sliced* by the subspace, the linear space (linear manifold) $\mathcal{R}(\mathbf{A})$, $\dim \mathcal{R}(\mathbf{A}) = \mathrm{rk}(\mathbf{A}) = r$, namely a straight line, a plane (here), a higher dimensional plane through the origin $\mathbf{O}$.

What is the geometric interpretation of the least-squares solution $\| \mathbf{i} \|_{\mathbf{I}}^2 =$ min?

With reference to Fig. 3.2 we additively decompose the observation vector accordingly to

Box 3.5:
Algorithm
Diagnostic algorithm for solving an overdetermined system of linear equations $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$, rk $\mathbf{A} = \dim \mathbf{X}$, $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$ by means of *rank partitioning*

Determine
the rank of the matrix $\mathbf{A}$
rk $\mathbf{A} = \dim \mathbb{X} = m$

↓

Compute
the "*vertical rank partitioning*"

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix},\ \mathbf{A}_1 \in \mathbb{R}^{r\times r} = \mathbb{R}^{m\times n},\ \mathbf{A}_2 \in \mathbb{R}^{(n-r)\times r} = \mathbb{R}^{(n-m)\times m}$$

"$n - r = n - m = d_s$ *is called left complementary index*"
"$\mathbf{A}$ as a linear operator is *not* surjective, *but* injective"

↓

Compute
the *range space* $\mathcal{R}(\mathbf{A})$

$$\mathcal{R}(\mathbf{A}) := \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}_2 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = 0\}$$

↓

Compute
the *inconsistency vector* of type LESS

$$\mathbf{i}_l = -\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{y} + \mathbf{y}$$

test : $\mathbf{A}'\mathbf{i}_l = 0$

↓

Compute
the *unknown parameter vector* of type LESS

$$\mathbf{x}_l = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}.$$

♣

$$\mathbf{y} = \mathbf{y}_{\mathcal{R}(\mathbf{A})} + \mathbf{y}_{\mathcal{R}(\mathbf{A})^\perp},$$

where $\mathbf{y}_{\mathcal{R}(\mathbf{A})} \in \mathcal{R}(\mathbf{A})$ is an element of the *range space* $\mathcal{R}(\mathbf{A})$, but the inconsistency vector $\mathbf{i}_l = \mathbf{i}_{\mathcal{R}(\mathbf{A})^\perp} \in \mathcal{R}(\mathbf{A})^\perp$ an element of its *orthogonal complement*, the *normal space* $\mathcal{R}(\mathbf{A})^\perp$. Here$\mathcal{R}(\mathbf{A})$ is the central plane $\mathbb{P}_0^2$, $\mathbf{y}_{\mathcal{R}(\mathbf{A})\in\mathbb{P}_0^2}$, but $\mathcal{R}(\mathbf{A})^\perp$ the straight line $\mathbb{L}^1$, $\mathbf{i}_l \in \mathcal{R}(\mathbf{A})^\perp$. $\|\ \mathbf{i}\ \|_{\mathbf{I}}^2 = \|\ \mathbf{y} - \mathbf{y}_{\mathcal{R}(\mathbf{A})^\perp}\ \|^2 =$ min can be understood as the *minimum distance mapping* of the observation point $\mathbf{y} \in \mathbb{Y}$ onto the range space $\mathcal{R}(\mathbf{A})$. Such a mapping is *minimal*, if and only if the inner product $\langle \mathbf{y}_{\mathcal{R}(\mathbf{A})} | \ \mathbf{i}_{\mathcal{R}(\mathbf{A})^\perp} \rangle = 0$ approaches zero, we say

"$\mathbf{y}_{\mathcal{R}}(\mathbf{A})$ and $\mathbf{i}_{\mathcal{R}}(\mathbf{A})^\perp$ are orthogonal"

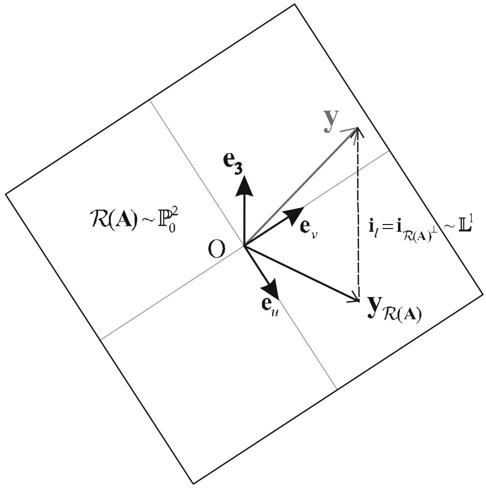


Fig. 3.2 Orthogonal projection of the *observation vector* $\mathbf{y} \in \mathbb{Y}$ onto the range space $\mathcal{R}(\mathbf{A})$, $\mathcal{R}(\mathbf{A}) := \mathbb{R}^n|\{\mathbf{y}_2 - \mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{y}_1 = 0\}$, $\mathbf{i}_l \in \mathcal{R}(\mathbf{A})^\perp$, here: $\mathbf{y}_{\mathcal{R}(\mathbf{A}} \in \mathbb{P}_0^2$ (*central plane*), $\mathbf{y} \in \mathbb{L}^1$ (*straight line*), representation of $\mathbf{y}_{\mathcal{R}(\mathbf{A})}$(LESS): $\mathbf{y} = \mathbf{e}_1 u + \mathbf{e}_2 v + \mathbf{e}_3(-u+2v) \in \mathbb{R}^3$, $\mathcal{R}(\mathbf{A}) = \mathrm{span}\{\mathbf{e}_u, \mathbf{e}_v\}$

The solution point $\mathbf{y}_{\mathcal{R}(\mathbf{A})}$ is the orthogonal projection of the observation point $\mathbf{y} \in \mathbb{Y}$ onto the range space $\mathcal{R}(\mathbf{A})$, an m-dimensional linear manifold, also called a *Grassmann manifold* $\mathrm{G}^{n,m}$.

$$\textit{Gram–Schmidt}: \left[\begin{array}{l} \mathbf{e}_u := \mathbf{D}_u\mathbf{y}_{\mathcal{R}(\mathbf{A})} / \parallel \mathbf{D}_u\mathbf{y}_{\mathcal{R}(\mathbf{A})} \parallel = (\mathbf{e}_1 - \mathbf{e}_3)/\sqrt{2} \\ \mathbf{e}_v := \dfrac{\mathbf{D}_v\mathbf{y}_{\mathcal{R}(\mathbf{A})} - < \mathbf{D}_v\mathbf{y}_{\mathcal{R}(\mathbf{A})}|\mathbf{e}_u > \mathbf{e}_u}{\parallel \mathbf{D}_v\mathbf{y}_{\mathcal{R}(\mathbf{A})} - < \mathbf{D}_v\mathbf{y}_{\mathcal{R}(\mathbf{A})}|\mathbf{e}_u > \mathbf{e}_u \parallel} = (\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3)/\sqrt{3} \\ < \mathbf{e}_u|\mathbf{e}_v > = 0, \ \mathbf{D}_v\mathbf{y}_{\mathcal{R}(\mathbf{A})} = \mathbf{e}_2 + 2\mathbf{e}_3 \end{array}\right.$$

As an "*intermezzo*" let us consider for a moment the *nonlinear model* by means of the nonlinear mapping

$$\text{“}\mathbb{X} \ni \mathbf{x} \mapsto f(\mathbf{x}) = \mathbf{y}_{R(\mathbf{A})}, \ \mathbf{y} \in \mathbb{Y}\text{''}.$$

In general, the observation space $\mathbb{Y}$ as well as the parameter space $\mathbb{X}$ may be considered as *differentiable manifolds*, for instance "*curved surfaces*". The range $\mathcal{R}(f)$ may be interpreted as the *differentiable manifolds*. $\mathbb{X}$ embedded or more generally *immersed*, in the *observation space* $\mathbb{Y} = \mathbb{R}^n$ for instance: $\mathbb{X} \subset \mathbb{Y}$. The parameters $[x_1, \ldots, x_m]$ constitute a chart of the differentiable manifolds $\mathbb{X} = \mathbb{M}^m \subset \mathbb{M}^m = \mathbb{Y}$. Let us assume that a point $p \in \mathcal{R}(f)$ is given and we are going to attach the *tangent space* $\mathbb{T}_p\mathbb{M}^m$ locally. Such a tangent space $\mathbb{T}_p\mathbb{M}^m$ at $p \in \mathcal{R}(f)$ may be constructed by means of the *Jacobi map*, parameterized by the *Jacobi matrix* $\mathbf{J}, rk\mathbf{J} = m$, a standard procedure in *Differential Geometry*. An *observation point* $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n$ is orthogonally projected onto the *tangent space* $\mathbb{T}_p\mathbb{M}^m$ at $p \in \mathcal{R}(f)$, namely by LESS as a *minimum distance mapping*. In a second step – in common use is the equidistant mapping – we bring the point which is located in the *tangent space* $\mathbb{T}_p\mathbb{M}^m$ at $p \in \mathcal{R}(f)$ back to the differentiable manifold, namely $\mathbf{y}_- \in \mathcal{R}(f)$. The *inverse map*

$$\text{“}\mathcal{R}(f) \ni \mathbf{y}_- \mapsto \; g(\mathbf{y}_-) = \mathbf{x}_l \in \mathbb{X}\text{''}$$

maps the point $\mathbf{y}_- \in \mathcal{R}(f)$ to the point $\mathbf{x}_l$ of the *chosen chart* of the parameter $\mathbb{X}$ as a differentiable manifold. Examples follow later on.

Let us continue with the geometric interpretation of the linear model of this paragraph. The *range space* $\mathcal{R}(\mathbf{A})$, $\dim \mathcal{R}(\mathbf{A}) = \mathrm{rk}(\mathbf{A}) = m$ is a *linear space* of dimension m, here $m = rk\mathbf{A}$, which *slices* $\mathbb{R}^n$. In contrast, the subspace $\mathcal{R}(\mathbf{A})^\perp$ corresponds to a $n - rk\mathbf{A} = d_s$ dimensional linear space $\mathbb{L}^{n-r}$ here $n - rk\mathbf{A} = n - m, r = rk\mathbf{A} = m$.

Let the algebraic partitioning and the geometric partitioning be merged to interpret the *least squares solution* of the inconsistent system of linear equations as a *generalized inverse* (g-inverse) of type LESS. As a summary of such a merger we take reference to Box 3.7.

The first condition:

$$\mathbf{AA}^-\mathbf{A} = \mathbf{A}$$

Let us depart from LESS of $\mathbf{y} = \mathbf{Ax} + \mathbf{i}$, namely

$$\mathbf{x}_l = \mathbf{A}_l^-\mathbf{y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}, \; \mathbf{i}_l = (\mathbf{I} - \mathbf{AA}_l^-)\mathbf{y} = [\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{y}.$$

$$\left.\begin{array}{r} \mathbf{Ax}_l = \mathbf{AA}_l^-\mathbf{y} = \mathbf{AA}_l^-(\mathbf{Ax}_l + \mathbf{i}_l) \\ \mathbf{A}'\mathbf{i}_l = \mathbf{A}'[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{y} = 0 \Rightarrow \mathbf{A}_l^-\mathbf{i}_l = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{i}_l = 0 \end{array}\right] \Rightarrow$$

$$\Rightarrow \quad \mathbf{Ax}_l = \mathbf{AA}_l^-\mathbf{Ax}_l \quad \Leftrightarrow \quad \mathbf{AA}^-\mathbf{A} = \mathbf{A}.$$

The second condition

$$\mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-$$

$$\left.\begin{array}{c} \mathbf{x}_l = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{A}_l^-\mathbf{y} = \mathbf{A}_l^-(\mathbf{Ax}_l + \mathbf{i}_l) \\ \mathbf{A}_l^-\mathbf{i}_l = 0 \end{array}\right] \Rightarrow$$

$$\mathbf{x}_l = \mathbf{A}_l^-\mathbf{y} = \mathbf{A}_l^-\mathbf{AA}_l^-\mathbf{y} \qquad \Rightarrow$$

$$\Rightarrow \quad \mathbf{A}_l^-\mathbf{y} = \mathbf{A}_l^-\mathbf{AA}_l^-\mathbf{y} \quad \Leftrightarrow \quad \mathbf{A}^-\mathbf{AA}^- = \mathbf{A}^-.$$

$rk\mathbf{A}_l^- = rk\mathbf{A}$ is interpreted as following: the g-inverse of type LESS is the generalized inverse of *maximal rank* since in general $rk\mathbf{A}^- \leq rk\mathbf{A}$ holds.

The third condition

$$\mathbf{AA}^- = \mathbb{P}_{\mathcal{R}(\mathbf{A}^-)}$$

$$\left.\begin{array}{c}\mathbf{y}=\mathbf{A}\mathbf{x}_l+\mathbf{i}_l=\mathbf{A}\mathbf{A}_l^-+(\mathbf{I}-\mathbf{A}\mathbf{A}_l^-)\mathbf{y}\\ \mathbf{y}=\mathbf{A}\mathbf{x}_l+\mathbf{i}_l=\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}+[\mathbf{I}-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{y}\\ \mathbf{y}=\mathbf{y}_{\mathcal{R}(\mathbf{A})}+\mathbf{i}_{\mathcal{R}(\mathbf{A})^\perp}\end{array}\right]\Rightarrow$$

$$\Rightarrow \mathbf{A}^-\mathbf{A}=\mathbb{P}_{\mathcal{R}(\mathbf{A}^-)},(\mathbf{I}-\mathbf{A}\mathbf{A}^-)=\mathbb{P}_{\mathcal{R}(\mathbf{A}^-)^\perp}.$$

Obviously $\mathbf{A}\mathbf{A}_l^-$ is an *orthogonal projection* onto $\mathcal{R}(\mathbf{A})$, but $\mathbf{I}-\mathbf{A}\mathbf{A}_l^-$ onto its *orthogonal complement* $\mathcal{R}(\mathbf{A})^\perp$.

Box 3.6. (The three condition of the generalized inverse mapping (generalized inverse matrix) LESS type).

Condition #1	*Condition* #1
$f(\mathbf{x})=f(g(\mathbf{y}))$	$\mathbf{A}\mathbf{x}=\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x}$
$\Leftrightarrow$	$\Leftrightarrow$
$f=f\circ g\circ f$	$\mathbf{A}\mathbf{A}^-\mathbf{A}=\mathbf{A}$
Condition #2	*Condition* #2
$\mathbf{x}=g(\mathbf{y})$	$\mathbf{x}_-=\mathbf{A}^-\mathbf{y}=\mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{y}$
$=g(f(\mathbf{x}))$	$\Leftrightarrow\ \mathbf{A}^-\mathbf{A}\mathbf{A}^-=\mathbf{A}^-$
Condition #3	*Condition* #3
(*reflexive g-inverse mapping*)	(*reflexive g-inverse*)
$f(g(\mathbf{y}))=\mathbf{y}_{\mathcal{R}(\mathbf{A})}$	$\mathbf{A}^-\mathbf{A}\mathbf{y}=\mathbf{y}_{\mathcal{R}(\mathbf{A})}$
$\Leftrightarrow$	$\Leftrightarrow$
$f\circ g=\mathrm{proj}_{\mathcal{R}(\mathrm{f})}$	$\mathbf{A}^-\mathbf{A}=\mathbf{P}_{\mathcal{R}(\mathbf{A})}$.

The *set-theoretical partitioning*, the fibering of the *set system of points* which constitute the *observation space* $\mathbb{Y}$, the range $\mathcal{R}(f)$, will be finally outlined. Since the set system $\mathbb{Y}$ (the observation space) is $\mathbb{R}^n$, the fibering is called "*trivial*". Non-trivial fibering is reserved for *nonlinear models* in which case we are dealing with a observation space as well as an range space which is a *differentiable manifold*. Here the fibering

$$\mathbb{Y}=\mathcal{R}(f)\cup\mathcal{R}(f)^\perp$$

produces the *trivial fibers* $\mathcal{R}(f)$ and $\mathcal{R}(f)^\perp$ where the *trivial fibers* $\mathcal{R}(f)^\perp$ is the *quotient set* $\mathbb{R}^n/\mathcal{R}(f)$. By means of a *Venn diagram* (John Venn 1834–1928) also called *Euler circles* (Leonhard Euler 1707–1783). Figure 3.3 illustrates the *trivial fibers of the set system* $\mathbb{Y}=\mathbb{R}^n$ generated by $\mathcal{R}(f)$ and $\mathcal{R}(f)^\perp$. The set system of points which constitute the *parameter space* $\mathbb{X}$ is not subject to *fibering* since all points of the set system $\mathcal{R}(f)$ are mapped *into* the domain $\mathfrak{D}(f)$.

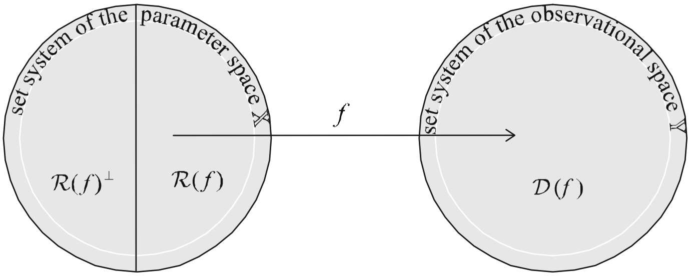


Fig. 3.3 *Venn diagram*, trivial fibering of the observation space $\mathbb{Y}$, trivial fibers $\mathcal{R}(f)$ and $\mathcal{R}(f)^{\perp}$, $f : \mathbb{R}^m = \mathbb{X} \longmapsto \mathbb{Y} = \mathcal{R}(f) \cup \mathcal{R}(f)^{\perp}$, $\mathbb{X}$ set system of the parameter space, $\mathbb{Y}$ set system of the observation space

3-2 The Least Squares Solution: "LESS"

The system of inconsistent linear equations $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$ subject to $\mathbf{A} \in \mathbb{R}^{n \times m}$. rk$\mathbf{A} = m < n$, allows certain solutions which we introduce by means of *Definition* 3.1 as a solution of a certain optimization problem. *Lemma* 3.2 contains the normal equations of the optimization problem. The solution of such a system of normal equations is presented in *Lemma 3.3* as the least squares solution with respect to the $\mathbf{G_y}$-norm. Alternatively *Lemma 3.4* shows the least squares solution generated by a constrained Lagrangean. Its normal equations are solved for (i) the *Lagrange multiplier*, (ii) the unknown vector of inconsistencies by *Lemma 3.5*. The unconstrained Lagrangean where the system of linear equations has been implemented as well as the constrained Lagrangean lead to the identical solution for (i) the vector of inconsistencies *and* (ii) the vector of unknown parameters. Finally we discuss the metric of the observation space and *alternative choices* of its *metric* before we identify the solution of the quadratic optimization problem by *Lemma 3.7* in terms of the (1, 2, 3)-*generalized inverse*.

Definition 3.1. (least squares solution w.r.t. the $\mathbf{G_y}$-seminorm):

A vector $\mathbf{x}_l \in \mathbb{X} = \mathbb{R}^m$ is called $\mathbf{G_y}$-LESS (LEast Squares Solution with respect to the $\mathbf{G_y}$-seminorm) of the inconsistent system of linear equations

$$\mathbf{Ax} + \mathbf{i} = \mathbf{y},\ \ \mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n, \left[\begin{array}{c} rk\,\mathbf{A} = \mathtt{dim}\mathbb{X} = m \\ or \\ y \notin \mathcal{R}(\mathbf{A}) \end{array}\right. \tag{3.1}$$

(the system of inverse linear equations $\mathbf{A}^{-}\mathbf{y} = \mathbf{x}$, $rk\mathbf{A}^{-} = \mathtt{dim}\mathbb{X} = m$ or $\mathbf{x} \in \mathcal{R}(\mathbf{A}^{-})$, is consistent) if in comparison to all other vectors $\mathbf{x} \in \mathbb{X} \equiv \mathbb{R}^m$, the inequality

$$\begin{aligned}\| \mathbf{y} - \mathbf{A}\mathbf{x}_l \|^2_{\mathbf{G_y}} &= (\mathbf{y} - \mathbf{A}\mathbf{x}_l)'\mathbf{G_y}(\mathbf{y} - \mathbf{A}\mathbf{x}_l) \\ &\leq (\mathbf{y} - \mathbf{A}\mathbf{x})'\mathbf{G_y}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \| \mathbf{y} - \mathbf{A}\mathbf{x} \|^2_{\mathbf{G_y}}\end{aligned} \tag{3.2}$$

holds, in particular if the vector of inconsistency $\mathbf{i}_l := \mathbf{y} - \mathbf{A}\mathbf{x}_l$ has the least $\mathbf{G_y}$-seminorm.

The solution of type $\mathbf{G_y}$-LESS can be computed as following

Lemma 3.2. (least squares solution with respect to the $\mathbf{G_y}$-seminorm):

A vector $\mathbf{x}_l \in \mathbb{X} \equiv \mathbb{R}^m$, is $\mathbf{G_y}$-LESS of (3.1) if and only if the system of normal equations

$$\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_l = \mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{3.3}$$

is fulfilled. $\mathbf{x}_l$ always exists and is in particular unique, if $\mathbf{A}'\mathbf{G_y}\mathbf{A}$ is regular.

Proof.

$\mathbf{G_y}$-LESS is constructed by means of the Lagrangean

$$\begin{aligned}\mathcal{L}(\mathbf{x}) :&= \| \mathbf{i} \|^2_{\mathbf{G_y}} = \| \mathbf{y} - \mathbf{A}\mathbf{x} \|^2_{\mathbf{G_y}} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \mathbf{x}'\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x} - 2\mathbf{y}'\mathbf{G_y}\mathbf{A}\mathbf{x} + \mathbf{y}'\mathbf{G_y}\mathbf{y} = \min_{\mathbf{x}}\end{aligned}$$

such that the *first derivatives*

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}_l) = \frac{\partial\, \mathbf{i}'\mathbf{G_y}\mathbf{i}}{\partial \mathbf{x}}(\mathbf{x}_l) = 2\mathbf{A}'\mathbf{G_y}(\mathbf{A}\mathbf{x}_l - \mathbf{y}) = 0$$

constitute the *necessary conditions*. The theory of *vector derivative* is presented in Appendix B. The *second derivatives*

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}\, \partial \mathbf{x}'}(\mathbf{x}_l) = \frac{\partial^2\, \mathbf{i}'\mathbf{G_y}\mathbf{i}}{\partial \mathbf{x}\, \partial \mathbf{x}'}(\mathbf{x}_l) = 2\mathbf{A}'\mathbf{G_y}\mathbf{A} \geq \mathbf{0}$$

due to the positive semidefiniteness of the matrix $\mathbf{A}'\mathbf{G_y}\mathbf{A}$ generate the *sufficiency condition* for obtaining the minimum of the *unconstrained Lagrangean*. Because of the $\mathcal{R}(\mathbf{A}'\mathbf{G_y}\mathbf{A}) = \mathcal{R}(\mathbf{A}'\mathbf{G_y})$ there always exists a solution $\mathbf{x}_l$ whose uniqueness is guaranteed by means of the regularity of the matrix $\mathbf{A}'\mathbf{G_y}\mathbf{A}$.

It is obvious that the matrix $\mathbf{A}'\mathbf{G_y}\mathbf{A}$ is in particular regular, if $rk\mathbf{A} = \mathtt{dim}\mathbb{X} = m$, but on the other side the matrix $\mathbf{G_y}$ is *positive definite*, namely $\| \mathbf{i} \|^2_{\mathbf{G_y}}$ is a $\mathbf{G_y}$-norm. The *linear form* $\mathbf{x}_l = \mathbf{L}\mathbf{y}$ which for arbitrary observation vectors $\mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n$, leads to $\mathbf{G_y}$-LESS of (3.1) can be represented as following.

Lemma 3.3. (least squares solution with respect to the $\mathbf{G_y}$- norm, $rk\mathbf{A} = \mathtt{dim}\mathbb{X} = m$ or ($\mathbf{x} \in \mathcal{R}(\mathbf{A}^-)$):

$\mathbf{x}_l = \mathbf{Ly}$ is $\mathbf{G_y}$-LESS of the inconsistent system of linear equations (3.1) $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$ restricted to $rk(\mathbf{A}'\mathbf{G}_{\mathbf{yA})=\mathbf{rkA}=\mathtt{dim}\mathbb{X}}$ (or $\mathcal{R}(\mathbf{A}'\mathbf{G_y}) = \mathcal{R}(\mathbf{A}')$ and $\mathbf{x} \in \mathcal{R}(\mathbf{A}^-)$) if and only if $\mathbb{L} \in \mathbb{R}^{n\times m}$ is represented by

Case (i) :

$$\mathbf{G_y} = \mathbf{I}$$

$$\hat{\mathbf{L}} = \mathbf{A}_L^- = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \; \mathtt{(left\ inverse)} \tag{3.4}$$

$$\mathbf{x}_l = \mathbf{A}_L^-\mathbf{y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}. \tag{3.5}$$

$$\mathbf{y} = \mathbf{y}_l + i_\mathbf{l} \tag{3.6}$$

is an orthogonal decomposition of the observation vector $\mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n$ into the $\mathbf{I}$-LESS vector $\mathbf{y}_l \in \mathbb{Y} \equiv \mathbb{R}^n$ and the $\mathbf{I}$-LESS vector of inconsistency $\mathbf{i}_l \in \mathbb{Y} \equiv \mathbb{R}^n$, subject to

$$\mathbf{y}_l = \mathbf{Ax}_l = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \tag{3.7}$$

$$\mathbf{i}_l = \mathbf{y} - \mathbf{y}_l = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\,\mathbf{y}. \tag{3.8}$$

Due to $\mathbf{y}_l = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$, I-LESS has the reproducing property. As projection matrices $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']$ are independent. The "*goodness of fit*" of I-LESS is

$$\|\mathbf{y} - \mathbf{Ax}_l\|_\mathbf{i}^2 = \|\mathbf{i}_l\|_\mathbf{i}^2 = \mathbf{y}'[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-\mathbf{1}}\mathbf{A}']\,\mathbf{y}. \tag{3.9}$$

Case (ii) : $\mathbf{G_y}$ positive definite, $rk\mathbf{A} = \mathtt{dim}\mathbb{X}$

$$\hat{\mathbf{L}} = (\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y} \; \mathtt{(weighted\ left\ inverse)} \tag{3.10}$$

$$\mathbf{x}_l = (\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}. \tag{3.11}$$

$$\mathbf{y} = \mathbf{y}_l + \mathbf{i}_l \tag{3.12}$$

is an orthogonal decomposition of the observation vector $\mathbf{y} \in \mathbb{Y} \equiv \mathbb{R}^n$ into the $\mathbf{G_y}$-LESS vector $\mathbf{y}_l \in \mathbb{Y} \equiv \mathbb{R}^n$ and the $\mathbf{G_y}$-LESS vector of inconsistency $\mathbf{i}_l \in \mathbb{Y} \equiv \mathbb{R}^n$ subject to

$$\mathbf{y}_l = \mathbf{Ax}_l = \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{3.13}$$

$$\mathbf{i}_l = \mathbf{y} - \mathbf{Ax}_l = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}]\,\mathbf{y} \tag{3.14}$$

Due to $\mathbf{y}_l = \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}$ $\mathbf{G_y}$-LESS has the reproducing property. As projection matrices $\mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}$ and $[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}]$ are independent. The "*goodness of fit*" of $\mathbf{G_y}$-LESS is

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}_l\|^2_{\mathbf{G_y}} = \|\mathbf{i}_l\|^2_{\mathbf{G_y}} = \mathbf{y}'[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}]\,\mathbf{y}. \tag{3.15}$$

The *third case* $\mathbf{G_y}$ positive semidefinite will be treated independently. The proof of *Lemma 3.1* is straightforward. The result that LESS generates the left inverse, $\mathbf{G_y}$-LESS the *weighted* left inverse will be proved later.

An alternative way of producing the *least squares solution* with respect to the $\mathbf{G_y}$-seminorm of the *linear model* is based upon the *constrained Lagrangean* (3.16), namely $\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda)$. Indeed $\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda)$ integrates the linear model (3.1) by a *vector valued Lagrange multiplier* to the objective function of type "*least squares*", namely the distance function in a finite dimensional *Hilbert space*. Such an approach will be useful when we apply "*total least squares*" to the mixed linear model (error-in-variable model).

Lemma 3.4. (least squares solution with respect to the $\mathbf{G_y}$-norm, $rk\mathbf{A} = \dim\mathbb{X}$, *constrained Lagrangean*):

$\mathbf{G_y}$-LESS is assumed to be defined with respect to the *constrained Lagrangean*

$$\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda) := \mathbf{i}'\mathbf{G_y}\mathbf{i} + 2\lambda'(\mathbf{A}\mathbf{x} + \mathbf{i} - \mathbf{y}) = \min_{\mathbf{i}, \mathbf{x}, \lambda}. \tag{3.16}$$

A vector $[\mathbf{i}'_l, \mathbf{x}'_l, \lambda'_l]' \in \mathbb{R}^{(n+m+n)\times 1}$ is $\mathbf{G_y}$-LESS of (3.1) in the sense of the *constrained Lagrangean* $\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda)$= min if and only if the system of normal equations

$$\begin{bmatrix} \mathbf{G_y} & 0 & \mathbf{I}_n \\ 0 & 0 & \mathbf{A}' \\ \mathbf{I}_n & \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{i}_l \\ \mathbf{x}_l \\ \lambda_l \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{y} \end{bmatrix} \tag{3.17}$$

with the vector $\boldsymbol{\lambda}_l \in \mathbb{R}^{n\times 1}$ of "*Lagrange multiplier*" is fulfilled. $(\mathbf{i}_l, \mathbf{x}_l, \lambda_l)$ exists and is in particular unique, if $\mathbf{G_y}$ is positive semidefinite. There holds

$$(\mathbf{i}_l, \mathbf{x}_l, \lambda_l) = \arg\{\texttt{user1}\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda) = \min\}. \tag{3.18}$$

Proof.

$\mathbf{G_y}$-LESS is based on the *constrained Lagrangean*

$$\mathcal{L}(\mathbf{i}, \mathbf{x}, \lambda) := \mathbf{i}'\mathbf{G_y}\mathbf{i} + 2\lambda'(\mathbf{A}\mathbf{x} + \mathbf{i} - \mathbf{y}) = \min_{\mathbf{i}, \mathbf{x}, \lambda}$$

such that the *first derivatives*

$$\frac{\partial \mathcal{L}}{\partial \mathbf{i}}(\mathbf{i}_l, \mathbf{x}_l, \lambda_l) = 2(\mathbf{G_y i}_l + \lambda_l) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{i}_l, \mathbf{x}_l, \lambda_l) = 2\mathbf{A}'\boldsymbol{\lambda}_l = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{i}_l, \mathbf{x}_l, \lambda_l) = 2(\mathbf{Ax}_l + \mathbf{i}_l - \mathbf{y}) = 0$$

or

$$\begin{bmatrix} \mathbf{G_y} & 0 & \mathbf{I}_n \\ 0 & 0 & \mathbf{A}' \\ \mathbf{I}_n & \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{i}_l \\ \mathbf{x}_l \\ \lambda_l \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mathbf{y} \end{bmatrix}$$

constitute the *necessary conditions*. (The theory of vector derivative is presented in Appendix B.) The *second derivatives*

$$\frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \mathbf{i}\,\partial \mathbf{i}'}(\mathbf{x}_l) = \mathbf{G_y} \geq \mathbf{0}$$

due to the positive semidefiniteness of the matrix $\mathbf{G_y}$ generate the *sufficiency condition* for obtaining the minimum of the *constrained Lagrangean*.

Lemma 3.5. (least squares solution with respect to the $\mathbf{G_y}$-norm, $rk\mathbf{A} = \mathtt{dim}\mathbb{X}$, *constrained Lagrangean*)

If $\mathbf{G_y}$-LESS of the linear equations (3.1) is generated by the *constrained Lagrangean* (3.16) with respect to a *positive definite* weight matrix $\mathbf{G_y}$, $rk\,\mathbf{G_y} = n$, then the normal equations (3.17) are *uniquely solved* by

$$\mathbf{x}_l = (\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y y}, \tag{3.19}$$

$$\mathbf{i}_l = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y}]\,\mathbf{y}, \tag{3.20}$$

$$\lambda_l = [\mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}' - \mathbf{I}_n]\,\mathbf{G_y y}. \tag{3.21}$$

Proof.

A basis of the proof could be *C.R. Raos Pandora Box*, the theory of *inverse partitioned matrices* (Appendix A: Fact: Inverse Partitioned Matrix /IPM/ of a symmetric matrix). Due to the rank identities $\text{rk}\,\mathbf{G_y} = n$, $\text{rk}\,\mathbf{A} = rk\,(\mathbf{A}'\mathbf{G_y}\mathbf{A}) = m < n$, the normal equations can be solved faster directly by *Gauss elimination*.

$$\begin{aligned}\mathbf{G_y}\mathbf{i}_l + \lambda_l &= 0\\ \mathbf{A}'\lambda_l &= 0\\ \mathbf{A}\mathbf{x}_l + \mathbf{i}_l - \mathbf{y} &= 0.\end{aligned}$$

Multiply the *third* normal equation by $\mathbf{A}'\mathbf{G_y}$, multiply the *first* normal equation by $\mathbf{A}'$ and substitute $\mathbf{A}'\,\lambda_l$ from the *second* normal equation in the modified first one.

$$\left.\begin{array}{c}\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_l + \mathbf{A}'\mathbf{G_y}\mathbf{i}_l - \mathbf{A}'\mathbf{G_y}\mathbf{y} = 0\\ \mathbf{A}'\mathbf{G_y}\mathbf{i}_l + \mathbf{A}'\lambda_l = 0\\ \mathbf{A}'\lambda_l = 0\end{array}\right] \Rightarrow$$

$$\Rightarrow \quad \left.\begin{array}{c}\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_l + \mathbf{A}'\mathbf{G_y}\mathbf{i}_l - \mathbf{A}'\mathbf{G_y}\mathbf{y} = 0\\ \mathbf{A}'\mathbf{G_y}\mathbf{i}_l = 0\end{array}\right] \Rightarrow$$

$$\Rightarrow \qquad \mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_l - \mathbf{A}'\mathbf{G_y}\mathbf{y} = 0, \qquad \Rightarrow$$

$$\mathbf{x}_l = (\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}.$$

Let us subtract the third normal equation and solve for $\mathbf{i}_l$.

$$\mathbf{i}_l = \mathbf{y} - \mathbf{A}\mathbf{x}_l,$$

$$\mathbf{i}_l = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}]\,\mathbf{y}.$$

Finally we determine the *Lagrange multiplier*: substitute $\mathbf{i}_l$ in the *first normal equation* in order to find

$$\lambda_l = -\mathbf{G_y}\mathbf{i}_l$$

$$\lambda_l = [\mathbf{G_y}\mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y} - \mathbf{G_y}]\,\mathbf{y}.$$

Of course the $\mathbf{G_y}$-LESS of type (3.2) and the $\mathbf{G_y}$-LESS solution of type *constrained Lagrangean* (3.16) are equivalent, namely (3.11)–(3.19) and (3.14)–(3.20).

In order to analyze the finite dimensional linear space $\mathbb{Y}$ called "*the observation space*", namely the case of a singular matrix of its metric, in more detail, let us take reference to the following.

Theorem 3.6. (bilinear form) :

Suppose that the *bracket* $\langle .|. \rangle$ or $g(.,.) : \mathbb{Y} \times \mathbb{Y} \longrightarrow \mathbb{R}$ is a bilinear form or a finite dimensional linear space $\mathbb{Y}$, *dim* $\mathbb{Y} = n$, for in-stance a vector space over the field of real numbers. There *exists* a basis $\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ such that

$$\text{(i)} \qquad \left\langle \mathbf{e_i} \,\middle|\, \mathbf{e_j} \right\rangle = 0 \texttt{ or } g(\mathbf{e_i}, \mathbf{e_j}) = 0 \texttt{ for } i \neq j$$

$$\text{(ii)} \qquad \begin{cases} \langle \mathbf{e}_{i_1} \mid \mathbf{e}_{i_1} \rangle = +1 \texttt{ or } g\left(\mathbf{e}_{i_1}, \mathbf{e}_{i_1}\right) = +1 \texttt{ for } 1 \leq i_1 \leq p \\ \langle \mathbf{e}_{i_2} \mid \mathbf{e}_{i_2} \rangle = -1 \texttt{ or } g\left(\mathbf{e}_{i_2}, \mathbf{e}_{i_2}\right) = -1 \texttt{ for } p+1 \leq i_2 \leq p+q = r \\ \langle \mathbf{e}_{i_3} \mid \mathbf{e}_{i_3} \rangle = 0 \texttt{ or } g\left(\mathbf{e}_{i_3}, \mathbf{e}_{i_3}\right) = 0 \texttt{ for } r+1 \leq i_3 \leq n \quad . \end{cases}$$

The numbers r and p are determined exclusively by the bilinear form. r is called the rank, $r - p = q$ is called the *relative index* and the *ordered pair* (p, q) the signature. The theorem states that any two spaces of the same dimension with *bilinear forms of the same signature* are isometrically isomorphic. A scalar product ("*inner product*") in this context is a *nondegenerate bilinear form*, for instance a form with rank equal to the dimension of $\mathbb{Y}$. When dealing with low dimensional spaces as we do, we will often indicate the *signature* with a series of plus and minus signs when appropriate. For instance the signature of $\mathbb{R}^4_1$ may be written $(+ + + -)$ instead of (3,1). Such an *observation space* $\mathbb{Y}$ is met when we are dealing with observations in *Special Relativity*.

For *instance*, let us summarize the peculiar LESS features if the matrix $\mathbf{G}_y \in \mathbb{R}^{n\times n}$ of the observation space is *semidefinite*, rk$\mathbf{G_y} := r_y < n$. By means of *Box 3.7* we have collected the essential items of the *eigenspace analysis* as well as the eigenspace synthesis $\mathbf{G}^*_\mathbf{y}$ versus $\mathbf{G_y}$ of the metric. $\boldsymbol{\Lambda}_\mathbf{y} = = \mathrm{Diag}(\lambda_1, \ldots, \lambda_{r_y})$ denotes the matrix of *non-vanishing eigenvalues* $\{\lambda_1, \ldots, \lambda_{r_y}\}$. Note the norm identity

$$||\mathbf{i}||^2_{\mathbf{G_y}} = ||\mathbf{i}||^2_{\mathbf{U_1}\boldsymbol{\Lambda}_\mathbf{y}\mathbf{U}'_\mathbf{1}} \tag{3.22}$$

which leads to the $\mathbf{U_1}\boldsymbol{\Lambda}_\mathbf{y}\mathbf{U}'_\mathbf{1}$-LESS normal equations

$$\mathbf{A}'\mathbf{U_1}\boldsymbol{\Lambda}_\mathbf{y}\mathbf{U}'_\mathbf{1}\mathbf{x}_\ell = \mathbf{A}'\mathbf{U_1}\boldsymbol{\Lambda}_\mathbf{y}\mathbf{U}'_\mathbf{1}\mathbf{y}. \tag{3.23}$$

Box 3.7. (Canonical representation of the rank deficient matrix of the matrix of the observation space $\mathbb{Y}$).

$$\mathrm{rk}\,\mathbf{G_y} =: r_y, \;\; \boldsymbol{\Lambda}_\mathbf{y} := \mathrm{Diag}(\lambda_1, \ldots, \lambda_{r_y}).$$

"*eigenspace analysis*" "*eigenspace synthesis*"

$$\mathbf{G}_\mathbf{y}^* = \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix} \mathbf{G}_\mathbf{y} [\mathbf{U}_1, \mathbf{U}_2] = \qquad \mathbf{G}_\mathbf{y} = [\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda}_\mathbf{y} & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_3 \end{bmatrix} \begin{bmatrix} \mathbf{U}_1' \\ \mathbf{U}_2' \end{bmatrix} \tag{3.24}$$

$$= \begin{bmatrix} \mathbf{\Lambda}_\mathbf{y} & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_3 \end{bmatrix} \in \mathbb{R}^{n \times n} \qquad \in \mathbb{R}^{n \times n}$$

subject to

$$\mathbf{U} \in \mathbf{SO}(n(n-1)/2) := \left\{ \mathbf{U} \in R^{n \times n} | \mathbf{U}'\mathbf{U} = \mathbf{I_n}, |\mathbf{U}| = +1 \right\}$$

$$\mathbf{U_1} \in R^{n \times r_y}, \ \mathbf{U_2} \in R^{n \times (n - r_y)}, \ \mathbf{\Lambda}_\mathbf{y} \in R^{r_y \times r_y}$$

$$\mathbf{0_1} \in R^{r_y \times n}, \ \mathbf{0_2} \in R^{(n-r_y) \times r_y}, \ \mathbf{0_3} \in R^{(n-r_y) \times (n-r_y)}$$

"*norms*"

$$||\mathbf{i}||^2_{\mathbf{G}_\mathbf{y}} = \ ||\mathbf{i}||^2_{\mathbf{U}_1 \mathbf{\Lambda}_\mathbf{y} \mathbf{U}_1'} \qquad \mathbf{i}'\mathbf{G}_\mathbf{y}\mathbf{i} = \mathbf{i}'\mathbf{U}_1\mathbf{\Lambda}_\mathbf{y}\mathbf{U}_1'\mathbf{i} \tag{3.25}$$

$$\text{LESS: } ||\mathbf{i}||^2_{\mathbf{G}_\mathbf{y}} = \min_\mathbf{x} \Leftrightarrow ||\mathbf{i}||^2_{\mathbf{U}_1 \mathbf{\Lambda}_\mathbf{y} \mathbf{U}_1'} = \min_\mathbf{x}$$

$$\Leftrightarrow \mathbf{A}'\mathbf{U}_1\mathbf{\Lambda}_\mathbf{y}\mathbf{U}_1'\mathbf{x}_\ell = \mathbf{A}'\mathbf{U}_1\mathbf{\Lambda}_\mathbf{y}\mathbf{U}_1'\mathbf{y}.$$

Another example relates to an observation space

$$\mathbb{Y} = \mathbb{R}^{2k}_1 \ (k \in \{1, \ldots, K\})$$

of *even dimension*, but one negative eigenvalue. In such a *pseudo-Euclidean space* of signature $(+ \cdots + -)$ the determinant of the matrix of metric $\mathbf{G}_\mathbf{y}$ is *negative*, namely $\det \mathbf{G}_\mathbf{y} = -\lambda_1 \ldots \lambda_{2K-1} \, |\lambda_{2K}|$. Accordingly

$$\mathbf{x}_{\max} = \arg\{||\mathbf{i}||^2_{\mathbf{G}_\mathbf{y}} = \max | \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}, \mathrm{rk}\mathbf{A} = m\}$$

is $\mathbf{G}_\mathbf{y}$-MORE (Maximal ObseRvational inconsistEncy solution), but *not* $\mathbf{G}_\mathbf{y}$-LESS. Indeed, the structure of the observational space, either *pseudo-Euclidean* or *Euclidean*, decides upon MORE or LESS.

3-21 A Discussion of the Metric of the Parameter Space $\mathbb{X}$

With the completion of the proof we *have to discuss* the basic results of *Lemma 3.3* in more detail. *At first* we have to observe that the matrix $\mathbf{G}_\mathbf{y}$ of the metric of the *observation space* $\mathbb{Y}$ *has to be given a priori*. We classified LESS according to (i) $\mathbf{G}_\mathbf{y} = \mathbf{I}_n$, (ii) $\mathbf{G}_\mathbf{y}$ positive definite and (iii) $\mathbf{G}_\mathbf{y}$ positive semidefinite. But how do we know the metric of the observation space $\mathbb{Y}$? Obviously we need *prior information* about the geometry of the observation space $\mathbb{Y}$, namely from the empirical sciences like physics, chemistry, biology, geosciences, social sciences. If the *observation space* $\mathbb{Y} \in \mathbb{R}^n$ is equipped with an inner product $\langle \mathbf{y}_1 | \mathbf{y}_2 \rangle = \mathbf{y'}_1 \mathbf{G_y y}_2,\ \mathbf{y}_1 \in \mathbb{Y},\ \mathbf{y}_2 \in \mathbb{Y}$ where the matrix $\mathbf{G_y}$ of the metric $\| \mathbf{y} \|^2 = \mathbf{y' G_y y}$ is *positive definite*, we refer to the metric space $\mathbb{Y} \in \mathbb{R}^n$ as *Euclidean* $\mathbb{E}^n$. In contrast, if the observation space is positive semidefinite we call the observation space *semi Euclidean* $\mathbb{E}^{n_1,n_2}$. n_1 is the number of *positive eigenvalues*, n_2 the number of *zero eigenvalues* of the positive semidefinite matrix $\mathbf{G_y}$ of the metric ($n = n_1 + n_2$). In various applications, namely in the adjustment of observations which refer to *Special Relativity* or *General Relativity* we have to generalize the metric structure of the observation space $\mathbb{Y}$: If the matrix $\mathbf{G_y}$ of the *pseudometric* $\| \mathbf{y} \|^2 = \mathbf{y' G_y y}$ is built on n_1 positive eigenvalues (signature +), n_2 zero eigenvalues and n_3 negative eigenvalues (signature -), we call the *pseudometric parameter space pseudo Euclidean* $\mathbb{E}^{n_1,n_2,n_3}$, $n = n_1 + n_2 + n_3$. For such an observation space LESS has to be generalized to $\| \mathbf{y} - \mathbf{Ax} \|^2_{\mathbf{G_y}} = extr$, for instance "*maximum norm solution*".

3-22 Alternative Choices of the Metric of the Observation $\mathbb{Y}$

Another problem associated with the *observation space* $\mathbb{Y}$ is the *norm choice problem*. Up to now we have used the ℓ_2-norm, for instance

$$\ell_2\text{-norm}: \quad \| \mathbf{y} - \mathbf{Ax} \|_2 := \sqrt{(\mathbf{y} - \mathbf{Ax})(\mathbf{y} - \mathbf{Ax})} = \sqrt{\mathbf{i'i}}$$

$$= \sqrt{i_1^2 + i_2^2 + \cdots + i_{n-1}^2 + i_n^2},$$

$$\ell_p\text{-norm}: \| \mathbf{y} - \mathbf{Ax} \|_p := \sqrt[p]{|i_1|^p + |i_2|^p + \cdots + |i_{n-1}|^p + |i_n|^p},$$

$$1 < p < \infty$$

$$\ell_\infty\text{-norm}: \| \mathbf{i} \|_\infty := \max_{1 \le i \le n} |i_i|$$

are alternative norms of choice.

Beside the choice of the matrix $\mathbf{G_y}$ of the metric within the weighted ℓ_2-norm we like to discuss the result of the LESS matrix $\mathbf{G}_l$ of the metric. Indeed we have constructed LESS from an a priori choice of the metric $\mathbf{G}$ called $\mathbf{G_y}$ and were led to the a posteriori choice of the metric $\mathbf{G}_l$ of type (3.9) and (3.15). The matrices

$$\text{(i)} \ \ \mathbf{G}_l = \mathbf{I}_n - \mathbf{A}(\mathbf{A'}_\mathbf{y}\mathbf{A})^{-1}\mathbf{A'}_\mathbf{y} \ (3.9)$$

$$\text{(ii)} \ \ \mathbf{G}_l = \mathbf{I}_n - \mathbf{A}(\mathbf{A'G_yA})^{-1}\mathbf{A'G_y} \ (3.15)$$

are (i) *idempotent* and (ii) $\mathbf{G}_\mathbf{y}^{-1}$ *idempotent*, in addition. There are various alternative *scales* or *objective functions* for projection matrices for substituting Euclidean metrics termed robustifying. In special cases those *objective functions* operate on

$$(3.11) \quad \mathbf{x}_l = \mathbf{H}\mathbf{y} \texttt{ subject to } \mathbf{H}_x = (\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A})^{-1}\mathbf{A}\mathbf{G}_\mathbf{y},$$

$$(3.13) \quad \mathbf{y}_\ell = \mathbf{H}_\mathbf{y}\mathbf{y} \texttt{ subject to } \mathbf{H}_\mathbf{y} = \mathbf{A}(\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A})^{-1}\mathbf{A}\mathbf{G}_\mathbf{y}$$

$$(3.14) \quad \mathbf{i}_\ell = \mathbf{H}_\ell\mathbf{y} \texttt{ subject to } \mathbf{H}_\ell = \left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A})^{-1}\mathbf{A}\mathbf{G}_\mathbf{y}\right]\mathbf{y},$$

where $\{\mathbf{H_x}, \mathbf{H_y}, \mathbf{H}_\ell\}$ are called "*hat matrices*". In other cases *analysts* have to accept that the *observation space* is *non-Euclidean*. For instance, direction observations in $\mathbb{R}^P$ locate points on the hypersphere $\mathbb{S}^{P-1}$. Accordingly we have to accept an *objective function* of *von Mises–Fisher* type which measures the *spherical distance* along a great circle between the measurement points on $\mathbb{S}^{P-1}$ and the *mean direction*. Such an alternative choice of a metric of a *non-Euclidean space* $\mathbb{Y}$ will be presented in Chap. 7. Here we discuss in some detail alternative *objective functions*, namely

- Optimal choice of the weight matrix $\mathbf{G_y}$: second order design SOD
- Optimal choice of the weight matrix $\mathbf{G_y}$ by means of condition equations
- Robustifying objective functions

3-221 Optimal Choice of Weight Matrix: SOD

The optimal choice of the weight matrix, also called *second order design* (SOD), is a traditional topic in the design of *geodetic networks*. Let us refer to the review papers by *A. A. Seemkooei* (2001), *W. Baarda* (1968, 1973), *P. Cross* (1985), *P. Cross and K. Thapa* (1979), *E. Grafarend* (1970, 1972, 1974, 1975), *E. Grafarend and B. Schaffrin* (1979), *B. Schaffrin* (1981, 1983, 1985), *F. Krumm* (1985), *S. L. Kuang* (1991), *P. Vanicek*, *K. Thapa* and *D. Schröder* (1981), *B. Schaffrin*, *E. Grafarend* and G. Schmitt (1977), *B. Schaffrin, F. Krumm* and *D. Fritsch* (1980), J. van Mierlo (1981), G. Schmitt (1980, 1985), *C. C. Wang* (1970), *P. Whittle* (1954, 1963), *H. Wimmer* (1982) and the textbooks by E. Grafarend, *H. Heister, R. Kelm, H. Knopff* and *B. Schaffrin* (1979) and *E. Grafarend* and *F. Sanso* (1985, editors).

What is an optimal choice of the weight matrix $\mathbf{G_y}$, what is "*a second order design problem*"?

Let us begin with Fisher's Information Matrix which agrees to the half of the *Hesse matrix*, the matrix of second derivatives of the *Lagrangean* $\mathcal{L}(\mathbf{x}) := ||\mathbf{i}||^2_{\mathbf{G_y}} = ||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathbf{G_y}}$, namely

$$\mathbf{G_x} = \mathbf{A}'(\mathbf{x})\mathbf{G_y}\mathbf{A}(\mathbf{x}) = \frac{1}{2}\frac{\partial^2\mathcal{L}}{\partial\mathbf{x}_\ell\partial\mathbf{x}'_\ell} =: \mathit{FISHER}$$

at the "point" $\mathbf{x}_\ell$ of type LESS. The first order design problem aims at determining those points $\mathbf{x}$ within the *Jacobi matrix* $\mathbf{A}$ by means of a properly chosen risk operating on "*FISHER*". Here, "*FISHER*" relates the *weight matrix of the observations* $\mathbf{G_y}$, previously called the matrix of the metric of the *observation space*, to the weight matrix $\mathbf{G_x}$ of the *unknown parameters*, previously called the matrix of the metric of the *parameter space*.

$\mathbf{G_x}$	$\mathbf{G_y}$
weight matrix of the unknown parameters or matrix of the metric of the parameter space $\mathbb{X}$	weight matrix of the observations or matrix of the metric of the observation space $\mathbb{Y}$.

Being properly prepared, we are able to outline the optimal choice of the weight matrix $\mathbf{G_y}$ or $\mathbf{X}$, also called the *second order design problem*, from a criterion matrix $\mathbf{Y}$, an *ideal weight matrix* $\mathbf{G_x}$ (*ideal*) of the unknown parameters, We hope that the translation of $\mathbf{G_x}$ and $\mathbf{G_y}$ "*from metric to weight*" does not cause any confusion. *Box 3.9* elegantly out-lines SOD.

Box 3.8. (Second order design SOD, optimal fit to a criterion matrix of weights).

"*weight matrix of the parameter space*" 3-21 weight matrix of *the observation space*"

$$\mathbf{Y} := \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}_\ell \partial \mathbf{x}'_\ell} = \mathbf{G_x} \qquad \mathbf{X} := \mathbf{G_y} = \mathrm{Diag}\left(g_1^{\mathbf{y}}, \ldots, g_n^{\mathbf{y}}\right) \tag{3.26}$$

$$\mathbf{x} := \left[g_1^{\mathbf{y}}, \ldots, g_n^{\mathbf{y}}\right]'$$

"*inconsistent matrix equation of the second order design problem*"

$$\mathbf{A}'\mathbf{X}\mathbf{A} + \mathbf{A} = \mathbf{Y} \tag{3.27}$$

"*optimal fit*"

$$||\Delta||^2 = \mathrm{tr}\Delta'\Delta = (\mathrm{vec}\Delta)'(\mathrm{vec}\Delta) = \min_{\mathbf{x}} \tag{3.28}$$

$$\mathbf{x}_S := \arg\{||\Delta||^2 = \min|\mathbf{A}'\mathbf{X}\mathbf{A} + \Delta = \mathbf{Y}, \mathbf{X} = \mathrm{Diag}\,\mathbf{x}\} \tag{3.29}$$

$$\mathrm{vec}\,\Delta =$$

$$= \mathrm{vec}\mathbf{Y} - \mathrm{vec}(\mathbf{A}'\mathbf{X}\mathbf{A}) = \mathrm{vec}\mathbf{Y} - (\mathbf{A}' \otimes \mathbf{A}')\mathrm{vec}\mathbf{X} \tag{3.30}$$

$$\mathrm{vec}\,\Delta = \mathrm{vec}\mathbf{Y} - (\mathbf{A}' \otimes \mathbf{A}')\mathbf{x}$$

$$\mathbf{x} \in \mathbb{R}^n, \quad \mathrm{vec}\mathbf{Y} \in \mathbb{R}^{n^2\times 1}, \quad \mathrm{vech}\mathbf{Y} \in \mathbb{R}^{n(n+1)/2\times 1}$$

$$\mathrm{vec}\Delta \in \mathbb{R}^{n^2\times 1}, \mathrm{vec}\mathbf{X} \in \mathbb{R}^{n^2\times 1}, (\mathbf{A}' \otimes \mathbf{A}') \in \mathbb{R}^{n^2\times n^2}, \mathbf{A}' \odot \mathbf{A}' \in \mathbb{R}^{n^2\times n}$$

$$\mathbf{x}_S = \left[(\mathbf{A}' \odot \mathbf{A}')'(\mathbf{A}' \odot \mathbf{A}')\right]^{-1}(\mathbf{A}' \odot \mathbf{A}')\mathrm{vec}\mathbf{Y}. \tag{3.31}$$

In general, the matrix equation $\mathbf{A}'\mathbf{X}\mathbf{A} + \Delta = \mathbf{Y}$ is *inconsistent*. Such a *matrix inconsistency* we have called $\Delta \in \mathbb{R}^{m \times m}$: For a given *ideal weight matrix* $\mathbf{G}_{\mathbf{x}}$(*ideal*), $\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{A}$ is only an approximation. The unknown weight matrix of the observations $\mathbf{G}_{\mathbf{y}}$, here called $X \in \mathbb{R}^{n \times n}$, *can only be designed* in its diagonal form. A general weight matrix $\mathbf{G}_{\mathbf{y}}$ does not make any sense since "oblique weights" cannot be associated to experiments. A *natural restriction* is therefore $\mathbf{X} = \text{Diag}\left(g_1^{\mathbf{y}}, \ldots, g_n^{\mathbf{y}}\right)$. The "diagonal weights" are collected in the *unknown vector of weights*

$$\mathbf{x} := \left[g_1^{\mathbf{y}}, \ldots, g_n^{\mathbf{y}}\right]' \in \mathbb{R}^n$$

The *optimal fit* "$\mathbf{A}'\mathbf{X}\mathbf{A}$ to $\mathbf{Y}$ " is achieved by the *Lagrangean* $||\Delta||^2 = \min$, the optimum of the *Frobenius norm of the inconsistency matrix* Δ. The vectorized form of the inconsistency matrix, vecΔ, leads us *first* to the matrix $(\mathbf{A}' \otimes \mathbf{A}')$, the *Zehfuss product* of $\mathbf{A}'$, second to the *Kronecker matrix* $(\mathbf{A}' \odot \mathbf{A}')$, the *Khatri–Rao product* of $\mathbf{A}'$, as soon as we implement the diagonal matrix $\mathbf{X}$. For a definition of the *Kronecker–Zehfuss product* as well as of the Khatri–Rao product and related laws we refer to Appendix A. The *unknown weight vector* $\mathbf{x}$ is LESS, if

$$\mathbf{x}_S = \left[(\mathbf{A}' \odot \mathbf{A}')'(\mathbf{A}' \odot \mathbf{A}')\right]^{-1} (\mathbf{A}' \odot \mathbf{A}')'\text{vec}\mathbf{Y}$$

Unfortunately, the weights $\mathbf{x}_S$ may come out *negative*. Accordingly we have to build in extra condition, $\mathbf{X} = \text{Diag}(x_1, \ldots, x_m)$ to be *positive definite*. The given references address this problem as well as the *datum problem* inherent in $\mathbf{G}_{\mathbf{x}}(ideal)$.

Example 3.2. (Second order design):

The introductory example we outline here may serve as a firsthand insight into the observational weight design, also known as *second order design*. According to *Fig. 3.4* we present you with the graph of a two-dimensional planar network. From three given points $\{P_\alpha, P_\beta, P_\gamma\}$ we measure distances to the unknown point P_δ, a

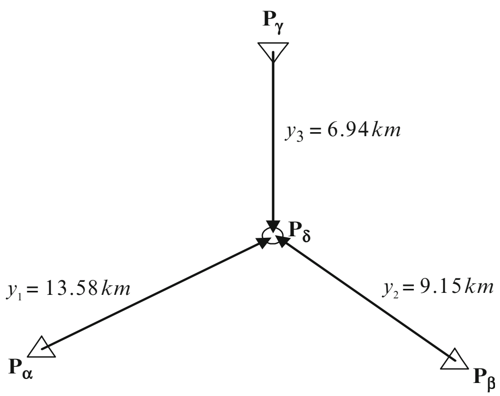


Fig. 3.4 Directed graph of a trilateration network, known points $\{P_\alpha, P_\beta, P_\gamma\}$, unknown point P_δ, distance observations $[y_1, y_2, y_3]' \in \mathbb{Y}$

typical problem in densifying a geodetic network. For the weight matrix $\mathbf{G_x} \sim \mathbf{Y}$ of the unknown point we postulate $\mathbf{I}_2$, unity. In contrast, we aim at an observational weight design characterized by a weight matrix

$$\mathbf{G_x} \sim \mathbf{X} = \mathrm{Diag}(x_1, x_2, x_3).$$

The second order design equation

$$\mathbf{A}'\mathrm{Diag}(x_1, x_2, x_3)\mathbf{A} + \Delta = \mathbf{I}_2$$

is supposed to supply us with a circular weight matrix $\mathbf{G_y}$ of the *Cartesian coordinates* (x_δ, y_δ) of P_δ. The observational equations for distances $(s_{\alpha\delta}, s_{\beta\delta}, s_{\gamma\delta}) = (13.58, 9.15, 6.94\,km)$ have already been derived in Sect. 1-4. Here we just take advantage of the *first design matrix* $\mathbf{A}$ as given in *Box 3.10* together with all further matrix operations.

A peculiar situation for the matrix equation $\mathbf{A}'\mathbf{X}\mathbf{A} + \Delta = \mathbf{I}_2$ is met: In the special configuration of the *trilateration network* the characteristic equation of the second order design problem is consistent. Accordingly we have no problem to get the weights

$$\mathbf{G_y} = \begin{bmatrix} 0.511 & 0 & 0 \\ 0 & 0.974 & 0 \\ 0 & 0 & 0.515 \end{bmatrix},$$

which lead us to the weight $\mathbf{G_x} = \mathbf{I}_2$ a posteriori. Note that the weights came out *positive*.

Box 3.9. (Example for a second order design problem, trilateration network).

$$\mathbf{A} = \begin{bmatrix} -0.454 & -0.891 \\ -0.809 & +0.588 \\ +0.707 & +0.707 \end{bmatrix}, \mathbf{X} = \mathrm{Diag}(x_1, x_2, x_3), \mathbf{Y} = \mathbf{I_2}$$

$$\mathbf{A}'\mathrm{Diag}(x_1, x_2, x_3)\mathbf{A} = \mathbf{I_2}$$

$$\Leftrightarrow \begin{bmatrix} 0.206x_1 + 0.654x_2 + 0.5x_3 & 0.404x_1 - 0.476x_2 + 0.5x_3 \\ 0.404x_1 - 0.476x_2 + 0.5x_3 & 0.794x_1 + 0.346x_2 + 0.5x_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

"*inconsistency* $\Delta = 0$"

$$(1st)\ 0.206x_1 + 0.654x_2 + 0.5x_3 = 1$$
$$(2nd)\ 0.404x_1 - 0.476x_2 + 0.5x_3 = 0$$
$$(3rd)\ 0.794x_1 + 0.346x_2 + 0.5x_3 = 1$$
$$x_1 = 0.511, x_2 = 0.974, x_3 = 0.515.$$

3-222 The Taylor Karman Criterion Matrix

? What is a proper choice of the ideal weight matrix $\mathbf{G_x}$?

There has been made a great variety of proposals.

First, $\mathbf{G_x}$(*ideal*) has been chosen *simple*: A weight matrix $\mathbf{G_x}$ is called *ideally simple* if $\mathbf{G_x}(ideal) = \mathbf{I_m}$. For such a *simple weight matrix* of the unknown parameters *Example* 3.2 is an illustration of SOD for a densification problem in a trilateration network.

Second, nearly all geodetic networks have been SOD optimized by a *criterion matrix* $\mathbf{G_x}$(*ideal*) which is *homogeneous* and *isotropic* in a two- or three-dimensional *Euclidean space*. In particular, the *Taylor–Karman structure* of a homogeneous and isotropic weight matrix $\mathbf{G_x}$(*ideal*) has taken over the SOD network design. *Box 3.11* summarizes the TK-$\mathbf{G_x}$(*ideal*) of a two-dimensional, planar network. Worth to be mentioned, TK-$\mathbf{G_x}$(*ideal*) has been developed in the *Theory of Turbulence*, namely in analyzing the *two-point correlation function* of the velocity field in a turbulent medium. (*G. I. Taylor* 1935, 1936, *T. Karman* (1937), *T. Karman and L. Howarth* (1936), *C. C. Wang* (1970), *P. Whittle* (1954, 1963)).

Box 3.10. (*Taylor–Karman* structure of a homogeneous and isotropic tensor-valued, two-point function, two-dimensional, planar network).

$$\mathbf{G_x} = \begin{bmatrix} \begin{bmatrix} g_{x_1x_1} & g_{x_1y_1} \\ g_{y_1x_1} & g_{y_1y_1} \end{bmatrix} & \begin{matrix} g_{x_1x_2} & g_{x_1y_2} \\ g_{y_1x_2} & g_{y_1y_2} \end{matrix} \\ \begin{matrix} g_{x_2x_1} & g_{x_2y_1} \\ g_{y_2x_1} & g_{y_2y_1} \end{matrix} & \begin{bmatrix} g_{x_2x_2} & g_{x_2y_2} \\ g_{y_2x_2} & g_{y_2y_2} \end{bmatrix} \end{bmatrix} \mathbf{G_x}(\mathbf{x}_\alpha, \mathbf{x}_\beta)$$

"*Euclidean distance function of points* $P_\alpha \sim (\mathbf{x}_\alpha, \mathbf{y}_\alpha)$ and $P_\beta \sim (\mathbf{x}_\beta, \mathbf{y}_\beta)$"

$$s_{\alpha\beta} := ||\mathbf{x}_\alpha - \mathbf{x}_\beta|| = \sqrt{(x_\alpha - x_\beta)^2 + (y_\alpha - y_\beta)^2}$$

"*decomposition of the tensor-valued, two-point weight function* $\mathbf{G_x}(\mathbf{x}_\alpha, \mathbf{x}_\beta)$ *into the longitudinal weight function* f_ℓ *and the transversal weight function* f_m"

$$\mathbf{G_x}(\mathbf{x}_\alpha, \mathbf{x}_\beta) = \left[g_{j_1 j_2}(\mathbf{x}_\alpha, \mathbf{x}_\beta)\right]$$
$$= f_m(s_{\alpha\beta})\delta_{j_1 j_2} + \left[f_\ell(s_{\alpha\beta}) - f_m(s_{\alpha\beta})\right] \frac{\left[x_{j_1}(P_\alpha) - x_{j_1}(P_\beta)\right]\left[x_{j_2}(P_\alpha) - x_{j_2}(P_\beta)\right]}{s^2_{\alpha\beta}}$$
$$\forall j_1, j_2 \in \{1, 2\}, (x_\alpha, y_\alpha) = (x_1, y_1), (x_\beta, y_\beta) = (x_2, y_2). \tag{3.32}$$

3-223 Optimal Choice of the Weight Matrix: The Space $\mathcal{R}$ (A) and $\mathcal{R}(\mathbf{A})^\perp$

In the introductory paragraph we already outlined the additive basic decomposition of the observation vector into

$$\mathbf{y} = \mathbf{y}_{\mathcal{R}(\mathbf{A})} + \mathbf{y}_{\mathcal{R}(\mathbf{A})^\perp} = \mathbf{y}_\ell + \mathbf{i}_\ell,$$
$$\mathbf{y}_{\mathcal{R}(\mathbf{A})} = \mathbf{P}_{\mathcal{R}(\mathbf{A})}\mathbf{y}, \mathbf{y}_{\mathcal{R}(\mathbf{A})^\perp} = \mathbf{P}_{\mathcal{R}(\mathbf{A})^\perp}\mathbf{y},$$

where $\mathbf{P}_{\mathcal{R}(\mathbf{A})}$ and $\mathbf{P}_{\mathcal{R}(\mathbf{A})^\perp}$ are *projectors* as well as

$\mathbf{y}_\ell \in \mathcal{R}\,(\mathbf{A})$ is an element of the range space $\mathcal{R}\,(\mathbf{A})^\perp$, in general the tangent space $\mathbb{T}_x\mathbb{M}$ of the mapping $f(\mathbf{x})$	versus	$\mathbf{i}_\ell \in \mathcal{R}\,(\mathbf{A})$ is an element of its orthogonal complement in general the normal space $\mathcal{R}\,(\mathbf{A})^\perp$.

$\mathbf{G_y}$-orthogonality $\langle \mathbf{y}_\ell \mid \mathbf{i}_\ell \rangle_{\mathbf{G_y}} = 0$ is proven in Box 3.12.

Box 3.11. ($\mathbf{G_y}$-Orthogonality of $\mathbf{y}_\ell = \mathbf{y}(LESS)$ and $\mathbf{i}_\ell = \mathbf{i}(LESS)$)

"$\mathbf{G_y}$-orthogonality"

$$\langle \mathbf{y}_\ell \mid \mathbf{i}_\ell \rangle_{\mathbf{G_y}} = 0 \tag{3.33}$$

$$\langle \mathbf{y}_\ell \mid \mathbf{i}_\ell \rangle_{\mathbf{G}_\ell} = \mathbf{y}'\left[\mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\right]\mathbf{G_y}\left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y}\right]\mathbf{y}$$
$$= \mathbf{y}'\mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y} - \mathbf{y}'\mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y y} = 0.$$

There is an alternative interpretation of the equations of $\mathbf{G_y}$-orthogonality $\langle \mathbf{i}_\ell \mid \mathbf{y}_\ell \rangle_{\mathbf{G_y}} = \mathbf{i}'_\ell \mathbf{G_y y}_\ell = 0$ of $\mathbf{i}_\ell$ *and* $\mathbf{y}_\ell$. *First*, replace $\mathbf{i}_\ell = \mathbf{P}_{\mathcal{R}(\mathbf{A})^+}\mathbf{y}$ where $\mathbf{P}_{\mathcal{R}(\mathbf{A})^+}$ is a characteristic projection matrix. *Second*, substitute $\mathbf{y}_\ell = \mathbf{A}\mathbf{x}_\ell$ where $\mathbf{x}_\ell$ is $\mathbf{G_y}$-LESS of $\mathbf{x}$. As outlined in *Box 3.13*, $\mathbf{G_y}$-*orthogonality* $\mathbf{i}'_\ell \mathbf{G_y y}_\ell$ of the vectors

$\mathbf{i}_\ell$ and $\mathbf{y}_\ell$ is transformed into the $\mathbf{G_y}$-*orthogonality of the matrices* $\mathbf{A}$ and $\mathbf{B}$. The columns of the matrices $\mathbf{A}$ and $\mathbf{B}$ are $\mathbf{G_y}$-orthogonal. Indeed we have derived the basic equations for transforming

parametric adjustment	into	*adjustment of conditional equations*
$\mathbf{y}_\ell = \mathbf{A}\mathbf{x}_\ell,$		$\mathbf{B}'\mathbf{G_y}\mathbf{y}_\ell = 0,$

by means of

$$\mathbf{B}'\mathbf{G_y}\mathbf{A} = 0.$$

Box 3.12. $\mathbf{G_y}$-orthogonality of $\mathbf{A}$ and $\mathbf{B}$

$$\mathbf{i}_\ell \in \mathcal{R}(\mathbf{A})^\perp, \dim \mathcal{R}(\mathbf{A})^\perp = n - \mathrm{rk}\mathbf{A} = n - m$$

$$\mathbf{y}_\ell \in \mathcal{R}(\mathbf{A}), \ \dim \mathcal{R}(\mathbf{A}) = \mathrm{rk}\mathbf{A} = m$$

$$\langle \mathbf{i}_\ell \mid \mathbf{y}_\ell \rangle_{\mathbf{G_y}} = 0 \Leftrightarrow \left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\right]' \mathbf{G_y}\mathbf{A} = 0 \tag{3.34}$$

$$\mathrm{rk}\left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\right] = n - \mathrm{rk}\mathbf{A} = n - m \tag{3.35}$$

"horizontal rank partitioning"

$$\left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\right] = [\mathbf{B}, \mathbf{C}] \tag{3.36}$$

$$\mathbf{B} \in \mathbb{R}^{n\times(n-m)}, \mathbf{C} \in \mathbb{R}^{n\times m}, rk\mathbf{B} = n - m$$

$$\langle \mathbf{i}_\ell \mid \mathbf{y}_\ell \rangle_{\mathbf{G_y}} = 0 \Leftrightarrow \mathbf{B}'\mathbf{G_y}\mathbf{A} = 0. \tag{3.37}$$

Example 3.3 finally illustrates $\mathbf{G_y}$-orthogonality of the matrices $\mathbf{A}$ and $\mathbf{B}$.

Example 3.3. (gravimetric leveling, $\mathbf{G_y}$-orthogonality of $\mathbf{A}$ and $\mathbf{B}$).

Let us consider a *triangular leveling network* $\{P_\alpha, P_\beta, P_\gamma\}$ which consists of three observations of height differences. These height differences are considered holonomic, determined from gravity potential differences, known as gravimetric leveling. Due to

$$h_{\alpha\beta} := h_\beta - h_\alpha, h_{\beta\gamma} := h_\gamma - h_\beta, h_{\gamma\alpha} := h_\alpha - h_\gamma$$

the *holonomity condition*

$$\int dh = 0 \ \ or \ \ h_{\alpha\beta} + h_{\beta\gamma} + h_{\gamma\alpha} = 0$$

applies. In terms of a *linear model* the observational equations can accordingly be established by

$$\begin{bmatrix} \alpha\beta \\ \beta\gamma \\ \gamma\alpha \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \end{bmatrix} + \begin{bmatrix} i_{\alpha\beta} \\ i_{\beta\gamma} \\ i_{\gamma\alpha} \end{bmatrix}$$

$$\mathbf{y} := \begin{bmatrix} \alpha\beta \\ \beta\gamma \\ \gamma\alpha \end{bmatrix}, \mathbf{A} := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \mathbf{x} := \begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \end{bmatrix}$$

$$\mathbf{y} \in \mathbb{R}^{3\times 1}, \mathbf{A} \in \mathbb{R}^{3\times 2}, \mathrm{rk}\mathbf{A} = 2, \mathbf{x} \in \mathbb{R}^{2\times 1}.$$

First, let us compute $(\mathbf{x}_\ell, \mathbf{y}_\ell, \mathbf{i}_\ell, ||\mathbf{i}_\ell||)$ **I**-LESS of $(\mathbf{x}, \mathbf{y}, \mathbf{i}, ||\mathbf{i}||)$. *A. Bjerhammar's left inverse* supplies us with

$$\mathbf{x}_\ell = \mathbf{A}_\ell^- \mathbf{y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$x_\ell = \begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \end{bmatrix}_\ell = \frac{1}{3}\begin{bmatrix} 2y_1 - y_2 - y_3 \\ -y_1 + 2y_2 - y_3 \end{bmatrix}$$

$$\mathbf{y}_\ell = \mathbf{A}\mathbf{x}_\ell = \mathbf{A}\mathbf{A}_\ell^- \mathbf{y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \mathbf{y}$$

$$\mathbf{y}_\ell = \frac{1}{3}\begin{bmatrix} 2y_1 - y_2 - y_3 \\ -y_1 + 2y_2 - y_3 \\ -y_1 - y_2 + 2y_3 \end{bmatrix}$$

$$\mathbf{i}_\ell = \mathbf{y} - \mathbf{A}\mathbf{x}_\ell = \left(\mathbf{I_n} - \mathbf{A}\mathbf{A}_\ell^-\right)\mathbf{y} = \left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{y}$$

$$\mathbf{i}_\ell = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} y = \frac{1}{3}\begin{bmatrix} y_1 + y_2 + y_3 \\ y_1 + y_2 + y_3 \\ y_1 + y_2 + y_3 \end{bmatrix}$$

$$||\mathbf{i}_\ell||^2 = \mathbf{y}'(\mathbf{I_n} - \mathbf{A}\mathbf{A}_\ell^-)\mathbf{y} = \mathbf{y}'\left[\mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{y}$$

$$||\mathbf{i}_\ell||^2 = [y_1, y_2, y_3]\,\frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$||\mathbf{i}_\ell||^2 = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1).$$

Second, we identify the orthogonality of **A** and **B**. **A** is given, finding **B** is the problem of *horizontal rank partitioning* of the projection matrix.

$$\mathbf{G}_\ell := \mathbf{I_n} - \mathbf{H_y} = \mathbf{I_n} - \mathbf{A}\mathbf{A}_\ell^- = \mathbf{I_n} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3},$$

with special reference to the "*hat matrix* $\mathbf{H_y}{:=}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$". The diagonal elements of $\mathbf{G}_\ell$ are of special interest for *robust approximation*. They amount to the uniform values

$$h_{ii} = \frac{1}{3}(2,2,2), (g_{ii})_\ell = (1-h_{ii}) = \frac{1}{3}(1,1,1).$$

Note

$$\det \mathbf{G}_\ell = \det\left(\mathbf{I_n} - \mathbf{A}\mathbf{A}_\ell^-\right) = 0, \operatorname{rk}(\mathbf{I_n} - \mathbf{A}\mathbf{A}_\ell^-) = n - m = 1$$

$$\mathbf{G}_\ell = \left[\mathbf{I_3} - \mathbf{A}\mathbf{A}_\ell^-\right] = [\mathbf{B}, \mathbf{C}] = \frac{1}{3}\left[\begin{array}{c|cc} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right].$$

$$\mathbf{B} \in \mathbb{R}^{3\times 1}, \mathbf{C} \in \mathbb{R}^{3\times 2}$$

The *holonomity condition* $h_{\alpha\beta} + h_{\beta\gamma} + h_{\gamma\alpha} = 0$ is reestablished by the orthogonality of $\mathbf{B}'\mathbf{A} = 0$.

$$\mathbf{B}'\mathbf{A} = 0 \Leftrightarrow \frac{1}{3}[1,1,1]\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} = [0,0].$$

The $\mathbf{G_y}$-orthogonality condition of the matrices **A** and **B** has been successfully used by *G. Kampmann* (1992, 1994, 1997), *G. Kampmann* and *B. Krause* (1996, 1997), *R. Jurisch*, *G. Kampmann* and *B. Krause* (1997), *R. Jurisch* and *G. Kampmann* (1997, 1998, 2001 a, b, 2002), G. Kampmann and B. Renner (1999), *R. Jurisch*, *G. Kampmann* and *J. Linke* (1999 a, b, c, 2000) in order to balance the observational weights, to robustify $\mathbf{G_y}$-LESS and to identify outliers. The *Grassmann–Plücker coordinates* which span the normal space $\mathcal{R}(\mathbf{A})^\perp$ will be discussed in Chap. 10 when we introduce *condition equations*.

3-224 Fuzzy Sets

How are *fuzzy sets* defined?

Let us start from uncertain quantities based on *interval calculus*. Inaccurate numbers are defined by convex sets demonstrated by Fig. 3.5. We define the determination of an interval by a lower and upper limit and its weighted center by its maximum of the property maximum. Let us call by

$$\mathbf{A} := [a^-, a^0, a^+] \text{ and } \mathbf{B} := [b^-, b^0, b^+]$$

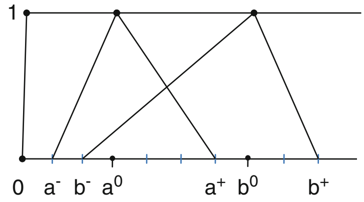


Fig. 3.5 Schematic representation of two inaccurate intervals $\mathbf{A} := [a^-, a^0, a^+]$ and $\mathbf{B} := [b^-, b^0, b^+]$ as well as the weighted centre, [0, 1] weight set

two inaccurate intervals. The lower and upper limit as well as the weighted center are called parameters. We need the following operations based on inaccurate intervals **A** and **B**:

$$(i)\ \mathit{sum}\ \mathbf{A} + \mathbf{B}$$
$$(ii)\ \mathit{difference}\ \mathbf{A} - \mathbf{B}$$
$$(iii)\ \mathit{product}\ \mathbf{A}.\mathbf{B}$$
$$(iv)\ \mathit{substruction}\ \mathbf{A}/\mathbf{B}$$

As a result of these operations is built a new interval $\mathbf{C} := [c^-, c^0, c^+]$, namely

$$(i)\ \mathbf{C} = \mathbf{A} + \mathbf{B}:\ c^- = a^- + b^-,\ c^0 = a^0 - b^0,\ c^+ = a^+ + b^+$$
$$(ii)\ \mathbf{C} = \mathbf{A} - \mathbf{B}:\ c^- = a^- - b^+,\ c^0 = a^0 - b^0,\ c^+ = a^+ - b^-$$
$$(iii)\ \mathbf{C} = \mathbf{A}.\mathbf{B}:\ c^- = min\{a^-b^-,\ a^-b^+,\ a^+b^+\}$$
$$c^0 = a^0b^0$$
$$c^+ = min\{a^-b^-,\ a^+b^-,\ a^-b^+,\ a^+b^+\}$$
$$(iv)\ \mathbf{C} = \mathbf{A}/\mathbf{B}:\ c^- = min\{a^-/b^-,\ a^+/b^-,\ a^-/b^+,\ a^+/b^+\}$$
$$c^0 = a^0/b^0$$
$$c^+ = max\{a^-/b^-,\ a^+/b^-,\ a^-/b^+,\ a^+/b^+\}$$
$$\mathit{if}\ b^- > 0\ \mathit{or}\ b^+ < 0$$

In addition, we define the square $\mathbf{C} := \mathbf{A}^2$ and the root $\mathbf{C} := \sqrt{\mathbf{A}}$ with respect to inaccurate intervals:

$$(v)\ \mathbf{C} := \mathbf{A}^2 = \mathbf{A}.\mathbf{A}$$
$$\begin{cases} c^- = min\{a^-a^-, a^+a^+\},\ c^0 = (c^0)^2,\ c^+ = max\{a^-a^-, a^+a^+\} \\ \mathit{if}\ a^- \geq 0\ \mathit{or}\ a^+ \leq 0 \\ c^- = 0,\ c^0 = (a^-)^2,\ c^+ = max\{a^-a^-, a^+a^+\} \\ \mathit{if}\ a^- < 0\ \mathit{and}\ a^+ > 0 \end{cases}$$
$$(vi)\ \mathbf{C} := \sqrt{\mathbf{A}}$$
$$c^- = \sqrt{a^-},\ c^0 = \sqrt{a^0},\ c^+ = \sqrt{a^+},\ \mathit{if}\ a^- \geq 0.$$

Example: Mathematical modelling of inaccurate surfaces

Let us consider inaccurate height measurement by the interval $\mathbf{z}_i^n := [z_i^-, z_i^0, z_i^+]$ for all $i \in [1, \ldots, n]$ with sharp coordinates (x_i, y_i) for all $i \in [1, \ldots, n]$. If these measurement data are given at a sharp grid (X_j, Y_k) for all $j \in [1, \ldots, N]$ and $k \in [1, \ldots, M]$ and followed by an interpolation, then the interpolated heights $\mathbf{Z}_{jk}^n := [Z_{jk}^-, Z_{jk}^0, Z_{jk}^+]$ at the nodal point of the grid are also inaccurate. The results depend on the method of interpolation which are functions of *inaccurate intervals*.

We will use *distance depending* weight functions such that

$$Z_{jk}^n = \alpha_1^{jk} z_1^n + \alpha_2^{jk} z_2^n + \cdots + \alpha_n^{jk} z_n^n$$

$$\textit{subject to}$$

$$\sum_{i=1}^n \alpha_i^{jk} = 1,\ \alpha_i^{jk} = \alpha_i^{jk}(d_i^{ij})$$

$$d_i^{ij} = (x_i - X_j)^2 + (y_i - Y_k)^2 \ \textit{for all } i \in [1, \ldots, n],\ j \in [1, \ldots, N], k \in [1, \ldots, M]$$

The assumption is the one that weights are inversely proportional to the square of distances:

$$Z_{jk}^n = \alpha_1^{jk} z_1^n + \cdots + \alpha_n^{jk} z_n^n$$

$$\alpha_i^{jk} = (w_i^{jk})/[\sum_{i=1}^n (w_i^{jk})]$$

$$\textit{subject to}$$

$$w_i^{jk} = 1/d_i^{jk} + \varepsilon \ \textit{for all } \varepsilon > 0.$$

By the choice of an artificial constant ε we are able to tune in such a way that weights gets inaccurate *if* the prediction point is a measurement point.

In summary, we can represent the interpolated height measurements for position weight functions by the intervals

$$\boxed{\begin{aligned} Z_{jk}^- &= \alpha_1^{jk} z_1^- + \cdots + \alpha_n^{jk} z_n^- = \sum_{i=1}^n \alpha_i^{jk} z_i^- \\ Z_{jk}^0 &= \alpha_1^{jk} z_1^0 + \cdots + \alpha_n^{jk} z_n^0 = \sum_{i=1}^n \alpha_i^{jk} z_i^0 \\ Z_{jk}^+ &= \alpha_1^{jk} z_1^+ + \cdots + \alpha_n^{jk} z_n^+ = \sum_{i=1}^n \alpha_i^{jk} z_i^+ \end{aligned}}$$

Please not our indices $i \in [1, \ldots, n]$, $j \in [1, \ldots, N]$ and $k \in [1, \ldots, M]$. In the next step, we relate the inaccurate interpolated measurements to an analytical surface. Again, such a resulting surface is inaccurate, too. Here we use a *two-dimensional Lagrange polynomial*, for instance

$$\phi_j(x) = \prod_{\substack{j \in [1,\dots,N] \\ i \neq j}} \frac{x - X_j}{(X_j - X_i)}$$

$$\phi_k(y) = \prod_{\substack{i \in [1,\dots,M] \\ i \neq k}} \frac{y - Y_i}{(Y_k - X_i)}.$$

These polynomial transfer the well known property at the nodal point $\{X_j, Y_k\}$ for $j \in [1,\dots,N], k \in [1,\dots,M]$:

$$\phi_j(x_p) = \delta_{jp} = \begin{cases} 1, p = j \\ 0, p \neq j \end{cases}, \ \phi_k(y_p) = \delta_{ip} = \begin{cases} 1, p = k \\ 0, p \neq k \end{cases}$$

Now we are able to represent the approximating polynomial by inaccurate surface data by

$$\tilde{p}(x,y) = \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M]}} \tilde{Z}_{jk}\phi_j(x)\phi_k(y) = \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M]}} \tilde{Z}_{jk} L_{jk}(x,y)$$

subject to

$$L_{jk}(x,y) := \phi_j(x)\phi_k(y)$$

and

$$\boxed{\begin{gathered} p^-(x,y) = \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M] \\ L_{jk} \geq 0}} Z^-_{jk} L_{jk}(x,y) + \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M] \\ L_{jk} < 0}} Z^+_{jk} L_{jk}(x,y) \\ p^0(x,y) = \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M]}} Z^0_{jk} L_{jk}(x,y) \\ p^+(x,y) = \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M] \\ L_{jk} \geq 0}} Z^+_{jk} L_{jk}(x,y) + \sum_{\substack{j \in [1,\dots,N] \\ k \in [1,\dots,M] \\ L_{jk} < 0}} Z^-_{jk} L_{jk}(x,y) \end{gathered}}$$

These three sets of surfaces relate to the lower, the middle and the upper limits of uncertain surfaces. *These surfaces are continuous and differentiable*. How to calculate inaccurate first differentials given inaccurate surfaces? We propose the following processes.

$$\tilde{\triangle}_x := \frac{d}{dx}\tilde{p}(x,y) = \frac{d}{dx}\sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]}} \tilde{Z}_{jk}\phi_j(x)\phi_k(y) = $$
$$= \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]}} \tilde{Z}_{jk}\phi_k(y)\frac{d}{dx}\phi_j(x)$$

$$\tilde{\triangle}_y := \frac{d}{dy}\tilde{p}(x,y) = \frac{d}{dy}\sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]}} \tilde{Z}_{jk}\phi_j(x)\phi_k(y) = $$
$$= \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]}} \tilde{Z}_{jk}\phi_j(x)\frac{d}{dy}\phi_k(y)$$

and

$$\boxed{\begin{aligned}
\triangle_x^- &:= \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]\\ \phi_k(y)\frac{d}{dx}\phi_j(x)\geq 0}} Z_{jk}^-\phi_k(y)\frac{d}{dx}\phi_j(x) + \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]\\ \phi_k(y)\frac{d}{dx}\phi_j(x)< 0}} Z_{jk}^+\phi_k(y)\frac{d}{dx}\phi_j(x)\\
\triangle_x^0 &:= \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]}} Z_{jk}^0\phi_k(y)\frac{d}{dx}\phi_j(x)\\
\triangle_x^+ &:= \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]\\ \phi_k(y)\frac{d}{dx}\phi_j(x)\geq 0}} Z_{jk}^+\phi_k(y)\frac{d}{dx}\phi_j(x) + \sum_{\substack{j\in[1,\dots,N]\\ k\in[1,\dots,M]\\ \phi_k(y)\frac{d}{dx}\phi_j(x)< 0}} Z_{jk}^-\phi_k(y)\frac{d}{dx}\phi_j(x)
\end{aligned}}$$

An analog formula exist for $\{\triangle_y^-, \triangle_y^0, \triangle_y^+\}$. The propagation of inaccurate data can be achieved by the following:

(i) Mean operation

Define the *mean* $\tilde{m}$ and separate the three parts belonging to *inaccurate means*.

$$m^+ := \sum_{\substack{j,k=1\\ \beta_{jk}\geq 0}}^{L} \beta_{jk} z_i^- + \sum_{\substack{j,k=1\\ \beta_{jk}<0}}^{L} \beta_{jk} z_i^+$$

$$m^0 := \sum_{j,k=1}^{N} \beta_{jk} z_i^0$$

$$m^- := \sum_{\substack{j,k=1 \\ \beta_{jk} \geq 0}}^{L} \beta_{jk} z_i^+ + \sum_{\substack{j,k=1 \\ \beta_{jk} < 0}}^{L} \beta_{jk} z_i^-$$

The window is chosen by $L = 3, 5, 7$ etc.

Partial derivatives

$$\tilde{\triangle}_x^{jk} := \frac{1}{x_{j+1} - x_{j-1}} [\tilde{Z}_{(j+1)k} - \tilde{Z}_{(j-1)k}]$$

$$\tilde{\triangle}_y^{jk} := \frac{1}{y_{j+1} - y_{j-1}} [\tilde{Z}_{j(k+1)} - \tilde{Z}_{j(k-1)}]$$

respectively

$$\tilde{\triangle}_x^{jk-} := \frac{1}{x_{j+1} - x_{j-1}} [\tilde{Z}^-_{(j-1)k} - \tilde{Z}^+_{(j-1)k}],$$

$$\tilde{\triangle}_x^{jk0} := \frac{1}{x_{j+1} - x_{j-1}} [\tilde{Z}^0_{(j-1)k} - \tilde{Z}^0_{(j-1)k}],$$

$$\tilde{\triangle}_x^{jk+} := \frac{1}{x_{j+1} - x_{j-1}} [\tilde{Z}^+_{(j-1)k} - \tilde{Z}^-_{(j-1)k}],$$

for all $j \in [2, \ldots, N-1]$ and $k \in [2, \ldots, M-1]$.

The partial derivatives with respect to y can be determined in an analog manner. Higher derivatives can be computed in a recursive way, for instance

$$\boxed{\begin{aligned} \tilde{\triangle}_x(p)^{jk-} &:= \frac{1}{x_{j+1} - x_{j-1}} [\tilde{\triangle}_{x(p-1)}^{(j+1)k-} - \tilde{\triangle}_{x(p-1)}^{(j-1)k+}] \\ \tilde{\triangle}_x(p)^{jk0} &:= \frac{1}{x_{j+1} - x_{j-1}} [\tilde{\triangle}_{x(p-1)}^{(j+1)0} - \tilde{\triangle}_{x(p-1)}^{(j-1)0}] \\ \tilde{\triangle}_x(p)^{jk+} &:= \frac{1}{x_{j+1} - x_{j-1}} [\tilde{\triangle}_{x(p-1)}^{(j+1)k+} - \tilde{\triangle}_{x(p-1)}^{(j-1)k-}] \end{aligned}}$$

In analogous manner, we compute the derivatives with respect to y.

Example: Deformation analysis of surfaces (E. Grafarend (2006), E, Grafarend and F. Krumm (2006), B. Voosoghi (2000), D. Waelder (2008))

Here we will apply the *theory of inaccurate data* to surface deformation, namely *surface dilatation and surface maximum shear*. In detail we refer to the sum of eigenvalues $\widetilde{DIL}$, namely $\lambda_1 + \lambda_2$, and its difference $\widetilde{\Gamma}^2$, namely $\tilde{\lambda}_1 + \tilde{\lambda}_2$ subject to the eigenvalue representation

$$\tilde{\lambda}_1 := \frac{1}{2}[tr(E_r^T G_r^{-1}) \pm \sqrt{[tr(E_r G_r^{-1})]^2 + 4\,det\,[E_r G_r^{-1}]}$$

$$\tilde{\lambda}_2 := \frac{1}{2}[tr(E_r^T G_r^{-1}) \pm \sqrt{[tr(E_r G_r^{-1})]^2 - 4\,det\,[E_r G_r^{-1}]}$$

The eigenvalues, namely in *E. Grafarend and F. Krumm* (2006 p.29), refer to the *Hilbert basis invariants* trace determinant. They are built on *Euler–Lagrange strain* $E_r^T G_r^{-1}$ where E_r^T is the "*right strain component*" and G_r is the matrix of "*right Gauss metric*". Indeed we map from *a left metric* G_l and a *right metric* G_r, for instance from an ellipsoid to a sphere. The *Hilbert basis invariants* relate to dilatation $\widetilde{DIL}$ and surface maximum shear $\widetilde{\Gamma}^2$ by the following representation,

$$\widetilde{DIL} = \tilde{\lambda}_1 + \tilde{\lambda}_2 \; and \; \widetilde{\Gamma}^2 = \tilde{\lambda}_1 - \tilde{\lambda}_2$$

$$\widetilde{DIL} = tr(E_r^T G_r^{-1}) \; and \; \widetilde{\Gamma}^2 = [tr(\tilde{E}_r^T \tilde{G}_r^{-1})]^2 - 4\,det\,(\tilde{E}_r \tilde{G}_r^{-1})$$

For detail computation we use *right Gauss metric* as well as the *left Gauss metric* and derive the corresponding operators $\widetilde{DIL}$ and $\widetilde{\Gamma}^2$, respectively.

Coordinates of the right metric matrix

$$\mathbf{G}_r := \begin{bmatrix} 1 + (\tilde{\Delta}_x^{jk})_r^2 & (\tilde{\Delta}_x^{jk})_r (\tilde{\Delta}_y^{jk})_r \\ (\tilde{\Delta}_x^{jk})_r (\tilde{\Delta}_y^{jk})_r & 1 + (\tilde{\Delta}_y^{jk})_r^2 \end{bmatrix}$$

$$\mathbf{G}_l := \begin{bmatrix} 1 + (\tilde{\Delta}_x^{jk})_l^2 & (\tilde{\Delta}_x^{jk})_l (\tilde{\Delta}_y^{jk})_l \\ (\tilde{\Delta}_x^{jk})_l (\tilde{\Delta}_y^{jk})_l & 1 + (\tilde{\Delta}_y^{jk})_l^2 \end{bmatrix}$$

In order to calculate the *Euler–Lagrange tensor* $\tilde{\mathbf{E}}_r := (\mathbf{G}_r - \mathbf{C}_r)/2$ based on the *Cauchy–Green tensor* $\mathbf{C}_r := \mathbf{U}^T \mathbf{G}_l \mathbf{U}$, $\mathbf{U} := [\partial u / \partial U]$ we need *mapping equations*, namely the partial derivatives $\partial u^\alpha / \partial U^A \; u^\alpha = u^\alpha(U^A)$, indeed the mapping equations *from the ellipsoid to the sphere*. Then finally we are able to calculate the right *Euler–Lagrange tensor* $\tilde{\mathbf{E}}_r := (\mathbf{G}_r - \mathbf{C}_r)/2$. We are ready to present you the test results.

Finally we refer to some recent publication on *fuzzy sets and systems*: *H. Bunke and O. Bunke* (1989), *J. C. Bezdek* (1973), *G. Alefeld and J. Herzberger* (1974,1983), H. *Bandemer and S. Gottwald* (1993), R. Fletling (2007), K. Grahisch (1998), O. Kaleva (1994), *B. F. Arnold* and *P. Stahlecker* (1999), *A. Chaturvedi* and *A. T. K. Wan* (1999), *S. M. Guu*, *Y. Y. Lur* and *C. T. Pang* (2001), *H. Jshibuchi*, *K. Nozaki* and *H. Tanaka* (1992), *H. Jshibuchi*, *K. Nozaki*, *N. Yamamoto* and *H. Tanaka* (1995), *B. Kosko* (1992), *H. Kutterer* (1994, 1999), *V. Ravi*, *P. J. Reddy* and *H. J. Zimmermann* (2000), *V. Ravi* and *H. J. Zimmermann* (2000), *S. Wang*, *T. Shi* and *C. Wu* (2001), *L. Zadch* (1965), *B. Voosoghi* (2000), and *H. J. Zimmermann* (1991).

Geodetic applications were reported by *H. Kutterer* (2002), *J. Neumann and H. Kutterer* (2007), *S. Schön* (2003), *S. Schön and H. Kutterer* (2006), *W. Näther and K. Wälder* (2003,2007), *O. Wälder* (2006, 2007a,b) and *O. Wälder and M. Buchroithner* (2004).

3-23 $\mathbf{G_x}$-*LESS and Its Generalized Inverse*

A more formal version of the *generalized inverse* which is characteristic for $\mathbf{G_y}$-LESS is presented by

Lemma 3.7. *(characterization of* $\mathbf{G_y}$*-LESS):*

$\mathbf{x}_\ell = \mathbf{Ly}$ is I-LESS of the inconsistent system of linear equations (3.1) $\mathbf{Ax}+\mathbf{i} = \mathbf{y}$, rk$\mathbf{A} = m$, (or $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$) if and only if the matrix $L \in \mathbb{R}^{m\times n}$ fulfils

$$\left[\begin{array}{l} \mathbf{ALA} = \mathbf{A} \\ \mathbf{AL} = (\mathbf{AL})'. \end{array}\right. \tag{3.38}$$

The matrix $\mathbf{L}$ is the unique $\mathbf{A}^{1,2,3}$ generalized inverse, also called *left inverse* $\mathbf{A}_L^-$.

$\mathbf{x}_\ell = \mathbf{Ly}$ is $\mathbf{G_y}$-LESS of the inconsistent system of linear equations (3.1) $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, rk$\mathbf{A} = m$ (or $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$) if and only if the matrix $\mathbf{L}$ fulfils

$$\left[\begin{array}{l} \mathbf{G_y ALA} = \mathbf{G_y A} \\ \mathbf{G_y AL} = (\mathbf{G_y AL})'. \end{array}\right. \tag{3.39}$$

The matrix $\mathbf{L}$ is the $\mathbf{G_y}$ weighted $\mathbf{A}^{1,2,3}$ generalized inverse, in short $\mathbf{A}_\ell^-$, also called *weighted left inverse*.

Proof.

According to the theory of the *generalized inverse* presented in Appendix A $\mathbf{x}_\ell = \mathbf{Ly}$ is $\mathbf{G_y}$-LESS of (3.1) if and only if $\mathbf{A'G_yAL} = \mathbf{A'G_y}$ is fulfilled. Indeed $\mathbf{A'G_yAL} = \mathbf{A'G_y}$ is equivalent to the two conditions $\mathbf{G_yALA} = \mathbf{G_yA}$ and $\mathbf{G_yAL} = (\mathbf{G_yAL})'$. For a proof of such a statement multiply $\mathbf{A'G_yAL} = \mathbf{A'G_y}$ *left* by $\mathbf{L}'$ *and* receive

$$\mathbf{L'A'G_yAL} = \mathbf{L'A'G_y}.$$

The *left-hand side* of such a matrix identity is a *symmetric matrix*. In consequence, the right-hand side has to be symmetric, too. When applying the central symmetry condition to

$$\mathbf{A'G_yAL} = \mathbf{A'G_y} \ or \ \mathbf{G_yA} = \mathbf{L'A'G_yA},$$

we are led to

$$\mathbf{G_yAL} = \mathbf{L'A'G_yAL} = (\mathbf{G_yAL}),$$

what had to be proven.

? How to prove *uniqueness* of $\mathbf{A}^{1,2,3} = \mathbf{A}_\ell$?

Let us fulfil $\mathbf{G_y A x}_\ell$ by

$$\mathbf{G_y A L_1 y} = \mathbf{G_y A L_1 A L_1 y} = \mathbf{L'_1 A' G_y A L_1 y} = \mathbf{L'_1 A' L'_1 A' G_y y} =$$

$$= \mathbf{L'_1 A' L'_2 A' G_y y} = \mathbf{L'_1 A' G_y L_2 y} = \mathbf{G_y A L_1 A L_2 y} = \mathbf{G_y A L_2 y},$$

in particular by two arbitrary matrices $\mathbf{L_1}$ and $\mathbf{L_2}$, respectively, which fulfil

$$(i)\ \mathbf{G_y ALA} = \mathbf{G_y A}\ \textit{as well as}$$

$$(ii)\ \mathbf{G_y AL} = (\mathbf{G_y AL})'.$$

Indeed we have derived *one result* irrespective of $\mathbf{L_1}$ or $\mathbf{L_2}$.

If the matrix of the metric $\mathbf{G_y}$ of the *observation space* is positive definite, we can prove the following *duality*

Theorem 3.8. (duality):

Let the matrix of the metric $\mathbf{G_x}$ of the observation space be positive definite. Then $\mathbf{x}_\ell = \mathbf{Ly}$ is $\mathbf{G_y}$-LESS of the linear model (3.1) for any observation vector $\mathbf{y} \in \mathbb{R}^n$, if $\mathbf{x}_m^\sim = \mathbf{L'y}^\sim$ is $\mathbf{G_y}^{-1}$-MINOS of the linear model $\mathbf{y}^\sim = \mathbf{A'x}$ for all $m \times 1$ columns $\mathbf{y}^\sim \in \mathcal{R}(\mathbf{A'})$.

Proof. If $\mathbf{G_y}$ is positive definite, there exists the inverse matrix $\mathbf{G_y}^{-1}$. Equation (3.39) can be transformed into the equivalent condition

$$\mathbf{A'} = \mathbf{A'L'A}\ \textit{and}\ \mathbf{G_y}^{-1}\mathbf{L'A'} = (\mathbf{G_y}^{-1}\mathbf{L'A'})',$$

which is equivalent to (1.81).

3-24 *Eigenvalue Decomposition of* $\mathbf{G_y}$*-LESS: Canonical LESS*

For the system analysis of an inverse problem the *eigenspace analysis* and *eigenspace synthesis* of $\mathbf{x}_\ell$ $\mathbf{G_y}$-LESS of $\mathbf{x}$ is very useful and gives some peculiar insight into a dynamical system. Accordingly we are confronted with the problem to construct "*canonical LESS*", also called the *eigenvalue decomposition* of $\mathbf{G_y}$-LESS.

First, we refer to the canonical representation of the parameter space $\mathbb{X}$ as well as the observation space introduced to you in the Chap. 1, *Box 1.8* and *Box 1.9*. But here we add by means of *Box 3.14* the comparison of the *general bases* versus the *orthonormal bases* spanning the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$. In addition, we refer to *Definition 1.5* and *Lemma 1.6* where the adjoint operator $\mathbf{A}^\#$ has been introduced and represented.

Box 3.13. (General bases versus orthonormal bases spanning the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$)

"*left*"	"*right*"
"*parameter space*"	"*observation space*"
"*general left base*"	"*general right base*"

$$\text{span}\{\mathbf{a_1}, \ldots, \mathbf{a_m}\} = \mathbb{X} \qquad \mathbb{Y} = \text{span}\{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$$

: *matrix of the metric* :   : *matrix of the metric* :

$$\mathbf{aa'} = \mathbf{G_x} \qquad \mathbf{bb'} = \mathbf{G_y} \tag{3.40}$$

"*orthonormal left base*"   "*orthonormal right base*"

$$\text{span}\{\mathbf{e_1^x}, \ldots, \mathbf{e_m^x}\} = \mathbb{X} \qquad \mathbb{Y} = \text{span}\{\mathbf{e_1^y}, \ldots, \mathbf{e_n^y}\}$$

: *matrix of the metric* :   : *matrix of the metric* :

$$\mathbf{e_x e'_x} = \mathbf{I_m} \qquad \mathbf{e_y e'_y} = \mathbf{I_n} \tag{3.41}$$

"*base transformation*"   "*base transformation*"

$$\mathbf{a} = \mathbf{\Lambda_x^{\frac{1}{2}} V e_x} \qquad \mathbf{b} = \mathbf{\Lambda_y^{\frac{1}{2}} U e_y} \tag{3.42}$$

versus   versus

$$\mathbf{e_x} = \mathbf{V' \Lambda_x^{-\frac{1}{2}} a} \qquad \mathbf{e_y} = \mathbf{U' \Lambda_y^{-\frac{1}{2}} b} \tag{3.43}$$

$$\text{span}\{\mathbf{e}_1^{\mathbf{x}}, \ldots, \mathbf{e}_m^{\mathbf{x}}\} = \mathbb{X} \qquad \mathbb{Y} = \text{span}\{\mathbf{e}_1^{\mathbf{y}}, \ldots, \mathbf{e}_n^{\mathbf{y}}\}$$

Second, we are going to solve the overdetermined system of

$$\{\mathbf{y} = \mathbf{Ax} | \mathbf{A} \in \mathbb{R}^{n \times m}, \text{rk}\mathbf{A} = n, n > m\}$$

by introducing

- The eigenspace of the rectangular matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ of rank $r := \text{rk}\mathbf{A} = m$, $n > m : \mathbf{A} \mapsto \mathbf{A}^*$
- The left and right canonical coordinates: $\mathbf{x} \to \mathbf{x}^*$, $\mathbf{y} \to \mathbf{y}^*$

as supported by *Box 3.15*. The transformations $\mathbf{x} \mapsto \mathbf{x}^*$, $\mathbf{y} \mapsto \mathbf{y}^*$ (3.44) from the original coordinates $(x_1, \ldots, x_m)$ to the canonical coordinates $(x_1^*, \ldots, x_m^*)$, the *left star coordinates*, as well as from the original coordinates $(y_1, \ldots, y_n)$ to the canonical coordinates $(y_1^*, \ldots, y_n^*)$, the *right star coordinates*, are *polar decompositions*: a *rotation* $\{\mathbf{U}, \mathbf{V}\}$ is followed by a *general stretch* $\{\mathbf{G_y^{\frac{1}{2}}}, \mathbf{G_x^{\frac{1}{2}}}\}$. Those root matrices are generated by *product decompositions* of type $\mathbf{G_y} = (\mathbf{G_y^{\frac{1}{2}}})'\mathbf{G_y^{\frac{1}{2}}}$ as well as $\mathbf{G_x} = (\mathbf{G_x^{\frac{1}{2}}})'\mathbf{G_x^{\frac{1}{2}}}$. Let us substitute the *inverse transformations* $\mathbf{x}^* \mapsto \mathbf{x} = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{Vx}^*$ and $\mathbf{y}^* \mapsto \mathbf{y} = \mathbf{G_y}^{-\frac{1}{2}}\mathbf{Uy}^*$ in (3.45) into the system of linear equations (3.1) $\mathbf{y} = \mathbf{Ax} + \mathbf{i}$ or its dual (3.46) $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$. Such an operation leads us to $\mathbf{y}^* = f(\mathbf{x}^*)$ as well as $\mathbf{y} = f(\mathbf{x})$ in (3.47). Subject to the orthonormality conditions $\mathbf{U'U} = \mathbf{I}_n$ and $\mathbf{V'V} = \mathbf{I}_m$ in (3.48) we have generated the *left-right eigenspace analysis* (3.49)

$$\mathbf{\Lambda}^* = \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix}$$

subject to the *horizontal rank partitioning* of the matrix $\mathbf{U} = [\mathbf{U_1}, \mathbf{U_2}]$. Alternatively, the *left-right eigenspace synthesis* (3.50)

$$\mathbf{A} = \mathbf{G_y}^{-\frac{1}{2}} [\mathbf{U_1}, \mathbf{U_2}] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{V}'\mathbf{G_x^{\frac{1}{2}}}$$

is based upon the *left matrix* $\mathbf{L} := \mathbf{G_y}^{-\frac{1}{2}}\mathbf{U}$ and the *right matrix* $\mathbf{R} := \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V}$ in (3.51). Indeed the *left matrix* $\mathbf{L}$ in (3.52) $\mathbf{LL}' = \mathbf{G_y}^{-1}$ reconstructs the inverse matrix of the metric of the observation space $\mathbb{Y}$. Similarly, the *right matrix* $\mathbf{R}$ by means of (3.52) $\mathbf{RR}' = \mathbf{G_x}^{-1}$ generates the inverse matrix of the metric of the parameter space $\mathbb{X}$. In terms of "$\mathbf{L}$, $\mathbf{R}$" we have summarized the *eigenvalue decompositions* (3.53)–(3.55). Such an eigenvalue decomposition helps us to *canonically invert* $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$ by means of (3.56), (3.57), namely the *rank partitioning* of the *canonical observation vector* $\mathbf{y}^*$ into $\mathbf{y}_1^* \in \mathbb{R}^{r\times 1}$ and $\mathbf{y}_2^* \in \mathbb{R}^{(n-r)\times 1}$ to determine $\mathbf{x}_\ell^* = \mathbf{\Lambda}^{-1}\mathbf{y}_1^*$ leaving $\mathbf{y}_2^*$ "*unrecognized*". Next we shall proof $\mathbf{i}_1^* = 0$ if $\mathbf{i}_1^*$ is LESS.

Box 3.14. (Canonical representation, overdetermined system of linear equations)

"*parameter space* $\mathbb{X}$" versus "*observation space* $\mathbb{Y}$"

$$\mathbf{x}^* = \mathbf{V}'\mathbf{G_x^{\frac{1}{2}}}\mathbf{x} \qquad \mathbf{y}^* = \mathbf{U}'\mathbf{G_y^{\frac{1}{2}}}\mathbf{y} \tag{3.44}$$

and *and*

$$\mathbf{x} = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{Vx}^* \qquad \mathbf{y} = \mathbf{G_y}^{-\frac{1}{2}}\mathbf{Uy}^* \tag{3.45}$$

"*overdetermined system of linear equations*"

$$\{\mathbf{y} = \mathbf{Ax} + \mathbf{i} | \mathbf{A} \in \mathbb{R}^{n\times m},\ rk\mathbf{A} = m, n > m\}$$

$$\mathbf{y} = \mathbf{Ax} + \mathbf{i} \qquad \textit{versus} \qquad \mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^* \tag{3.46}$$

$$\mathbf{G_y^{-\frac{1}{2}} U y^* = A G_x^{-\frac{1}{2}} V x^* + G_y^{-\frac{1}{2}} U i^*} \qquad \mathbf{U' G_y^{\frac{1}{2}} y = A^* V' G_x^{\frac{1}{2}} x + U' G_y^{\frac{1}{2}} i}$$

$$\mathbf{y^*} = \left(\mathbf{U' G_y^{\frac{1}{2}} A G_x^{-\frac{1}{2}} V}\right)\mathbf{x^* + i^*} \qquad \mathbf{y} = \left(\mathbf{G_y^{-\frac{1}{2}} U A^* V' G_x^{\frac{1}{2}}}\right)\mathbf{x + i} \quad (3.47)$$

subject to *subject to*

$$\mathbf{U'U = UU' = I_n} \quad \textit{versus} \quad \mathbf{V'V = VV' = I_m} \quad (3.48)$$

"*left and right eigenspace*"

$$\text{"}\textit{left eigenspace analysis}\text{''}: \mathbf{A^*} = \begin{bmatrix} \mathbf{U'_1} \\ \mathbf{U'_2} \end{bmatrix} \mathbf{G_y^{\frac{1}{2}} A G_x^{-\frac{1}{2}} V} = \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \quad (3.49)$$

$$\text{"}\textit{right eigenspace analysis}\text{''}: \mathbf{A = G_y^{-\frac{1}{2}}} [\mathbf{U_1}, \mathbf{U_2}] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{V' G_x^{\frac{1}{2}}} \quad (3.50)$$

"*dimension identities*"

$$\mathbf{\Lambda} \in \mathbb{R}^{r\times r}, \mathbf{U_1} \in \mathbb{R}^{n\times r}$$
$$\mathbf{0} \in \mathbb{R}^{(n-r)\times r}, \mathbf{U_2} \in \mathbb{R}^{n\times(n-r)}, \mathbf{V} \in \mathbb{R}^{r\times r}$$
$$r := \mathrm{rk}\mathbf{A} = m, n > m$$

"*left eigenspace*" "*right eigenspace*"

$$\mathbf{L := G_y^{-\frac{1}{2}} U \Rightarrow L^{-1} = U' G_y^{\frac{1}{2}}} \quad \textit{versus} \quad \mathbf{R := G_x^{-\frac{1}{2}} V \Rightarrow R^{-1} = V' G_x^{\frac{1}{2}}} \quad (3.51)$$

$$\mathbf{L_1 := G_y^{-\frac{1}{2}} U_1, L_2 := G_y^{-\frac{1}{2}} U_2} \Rightarrow$$

$$\mathbf{LL' = G_y^{-1} \Rightarrow (L^{-1})'L^{-1} = G_y} \quad \textit{versus} \quad \mathbf{RR' = G_x^{-1} \Rightarrow (R^{-1})'R^{-1} = G_x} \quad (3.52)$$

$$\mathbf{L^{-1}} = \begin{bmatrix} \mathbf{U'_1} \\ \mathbf{U'_2} \end{bmatrix} \mathbf{G_y^{\frac{1}{2}}} =: \begin{bmatrix} \mathbf{L_1^-} \\ \mathbf{L_2^-} \end{bmatrix}$$

$$\mathbf{A = L A^* R^{-1}} \quad \textit{versus} \quad \mathbf{A^* = L^{-1} A R} \quad (3.53)$$

$$\mathbf{A} = [\mathbf{L_1}, \mathbf{L_2}]\, \mathbf{A^\# R^{-1}} \quad \textit{versus} \quad \mathbf{A^*} = \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{L_1^-} \\ \mathbf{L_2^-} \end{bmatrix} \mathbf{AR} \quad (3.54)$$

$$\left[\begin{array}{l} \mathbf{A^\# A L_1 = L_1 \Lambda^2} \\ \mathbf{A^\# A L_2 = 0} \end{array}\right. \quad \textit{versus} \quad \mathbf{A A^\# R = R \Lambda^2} \quad (3.55)$$

"*overdetermined system of linear equations solved in canonical coordinates*"

$$\begin{tikzcd}
\mathbb{X} \ni \mathbf{x} \arrow[r, "\mathbf{A}"] \arrow[d, "\mathbf{V}'\mathbf{G}_{\mathbf{y}}^{\frac{1}{2}}"'] & \mathbf{y} \in \mathcal{R}(\mathbf{A}) \subset \mathbb{Y} \arrow[d, "\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{\frac{1}{2}}"] \\
\mathbb{X} \ni \mathbf{x}^* \arrow[r, "\mathbf{A}^*"'] & \mathbf{y}^* \in \mathcal{R}(\mathbf{A}^*) \subset \mathbb{Y}
\end{tikzcd}$$

Fig. 3.6 Commutative diagram of coordinate transformations

$$\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^* = \begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{x}^* + \begin{bmatrix} \mathbf{i}_1^* \\ \mathbf{i}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} \tag{3.56}$$

"*dimension identities*"

$$\begin{gathered} \mathbf{y}_1^* \in \mathbb{R}^{r\times 1}, \mathbf{y}_2^* \in \mathbb{R}^{(n-r)\times 1}, \mathbf{i}_1^* \in \mathbb{R}^{r\times 1}, \mathbf{i}_2^* \in \mathbb{R}^{(n-r)\times 1} \\ \mathbf{y}_1^* = \boldsymbol{\Lambda} \boldsymbol{x}^* + \mathbf{i}_1^* \Rightarrow \mathbf{x}^* = \boldsymbol{\Lambda}^{-1}(\mathbf{y}_1^* - \mathbf{i}_1^*) \\ \text{“}\textit{if } \mathbf{i}^* \textit{ is LESS, then } \mathbf{x}_\ell^* = \boldsymbol{\Lambda}^{-1}\mathbf{y}_1^*, \mathbf{i}_1^* = 0 \end{gathered} \tag{3.57}$$

Consult the commutative diagram of *Fig. 3.6* for a shorthand summary of the newly introduced transformations of coordinates, both of the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$. *Third*, we prepare ourselves for LESS of the overdetermined system of linear equations

$$\{\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} | \mathbf{A} \in \mathbb{R}^{n\times m}, \mathrm{rk}\mathbf{A} = m, n > m, ||\mathbf{i}||_{\mathbf{G_y}}^2 = \min\}$$

by introducing *Lemma 3.9*, namely the *eigenvalue-eigencolumn equations* of the matrices $\mathbf{A}^\#\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\#$, respectively, as well as *Lemma 3.11*, our basic result of "*canonical LESS*", subsequently completed by proofs. Throughout we refer to the adjoint operator which has been introduced by *Definition 1.5* and *Lemma 1.6*.

Lemma 3.9. *(eigenspace analysis versus eigenspace synthesis of the matrix* $\mathbf{A} \in \mathbb{R}^{n\times m}, r := \mathrm{rk}\mathbf{A} = m < n$*)*

The pair of matrices $\{\mathbf{L}, \mathbf{R}\}$ for the eigenspace analysis and the eigenspace synthesis of the rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ of rank $r := \mathrm{rk}\mathbf{A} = m < n$, namely

$$\mathbf{A}^* = \mathbf{L}^{-1}\mathbf{A}\mathbf{R} \quad versus \quad \mathbf{A} = \mathbf{L}\mathbf{A}^*\mathbf{R}^{-1}$$

or

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1^- \\ \mathbf{L}_2^- \end{bmatrix} \mathbf{AR} \quad versus \quad \mathbf{A} = [\mathbf{L}_1, \mathbf{L}_2] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{R}$$

are determined by the eigenvalue-eigencolumn equations (eigenspace equations) of the matrices $\mathbf{A}^{\#}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{\#}$, respectively, namely

$$\mathbf{A}^{\#}\mathbf{AR} = \mathbf{R}\mathbf{\Lambda}^2 \quad versus \quad \begin{bmatrix} \mathbf{AA}^{\#}\mathbf{L}_1 = \mathbf{L}_1\mathbf{\Lambda}^2 \\ \mathbf{AA}^{\#}\mathbf{L}_2 = \mathbf{0} \end{bmatrix}$$

subject to

$$\mathbf{\Lambda}^2 = \begin{bmatrix} \lambda_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r^2 \end{bmatrix}, \mathbf{\Lambda} = \mathrm{Diag}\left(+\sqrt{\lambda_1^2}, \dots, +\sqrt{\lambda_r^2}\right).$$

Let us prove *first* $\mathbf{A}^{\#}\mathbf{AR} = \mathbf{R}\mathbf{\Lambda}^2$, *second* $\mathbf{A}^{\#}\mathbf{AL}_1 = \mathbf{L}_1\mathbf{\Lambda}^2$, $\mathbf{AA}^{\#}\mathbf{L}_2 = \mathbf{0}$.

(i) $\mathbf{A}^{\#}\mathbf{AR} = \mathbf{R}\mathbf{\Lambda}^2$

$$\mathbf{A}^{\#}\mathbf{AR} = \mathbf{G}_\mathbf{x}^{-1}\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{AR}$$

$$= \mathbf{G}_\mathbf{x}^{-1}\mathbf{G}_\mathbf{x}^{\frac{1}{2}}\mathbf{V}\,[\mathbf{\Lambda}, \mathbf{0}'] \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix} (\mathbf{G}_\mathbf{y}^{-\frac{1}{2}})'\mathbf{G}_\mathbf{y}\mathbf{G}_\mathbf{y}^{-\frac{1}{2}}\,[\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{V}'\mathbf{G}_\mathbf{x}^{\frac{1}{2}}\mathbf{G}_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}$$

$$\mathbf{A}^{\#}\mathbf{AR} = \mathbf{G}_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}\left[\mathbf{\Lambda}, \mathbf{0}'\right] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} = \mathbf{G}_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}\mathbf{\Lambda}^2$$

$$\mathbf{A}^{\#}\mathbf{AR} = \mathbf{R}\mathbf{\Lambda}^2.$$

(ii) $\mathbf{AA}^{\#}\mathbf{L}_1 = \mathbf{L}_1\mathbf{\Lambda}^2, \mathbf{AA}^{\#}\mathbf{L}_2 = \mathbf{0}$

$$\mathbf{AA}^{\#}\mathbf{L} = \mathbf{AG}_\mathbf{x}^{-1}\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{L}$$

$$= \mathbf{G}_\mathbf{y}^{-\frac{1}{2}}\,[\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{V}'\mathbf{G}_\mathbf{x}^{\frac{1}{2}}\mathbf{G}_\mathbf{x}^{-1}\mathbf{G}_\mathbf{x}^{\frac{1}{2}}\mathbf{V}\,[\mathbf{\Lambda}, \mathbf{0}'] \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix} (\mathbf{G}_\mathbf{y}^{-\frac{1}{2}})'\mathbf{G}_\mathbf{y}\mathbf{G}_\mathbf{y}^{-\frac{1}{2}}\,[\mathbf{U}_1, \mathbf{U}_2]$$

$$\mathbf{AA}^{\#}\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2] \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{bmatrix} \left[\mathbf{\Lambda}, \mathbf{0}'\right] \begin{bmatrix} \mathbf{U}'_1\mathbf{U}_1 & \mathbf{U}'_1\mathbf{U}_2 \\ \mathbf{U}'_2\mathbf{U}_1 & \mathbf{U}'_2\mathbf{U}_2 \end{bmatrix}$$

$$\mathbf{AA}^{\#}\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2] \begin{bmatrix} \mathbf{\Lambda}^2 & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I}_\mathbf{r} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathbf{n-r}} \end{bmatrix}$$

$$\mathbf{AA}^{\#}\,[\mathbf{L}_1, \mathbf{L}_2] = \left[\mathbf{L}_1\mathbf{\Lambda}^2, \mathbf{0}\right], \mathbf{AA}^{\#}\mathbf{L}_1 = \mathbf{L}_1\mathbf{\Lambda}^2, \mathbf{AA}^{\#}\mathbf{L}_2 = \mathbf{0}.$$

The *pair of eigensystems* $\{\mathbf{A}^{\#}\mathbf{AR} = \mathbf{R}\mathbf{\Lambda}^2, \mathbf{AA}^{\#}\,[\mathbf{L}_1, \mathbf{L}_2] = \left[\mathbf{L}_1\mathbf{\Lambda}^2, \mathbf{0}\right]\}$ is unfortunately based upon *non-symmetric matrices* $\mathbf{AA}^{\#} = \mathbf{AG}_\mathbf{x}^{-1}\mathbf{A}'\mathbf{G}_\mathbf{y}$ and $\mathbf{A}^{\#}\mathbf{A} = \mathbf{G}_\mathbf{x}^{-1}\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A}$ which make the left and right eigenspace analysis numerically more complex. It appears that we are forced to use the *Arnoldi method* rather than the more efficient *Lanczos method* used for symmetric matrices. In this situation we look out for an alternative. Actually as soon as we *substitute*

$$\{\mathbf{L}, \mathbf{R}\}\ by\ \{\mathbf{G}_\mathbf{y}^{-\frac{1}{2}}\mathbf{U}, \mathbf{G}_\mathbf{x}^{-\frac{1}{2}}\mathbf{V}\}$$

into the pair of eigensystems and *consequently* multiply $\mathbf{AA}^{\#}\mathbf{L}$ by $\mathbf{G}_\mathbf{x}^{-\frac{1}{2}}$, we achieve a pair of eigensystems identified in *Corollary 3.10* relying on symmetric matrices. In addition, such a pair of eigensystems produces the *canonical base*, namely orthonormal eigencolumns.

Corollary 3.10. *(symmetric pair of eigensystems):*

The pair of eigensystems

$$\mathbf{G}_\mathbf{y}^{\frac{1}{2}}\mathbf{A}\mathbf{G}_\mathbf{x}^{-1}\mathbf{A}'(\mathbf{G}_\mathbf{y}^{\frac{1}{2}})'\mathbf{U}_1 = \mathbf{\Lambda}^2\mathbf{U}_1 \quad versus \quad (\mathbf{G}_\mathbf{x}^{-\frac{1}{2}})'\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A}\mathbf{G}_\mathbf{x}^{-\frac{1}{2}}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^2 \tag{3.58}$$

$$|\mathbf{G}_\mathbf{y}^{\frac{1}{2}}\mathbf{A}\mathbf{G}_\mathbf{x}^{-1}\mathbf{A}'(\mathbf{G}_\mathbf{y}^{-\frac{1}{2}})' - \lambda_\mathbf{i}^2\mathbf{I}_\mathbf{n}| = 0 \quad versus \quad |(\mathbf{G}_\mathbf{x}^{-\frac{1}{2}})'\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A}\mathbf{G}_\mathbf{x}^{-\frac{1}{2}} - \lambda_\mathbf{j}^2\mathbf{I}_\mathbf{r}| = 0 \tag{3.59}$$

is based upon symmetric matrices. The left and right eigencolumns are orthogonal.

Such a procedure requires two factorizations,

$$\mathbf{G}_\mathbf{x} = (\mathbf{G}_\mathbf{x}^{\frac{1}{2}})'\mathbf{G}_\mathbf{x}^{\frac{1}{2}}, \mathbf{G}_\mathbf{x}^{-1} = \mathbf{G}_\mathbf{x}^{-\frac{1}{2}}(\mathbf{G}_\mathbf{x}^{-\frac{1}{2}})' \quad and \quad \mathbf{G}_\mathbf{y} = (\mathbf{G}_\mathbf{y}^{\frac{1}{2}})'\mathbf{G}_\mathbf{y}^{\frac{1}{2}}, \mathbf{G}_\mathbf{y}^{-1} = \mathbf{G}_\mathbf{y}^{-\frac{1}{2}}(\mathbf{G}_\mathbf{y}^{-\frac{1}{2}})'$$

via *Choleski factorization or eigenvalue decomposition* of the matrices $\mathbf{G}_\mathbf{x}$ and $\mathbf{G}_\mathbf{y}$.

Lemma 3.11. (canonical LESS):

Let $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$ be a canonical representation of the *overdetermined system of linear equations*

$$\{\mathbf{y} = \mathbf{Ax} + \mathbf{i} | \mathbf{A} \in R^{n \times m}, r := \text{rk}\mathbf{A} = m, n > m\}.$$

Then the rank partitioning of $\mathbf{y}^* = \left[(\mathbf{y}_1^*)', (\mathbf{y}_2^*)'\right]'$ leads to the *canonical unknown vector*

$$\mathbf{x}_\ell^* = \left[\mathbf{\Lambda}^{-1}, \mathbf{0}\right]\begin{bmatrix}\mathbf{y}_1^* \\ \mathbf{y}_2^*\end{bmatrix} = \mathbf{\Lambda}^{-1}\mathbf{y}_1^*, \mathbf{y}^* = \begin{bmatrix}\mathbf{y}_1^* \\ \mathbf{y}_2^*\end{bmatrix}, \begin{array}{l}\mathbf{y}_1^* \in \mathbb{R}^{r \times 1} \\ \mathbf{y}_2^* \in \mathbb{R}^{(n-r)\times 1}\end{array} \tag{3.60}$$

and to the canonical vector of inconsistency

$$\mathbf{i}_\ell^* = \begin{bmatrix}\mathbf{i}_1^* \\ \mathbf{i}_2^*\end{bmatrix}_\ell := \begin{bmatrix}\mathbf{y}_1^* \\ \mathbf{y}_2^*\end{bmatrix}_\ell - \begin{bmatrix}\mathbf{\Lambda} \\ \mathbf{0}\end{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{y}_1^* \quad or \quad \begin{array}{l}\mathbf{i}_1^* = \mathbf{0} \\ \mathbf{i}_2^* = \mathbf{y}_2^*\end{array} \tag{3.61}$$

of type $\mathbf{G_y}$-LESS. In terms if the original coordinates a canonical representation of $\mathbf{G_y}$-LESS is

$$\mathbf{x}_\ell = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V}\left[\mathbf{\Lambda}^{-1},\mathbf{0}\right]\begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}\mathbf{G_y^{\frac{1}{2}}y} \tag{3.62}$$

$$\mathbf{x}_\ell = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V\Lambda}^{-1}\mathbf{U'}_1\mathbf{G_y^{\frac{1}{2}}y} = \mathbf{R\Lambda}^{-1}\mathbf{L}_1^-\mathbf{y}. \tag{3.63}$$

$\mathbf{x}_\ell = \mathbf{A}_\ell^-\mathbf{y}$ is built on the canonical $(\mathbf{G_x},\mathbf{G_y})$ weighted right inverse. For the proof we depart from $\mathbf{G_y}$-LESS (3.11) and replace the matrix by its canonical representation, namely by eigenspace synthesis.

$$\left.\begin{array}{c}\mathbf{x}_\ell = \left(\mathbf{A'G_yA}\right)^{-1}\mathbf{A'G_yy}\\ \mathbf{A} = \mathbf{G_y}^{-\frac{1}{2}}\left[\mathbf{U}_1,\mathbf{U}_2\right]\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}\mathbf{V'G_x^{\frac{1}{2}}}\end{array}\right]$$

$$\Rightarrow \mathbf{A'G_yA} = (\mathbf{G_x^{\frac{1}{2}}})'\mathbf{V}\left[\mathbf{\Lambda},\mathbf{0}\right]\begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}(\mathbf{G_y}^{-\frac{1}{2}})'\mathbf{G_yG_y}^{-\frac{1}{2}}\left[\mathbf{U}_1,\mathbf{U}_2\right]\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}\mathbf{V'G_x^{\frac{1}{2}}}$$

$$\mathbf{A'G_yA} = (\mathbf{G_x^{\frac{1}{2}}})'\mathbf{V\Lambda^2V'G_x^{\frac{1}{2}}},\ \left(\mathbf{A'G_yA}\right)^{-1} = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V\Lambda}^{-2}\mathbf{V'}(\mathbf{G_x}^{-\frac{1}{2}})'$$

$$\Rightarrow \mathbf{x}_\ell = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V\Lambda^2V'}(\mathbf{G_x}^{-\frac{1}{2}})'(\mathbf{G_x^{\frac{1}{2}}})'\mathbf{V}\left[\mathbf{\Lambda},\mathbf{0}\right]\begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}(\mathbf{G_y}^{-\frac{1}{2}})'\mathbf{G_yy}$$

$$\mathbf{x}_\ell = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V}\left[\mathbf{\Lambda}^{-1},\mathbf{0}\right]\begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}\mathbf{G_y^{\frac{1}{2}}y}$$

$$\mathbf{x}_\ell = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V\Lambda}^{-1}\mathbf{U'}_1\mathbf{G_y^{\frac{1}{2}}y} = \mathbf{A}_\ell^-\mathbf{y}$$

$$\mathbf{A}_\ell^- = \mathbf{G_x}^{-\frac{1}{2}}\mathbf{V\Lambda}^{-1}\mathbf{U'}_1\mathbf{G_y^{\frac{1}{2}}} \in \mathbf{A}_{\mathbf{G_y}}^{\mathbf{1,2,3}}$$

($\mathbf{G_y}$ weighted reflexive inverse)

$$\left.\begin{array}{c}\mathbf{x}_\ell^* = \mathbf{V'G_x^{\frac{1}{2}}x}_\ell = \mathbf{\Lambda}^{-1}\mathbf{U'}_1\mathbf{G_y^{\frac{1}{2}}y} = \left[\mathbf{\Lambda}^{-1},\mathbf{0}\right]\begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}\mathbf{G_y^{\frac{1}{2}}y}\\ \mathbf{y}^* = \begin{bmatrix}\mathbf{y}_1^*\\ \mathbf{y}_2^*\end{bmatrix} = \begin{bmatrix}\mathbf{U'}_1\\ \mathbf{U'}_2\end{bmatrix}\mathbf{G_y^{\frac{1}{2}}y}\end{array}\right] \Rightarrow$$

$$\Rightarrow \mathbf{x}_\ell^* = \left[\mathbf{\Lambda}^{-1},\mathbf{0}\right]\begin{bmatrix}\mathbf{y}_1^*\\ \mathbf{y}_2^*\end{bmatrix} = \mathbf{\Lambda}^{-1}\mathbf{y}_1^*.$$

Thus we have proven the canonical inversion formula. The proof for the canonical representation of the vector of inconsistency is a consequence of the rank partitioning

$$\mathbf{i}_l^* = \begin{bmatrix} \mathbf{i}_1^* \\ \mathbf{i}_2^* \end{bmatrix}_\ell := \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0} \end{bmatrix} \mathbf{x}_\ell^*, \quad \begin{matrix} \mathbf{i}_1^*, \mathbf{y}_1^* \in \mathbb{R}^{r \times 1} \\ \mathbf{i}_2^*, \mathbf{y}_2^* \in \mathbb{R}^{(n-r) \times 1} \end{matrix},$$

$$\mathbf{i}_\ell^* = \begin{bmatrix} \mathbf{i}_1^* \\ \mathbf{i}_2^* \end{bmatrix}_\ell = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Lambda} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\Lambda}^{-1} \mathbf{y}_1^* = \begin{bmatrix} \mathbf{0} \\ \mathbf{y}_2^* \end{bmatrix}.$$

The *important result* of $\mathbf{x}_\ell^*$ based on the canonical $\mathbf{G_y}$-LESS of $\{\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^* | \mathbf{A}^* \in \mathbb{R}^{n \times m}, \text{rk}\mathbf{A}^* = \text{rk}_\mathbf{A} = m, n > m\}$ needs a comment. The rank partitioning of the canonical observation vector $\mathbf{y}^*$, namely $\mathbf{y}_1^* \in \mathbb{R}^r$, $\mathbf{y}_2^* \in \mathbb{R}^{(n-r)}$ again paved the way for an interpretation. *First*, we appreciate the simple "*direct inversion*" $\mathbf{x}_\ell^* = \boldsymbol{\Lambda}^{-1}\mathbf{y}_1^*$, $\boldsymbol{\Lambda} = \text{Diag}\left(+\sqrt{\lambda_1^2}, \ldots, +\sqrt{\lambda_r^2}\right)$, for instance

$$\begin{bmatrix} \mathbf{x}_1^* \\ \ldots \\ \mathbf{x}_m^* \end{bmatrix}_\ell = \begin{bmatrix} \lambda_1^{-1}\mathbf{y}_1^* \\ \ldots \\ \lambda_r^{-1}\mathbf{y}_r^* \end{bmatrix}.$$

Second, $\mathbf{i}_1^* = \mathbf{0}$, eliminates all elements of the vector of canonical inconsistencies, for instance $\left[i_1^*, \ldots, i_r^*\right]_\ell' = \mathbf{0}$, while $\mathbf{i}_2^* = \mathbf{y}_2^*$ identifies the deficient elements of the vector of canonical inconsistencies with the vector of canonical observations for instance $\left[i_{r+1}^*, \ldots, i_n^*\right]_\ell' = \left[y_{r+1}^*, \ldots, y_n^*\right]_\ell'$. *Finally*, enjoy the commutative diagram of *Fig. 3.6* illustrating our previously introduced transformations of type LESS and canonical LESS, by means of $\mathbf{A}_\ell^-$ and $(\mathbf{A}^*)_\ell^-$, respectively (Fig. 3.7). A first example is *canonical LESS* of the *Front Page Example* by $\mathbf{G_y} = \mathbf{I_3}$, $\mathbf{G_x} = \mathbf{I_2}$.

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} : \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}, r := \text{rk}\mathbf{A} = 2$$

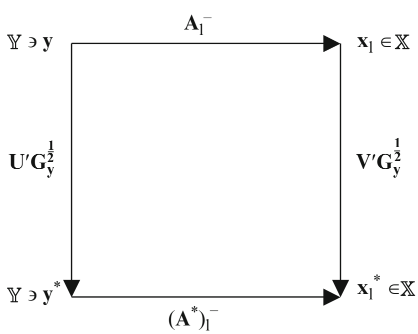


Fig. 3.7 Commutative diagram of inverse coordinate transformations

left eigenspace | *right eigenspace*

$$\mathbf{AA}^{\#}\mathbf{U}_1 = \mathbf{AA}'\mathbf{U}_1 = \mathbf{U}_1\boldsymbol{\Lambda}^{\mathbf{2}} \qquad \mathbf{A}^{\#}\mathbf{AV} = \mathbf{A}'\mathbf{AV} = \mathbf{V}\boldsymbol{\Lambda}^2$$

$$\mathbf{AA}^{\#}\mathbf{U}_2 = \mathbf{AA}'\mathbf{U}_2 = \mathbf{0}$$

$$\mathbf{AA}' = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 5 & 7 \\ 4 & 7 & 10 \end{bmatrix} \qquad \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} = \mathbf{A}'\mathbf{A}$$

eigenvalues

$$|\mathbf{AA}' - \lambda_i^2 \mathbf{I}_3| = 0 \Leftrightarrow \qquad |\mathbf{A}'\mathbf{A} - \lambda_j^2 \mathbf{I}_2| = 0 \Leftrightarrow$$

$$\forall i \in \{1,2,3\} \qquad \forall j \in \{1,2\}$$

$$\Leftrightarrow \lambda_1^2 = \frac{17}{2} + \frac{1}{2}\sqrt{265}, \lambda_2^2 = \frac{17}{2} - \frac{1}{2}\sqrt{265}, \lambda_3^2 = 0$$

left eigencolumns | *right eigencolumns*

$$(1\text{st}) \begin{bmatrix} 2-\lambda_1^2 & 3 & 4 \\ 3 & 5-\lambda_1^2 & 7 \\ 4 & 7 & 10-\lambda_1^2 \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{21} \\ u_{31} \end{bmatrix} = 0 \quad (1\text{st}) \begin{bmatrix} 3-\lambda_1^2 & 6 \\ 6 & 14-\lambda_1^2 \end{bmatrix} \begin{bmatrix} \mathrm{v}_{11} \\ \mathrm{v}_{21} \end{bmatrix} = 0$$

subject to | *subject to*

$$u_{11}^2 + u_{21}^2 + u_{31}^2 = 1 \qquad \mathrm{v}_{11}^2 + \mathrm{v}_{21}^2 = 1$$

$$\begin{bmatrix} (2-\lambda_1^2)u_{11} + 3u_{21} + 4u_{31} = 0 \\ 3u_{11} + (5-\lambda_1^2)u_{21} + 7u_{31} = 0 \end{bmatrix} \quad \textit{versus} \quad (3-\lambda_1^2)\mathrm{v}_{11} + 6\mathrm{v}_{21} = 0$$

$$\begin{bmatrix} \mathrm{v}_{11}^2 = \dfrac{36}{36 + (3-\lambda_1^2)^2} = \dfrac{72}{265 + 11\sqrt{265}} \\ \mathrm{v}_{21}^2 = \dfrac{(3-\lambda_1^2)^2}{36 + (3-\lambda_1^2)^2} = \dfrac{193 + 11\sqrt{265}}{265 + 11\sqrt{265}} \end{bmatrix}$$

$$\begin{bmatrix} u_{11}^2 \\ u_{21}^2 \\ u_{31}^2 \end{bmatrix} = \frac{1}{(1+4\lambda_1^2)^2 + (2-7\lambda_1^2)^2 + (1-7\lambda_1^2+\lambda_1^4)^2} \begin{bmatrix} (1+4\lambda_1^2)^2 \\ (2-7\lambda_1^2)^2 \\ (1-7\lambda_1^2+\lambda_1^4)^2 \end{bmatrix}$$

$$\begin{bmatrix} u_{11}^2 \\ u_{21}^2 \\ u_{31}^2 \end{bmatrix} = \frac{2}{43725 + 2685\sqrt{265}} \begin{bmatrix} \left(35 + 2\sqrt{265}\right)^2 \\ \left(\dfrac{115}{2} + \dfrac{7}{2}\sqrt{265}\right)^2 \\ \left(80 + 5\sqrt{265}\right)^2 \end{bmatrix}$$

$$
(2\text{nd})\begin{bmatrix} 3-\lambda_2^2 & 7 \\ 7 & 21-\lambda_2^2 \end{bmatrix}\begin{bmatrix} u_{12} \\ u_{22} \end{bmatrix} = 0 \quad (2\text{nd})\begin{bmatrix} 2-\lambda_2^2 & 3 & 5 \\ 3 & 5-\lambda_2^2 & 9 \\ 5 & 9 & 17-\lambda_2^2 \end{bmatrix}\begin{bmatrix} \mathrm{v}_{12} \\ \mathrm{v}_{22} \\ \mathrm{v}_{32} \end{bmatrix} = 0
$$

subject to *subject to*

$$
u_{12}^2 + u_{22}^2 + u_{32}^2 = 1 \qquad \mathrm{v}_{12}^2 + \mathrm{v}_{22}^2 = 1
$$

$$
\begin{bmatrix} (2-\lambda_2^2)u_{12} + 3u_{22} + 4u_{32} = 0 \\ 3u_{12} + (5-\lambda_2^2)u_{22} + 7u_{32} = 0 \end{bmatrix} \quad \textit{versus} \quad (3-\lambda_2^2)v_{12} + 6v_{22} = 0
$$

$$
\begin{bmatrix} \mathrm{v}_{12}^2 = \dfrac{36}{36+(3-\lambda_2^2)^2} = \dfrac{72}{265-11\sqrt{265}} \\ v_{22}^2 = \dfrac{(3-\lambda_2^2)^2}{36+(3-\lambda_2^2)^2} = \dfrac{193-11\sqrt{265}}{265-11\sqrt{265}} \end{bmatrix}
$$

$$
\begin{bmatrix} u_{12}^2 \\ u_{22}^2 \\ u_{32}^2 \end{bmatrix} = \frac{1}{(1+4\lambda_2^2)^2 + (2-7\lambda_2^2)^2 + (1-7\lambda_2^2+\lambda_2^4)^2}\begin{bmatrix} (1+4\lambda_2^2)^2 \\ (2-7\lambda_2^2)^2 \\ (1-7\lambda_2^2+\lambda_2^4)^2 \end{bmatrix}
$$

$$
\begin{bmatrix} u_{12}^2 \\ u_{22}^2 \\ u_{32}^2 \end{bmatrix} = \frac{2}{43725-2685\sqrt{265}}\begin{bmatrix} (35-2\sqrt{265})^2 \\ \left(\dfrac{115}{2} - \dfrac{7}{2}\sqrt{265}\right)^2 \\ (80-5\sqrt{265})^2 \end{bmatrix}
$$

$$
(3\text{rd})\begin{bmatrix} 2 & 3 & 4 \\ 3 & 5 & 7 \\ 4 & 7 & 10 \end{bmatrix}\begin{bmatrix} u_{13} \\ u_{23} \\ u_{33} \end{bmatrix} = 0 \quad \textit{subject to} \;\; u_{13}^2 + u_{23}^2 + u_{33}^2 = 1
$$

$$
\begin{aligned} 2u_{13} + 3u_{23} + 4u_{33} &= 0 \\ 3u_{13} + 5u_{23} + 7u_{33} &= 0 \end{aligned}
$$

$$
\begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} u_{13} \\ u_{23} \end{bmatrix} = \begin{bmatrix} -4 \\ -7 \end{bmatrix} u_{33} \Leftrightarrow \begin{bmatrix} u_{13} \\ u_{23} \end{bmatrix} = -\begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 4 \\ 7 \end{bmatrix} u_{33}
$$

$$
u_{13} = +u_{33}, u_{23} = -2u_{33}
$$

$$
u_{13}^2 = \frac{1}{6}, u_{23}^2 = \frac{2}{3}, u_{33}^2 = \frac{1}{6}.
$$

There are four combinatorial solutions to generate square roots.

$$
\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{u_{11}^2} & \pm\sqrt{u_{12}^2} & \pm\sqrt{u_{13}^2} \\ \pm\sqrt{u_{21}^2} & \pm\sqrt{u_{22}^2} & \pm\sqrt{u_{23}^2} \\ \pm\sqrt{u_{31}^2} & \pm\sqrt{u_{32}^2} & \pm\sqrt{u_{33}^2} \end{bmatrix}
$$

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} \pm\sqrt{v_{11}^2} & \pm\sqrt{v_{12}^2} \\ \pm\sqrt{v_{21}^2} & \pm\sqrt{v_{22}^2} \end{bmatrix}.$$

Here we have chosen the one with the positive sign exclusively. In summary, the eigenspace analysis gave the result as follows.

$$\mathbf{\Lambda} = \mathrm{Diag}\left(\sqrt{\frac{17+\sqrt{265}}{2}}, \sqrt{\frac{17-\sqrt{265}}{2}}\right)$$

$$\mathbf{U} = \left[\begin{array}{cc|c} \sqrt{2}\dfrac{35+2\sqrt{265}}{\sqrt{43725+2685\sqrt{265}}} & \sqrt{2}\dfrac{35-2\sqrt{265}}{\sqrt{43725-2685\sqrt{265}}} & \dfrac{1}{6}\sqrt{6} \\ \dfrac{\sqrt{2}}{2}\dfrac{115+7\sqrt{265}}{\sqrt{43725+2685\sqrt{265}}} & \dfrac{\sqrt{2}}{2}\dfrac{115-7\sqrt{265}}{\sqrt{43725-2685\sqrt{265}}} & \dfrac{1}{3}\sqrt{6} \\ \sqrt{2}\dfrac{80+5\sqrt{265}}{\sqrt{43725+2685\sqrt{265}}} & \sqrt{2}\dfrac{80-5\sqrt{265}}{\sqrt{43725-2685\sqrt{265}}} & \dfrac{1}{6}\sqrt{6} \end{array}\right] = [\mathbf{U}_1, \mathbf{U}_2]$$

$$\mathbf{V} = \begin{bmatrix} \dfrac{72}{\sqrt{265+11\sqrt{265}}} & \dfrac{72}{\sqrt{265-11\sqrt{265}}} \\ \dfrac{\sqrt{193+11\sqrt{265}}}{\sqrt{265+11\sqrt{265}}} & \dfrac{\sqrt{193-11\sqrt{265}}}{\sqrt{265-11\sqrt{265}}} \end{bmatrix}.$$

3-3 Case Study

Partial redundancies, latent conditions, high leverage points versus break points, direct and inverse Grassmann coordinates, Plücker coordinates

This case study has various targets. *First* we aim at a canonical analysis of the hat matrices $\mathbf{H_x}$ and $\mathbf{H_y}$ for a simple linear model with a leverage point. The impact of a high leverage point is studied in all detail. Partial redundancies are introduced and interpreted in their peculiar role of weighting observations. *Second*, preparatory in nature, we briefly introduce multilinear algebra, the operations "*join and meet*", namely the *Hodge star operator*. *Third*, we go "from A to B": Given the *columns space* $\mathcal{R}(\mathbf{A}) = \mathbf{G}^{m,n}(\mathbf{A})$, identified as the *Grassmann space* $\mathbf{G}^{m,n} \subset \mathbb{R}^n$ of the matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, $n > m$, $\mathrm{rk}\mathbf{A} = m$, we construct the column space $\mathcal{R}(\mathbf{B}) = \mathcal{R}^{\perp}(\mathbf{A}) = \mathbf{G}^{n-m,n} \subset \mathbb{R}^n$ of the matrix $\mathbf{B}$ which agrees to the *orthogonal column space* $\mathcal{R}^{\perp}(\mathbf{A})$ of the matrix $\mathbf{A}$. $\mathcal{R}^{\perp}(\mathbf{A})$ is identified as *Grassmann space* $\mathbf{G}^{n-m,n} \subset \mathbb{R}^n$ and is covered by *Grassmann coordinates*, also called *Plücker coordinates* p_{ij}. The matrix $\mathbf{B}$, alternatively the *Grassmann coordinates* (*Plücker coordinates*), constitute the *latent restrictions*, also called *latent condition equations*, which control parameter adjustment and lead to a proper choice of observational weights. *Fourth*, we reverse our path: we go "from B to A": Given the column space $\mathcal{R}(\mathbf{B})$ of the matrix of restrictions $\mathbf{B} \in \mathbb{R}^{\ell \times n}$, $\ell < n$, $\mathrm{rk}\mathbf{B} = \ell$ we construct the column space $\mathcal{R}^{\perp}(\mathbf{B}) = \mathcal{R}(\mathbf{A}) \subset \mathbb{R}^n$, the *orthogonal column space* of the matrix $\mathbf{B}$ which is apex to the column space $\mathcal{R}(\mathbf{A})$ of the matrix $\mathbf{A}$. The matrix $\mathbf{A}$, alternatively the *Grassmann coordinates* (*Plücker coordinates*) of the matrix $\mathbf{B}$ constitute the *latent parametric equations* which are "*behind a conditional adjustment*". *Fifth*, we break-up the linear model into pieces, and introduce the notion of *break points* and their determination.

The present analysis of partial redundancies and latent restrictions has been pioneered by Kampmann (1992), as well as *R. Jurisch and G. Kampmann* (2002 a, b). Additional useful references are *D. W. Behmken and N. R. Draper* (1972), *S. Chatterjee and A. S. Hadi* (1988), *R. D. Cook and S. Weisberg* (1982). Multilinear algebra, the operations "*join and meet*" and the *Hodge star operator* are reviewed in *W. Hodge and D. Pedoe* (1968), *C. Macinnes* (1999), *S. Morgera* (1992), *W. Neutsch* (1995), *B. F. Doolin and C. F. Martin* (1990). A sample reference for *break point* synthesis is *C. H. Mueller* (1998), *N. M. Neykov and C. H. Mueller* (2003) and *D. Tasche* (2003).

3-31 Canonical Analysis of the Hat Matrix, Partial Redundancies, High Leverage Points

A beautiful example for the power of *eigenspace synthesis* is the least squares fit of a straight line to a set of observation: Let us assume that we have observed a dynamical system $y(t)$ which is represented by a *polynomial of degree one* with respect to time t.

$$y(t_i) = 1_i x_1 + t_i x_2 \ \forall\, i \in \{1, \cdots, n\}.$$

Due to $y^{\bullet}(t) = x_2$ it is a dynamical system with constant velocity or constant first derivative with result to time t_0. The *unknown* polynomial coefficients are collected in the column array $\mathbf{x} = [x_1, x_2]'$, $\mathbf{x} \in \mathbb{X} = \mathbb{R}^2$, $\dim \mathbb{X} = 2$ and constitute the coordinates of the two-dimensional parameter space $\mathbb{X}$. For this example we choose $n = 4$ observations, namely $\mathbf{y} = [y(t_1), y(t_2), y(t_3), y(t_4)]'$, $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^4$, $dim\, \mathbb{Y} = 4$. The samples of the polynomial are taken at $t_1 = 1, t_2 = 2, t_3 = 3\, and\, t_4 = a$. With such a choice of t_4 we aim at modeling the behavior of *high leverage points*, e.g. $a >> (t_1, t_2, t_3)$ or $a \rightarrow \infty$, illustrated by Fig. 3.8. *Box 3.16* summarizes the right eigenspace analysis of the hat matrix $\mathbf{H_y} := \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$. First, we have computed the spectrum of $\mathbf{A}'\mathbf{A}$ and $(\mathbf{A}'\mathbf{A})^{-1}$ for the given matrix $\mathbf{A} \in \mathbb{R}^{4\times 2}$, namely the eigenvalues squared $\lambda^2_{1,2} = 59 \pm \sqrt{3261}$. Note the leverage point $t_4 = a = 10$. Second, we computed the right eigencolumns $\mathbf{v}_1$ and $\mathbf{v}_2$ which constitute the orthonormal matrix $\mathbf{V} \in \mathbf{SO}(2)$. The angular representation of the orthonormal matrix $\mathbf{V} \in \mathbf{SO}(2)$ follows: Third, we take advantage of the sine-cosine representation (3.64) $\mathbf{V} \in \mathbf{SO}(2)$, the special orthonormal group over $\mathbb{R}^2$.

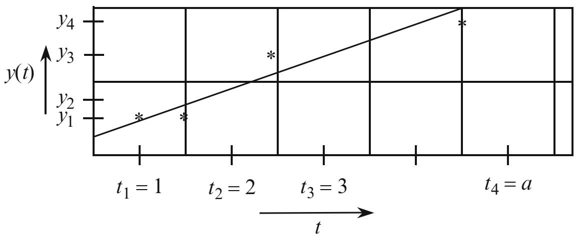


Fig. 3.8 Graph of the function $y(t)$, high leverage point $t_4 = a$

Indeed, we find the angular parameter $\gamma = 81°53'25.4''$. Fourth, we are going to represent the hat matrix $\mathbf{H_y}$ in terms of the angular parameter namely (3.65)–(3.68). In this way, the general representation (3.69) is obtained, illustrated by four cases. Equation (3.65) is a special case of the general angular representation (3.69) of the hat matrix $\mathbf{H_y}$. Five, we sum up the canonical representation $\mathbf{AV'\Lambda^{-2}V'A'}$ (3.70), of the hat matrix $\mathbf{H_y}$, also called right eigenspace synthesis. Note the rank of the hat matrix, namely $\mathrm{rk}\mathbf{H_y} = \mathrm{rk}\mathbf{A} = m = 2$, as well as the peculiar fourth adjusted observation

$$\hat{y}_4 = y_4(\mathbf{I} - \mathrm{LESS}) = \frac{1}{100}(-11y_1 + y_2 + 13y_3 + 97y_4),$$

which highlights the weight of the leverage point t_4: This analysis will be more pronounced if we go through the same type of *right eigenspace synthesis* for the leverage point $t_4 = a$, $a\infty$, outlined in *Box 3.19*.

Box 3.15. Right eigenspace analysis of a linear model of an univariate polynomial of degree one – high leverage point $a = 10$ –

$$\text{“}\textit{Hat matrix}\ \mathbf{H_y} = \mathbf{A(A'A)^{-1}A} = \mathbf{AV\Lambda^{-2}V'A'}\text{”}$$

$$\textit{right eigenspace analysis}: \left[\begin{array}{l} \mathbf{A'AV} = \mathbf{V\Lambda}^2 \\ \textit{subject to} \\ \mathbf{VV'} = \mathbf{I}_2 \end{array}\right.$$

$$\mathbf{A} := \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 10 \end{bmatrix},\ \mathbf{A'A} = \begin{bmatrix} 4 & 16 \\ 16 & 114 \end{bmatrix},\ (\mathbf{AA})^{-1} = \frac{1}{100}\begin{bmatrix} 57 & -8 \\ -8 & 2 \end{bmatrix}$$

$$spec(\mathbf{A'A}) = \{\lambda_1^2, \lambda_2^2\} : \left|\mathbf{A'A} - \lambda_j^2\mathbf{I}_2\right| = 0,\ \forall\, j \in \{1,2\} \Leftrightarrow$$

$$\begin{vmatrix} 4-\lambda^2 & 16 \\ 16 & 114-\lambda^2 \end{vmatrix} = 0 \Leftrightarrow \lambda^4 - 118\lambda^2 + 200 = 0$$

$$\lambda_{1,2}^2 = 59 \pm \sqrt{3281} = 59 \pm 57.26 = 0$$

$$spec(\mathbf{A}'\mathbf{A}) = \{\lambda_1^2, \lambda_2^2\} = \{116.28,\ 1.72\}$$

versus

$$spec(\mathbf{A}'\mathbf{A})^{-1} = \{\frac{1}{\lambda_1^2}, \frac{1}{\lambda_2^2}\} = \{8.60 * 10^{-3},\ 0.58\}$$

$$\textit{right eigencolumn analysis}: \left[\begin{array}{l} (\mathbf{A}'\mathbf{A} - \lambda_j^2 \mathbf{I}_2)\mathbf{V} = 0 \\ \textit{subject to} \\ \mathbf{V}\mathbf{V}' = \mathbf{I}_2 \end{array}\right.$$

$$(\texttt{1st})\ (\mathbf{A}'\mathbf{A} - \lambda_1^2 \mathbf{I}) \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = 0 \quad \textit{subject to} \quad v_{11}^2 + v_{21}^2 = 1$$

$$\left.\begin{array}{c} (4 - \lambda_1^2)v_{11} + 16 v_{21} = 0 \\ v_{11}^2 + v_{21}^2 = 1 \end{array}\right] \Rightarrow$$

$$v_{11} = +\sqrt{v_{11}^2} = \frac{16}{\sqrt{256 + (4 - \lambda_1^2)^2}} = 0.141$$

$$v_{21} = +\sqrt{v_{21}^2} = \frac{4 - \lambda_1^2}{\sqrt{256 + (4 - \lambda_1^2)^2}} = -0.990$$

$$(\texttt{2nd})\ (\mathbf{A}'\mathbf{A} - \lambda_2^2 \mathbf{I}_2) \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} = 0 \ \textit{subject to}\ v_{12}^2 + v_{22}^2 = 1$$

$$\left.\begin{array}{c} (4 - \lambda_2^2)v_{12} + 16 v_{22} = 0 \\ v_{12}^2 + v_{22}^2 = 1 \end{array}\right] \Rightarrow$$

$$v_{12} = +\sqrt{v_{12}^2} = \frac{16}{\sqrt{256 + (4 - \lambda_2^2)^2}} = 0.990$$

$$v_{22} = +\sqrt{v_{22}^2} = \frac{4 - \lambda_2^2}{\sqrt{256 + (4 - \lambda_2^2)^2}} = 0.141$$

$$\textit{right eigenspace}: \left[\begin{array}{l} spec\,(\mathbf{A}'\mathbf{A}) = \{116.28,\ 1.72\} \\ spec\,(\mathbf{A}'\mathbf{A})^{-1} = \{8.60 * 10^{-3},\ 0.58\} \\ \mathbf{V} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} 0.141 & 0.990 \\ -0.990 & 0.141 \end{bmatrix} \in \mathbf{SO}(2) \end{array}\right.$$

"Angular representation of $\mathbf{V} \in \mathbf{SO}(2)$"

$$\mathbf{V} = \begin{bmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{bmatrix} = \begin{bmatrix} 0.141 & 0.990 \\ -0.990 & 0.141 \end{bmatrix} \tag{3.64}$$

$$\sin\gamma = 0.990,\ \cos\gamma = 0.141,\ \tan\gamma = 7.021$$

$$\gamma = 81^\circ.890,386 = 81^\circ 53' 25.4"$$

hat matrix $\mathbf{H_y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}'\mathbf{A}'$

$$(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{V}\Lambda^{-2}\mathbf{V} = \begin{bmatrix} \frac{1}{\lambda_1^2}\cos^2\gamma + \frac{1}{\lambda_2^2}\sin^2\gamma & \left(-\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}\right)\sin\gamma\cos\gamma \\ \left(-\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}\right)\sin\gamma\cos\gamma & \frac{1}{\lambda_1^2}\sin^2\gamma + \frac{1}{\lambda_2^2}\cos^2\gamma \end{bmatrix} \tag{3.65}$$

$$(\mathbf{A}'\mathbf{A})^{-1}_{j_1 j_2} = \sum_{j_3=1}^{m=2} \frac{1}{\lambda_{j_3}^2}\cos\gamma_{j_1 j_3}\cos\gamma_{j_2 j_3} \tag{3.66}$$

subject to

$$\mathbf{V}\mathbf{V}' = \mathbf{I}_2 \quad \sum_{j_3=1}^{m=2}\cos\gamma_{j_1 j_3}\cos\gamma_{j_2 j_3} = \delta_{j_1 j_2} \tag{3.67}$$

case 1: $j_1 = 1, j_2 = 1$:

$$\cos^2\gamma_{11} + \cos^2\gamma_{12} = 1$$
$$(\cos^2\gamma + \sin^2\gamma = 1)$$

case 2: $j_1 = 1, j_2 = 1$:

$$\cos\gamma_{11}\cos\gamma_{21} + \cos\gamma_{12}\cos\gamma_{22} = 0$$
$$(-\cos\gamma\sin\gamma + \sin\gamma\cos\gamma = 0)$$

case 3: $j_1 = 1, j_2 = 1$:

$$\cos\gamma_{21}\cos\gamma_{11} + \cos\gamma_{22}\cos\gamma_{12} = 0$$
$$(-\sin\gamma\cos\gamma + \cos\gamma\sin\gamma = 0)$$

case 4: $j_1 = 1, j_2 = 1$:

$$\cos^2\gamma_{21} + \cos^2\gamma_{22} = 1$$
$$(\sin^2\gamma + \cos^2\gamma = 1)$$

$$(\mathbf{A}'\mathbf{A})^{-1} = \begin{bmatrix} \lambda_1^{-2}\cos^2\gamma_{11} + \lambda_2^{-2}\cos^2\gamma_{12} & \lambda_1^{-2}\cos\gamma_{11}\cos\gamma_{21} + \lambda_2^{-2}\cos\gamma_{12}\cos\gamma_{22} \\ \lambda_1^{-2}\cos\gamma_{21}\cos\gamma_{11} + \lambda_2^{-2}\cos\gamma_{22}\cos\gamma_{12} & \lambda_1^{-2}\cos^2\gamma_{21} + \lambda_2^{-2}\cos^2\gamma_{22} \end{bmatrix} \tag{3.68}$$

$$\mathbf{H_y} = \mathbf{A}\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}'\mathbf{A}' \quad h_{i_1 i_2} = \sum_{j_1, j_2, j_3 = 1}^{m=2} a_{i_1 j_1} a_{i_2 j_2} \frac{1}{\lambda_{j_3}^2}\cos\gamma_{j_1 j_3}\cos\gamma_{j_2 j_3} \tag{3.69}$$

$$\mathbf{H_y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}'\mathbf{A}' \tag{3.70}$$

$$\mathbf{A} := \mathbf{A}\mathbf{V} = \begin{bmatrix} -0.849 & 1.131 \\ -1.839 & 1.272 \\ -2.829 & 1.413 \\ -9.759 & 2.400 \end{bmatrix},\ \boldsymbol{\Lambda}^{-2} = \mathrm{Diag}(8.60 \times 10^{-3},\ 0.58)$$

$$\mathbf{H_y} = \mathbf{A}\ \mathbf{\Lambda}^{-2}(\mathbf{A}\ )' = \frac{1}{100}\begin{bmatrix} 43 & 37 & 31 & -11 \\ 37 & 33 & 29 & 1 \\ 31 & 29 & 27 & 13 \\ -11 & 1 & 13 & 97 \end{bmatrix}$$

$$\mathrm{rk}\mathbf{H_y} = \mathrm{rk}\mathbf{A} = m = 2$$

$$\hat{y}_4 = y_4(\mathbf{I}\text{-LESS}) = \frac{1}{100}(-11y_1 + y_2 + 13y_3 + 97y_4).$$

By means of *Box 3.17* we repeat the *right eigenspace analysis* for one leverage point $t_4 = a$, later on $a\infty$, for both the *hat matrix* $\mathbf{H_x} := (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$and $\mathbf{H_y} := \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. *First*, $\mathbf{H_x}$ is the linear operator producing $\hat{\mathbf{x}} = \mathbf{x}_\ell(\mathbf{I}\text{-}LESS)$. *Second*, $\mathbf{H_y}$ as linear operator generates $\hat{\mathbf{y}} = \mathbf{y}_\ell(\mathbf{I}\text{-}LESS)$. *Third*, the complementary operator $\mathbf{I}_4 - \mathbf{H_y} =: \mathbf{R}$ as the matrix of *partial redundancies* leads us to the inconsistency vector $\hat{\mathbf{i}} = \mathbf{i}_\ell(\mathbf{I}\text{-}LESS)$. The structure of the *redundancy matrix* $\mathbf{R}$, $rk\mathbf{R} = n - m$, is most remarkable. Its *diagonal elements* will be interpreted soonest. *Fourth*, we have computed the length of the inconsistency vector $||\hat{\mathbf{i}}||^2$, the quadratic form $\mathbf{y}'\mathbf{R}\mathbf{y}$.

The highlight of the analysis of hat matrices is set by computing

$$1st:\ \mathbf{H_x}(a \to \infty)\ versus\ 2nd:\ \mathbf{H_y}(a \to \infty)$$

$$3rd:\ \mathbf{R}(a \to \infty)\ versus\ 4th:\ ||\hat{\mathbf{i}}(a \to \infty)||^2$$

for "*highest leverage point*" $a \to \infty$, in detail reviewed *Box 3.18*. Please, notice the two unknowns $\hat{x}_1$ and $\hat{x}_2$ as best approximations of type **I**-LESS. $\hat{x}_1$ resulted in the *arithmetic mean* of the first three measurements. The point $y_4,\ t_4 = a \to \infty$, had no influence at all. Here, $\hat{x}_2 = 0$ was found. The hat matrix $\mathbf{H_y}(a \to \infty)$ has produced *partial hats* $h_{11} = h_{22} = h_{33} = 1/3$, but $h_{44} = 1$ if $a \to \infty$. The best approximation of the **I**-LESS observations were $\hat{y}_1 = \hat{y}_2 = \hat{y}_3$ as the arithmetic mean of the first three observations but $\hat{y}_4 = y_4$ has been a reproduction of the fourth observation. *Similarly* the *redundancy matrix* $\mathbf{R}(a \to \infty)$ produced the weighted means $\hat{i}_1$, $\hat{i}_2$ and $\hat{i}_3$. The *partial redundancies* $r_{11} = r_{22} = r_{33} = 2/3,\ r_{44} = 0$, sum up to $r_{11} + r_{22} + r_{33} + r_{44} = n - m = 2$. Notice the value $\hat{i}_4 = 4$: The observation indexed four is left uncontrolled.

Box 3.16. The linear model of a univariate polynomial of degree one – one high leverage point –

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} \quad \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & a \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix}$$

$$\mathbf{x} \in \mathbb{R}^2,\ \mathbf{y} \in \mathbb{R}^4,\ \mathbf{A} \in \mathbb{R}^{4\times 2},\ rk\mathbf{A} = m = 2$$

$$dim\,\mathbb{X} = m = 2 \;\; versus \;\; dim\,\mathbb{Y} = n = 4$$

$$(1st)\hat{\mathbf{x}} = \mathbf{x}_\ell(\mathbf{I}\text{-}LESS) = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{H}_\mathbf{x}\mathbf{y} \quad (3.71)$$

$$\mathbf{H}_\mathbf{x} = \frac{1}{18 - 12a + 3a^2}\begin{bmatrix} 8 - a + a^2 & 2 - 2a + a^2 & -4 - 3a + a^2 & 14 - 6a \\ -2 - a & 2 - a & 6 - a & -6 + 3a \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$$(2nd)\hat{\mathbf{y}} = \mathbf{y}_\ell(\mathbf{I}\text{-}LESS) = \mathbf{A}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{H}_\mathbf{y}\mathbf{y} \quad (3.72)$$

$$\text{“}hat\ matrix\text{”: } \mathbf{H}_\mathbf{y} = \mathbf{A}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}', \;\; \mathrm{rk}\mathbf{H}_\mathbf{y} = m = 2$$

$$\mathbf{H}_\mathbf{y} = \frac{1}{18{-}18a{+}3a^2} \times \begin{bmatrix} 6 - 2a + a^2 & 4 - 3a + a^2 & 2 - 4a + a^2 & 8 - 3a \\ 4 - 3a + a^2 & 6 - 4a + a^2 & 8 - 5a + a^2 & 2 \\ 2 - 4a + a^2 & 6 - 5a + a^2 & 14 - 6a + a^2 & -4 + 3a \\ 8 - 3a & 2 & -4 + 3a & 14 - 12a + 3a^2 \end{bmatrix}$$

$$(3\text{rd})\,\hat{\mathbf{i}} = \mathbf{i}_\ell(\mathbf{I}\text{-}LESS) = (\mathbf{I}_4 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} = \mathbf{R}\mathbf{y}$$

$$\text{“}redundancy\ matrix\text{”:} \mathbf{R} = \mathbf{I}_4 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}', \;\; \mathrm{rk}\mathbf{R} = n - m = 2$$

$$\text{“}redundancy\text{”: n - rk A = n - m = 2}$$

$$\mathbf{R} = \frac{1}{18 - 12a + 3a^2} \times \begin{bmatrix} 12 - 10a + 2a^2 & -4 + 3a - a^2 & -2 + 4a - a^2 & -8 + 3a \\ -4 + 3a - a^2 & 12 - 6a + 2a^2 & -8 + 5a - a^2 & -2 \\ -2 + 4a - a^2 & -8 + 5a - a^2 & 4 - 6a + 2a^2 & 4 - 3a \\ -8 + 3a & -2 & 4 - 3a & 4 \end{bmatrix}$$

$$(4th)||\hat{\mathbf{i}}||^2 = ||\mathbf{i}_\ell(\mathbf{I}\text{-LESS})||^2 = \mathbf{y}'\mathbf{R}\mathbf{y}.$$

At this end we shall compute the LESS fit

$$\lim_{a\to\infty} ||\hat{i}(a)||^2,$$

which turns out to be *independent of the fourth observation*.

Box 3.17. The linear model of a univariate polynomial of degree one – extreme leverage point $a \to \infty$ –

$$(1\text{st})\; \mathbf{H}_\mathbf{x}(a \to \infty)$$

$$\mathbf{H_x} = \frac{1}{\frac{18}{a^2} - \frac{12}{a} + 3} \begin{bmatrix} \frac{8}{a^2} - \frac{1}{a} + 1 & \frac{2}{a^2} - \frac{2}{a} + 1 & \frac{4}{a^2} - \frac{3}{a} + 1 & \frac{14}{a^2} - \frac{6}{a} \\ -\frac{2}{a^2} - \frac{1}{a} & +\frac{2}{a^2} - \frac{1}{a} & +\frac{6}{a^2} - \frac{1}{a} & -\frac{6}{a^2} + \frac{3}{a} \end{bmatrix}$$

$$\lim_{a\to\infty} \mathbf{H_x} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow$$

$$\hat{x}_1 = \frac{1}{3}(y_1 + y_2 + y_3), \ \hat{x}_2 = 0$$

(2nd) $\mathbf{H_y}(a \to \infty)$

$$\mathbf{H_y} = \frac{1}{\frac{18}{a^2} - \frac{12}{a} + 3} \begin{bmatrix} \frac{6}{a^2} - \frac{2}{a} + 1 & \frac{4}{a^2} - \frac{3}{a} + 1 & \frac{2}{a^2} - \frac{4}{a} + 1 & \frac{8}{a^2} - \frac{3}{a} \\ \frac{4}{a^2} - \frac{3}{a} + 1 & \frac{6}{a^2} - \frac{4}{a} + 1 & \frac{8}{a^2} - \frac{5}{a} + 1 & \frac{2}{a^2} \\ \frac{2}{a^2} - \frac{4}{a} + 1 & \frac{8}{a^2} - \frac{5}{a} + 1 & \frac{14}{a^2} - \frac{6}{a} + 1 & -\frac{4}{a^2} + \frac{3}{a} \\ \frac{8}{a^2} - \frac{3}{a} & \frac{2}{a^2} & -\frac{4}{a^2} + \frac{3}{a} & \frac{14}{a^2} - \frac{12}{a} + 3 \end{bmatrix}$$

$$\lim_{a\to\infty} \mathbf{H_y} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \ \lim_{a\to\infty} h_{44} = 1 \Rightarrow$$

$$\hat{y}_1 = \hat{y}_2 = \hat{y}_3 = \frac{1}{3}(y_1 + y_2 + y_3), \ \hat{y}_4 = y_4$$

(3rd) $\mathbf{R}(a \to \infty)$

$$\mathbf{R} = \frac{1}{\frac{18}{a^2} - \frac{12}{a} + 3} \begin{bmatrix} \frac{10}{a^2} - \frac{10}{a} + 2 & -\frac{4}{a^2} + \frac{3}{a} - 1 & -\frac{2}{a^2} + \frac{4}{a} - 1 & -\frac{8}{a^2} + \frac{3}{a} \\ -\frac{4}{a^2} + \frac{3}{a} - 1 & \frac{12}{a^2} - \frac{8}{a} + 2 & -\frac{8}{a^2} + \frac{5}{a} - 1 & -\frac{2}{a^2} \\ -\frac{2}{a^2} + \frac{4}{a} - 1 & -\frac{8}{a^2} + \frac{5}{a} - 1 & \frac{4}{a^2} - \frac{6}{a} + 2 & \frac{4}{a^2} - \frac{3}{a} \\ -\frac{8}{a^2} + \frac{3}{a} & -\frac{2}{a^2} & \frac{4}{a^2} - \frac{3}{a} & \frac{4}{a^2} \end{bmatrix}$$

$$\lim_{a\to\infty} \mathbf{R}(a) = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$$\hat{i}_1 = \frac{1}{3}(2y_1 - y_2 - y_3),\ \hat{i}_2 = \frac{1}{3}(-y_1 + 2y_2 - y_3),\ \hat{i}_3 = \frac{1}{3}(-y_1 - y_2 + 2y_3),\ \hat{i}_4 = 0$$

(4th) LESS fit $:||\hat{i}||^2$

$$\lim_{a\to\infty} ||\hat{i}(a)||^2 = \frac{1}{3}\mathbf{y}' \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{y}$$

$$\lim_{a\to\infty} ||\hat{i}(a)||^2 = \frac{1}{3}(2y_1^2 + 2y_2^2 + 2y_3^2 - 2y_1y_2 - 2y_2y_3 - 2y_3y_1).$$

A fascinating result is achieved upon analyzing (the right eigenspace of the *hat matrix* $\mathbf{H_y}(a \to \infty)$. *First*, we computed the spectrum of the matrices $\mathbf{A'A}$ and $(\mathbf{A'A})^{-1}$. *Second*, we proved $\lambda_1(a \to \infty) = \infty$, $\lambda_2(a \to \infty) = 3$ or $\lambda_1^{-1}(a \to \infty) = 0$, $\lambda_2^{-1}(a \to \infty) = 1/3$.

Box 3.18. Right eigenspace analysis of a linear model of a univariate polynomial of degree one – extreme leverage point $a \to \infty$-.

"*Hat matrix* $\mathbf{H_y} = \mathbf{A'(A'A)^{-1}A'}$"

$$\textit{right eigenspace analysis}: \left[\begin{array}{l} \mathbf{A'AV} = \mathbf{V\Lambda}^2 \\ \text{subject to} \\ \mathbf{VV'} = \mathbf{I}_{m'} \end{array} \right.$$

$$\text{spec}(\mathbf{A'A}) = \{\lambda_1^2, \lambda_2^2\} : \left|\mathbf{A'A} - \lambda_j^2\mathbf{I}\right| = 0\ \forall\ j \in \{1,2\} \Leftrightarrow$$

$$\begin{vmatrix} 4-\lambda^2 & 6+a \\ 6+a & 14+a^2-\lambda^2 \end{vmatrix} = 0 \Leftrightarrow \lambda^4 - \lambda^2(18+a^2) + 20 - 12a + 3a^2 = 0$$

$$\lambda_{1,2}^2 = \frac{1}{2}\text{tr}\,(\mathbf{A'A}) \pm \sqrt{(\text{tr}\mathbf{A'A})^2 - 4\det \mathbf{A'A}}$$

$$\text{tr}\mathbf{A'A} = 18 + a^2,\ \det \mathbf{A'A} = 20 - 12a + 3a^3$$

$$(\text{tr}\mathbf{A'A})^2 - 4\det\mathbf{A'A} = 244 + 46a + 25a^2 + a^4$$

$$\lambda_{1,2}^2 = 9 + \frac{a^2}{2} \pm \sqrt{61 + 12a + 6a^2 + \frac{a^4}{4}}$$

$$\text{spec}(\mathbf{A'A}) = \{\lambda_1^2, \lambda_2^2\} =$$

$$= \left\{ 9 + \frac{a^2}{2} + \sqrt{61 + 12a + 6a^2 + \frac{a^4}{4}},\ 9 + \frac{a^2}{2} - \sqrt{61 + 12a + 6a^2 + \frac{a^4}{4}} \right\}$$

"*inverse spectrum*"

$$\text{spec}(\mathbf{A'A}) = \{\lambda_1^2, \lambda_2^2\} \Leftrightarrow \text{spec}(\mathbf{A'A})^{-1} = \left\{ \frac{1}{\lambda_1^2}, \frac{1}{\lambda_2^2} \right\}$$

$$\frac{1}{\lambda_2^2} = \frac{9 + \frac{a^2}{2} - \sqrt{61 + 12a + 6a^2 + \frac{a^4}{4}}}{20 - 12a + 3a^2} = \frac{\frac{9}{a^2} + \frac{1}{2} - \sqrt{\frac{61}{a^4} + \frac{12}{a^3} + \frac{6}{a^2} + \frac{1}{4}}}{\frac{20}{a^2} - \frac{12}{a} + 3}$$

$$\lim_{a\to\infty} \frac{1}{\lambda_1^2} = 0$$

$$\frac{1}{\lambda_1^2} = \frac{9 + \frac{a^2}{2} + \sqrt{61 + 12a + 6a^2 + \frac{a^4}{4}}}{20 - 12a + 3a^2} = \frac{\frac{9}{a^2} + \frac{1}{2} + \sqrt{\frac{61}{a^4} + \frac{12}{a^3} + \frac{6}{a^2} + \frac{1}{4}}}{\frac{20}{a^2} - \frac{12}{a} + 3}$$

$$\lim_{a\to\infty} \frac{1}{\lambda_2^2} = \frac{1}{3}$$

$$\lim_{a\to\infty} \mathrm{spec}(\mathbf{A}'\mathbf{A})(a) = \{\infty, 3\} \Leftrightarrow \lim_{a\to\infty} \mathrm{spec}(\mathbf{A}'\mathbf{A})^{-1} = \{0, \frac{1}{3}\}$$

$$\left.\begin{array}{l}\mathbf{A}'\mathbf{A}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda}^2 \\ \mathbf{V}\mathbf{V}' = \mathbf{I}_m\end{array}\right] \Rightarrow \mathbf{A}'\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}' \Leftrightarrow (\mathbf{A}'\mathbf{A})^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}'$$

$$\text{“}\textit{Hat matrix } \mathbf{H_y} = \mathbf{A}\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}'\mathbf{A}'\text{”}.$$

3-32 Multilinear Algebra, “Join” *and* “Meet”*, the Hodge Star Operator*

Before we can analyze the matrices “*hat* $\mathbf{H}_y$” and “*red* $\mathbf{R}$” in more detail, we have to listen to an “*intermezzo*” entitled multilinear algebra, “*join*” and “*meet*” as well as the *Hodge star operator*. The *Hodge star operator* will lay down the foundation of “*latent restrictions*” within our linear model and of *Grassmann coordinates*, also referred to as *Plücker coordinates*.

Box 3.20 summarizes the definitions of multilinear algebra, the relations “join and meet”, denoted by “$\wedge$ ” and “*”, respectively. In terms of orthonormal base vectors $\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}$, we introduce by (3.73) the exterior product $\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_m}$ also known as “*join*”, “*skew product*” or *1st Grassmann relation*. Indeed, such an exterior product is antisymmetric as defined by (3.74), (3.75), (3.76) and (3.77). The examples show $\mathbf{e}_1 \wedge \mathbf{e}_2 = -\mathbf{e}_2 \wedge \mathbf{e}_1$ and $\mathbf{e}_1 \wedge \mathbf{e}_1 = 0$, $\mathbf{e}_2 \wedge \mathbf{e}_2 = 0$. Though the operations“*join*”, namely the exterior product, can be digested without too much of an effort, the operation “*meet*”, namely the *Hodge star operator*, needs much more attention. Loosely speaking the *Hodge star operator* or *2nd Grassmann relation* is a generalization of the conventional “*cross product*” symbolized by “$\times$”. Let there be given an *exterior form of degree k* as an element of $\Lambda^k(\mathbb{R}^n)$ over the field of real numbers $\mathbb{R}^n$. Then the “*Hodge* *” transforms the input exterior form of degree m to the output exterior form of degree $n - m$, namely an element of $\Lambda^{n-k}(\mathbb{R}^n)$.

$$Input: \ \mathbf{X} \in \Lambda^m(\mathbb{R}^n) \to Output: \ *\mathbf{X} \in \Lambda^{n-m}.$$

Applying the summation convention over repeated indices, (3.79) introduces the input operation "*join*", while (3.80) provides the output operation "*meet*". We say that $*\mathbf{X}$, (3.80) is a representation of the *adjoint form* based on the *original form* $\mathbf{X}$, (3.79). The *Hodge dualizer* is a complicated exterior form (3.80) which is based upon *Levi–Civita's symbol of antisymmetry* (3.81) which is illustrated by 3 examples. $\varepsilon_{k_1 \cdots k_\ell}$ is also known as the *permutation operator*. Unfortunately, we have *no space and time* to go deeper into "*join and meet*". Instead we refer to those excellent text-books on exterior algebra and exterior analysis, differential topology, in short *exterior calculus*.

Box 3.19. ("join and meet"):

"*Hodge star operator*"$\wedge$, $*$

$$\mathbf{I} := \{i_1, \ldots, i_k, i_{k+1}, \ldots, i_n\} \subset \{1, \ldots, n\}$$

"*join*": exterior product, skew product,
1st Grassmann relation

$$\mathbf{e}_{i_1 \cdots i_m} := \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_j \wedge \mathbf{e}_{j+1} \wedge \cdots \wedge \mathbf{e}_{i_m} \tag{3.73}$$

"*antisymmetry*":

$$\mathbf{e}_{\mathbf{i_1} \cdots \mathbf{ij} \cdots \mathbf{i_m}} = -\mathbf{e}_{\mathbf{i_1} \cdots \mathbf{ji} \cdots \mathbf{i_m}} \ \forall \mathbf{i} \neq \mathbf{j} \tag{3.74}$$

$$\mathbf{e}_{i_1} \wedge \ldots \wedge \mathbf{e}_{j_k} \wedge \mathbf{e}_{j_{k+1}} \wedge \ldots \wedge \mathbf{e}_{i_m} = -\mathbf{e}_{i_1} \wedge \ldots \wedge \mathbf{e}_{j_{k+1}} \wedge \mathbf{e}_{j_k} \wedge \ldots \wedge \mathbf{e}_{i_m} \tag{3.75}$$

$$\mathbf{e}_{i_1 \cdots i_i i_j \cdots i_m} = 0 \quad \forall i = j \tag{3.76}$$

$$\mathbf{e}_{i_1} \wedge \cdots \mathbf{e}_i \wedge \mathbf{e}_j \wedge \cdots \wedge \mathbf{e}_{i_m} = 0 \ \ \forall i = j \tag{3.77}$$

$$Example: \mathbf{e}_1 \wedge \mathbf{e}_2 = -\mathbf{e}_2 \wedge \mathbf{e}_1 \quad or \quad \mathbf{e}_i \wedge \mathbf{e}_j = -\mathbf{e}_j \wedge \mathbf{e}_i \ \ \forall i \neq j$$

$$Example: \mathbf{e}_1 \wedge \mathbf{e}_1 = 0, \ \mathbf{e}_2 \wedge \mathbf{e}_2 = 0 \ \ or \ \ \mathbf{e}_i \wedge \mathbf{e}_j = 0 \quad \forall i = j$$

"meet": Hodge star operator, Hodge dualizer 2nd Grassmann relation

$$* \ : \ \mathbf{\Lambda}^m(\mathbb{R}^n) \to {}^{n-m}\mathbf{\Lambda}(\mathbb{R}^n) \tag{3.78}$$

"a m degree exterior form $\mathbf{X} \in \mathbf{\Lambda}^m(\mathbb{R}^n)$ *over* $\mathbb{R}^n$ is related to a $n - m$ degree exterior form $*\mathbf{X}$ called the *adjoint form*"

:*summation convention*:
"sum up over repeated indices" *input*: "join"

$$\mathbf{X} = \frac{1}{m!}\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_m} \mathbf{X}^{i_1 \cdots i_m} \tag{3.79}$$

output: "meet"

$$* \, \mathbf{X} := \frac{1}{m!(n-m)!}\sqrt{g}\, \mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_{n-m}} \quad \varepsilon_{i_1 \cdots i_m j_1 \cdots j_{n-m}} \mathbf{X}^{i_1 \cdots i_m} \tag{3.80}$$

antisymmetry operator ("*Eddington's epsilons*"):

$$\varepsilon_{k_1 \cdots k_\ell} := \left[\begin{array}{ll} +1 & \textit{for an even permutation of the indices } k_1 \cdots k_\ell \\ -1 & \textit{for an oded permutation of the indices } k_1 \cdots k_\ell \\ 0 & \textit{otherwise (for a repetition of the indices)}. \end{array} \right. \tag{3.81}$$

$$\textit{Example}: \ \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} = +1$$

$$\textit{Example}: \ \varepsilon_{213} = \varepsilon_{321} = \varepsilon_{132} = -1$$

$$\textit{Example}: \ \varepsilon_{112} = \varepsilon_{223} = \varepsilon_{331} = \ 0.$$

For our purposes two examples on "*Hodge's star*" will be sufficient for the following analysis of latent restrictions in our linear model. In all detail, *Box 3.21* illustrates "*join and meet*" for

$$\mathbf{\Lambda}^2(\mathbb{R}^3) \to \mathbf{\Lambda}^1(\mathbb{R}^3).$$

Given the exterior product $\mathbf{a} \wedge \mathbf{b}$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$ with

$$a_{i_1 1} = \mathrm{col}_1 \mathbf{A}, \ \ a_{i_2 2} = \mathrm{col}_2 \mathbf{A}$$

as their *coordinates*, the columns of the matrix $\mathbf{A}$ with respect to the *orthonormal frame of reference* $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 | \, 0\}$ at the origin 0.

$$\mathbf{a} \wedge \mathbf{b} = \sum_{i_1, i_2 = 1}^{n=3} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \ \ a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(\mathbb{R}^3)$$

is the representation of the exterior form $\mathbf{a} \wedge \mathbf{b} =: \mathbf{X}$ in the multibasis $\mathbf{e}_{i_1 i_2} = \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2}$. By *cyclic ordering*, (3.105) is an explicit write-up of $\mathbf{a} \wedge \mathbf{b} \in \mathcal{R}(\mathbf{A})$. *Please, notice* that there are

$$\binom{n}{m} = \binom{3}{2} = 3$$

subdeterminants of $\mathbf{A}$. If the determinant of the matrix $\mathbf{G} = \mathbf{I}_4$, $\sqrt{\det \mathbf{G}} = 1$ $\sqrt{g} = 1$, then according to (16.80), (3.86)

$$*(\mathbf{a} \wedge \mathbf{b}) \in \mathcal{R}(\mathbf{A})^\perp = \mathbf{G}^{1,3}$$

represent the exterior form $*\mathbf{X}$, which is an element of $\mathcal{R}(\mathbf{A})$ called *Grassmann space* $\mathbf{G}^{1,3}$. Notice that $*(\mathbf{a} \wedge \mathbf{b})$ is a vector whose *Grassmann coordinate* (*Plücker* *coordinate*) are

$$\binom{n}{m} = \binom{3}{2} = 3$$

subdeterminants of the matrix A, namely

$$a_{21}a_{32} - a_{31}a_{22}, a_{31}a_{12} - a_{11}a_{32}, a_{11}a_{23} - a_{21}a_{12}.$$

Finally, (3.87) $*(\mathbf{e}_2 \wedge \mathbf{e}_3) = \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$ for instance demonstrates the relation between called "*join, meet*" and the "*cross product*".

Box 3.20. (The first example:)

"join and meet"

$$* : \Lambda^2(\mathbb{R}^3) \to \Lambda^2(\mathbb{R}^3)$$

Input: "*join*"

$$\mathbf{a} = \sum_{i=1}^{n=3} \mathbf{e}_{i_1} a_{i_1 1}, \quad \mathbf{b} = \sum_{i=1}^{n=3} \mathbf{e}_{i_2} a_{i_2 2} \tag{3.82}$$

$$a_{i_1 1} = \mathrm{col}_1 \mathbf{A}; \quad a_{i_2 2} = \mathrm{col}_2 \mathbf{A}$$

$$\mathbf{a} \wedge \mathbf{b} = \frac{1}{2!} \sum_{i_1, i_2 = 1}^{n=3} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \quad a_{i_1 1} a_{i_2 2} \in \boldsymbol{\Lambda}^2(\mathbb{R}^3) \tag{3.83}$$

"*cyclic order*"

$$\begin{aligned} \mathbf{a} \wedge \mathbf{b} = & \frac{1}{2!} \mathbf{e}_2 \wedge \mathbf{e}_3 (a_{21}a_{32} - a_{31}a_{22}) \\ & + \frac{1}{2!} \mathbf{e}_3 \wedge \mathbf{e}_1 (a_{31}a_{12} - a_{11}a_{32}) \\ & + \frac{1}{2!} \mathbf{e}_1 \wedge \mathbf{e}_2 (a_{11}a_{23} - a_{21}a_{12}) \in \mathcal{R}(\mathbf{A}) = \mathbf{G}^{2,3}. \end{aligned} \tag{3.84}$$

Output: "*meet*" $(\sqrt{g} = 1, \; \mathbf{G_y} = \mathbf{I}_3, \; m = 2, \; n = 3, \; n - m = 1)$

$$* (\mathbf{a} \wedge \mathbf{b}) = \sum_{i_1, i_2, j = 1}^{n=2} \frac{1}{2!} \mathbf{e}_j \varepsilon_{i_1, i_2, j} \quad a_{i_1 1} a_{i_2 2} \tag{3.85}$$

$$
\begin{aligned}
*(\mathbf{a}\wedge\mathbf{b}) = & \frac{1}{2!}\mathbf{e}_1(a_{21}a_{32}-a_{31}a_{22}) \\
& + \frac{1}{2!}\mathbf{e}_2(a_{31}a_{12}-a_{11}a_{32}) \\
& + \frac{1}{2!}\mathbf{e}_3(a_{11}a_{23}-a_{21}a_{12}) \in \mathcal{R}^{\perp}(\mathbf{A}) = \mathbf{G}^{1,3}
\end{aligned} \tag{3.86}
$$

$$\binom{n}{m} = \binom{3}{2} \texttt{ subdeterminant of } \mathbf{A}$$

Grassmann coordinates (Plücker coordinates)

$$*\,(\mathbf{e}_2\wedge\mathbf{e}_3) = \mathbf{e}_1, \quad *(\mathbf{e}_3\wedge\mathbf{e}_1) = \mathbf{e}_2, \quad *(\mathbf{e}_1\wedge\mathbf{e}_2) = \mathbf{e}_3 \tag{3.87}$$

Alternatively, *Box 3.22* illustrates "*join and meet*" for selfduality

$$* : \Lambda^2(\mathbb{R}^4) \to \Lambda^2(\mathbb{R}^4).$$

Given the exterior product $\mathbf{a}\wedge\mathbf{b}$ of two vectors $\mathbf{a}\in\mathbb{R}^4$ and $\mathbf{b}\in\mathbb{R}^4$, namely the two column vectors of the matrix $\mathbf{A}\in\mathbb{R}^{4\times 2}$

$$a_{i_1 1} = \mathrm{col}_1\mathbf{A}, \quad a_{i_2 2} = \mathrm{col}_2\mathbf{A}$$

as their coordinates with respect to the *orthonormal frame of reference* $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4\,|\,\mathbf{0}\}$ at the origin $\mathbf{0}$. is the representation of the *exterior form* $\mathbf{a}\wedge\mathbf{b} := \mathbf{X}$ in the multibasis $\mathbf{e}_{i_1 i_2} = \mathbf{e}_{i_1}\wedge\mathbf{e}_{i_2}$. By *lexicographic ordering*, (3.90) is an explicit write-up of $\mathbf{a}\wedge\mathbf{b}$ $(\in\mathcal{R}(\mathbf{A}))$. Notice that these are

$$\binom{n}{m} = \binom{4}{2} = 6$$

subdeterminants of $\mathbf{A}$. If the determinant of the matrix G of the metric is one $\mathbf{G} = \mathbf{I}_4$, $\sqrt{\det\mathbf{G}} = \sqrt{g} = 1$, then according to (3.91), (3.92)

$$*(\mathbf{a}\wedge\mathbf{b}) \in \mathcal{R}(\mathbf{A})^{\perp} =: \mathbf{G}^{2,4}$$

represents the exterior form $*\mathbf{X}$, an element of $\mathcal{R}(\mathbf{A})^{\perp}$, called *Grassmann space* $\mathbf{G}^{2,4}$. Notice that $*(\mathbf{a}\wedge\mathbf{b})$ is an exterior 2-form which has been generated by an exterior 2-form, *too*. Such a relation is called "selfdual". Its *Grassmann coordinates* (*Plücker coordinates*) are

$$\binom{n}{m} = \binom{4}{2} = 6$$

subdeterminants of the matrix $\mathbf{A}$, namely

$$a_{11}a_{12}-a_{21}a_{12},\; a_{11}a_{32}-a_{31}a_{22},\; a_{11}a_{42}-a_{41}a_{12},$$

$$a_{21}a_{32}-a_{31}a_{22},\; a_{21}a_{42}-a_{41}a_{22},\; a_{31}a_{41}-a_{41}a_{32}.$$

Finally, (3.92), for instance $*(\mathbf{e}_1 \wedge \mathbf{e}_2) = \mathbf{e}_3 \wedge \mathbf{e}_4$, demonstrates the operation called "*join and meet*", indeed quite a generalization of the "*cross product*".

Box 3.21. (The second example:)

"join and meet"

$$* \, : \, \Lambda^2(\mathbb{R}^4) \to \Lambda^2(\mathbb{R}^4)$$

"*selfdual*"

Input: "*join*"

$$\mathbf{a} = \sum_{i_1=1}^{n=4} \mathbf{e}_{i_1} a_{i_1 1}, \;\; \mathbf{b} = \sum_{i_2=1}^{n=4} \mathbf{e}_{i_2} a_{i_2 2} \tag{3.88}$$

$$(a_{i_1 1} = \mathrm{col}_1(\mathbf{A}), \;\; a_{i_2 2} = \mathrm{col}_2(\mathbf{A}))$$

$$\mathbf{a} \wedge \mathbf{b} = \frac{1}{2!} \sum_{i_1, i_2=1}^{n=4} \mathbf{e}_{\mathbf{i}_1} \wedge \mathbf{e}_{\mathbf{i}_2} \, a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(R^4) \tag{3.89}$$

"*lexicographical order*"

$$\begin{aligned} \mathbf{a} \wedge \mathbf{b} = & \frac{1}{2!} \mathbf{e}_1 \wedge \mathbf{e}_2 (a_{11} a_{22} - a_{21} a_{12}) \\ & + \frac{1}{2!} \mathbf{e}_1 \wedge \mathbf{e}_3 (a_{11} a_{32} - a_{31} a_{22}) \\ & + \frac{1}{2!} \mathbf{e}_1 \wedge \mathbf{e}_4 (a_{11} a_{42} - a_{41} a_{12}) \\ & + \frac{1}{2!} \mathbf{e}_2 \wedge \mathbf{e}_3 (a_{21} a_{32} - a_{31} a_{22}) \\ & + \frac{1}{2!} \mathbf{e}_2 \wedge \mathbf{e}_4 (a_{21} a_{42} - a_{41} a_{22}) \\ & + \frac{1}{2!} \mathbf{e}_3 \wedge \mathbf{e}_4 (a_{31} a_{42} - a_{41} a_{32}) \in \mathcal{R}(\mathbf{A})^{\perp} = \mathbf{G}^{2,4} \end{aligned} \tag{3.90}$$

$$\binom{n}{m} = \binom{4}{2} \;\; \textit{subdeterminants of } \mathbf{A} :$$

Grassmann coordinates (Plücker coordinates).

Output:"*meet*" $\sqrt{g} = 1, \;\; \mathbf{G}_{\mathbf{y}} = \mathbf{I}_4, \;\; m = 2, \;\; n = 4, \;\; n - m = 2$

$$
\begin{aligned}
*(\mathbf{a} \wedge \mathbf{b}) &= \frac{1}{2!} \sum_{i_1,i_2,j_1,j_2=1}^{n=4} \frac{1}{2!} \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \;\; \varepsilon_{i_1 i_2 j_1 j_2} \;\; a_{i_1 1} a_{i_2 2} \\
&= \frac{1}{4} \mathbf{e}_3 \wedge \mathbf{e}_4 (a_{11} a_{22} - a_{21} a_{12}) \\
&\quad + \frac{1}{4} \mathbf{e}_2 \wedge \mathbf{e}_4 (a_{11} a_{32} - a_{31} a_{22}) \\
&\quad + \frac{1}{4} \mathbf{e}_3 \wedge \mathbf{e}_2 (a_{11} a_{42} - a_{41} a_{12}) \\
&\quad + \frac{1}{4} \mathbf{e}_4 \wedge \mathbf{e}_1 (a_{21} a_{32} - a_{31} a_{22}) \\
&\quad + \frac{1}{4} \mathbf{e}_3 \wedge \mathbf{e}_1 (a_{21} a_{22} - a_{41} a_{22}) \\
&\quad + \frac{1}{4} \mathbf{e}_1 \wedge \mathbf{e}_2 (a_{31} a_{42} - a_{41} a_{32}) \in \mathcal{R}(\mathbf{A})^{\perp} = \mathbf{G}^{2,4}
\end{aligned}
\tag{3.91}
$$

$$\binom{n}{m} = \binom{4}{2} \; \textit{subdeterminants of } \mathbf{A}:$$

Grassmann coordinates (*Plücker coordinates*).

$$
\begin{aligned}
&*(\mathbf{e}_1 \wedge \mathbf{e}_2) = \mathbf{e}_3 \wedge \mathbf{e}_4, \; *(\mathbf{e}_1 \wedge \mathbf{e}_3) = \mathbf{e}_2 \wedge \mathbf{e}_4, \; *(\mathbf{e}_1 \wedge \mathbf{e}_4) = \mathbf{e}_3 \wedge \mathbf{e}_2, \\
&*(\mathbf{e}_2 \wedge \mathbf{e}_3) = \mathbf{e}_4 \wedge \mathbf{e}_1, \; *(\mathbf{e}_2 \wedge \mathbf{e}_4) = \mathbf{e}_3 \wedge \mathbf{e}_1, \\
&*(\mathbf{e}_3 \wedge \mathbf{e}_4) = \mathbf{e}_1 \wedge \mathbf{e}_2.
\end{aligned}
\tag{3.92}
$$

3-33 From A to B: Latent Restrictions, Grassmann Coordinates, Plücker Coordinates

Before we return to the matrix $\mathbf{A} \in \mathbb{R}^{4\times 2}$ of our case study, let us analyze the matrix $\mathbf{A} \in \mathbb{R}^{2\times 3}$ of *Box 3.23* for simplicity. In the perspective of the example of our case study we may say that we have eliminated the third observation, *but kept* the leverage point. *First*, let us go through the routine to compute the hat matrices $\mathbf{H_x} = (\mathbf{A}'\,\mathbf{A})^{-1}\mathbf{A}'$ *and* $\mathbf{H_y} = \mathbf{A}(\mathbf{A}'\,\mathbf{A})^{-1}\mathbf{A}'$, to be identified by (3.93) and (3.94). The corresponding estimations $\hat{\mathbf{x}} = \mathbf{x}_{\ell}(\mathbf{I}\text{-LESS})$, (3.94), and $\mathbf{y} = \mathbf{y}_{\ell}(\mathbf{I}\text{-LESS})$, (3.96), prove the different weights of the observations (y_1, y_2, y_3) influencing $\hat{x}_1$ and $\hat{x}_2$ as well as $(\hat{y}_1, \hat{y}_2, \hat{y}_3)$. Notice the great weight of the leverage point $t_3 = 10$ on $\hat{y}_3$.

Second, let us interpret the redundancy matrix $\mathbf{R} = \mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, in particular the diagonal elements.

$$
\begin{aligned}
&r_{11} = \frac{\mathbf{A}'\mathbf{A}_{(1)}}{\det \mathbf{A}'\mathbf{A}} = \frac{64}{146}, \; r_{22} = \frac{\mathbf{A}'\mathbf{A}_{(2)}}{\det \mathbf{A}'\mathbf{A}} = \frac{81}{146}, \; r_{33} = \frac{\mathbf{A}'\mathbf{A}_{(3)}}{\det \mathbf{A}'\mathbf{A}} = \frac{1}{146}, \\
&\mathrm{tr}\mathbf{R} = \frac{1}{\det \mathbf{A}'\mathbf{A}} \sum_{i=1}^{n=3} (\mathbf{A}'\mathbf{A})_{(i)} = n - \mathrm{rk}\mathbf{A} = n - m = 1,
\end{aligned}
$$

the degrees of freedom of the $\mathbf{I}_3$-$LESS$ problem. There, for the *first time*, we meet the *subdeterminants* $(\mathbf{A}'\mathbf{A})_{(i)}$ which are generated in a two step procedure.

"First step"	"Second step"
eliminate the ith row from A as well as the ith column of A.	compute the determinant $\mathbf{A}'_{(i)}\mathbf{A}_{(i)}$.

Example :$(\mathbf{A}'\mathbf{A})_1$

$\mathbf{A}'_{(1)}\mathbf{A}_{(1)}$	$\cancel{1}$ $\cancel{1}$ 1 2 1 10
$\cancel{1}$ 1 1 $\cancel{1}$ 2 10	2 12 12 104

$(\mathbf{A}'\mathbf{A})_{(1)} = \det \mathbf{A}'_{(1)}\mathbf{A}_{(1)} = 64$
$\det \mathbf{A}'\mathbf{A} = 146$

Example:$(\mathbf{A}'\mathbf{A})_2$

$\mathbf{A}'_{(2)}\mathbf{A}_{(2)}$	1 1 $\cancel{1}$ $\cancel{2}$ 1 10
1 $\cancel{1}$ 1 1 $\cancel{2}$ 10	2 11 11 101

$(\mathbf{A}'\mathbf{A})_{(2)} = \det \mathbf{A}'_{(2)}\mathbf{A}_{(2)} = 81$
$\det \mathbf{A}'\mathbf{A} = 146$

Example:$(\mathbf{A}'\mathbf{A})_3$

$\mathbf{A}'_{(3)}\mathbf{A}_{(3)}$	1 1 1 2 $\cancel{1}$ $\cancel{10}$
1 1 $\cancel{1}$ 1 2 $\cancel{10}$	2 3 3 5

$(\mathbf{A}'\mathbf{A})_{(3)} = \det \mathbf{A}'_{(3)}\mathbf{A}_{(3)} = 1$
$\det \mathbf{A}'\mathbf{A} = 146$

Obviously, the *partial redundancies* (r_{11}, r_{22}, r_{33}) are associated with the influence of the observation $y_1, y_2 or y_3$ on the total degree of freedom. Here the observation y_1 and y_2 had the greatest contribution, the observation y_3 at a *leverage point* a *very small influence*.

The redundancy matrix $\mathbf{R}$, properly analyzed, will lead us to the *latent restrictions* or "from **A** to **B**". *Third*, we introduce the *rank partitioning* $\mathbf{R} = [\mathbf{B}, \mathbf{C}]$, $\mathrm{rk}\mathbf{R} = \mathrm{rk}\,\mathbf{B} = n - m = 1$, (3.98), of the matrix $\mathbf{R}$ of *spatial redundancies*. Here, $\mathbf{b} \in \mathbb{R}^{3\times 1}$, (3.100), is normalized to generate $\mathbf{b}* = \mathbf{b}/||\mathbf{b}||_2$, (3.101). Note, $\mathbf{C} \in \mathbb{R}^{3\times 2}$ is a dimension identity. We already introduced the *orthogonality condition*

$$\mathbf{b}'\mathbf{A} = 0 \quad or \quad \mathbf{b}'\mathbf{A}\mathbf{x}_\ell = \mathbf{b}'\hat{\mathbf{y}} = 0$$
$$(\mathbf{b}^*)'\mathbf{A} = 0 \quad or \quad (\mathbf{b}^*)'\mathbf{A}\mathbf{x}_\ell = (\mathbf{b}^*)'\hat{\mathbf{y}} = 0\,,$$

which establishes the latent restrictions (3.127)

$$8\hat{y}_1 - 9\hat{y}_2 + \hat{y}_3 = 0\,.$$

We shall geometrically interpret this *essential result* as soon as possible. *Fourth*, we aim at identifying $\mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})^{\perp}$ for the linear model $\{\mathbf{Ax} + \mathbf{i} = \mathbf{y}, \mathbf{A} \in \mathbb{R}^{n\times m}, rk\mathbf{A} = m = 2\}$

$$t_1 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y]\begin{bmatrix}1\\1\\1\end{bmatrix},$$

$$t_2 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_2} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y]\begin{bmatrix}1\\2\\10\end{bmatrix},$$

as derivatives of the observation functional $\mathbf{y} = f(\mathbf{x}_1, \mathbf{x}_2)$ establish the tangent vectors which span a linear manifold called *Grassmann space*.

$$\mathbf{G}^{2,3} = \text{span}\{t_1, t_2\} \subset \mathbb{R}^3,$$

in short GRASSMANN (A). Such a notation becomes more obvious if we compute

$$\mathbf{y} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y]\begin{bmatrix}a_{11}x_1 + a_{12}x_2\\a_{21}x_1 + a_{22}x_2\\a_{31}x_1 + a_{32}x_2\end{bmatrix} = \sum_{i=1}^{n=3}\sum_{j=1}^{m=2}\mathbf{e}_i^y a_{ij}x_j,$$

$$\frac{\partial \mathbf{y}}{\partial x_1}(x_1, x_2) = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y]\begin{bmatrix}a_{11}\\a_{21}\\a_{31}\end{bmatrix} = \sum_{i=1}^{n=3}\mathbf{e}_i^y a_{i1}$$

$$\frac{\partial \mathbf{y}}{\partial x_2}(x_1, x_2) = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y]\begin{bmatrix}a_{12}\\a_{22}\\a_{32}\end{bmatrix} = \sum_{i=1}^{n=3}\mathbf{e}_i^y a_{i2}.$$

Indeed, the columns of the matrix $\mathbf{A}$ lay the foundation of GRASSMANN (A). *Five*, let us turn to GRASSMANN (B) which is based on the normal space $\mathcal{R}(\mathbf{A})^{\perp}$. The normal vector $\mathbf{n} = \mathbf{t}_1 \times \mathbf{t}_2 = *(\mathbf{t}_1 \wedge \mathbf{t}_2)$ which spans GRASSMANN (B) is defined by the "*cross product*" identified by "$\wedge, *$", the *skew product symbol* as well as the *Hodge star symbol*. Alternatively, we are able to represent the normal vector $\mathbf{n}$, (3.107), (3.109), (3.112), constituted by the columns col1$\mathbf{A}$, col2$\mathbf{A}$ of the matrix, in terms of the *Grassmann coordinates* (*Plücker coordinates*).

$$p_{23} = \begin{vmatrix}a_{21} & a_{22}\\a_{31} & a_{32}\end{vmatrix} = 8, \quad p_{31} = \begin{vmatrix}a_{31} & a_{32}\\a_{11} & a_{12}\end{vmatrix} = -9, \quad p_{12} = \begin{vmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\end{vmatrix} = 1,$$

identified as the subdeterminants of the matrix $\mathbf{A}$, generated by

$$\sum_{i_1,i_2=1}^{n=3} *(\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2}) a_{i_1 1} a_{i_2 2}.$$

If we normalize the vector b to $\mathbf{b}^* = \mathbf{b}/||\mathbf{b}||_2$ and the vector n to $\mathbf{n}^* = \mathbf{n}/||\mathbf{n}||_2$, we are led to the first corollary $\mathbf{b}^* = \mathbf{n}^*$. The space *spanned by the normal vector* $\mathbf{n}$, namely the linear manifold $\mathbf{G}^{1,3} \subset \mathbb{R}^3$ defines GRASSMANN (B). In exterior calculus, the vector built on *Grassmann coordinates* (*Plücker coordinates*) is called *Grassmann vector* $\mathbf{g}$ or *normalized Grassmann vector* $\mathbf{g}^*$, here

$$\mathbf{g} := \begin{bmatrix} p_{23} \\ p_{31} \\ p_{12} \end{bmatrix} = \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix}, \ \mathbf{g}^* := \frac{\mathbf{g}}{\parallel \mathbf{g} \parallel_2} = \frac{1}{146} \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix}.$$

The *second* corollary identifies $\mathbf{b}^* = \mathbf{n}^* = \mathbf{g}^*$.

"The vector $\mathbf{b}^*$ which constitutes the latent restriction (latent condition equation) coincides with the normalized normal vector $\mathbf{n}^* \in \mathcal{R}(\mathbf{A})^\perp$, an element of the space $\mathcal{R}(\mathbf{A})^\perp$, which is normal to the column space $\mathcal{R}(\mathbf{A})$ of the matrix $\mathbf{A}$. The vector $\mathbf{b}^*$ is built on the *Grassmann coordinates* (*Plücker coordinates*), $[p_{23}, p_{31}, p_{12}]'$, subdeterminant of *vector* $\mathbf{g}^*$ in agreement with $\mathbf{b}^*$."

Box 3.22. (Latent restrictions *Grassmann coordinates* (*Plücker coordinates*) the second example):

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Leftrightarrow \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 10 \end{bmatrix}, \ \text{rk}\mathbf{A} = 2$$

$$(1st)\ \mathbf{H_x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

$$\mathbf{H_x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{146} \begin{bmatrix} 92 & 79 & -25 \\ -10 & -7 & 17 \end{bmatrix} \tag{3.93}$$

$$\hat{\mathbf{x}} = \mathbf{x}_\ell(\mathbf{I}-\text{LESS}) = \frac{1}{146} \begin{bmatrix} 92y_1 + 79y_2 - 25y_3 \\ -10y_1 - 7y_2 + 17y_3 \end{bmatrix} \tag{3.94}$$

$$(2nd)\ \mathbf{H_y} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

$$\mathbf{H_y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{146} \begin{bmatrix} 82 & 72 & -8 \\ 72 & 65 & 9 \\ -8 & 9 & 145 \end{bmatrix}, \ \text{rk}\mathbf{H_y} = \text{rk}\mathbf{A} = 2 \tag{3.95}$$

$$\hat{\mathbf{y}} = \mathbf{y}_\ell(\mathbf{I}-\text{LESS}) = \frac{1}{146} \begin{bmatrix} 82y_1 + 72y_2 - 8y_3 \\ 72y_1 + 65y_2 + 3y_3 \\ -8y_1 + 9y_2 + 145y_3 \end{bmatrix} \tag{3.96}$$

$$\hat{y}_3 = \frac{1}{146}(-8y_1 + 9y_2 + 145y_3) \tag{3.97}$$

$$(3rd)\ \mathbf{R} = \mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

$$\mathbf{R} = \mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{146}\begin{bmatrix} 64 & -72 & 8 \\ -72 & 81 & -9 \\ 8 & -9 & 1 \end{bmatrix} \tag{3.98}$$

$$r_{11} = \frac{64}{146} = \frac{\mathbf{A}'\mathbf{A}_{(1)}}{\det \mathbf{A}'\mathbf{A}},\ \ r_{22} = \frac{81}{146} = \frac{\mathbf{A}'\mathbf{A}_{(2)}}{\det \mathbf{A}'\mathbf{A}},\ \ r_{33} = \frac{1}{146} = \frac{\mathbf{A}'\mathbf{A}_{(3)}}{\det \mathbf{A}'\mathbf{A}}$$

$$\mathrm{tr}\mathbf{R} = \frac{1}{\det \mathbf{A}'\mathbf{A}} \sum_{i=1}^{n=3} (\mathbf{A}'\mathbf{A})_{(i)} = n - \mathrm{rk}\mathbf{A} = n - m = 1$$

latent restriction

$$\mathbf{R} = [\mathbf{B}, \mathbf{C}] = \frac{1}{146}\begin{bmatrix} 64 & -72 & 8 \\ -72 & 81 & -9 \\ 8 & -9 & 1 \end{bmatrix}, \mathrm{rk}\mathbf{R} = 1 \tag{3.99}$$

$$\mathbf{b} := \frac{1}{146}\begin{bmatrix} 64 \\ -72 \\ 8 \end{bmatrix} = \begin{bmatrix} 0.438 \\ -0.493 \\ 0.053 \end{bmatrix} \tag{3.100}$$

$$\mathbf{b}^* := \frac{\mathbf{b}}{\parallel \mathbf{b} \parallel} = \frac{1}{\sqrt{146}}\begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.662 \\ -0.745 \\ 0.083 \end{bmatrix} \tag{3.101}$$

$$\mathbf{b}'\mathbf{A} = 0 \quad \Leftrightarrow \quad (\mathbf{b}^*)'\mathbf{A} = 0 \tag{3.102}$$

$$\mathbf{b}'\hat{\mathbf{y}} = 0 \Leftrightarrow (\mathbf{b}^*)'\hat{\mathbf{y}} = 0 \tag{3.103}$$

$$8\hat{y}_1 - 9\hat{y}_2 + \hat{y}_3 = 0 \tag{3.104}$$

"$\mathcal{R}(\mathbf{A})$ *and* $\mathcal{R}(\mathbf{A})^\perp$

Tangent space $\mathbb{T}_x\mathbb{M}^2$ versus normal space $\mathbb{N}_x\mathbb{M}^2$
Grassmann manifold $\mathbf{G}^{2,3}_m \subset \mathbb{R}^3$ versus *Grassmann manifold* $\mathbf{G}^{1,3}_{n-m} \subset \mathbb{R}^3$"

$$\text{“}\textit{the first tangent vector''}: \mathbf{t}_1 := \frac{\partial \mathbf{y}}{\partial \mathrm{x}_1} = [\mathbf{e}^{\mathrm{y}}_1, \mathbf{e}^{\mathrm{y}}_2, \mathbf{e}^{\mathrm{y}}_3]\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{3.105}$$

$$\text{“}\textit{the second tangent vector''}: \mathbf{t}_2 := \frac{\partial \mathbf{y}}{\partial \mathrm{x}_2} = [\mathbf{e}^{\mathrm{y}}_1, \mathbf{e}^{\mathrm{y}}_2, \mathbf{e}^{\mathrm{y}}_3]\begin{bmatrix} 1 \\ 2 \\ 10 \end{bmatrix} \tag{3.106}$$

$$\mathbf{G}^{m,n}$$

$$\mathbf{G}^{2,3} = span\{\mathbf{t}_1, \mathbf{t}_2\} \in R^3 : Grassman(A)$$

"*the normal vector*"

$$\mathbf{n} := \mathbf{t}_1 \times \mathbf{t}_2 = *(\mathbf{t}_1 \wedge \mathbf{t}_2) \tag{3.107}$$

$$\mathbf{t}_1 = \sum_{i=1}^{n=3} \mathbf{e}_{i_1} a_{i_1 1} \quad and \quad \mathbf{t}_1 = \sum_{i=1}^{n=3} \mathbf{e}_{i_2} a_{i_2 2} \tag{3.108}$$

$$\mathbf{n} = \sum_{i_1, i_2 = 1}^{n=3} \mathbf{e}_{i_1} \mathbf{e}_{i_2} a_{i_1 1} a_{i_2 2} = \sum_{i_1, i_2 = 1}^{n=3} (\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2}) a_{i_1 1} a_{i_2 2} \tag{3.109}$$

$$\forall i, i_1, i_2 \in \{1, \ldots, n = 3\}$$

$$n = \qquad versus \qquad n$$

$$\begin{array}{ll} = \mathbf{e}_2 \times \mathbf{e}_3 (a_{21} a_{32} - a_{31} a_{22}) & = *(\mathbf{e}_2 \times \mathbf{e}_3)(a_{21} a_{32} - a_{31} a_{22}) \\ + \mathbf{e}_3 \times \mathbf{e}_1 (a_{31} a_{12} - a_{11} a_{32}) & + *(\mathbf{e}_3 \times \mathbf{e}_1)(a_{31} a_{12} - a_{11} a_{32}) \\ + \mathbf{e}_1 \times \mathbf{e}_2 (a_{11} a_{22} - a_{21} a_{12}) & + *(\mathbf{e}_1 \times \mathbf{e}_2)(a_{11} a_{22} - a_{21} a_{12}) \end{array} \tag{3.110}$$

$$Hodge\ star\ operator : \left[\begin{array}{l} *(\mathbf{e}_2 \wedge \mathbf{e}_3) = \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1 \\ *(\mathbf{e}_3 \wedge \mathbf{e}_1) = \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2 \\ *(\mathbf{e}_1 \wedge \mathbf{e}_2) = \mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3 \end{array} \right. \tag{3.111}$$

$$\mathbf{n} = \mathbf{t}_1 \times \mathbf{t}_2 = *(\mathbf{t}_1 \times \mathbf{t}_2) = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y] \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix} \tag{3.112}$$

$$\mathbf{n}^* := \frac{\mathbf{n}}{||\mathbf{n}||} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y] \frac{1}{\sqrt{146}} \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix} \tag{3.113}$$

Corollary: $\mathbf{b}^* = \mathbf{n}^*$

"Grassmann manifold $\mathbf{G}^{n-m,n}$"

$$\mathbf{G}^{1,3} = span\, \mathbf{n} \subset \mathbb{R}^3 : Grassman(\mathbf{B})$$

Grassmann coordinates (*Plücker coordinates*)

$$\begin{aligned} \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 10 & 10 \end{bmatrix}, g(\mathbf{A}) &:= \left\{ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, \begin{vmatrix} a_{31} & a_{32} \\ a_{11} & a_{12} \end{vmatrix}, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \right\} \\ &= \left\{ \begin{vmatrix} 1 & 2 \\ 1 & 10 \end{vmatrix}, \begin{vmatrix} 1 & 10 \\ 1 & 1 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \right\} = \{8, -9, 1\} \end{aligned} \tag{3.114}$$

(cyclic order)

$$g(\mathbf{A}) = \{p_{23}, p_{31}, p_{12}\}$$

$$p_{23} = 8, \;\; p_{31} = -9, \;\; p_{12} = 1$$

$$\textit{Grassmann vector} : \mathbf{g} := \begin{bmatrix} p_{23} \\ p_{31} \\ p_{12} \end{bmatrix} = \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix} \tag{3.115}$$

$$\textit{normalized Grassmann vector} : \mathbf{g}^* := \frac{\mathbf{g}}{||\mathbf{g}||} = \frac{1}{\sqrt{146}} \begin{bmatrix} 8 \\ -9 \\ 1 \end{bmatrix} \tag{3.116}$$

$$\text{Corollary} : \mathbf{b}^* = \mathbf{n}^* = \mathbf{g}^*$$

Now we are prepared to analyze the matrix $\mathbf{A} \in \mathbb{R}^{2\times 4}$ of our case study. *Box 3.24* outlines *first* the redundancy matrix $\mathbf{R} \in \mathbb{R}^{2\times 4}$ (3.117) used for computing the inconsistency coordinates $\hat{i}_4 = i_4(\mathbf{I} - \text{LESS})$, in particular. Again it is proven that the *leverage point* $t_4 = 10$ has little influence on this fourth coordinate of the inconsistency vector. The diagonal elements $(r_{11}, r_{22}, r_{33}, r_{44})$ of the redundancy matrix are of focal interest. As *partial redundancy numbers* in (3.123) $r_{11} = \frac{\mathbf{AA}_{(1)}}{\det \mathbf{A}'\mathbf{A}} = \frac{57}{100}$, $r_{22} = \frac{\mathbf{AA}_{(2)}}{\det \mathbf{A}'\mathbf{A}} = \frac{67}{100}$, $r_{33} = \frac{\mathbf{AA}_{(3)}}{\det \mathbf{A}'\mathbf{A}} = \frac{73}{100}$, $r_{44} = \frac{\mathbf{AA}_{(4)}}{\det \mathbf{A}'\mathbf{A}} = \frac{3}{100}$, they sum up to

$$\text{tr}\mathbf{R} = \frac{1}{\det \mathbf{A}'\mathbf{A}} \sum_{i=1}^{n=4} (\mathbf{A}'\mathbf{A})_{(i)} = n - \text{rk}, \mathbf{A} = n - m = 2$$

the degree of freedom of the $\mathbf{I}_4$-LESS problem. Here for the *second time* we meet the *subdeterminants* $(\mathbf{A}'\mathbf{A})_{(i)}$ which are generated in a two-step procedure. "First step" "Second step" eliminate the ith row from A as well ascompute the determinant of the ith column of $\mathbf{A}'\mathbf{A}'_{(i)}\mathbf{A}_{(i)}$.

Box 3.23. (Redundancy matrix of a linear model of a uninvariant polynomial of degree one)

– light leverage point $a = 10$ –

"Redundancy matrix $\mathbf{R} = (\mathbf{I}_4 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')$"

$$\mathbf{I}_4 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{100} \begin{bmatrix} 57 & -37 & -31 & 11 \\ -37 & 67 & -29 & -1 \\ -31 & -29 & 73 & -13 \\ 11 & -1 & -13 & 3 \end{bmatrix} \tag{3.117}$$

$$\hat{i}_4 = i_4(\mathbf{I}\text{-LESS}) = \mathbf{R}\mathbf{y} \tag{3.118}$$

$$\hat{i}_4 = i_4(\mathbf{I}\text{-LESS}) = \frac{1}{100}(11y_1 - y_2 - 13y_3 + 3y_4) \tag{3.119}$$

$$r_{11} = \frac{57}{100},\ r_{22} = \frac{67}{100},\ r_{33} = \frac{73}{100},\ r_{44} = \frac{3}{100} \tag{3.120}$$

"rank partitioning"

$$\mathbf{R} \in R^{4\times 4}, \mathrm{rk}\mathbf{R} = n - \mathrm{rk}\mathbf{A} = n - m = 2, \mathbf{B} \in R^{4\times 2}, \mathbf{C} \in R^{4\times 2}$$

$$\mathbf{R} = \mathbf{I}_4 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = [\mathbf{B}, \mathbf{C}] \tag{3.121}$$

$$\text{"}if\ \mathbf{B} := \frac{1}{100}\begin{bmatrix} 57 & -37 \\ -37 & 67 \\ -31 & -29 \\ 11 & -1 \end{bmatrix}, \ then\ \mathbf{B}'\mathbf{A} = 0'' \tag{3.122}$$

$$r_{11} = \frac{\mathbf{A}'\mathbf{A}_{(1)}}{\det \mathbf{A}'\mathbf{A}},\ r_{22} = \frac{\mathbf{A}'\mathbf{A}_{(2)}}{\det \mathbf{A}'\mathbf{A}},\ r_{33} = \frac{\mathbf{A}'\mathbf{A}_{(3)}}{\det \mathbf{A}'\mathbf{A}},\ r_{44} = \frac{\mathbf{A}'\mathbf{A}_{(4)}}{\det \mathbf{A}'\mathbf{A}} \tag{3.123}$$

$$\mathrm{tr}\mathbf{R} = \frac{1}{\det \mathbf{A}'\mathbf{A}} \sum_{i=1}^{n=4} (\mathbf{A}'\mathbf{A})_{(i)} = n - \mathrm{rk}\mathbf{A} = n - m = 2 \tag{3.124}$$

Example: $(\mathbf{A}'\mathbf{A})_{(1)}$:

$\mathbf{A}'_{(1)}\mathbf{A}_{(1)}$	$\cancel{1}\ \cancel{1}$ $1\ 2$ $1\ 3$ $1\ 10$
$\cancel{1}\ 1\ 1\ 1$ $\cancel{1}\ 2\ 3\ 10$	$3\ \ 15$ $15\ 113$

$$(\mathbf{A}'\mathbf{A})_{(1)} = \det(\mathbf{A}'_{(1)}\mathbf{A}_{(1)}) = 114, \det \mathbf{A}'\mathbf{A} = 200$$

Example: $(\mathbf{A}'\mathbf{A})_{(2)}$:

$\mathbf{A}'_{(2)}\mathbf{A}_{(2)}$	$1\ 1$ $\cancel{1}\ \cancel{2}$ $1\ 3$ $1\ 10$
$1\ \cancel{1}\ 1\ 1$ $1\ \cancel{2}\ 3\ 10$	$3\ \ 14$ $14\ 110$

$$(\mathbf{A}'\mathbf{A})_{(2)} = \det(\mathbf{A}'_{(2)}\mathbf{A}_{(2)}) = 134, \det \mathbf{A}'\mathbf{A} = 200$$

Example: $(\mathbf{A}'\mathbf{A})_{(3)}$

$\mathbf{A}'_{(3)}\mathbf{A}_{(3)}$	$1\ 1$ $1\ 2$ $\cancel{1}\ \cancel{3}$ $1\ 10$
$1\ 1\ \cancel{1}\ 1$ $1\ 2\ \cancel{3}\ 10$	$3\ \ 13$ $13\ 105$

$$(\mathbf{A}'\mathbf{A})_{(3)} = \det(\mathbf{A}'_{(3)}\mathbf{A}_{(3)}) = 146, \det \mathbf{A}'\mathbf{A} = 200$$

Example: $(\mathbf{A}'\mathbf{A})_{(4)}$

$\mathbf{A}'_{(2)}\mathbf{A}_{(2)}$	1 1 1 2 1 3 $\cancel{1}$ $\cancel{10}$
1 1 1 $\cancel{1}$ 1 2 3 $\cancel{10}$	3 6 6 10

$(\mathbf{A}'\mathbf{A})_{(4)} = \det(\mathbf{A}'_{(4)}\mathbf{A}_{(4)}) = 6, \det \mathbf{A}'\mathbf{A} = 200$ d

Again, the *partial redundancies* $(r_{11}, \ldots, r_{44})$ are associated with the influence of the observation y_1, y_2, y_3 or y_4 on the total degree of freedom. Here the observations y_1, y_2 and y_3 had the *greatest influence*, in contrast the observation y_4 at the leverage point a *very small impact*.

The *redundancy matrix* $\mathbf{R}$ will be properly analyzed in order to supply us with the *latent restrictions* or the details of "from $\mathbf{A}$ to $\mathbf{B}$". The *rank partitioning* $\mathbf{R} = [\mathbf{B}, \mathbf{C}]$, $\text{rk}\mathbf{R} = \text{rk}\,\mathbf{B} = n - m = 2$, leads us to (3.22) of the matrix $\mathbf{R}$ of *partial redundancies*. Here, $\mathbf{B} \in \mathbb{R}^{4\times 2}$, with two column vectors is established. Note $\mathbf{C} \in \mathbb{R}^{4\times 2}$ is a dimension identity. We already introduced the *orthogonality conditions* in (3.22)

$$\mathbf{B}'\mathbf{A} = 0 \; or \; \mathbf{B}'\mathbf{A}\mathbf{x}_\ell = \mathbf{B}'\mathbf{y}_\ell = 0,$$

which establish the *two latent conditions*

$$\frac{57}{100}\hat{y}_1 - \frac{37}{100}\hat{y}_2 - \frac{31}{100}\hat{y}_3 + \frac{11}{100}\hat{y}_4 = 0$$

$$-\frac{37}{100}\hat{y}_1 + \frac{67}{100}\hat{y}_2 - \frac{29}{100}\hat{y}_3 - \frac{1}{100}\hat{y}_4 = 0.$$

Let us identify in the context of this paragraph $\mathcal{R}(\mathbf{A})\, and\, \mathcal{R}(\mathbf{A})^\perp$ for the linear model

$$\{\mathbf{A}\mathbf{x} + \mathbf{i} := \mathbf{y}, \mathbf{A} \in \mathbb{R}^{m\times n},\ rk\mathbf{A} = m = 2\}.$$

The derivatives

$$\mathbf{t}_1 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y]\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{t}_2 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_2} = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y]\begin{bmatrix} 1 \\ 2 \\ 3 \\ 10 \end{bmatrix},$$

of the observational functional $\mathbf{y} = f(\mathbf{x}_1, \mathbf{x}_2)$ generate the tangent vectors which span a linear manifold called *Grassmann space*

$$\mathbf{G}^{2,4} = span\{\mathbf{t_1}, \mathbf{t_2}\} \subset \mathbb{R}^{\mathbf{4}},$$

in short GRASSMANN (A). An illustration of such a linear manifold is

$$y = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \\ a_{41}x_1 + a_{42}x_2 \end{bmatrix} = \sum_{i=1}^{n=4} \sum_{j=1}^{n=2} \mathbf{e}_i^y a_{ij} x_j,$$

$$\frac{\partial \mathbf{y}}{\partial x_1}(x_1, x_2) = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} = \sum_{i=1}^{n=4} \mathbf{e}_i^y a_{i1},$$

$$\frac{\partial \mathbf{y}}{\partial x_2}(x_1, x_2) = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} = \sum_{i=1}^{n=4} \mathbf{e}_i^y a_{i2}.$$

Box 3.24. (Latent restrictions *Grassmann coordinates* (*Plücker coordinates*): the first example)

$$\mathbf{B}'\mathbf{A} = 0 \Leftrightarrow \mathbf{B}'\mathbf{y} = 0 \tag{3.125}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 10 \end{bmatrix} \Rightarrow \mathbf{B} = \frac{1}{100} \begin{bmatrix} 57 & -37 \\ -37 & 67 \\ -31 & -29 \\ 11 & -1 \end{bmatrix} \tag{3.126}$$

"*latent restriction*"

$$57\hat{y}_1 - 37\hat{y}_2 - 31\hat{y}_3 + 11\hat{y}_4 = 0 \tag{3.127}$$

$$-37\hat{y}_1 + 67\hat{y}_2 - 29\hat{y}_3 - \hat{y}_4 = 0 \tag{3.128}$$

$\mathcal{R}(\mathbf{A})$: `the tangent space` $\mathbb{T}_x\mathbb{M}^2$

`the Grassmann manifold` $\mathbf{G^{2,4}}$

$$\text{“}\textit{the first tangent vector}\text{''} : \mathbf{t}_1 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_1}[\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{3.129}$$

$$\text{“}\textit{the second tangent vector}\text{''} : \mathbf{t}_2 := \frac{\partial \mathbf{y}}{\partial \mathbf{x}_2}[\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} 1 \\ 2 \\ 3 \\ 10 \end{bmatrix} \tag{3.130}$$

$$\mathbf{G}^{2,4} = span\{\mathbf{t_1}, \mathbf{t_2}\} \subset \mathbb{R}^{\mathbf{4}},\ Grassman(\mathbf{A})$$

$$\text{“}\textit{the first normal vector}\text{”} : \mathbf{n}_1^* := \frac{\mathbf{b}_1}{||\mathbf{b}_1||} \tag{3.131}$$

$$||\mathbf{b}_1||^2 = 10^{-4}(57^2 + 37^2 + 31^2 + 11^2) = 57 * 10^{-2} \tag{3.132}$$

$$\mathbf{n}_1^* = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} 0.755 \\ -0.490 \\ -0.411 \\ -0.146 \end{bmatrix} \tag{3.133}$$

$$\text{“}\textit{the second normal vector}\text{''} : \mathbf{n}_2^* := \frac{\mathbf{b}_2}{||\mathbf{b}_2||} \tag{3.134}$$

$$||\mathbf{b}_2||^2 = 10^{-4}(37^2 + 67^2 + 29^2 + 1^2) = 67 * 10^{-2} \tag{3.135}$$

$$\mathbf{n}_2^* = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} -0.452 \\ 0.819 \\ -0.354 \\ -0.012 \end{bmatrix} \tag{3.136}$$

Grassmann coordinates (Plücker coordinates)

$$\begin{gathered}\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 10 \end{bmatrix} \Rightarrow g(\mathbf{A}) := \left\{ \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 10 \end{vmatrix}, \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix}, \begin{vmatrix} 1 & 2 \\ 1 & 10 \end{vmatrix}, \begin{vmatrix} 1 & 3 \\ 1 & 10 \end{vmatrix} \right\} = \\ = \{p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}\} \\ p_{12} = 1,\ p_{13} = 2,\ p_{14} = 9,\ p_{23} = 1,\ p_{24} = 8, p_{34} = 7. \end{gathered} \tag{3.137}$$

Again, the columns of the matrix **A** lay the foundation of GRASSMANN (A). Next we turn to GRASSMANN (B) to be identified as the normal space $\mathcal{R}(\mathbf{A})^{\perp}$. The normal vectors

$$\mathbf{n}_1^* = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ b_{41} \end{bmatrix} \div ||\mathrm{col}_1\mathbf{B}||,\ \mathbf{n}_2^* = [\mathbf{e}_1^y, \mathbf{e}_2^y, \mathbf{e}_3^y, \mathbf{e}_4^y] \begin{bmatrix} b_{21} \\ b_{22} \\ b_{32} \\ b_{42} \end{bmatrix} \div ||\mathrm{col}_2\mathbf{B}||$$

are computed from the *normalized column vectors* of the matrix $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2]$. The normal vectors $\{\mathbf{n}_1, \mathbf{n}_2\}$ span the normal space $\mathcal{R}(\mathbf{A})^{\perp}$, also called GRASSMANN(B). Alternatively, we may substitute the normal vectors $\mathbf{n}_1$ *and* $\mathbf{n}_2$ by the *Grassmann coordinates* (*Plücker coordinates*) of the matrix **A**, namely by the *Grassmann column vector*.

$$p_{12} = \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1,\ p_{13} = \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = 2,\ p_{14} = \begin{vmatrix} 1 & 1 \\ 1 & 10 \end{vmatrix} = 9$$
$$p_{23} = \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} = 1,\ p_{24} = \begin{vmatrix} 1 & 2 \\ 1 & 10 \end{vmatrix} = 8,\ p_{34} = \begin{vmatrix} 1 & 3 \\ 1 & 10 \end{vmatrix} = 7$$

$$n = 4, m = 2, n - m = 2$$

$$\sum_{i_1,i_2=1}^{n=4} *(\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2}) a_{i_1 1} a_{i_2 2} = \frac{1}{2!} \sum_{i_1,i_2,j_1,j_2=1}^{n=4} \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \varepsilon_{i_1,i_2,j_1,j_2} a_{i_1 1} a_{i_2 2}$$

$$\mathbf{g} := \begin{bmatrix} p_{12} \\ p_{13} \\ p_{14} \\ p_{23} \\ p_{24} \\ p_{34} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 9 \\ 1 \\ 8 \\ 7 \end{bmatrix} \in \mathbb{R}^{6\times 1}.$$

?How do the vectors b1, b2,n1, n2 and g relate to each other?

Earlier we already normalized, $\{\mathbf{b}_1^*, \mathbf{b}_2^*\}$ to $\{\mathbf{b}_1^*, \mathbf{b}_2^*\}$, when we constructed $\{\mathbf{n}_1^*, \mathbf{n}_2^*\}$. Then we are left with the question how to relate $\{\mathbf{b}_1^*, \mathbf{b}_2^*\}$ and $\{\mathbf{n}_1^*, \mathbf{n}_2^*\}$ to the *Grassmann column vector* $\mathbf{g}$. The elements of the *Grassmann column vector* $\mathbf{g}(\mathbf{A})$ associated with matrix $\mathbf{A}$ are the *Grassmann coordinates* (*Plücker coordinates*)$\{p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}\}$ in lexicographical order. They originate from the dual exterior form $*\alpha_m = \beta_{n-m}$ where α_m is the original m-exterior form associated with the matrix $\mathbf{A}$.

$$n = 4, n - m = 2$$

$$\begin{aligned}
\alpha_2 &:= \frac{1}{2!} \sum_{i_1,i_2=1}^{n=4} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} a_{i_1} a_{i_2} \\
&= \frac{1}{2!}\mathbf{e}_1 \wedge \mathbf{e}_2 (a_{11}a_{22} - a_{21}a_{12}) + \frac{1}{2!}\mathbf{e}_1 \wedge \mathbf{e}_3 (a_{11}a_{32} - a_{31}a_{22}) \\
&\quad + \frac{1}{2!}\mathbf{e}_1 \wedge \mathbf{e}_4 (a_{11}a_{42} - a_{41}a_{12}) + \frac{1}{2!}\mathbf{e}_2 \wedge \mathbf{e}_3 (a_{21}a_{32} - a_{31}a_{22}) \\
&\quad + \frac{1}{2!}\mathbf{e}_2 \wedge \mathbf{e}_4 (a_{21}a_{42} - a_{41}a_{22}) + \frac{1}{2!}\mathbf{e}_3 \wedge \mathbf{e}_4 (a_{31}a_{42} - a_{41}a_{32})
\end{aligned}$$

$$\beta := *\alpha_2(\mathbf{R}^4) = \frac{1}{4} \sum_{i_1,i_2,j_1,j_2=1}^{n=4} \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \varepsilon_{i_1 i_2 j_1 j_2} a_{i_1 1} a_{i_1 2}$$

$$= \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_4 p_{12} + \frac{1}{4}\mathbf{e}_2 \wedge \mathbf{e}_4 p_{13} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_2 p_{14}$$
$$+ \frac{1}{4}\mathbf{e}_4 \wedge \mathbf{e}_1 p_{23} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_1 p_{24} + \frac{1}{4}\mathbf{e}_1 \wedge \mathbf{e}_2 p_{34}.$$

The *Grassmann coordinates* (*Plücker coordinates*) $\{p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}\}$ refer to the *basis*

$$\{\mathbf{e}_3 \wedge \mathbf{e}_4, \mathbf{e}_2 \wedge \mathbf{e}_4, \mathbf{e}_3 \wedge \mathbf{e}_2, \mathbf{e}_4 \wedge \mathbf{e}_1, \mathbf{e}_3 \wedge \mathbf{e}_1, \mathbf{e}_1 \wedge \mathbf{e}_2\}.$$

Indeed the *Grassmann space* $\mathbf{G}^{2,4}$ spanned by such a basis can be alternatively covered by the chart generated by the column vectors of the matrix $\mathbf{B}$,

$$\gamma_2 := \sum_{j_1, j_2}^{n=4} \mathbf{e}_{j_1}\mathbf{e}_{j_2}\mathbf{b}_{j_1}\mathbf{b}_{j_2} \in \text{GRASSMANN(B)},$$

a result which is independent of the normalisation of $\{\mathbf{b}_{j_1 1}, \mathbf{b}_{j_2 2}\}$.

As a summary of the result of the two examples (i) $\mathbf{A} \in \mathbb{R}^{3\times 2}$ and (ii) $\mathbf{A} \in \mathbb{R}^{4\times 2}$ for a general rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is needed.

"The matrix B constitutes the *latent restrictions* also called *latent condition equations*. The column space $\mathcal{R}(\mathbf{B})$ of the matrix $\mathbf{B}$ coincides with complementary column space $\mathcal{R}(\mathbf{A})^\perp$ orthogonal to column space $\mathcal{R}(\mathbf{A})$ of the matrix $\mathbf{A}$. The elements of the matrix $\mathbf{B}$ are the *Grassmann coordinates*, also called *Plücker coordinates*, special sub determinants of the matrix $\mathbf{A} = [a_{i1}, \ldots, a_{im}]$

$$p_{j_1 j_2} := \sum_{i_1,\ldots,i_m=1}^{n} \varepsilon_{i_1 \cdots i_m j_1 \cdots j_{n-m}} a_{i_1 1} \cdots a_{i_m m}.$$

The *latent restrictions* control the parameter adjustment in the sense of *identifying outliers or blunders* in observational data."

3-34 From B to A: Latent Parametric Equations, Dual Grassmann *Coordinates, Dual* Plücker *Coordinates*

While in the previous paragraph we started from a given matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, $n > m$, $\text{rk}\mathbf{A} = m$ representing a special inconsistent systems of linear equations $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$, namely in order to construct the orthogonal complement $\mathcal{R}(\mathbf{A})^\perp$ of $\mathcal{R}(\mathbf{A})$, we now reverse the problem. Let us assume that a matrix $\mathbf{B} \in \mathbb{R}^{l\times n}$, $\ell < n$, $\text{rk}\mathbf{B} = \ell$ is given which represents a special inconsistent system of linear homogeneous *condition equations* $\mathbf{B}'\mathbf{y} = \mathbf{B}'\mathbf{i}$. How can we construct the orthogonal complement $\mathcal{R}(\mathbf{A})^\perp$ of $\mathcal{R}(\mathbf{B})$ and how can we relate the elements of $\mathcal{R}(\mathbf{B})^\perp$ to the matrix $\mathbf{A}$ of parametric adjustment?

First, let us depart from the orthogonality condition $\mathbf{B}'\mathbf{A} = 0$ *or* $\mathbf{A}'\mathbf{B} = 0$ we already introduced and discussed at length. Such an orthogonality condition had been the result of the orthogonality of the vectors $\mathbf{y}_\ell = \hat{\mathbf{y}}(LESS)$ and $\mathbf{i}_\ell = \hat{\mathbf{i}}(LESS)$. We recall the general condition of the homogeneous matrix equation.

$$\mathbf{B}'\mathbf{A} = 0 \;\Leftrightarrow\; \mathbf{A} = [\mathbf{I}_\ell - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}']\mathbf{Z},$$

which is, of course, *not unique* since the matrix $\mathbf{Z} \in \mathbb{R}^{l\times l}$ is left undetermined. Such a result is typical for an orthogonality conditions.

Second, let us construct the *Grassmann space* $\mathbf{G}^{\ell,n}$, in short GRASSMANN (B) as well as the *Grassmann space* $\mathbf{G}^{n-\ell,n}$, in short GRASSMANN (A) representing $\mathcal{R}(\mathbf{B})$ and $\mathcal{R}(\mathbf{B})^\perp$, respectively.

$$\gamma_\ell = \frac{1}{\ell!}\sum_{j_1\cdots j_\ell=1}^{n} \mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_\ell} b_{j_1 1}\cdots b_{j_\ell \ell} \tag{3.138}$$

$$\delta_{n-\ell} := *\gamma_\ell = \frac{1}{(n-\ell)!}\sum_{i_1,\ldots,i_{n-\ell},j_1,\ldots,j_\ell=1}^{n} \frac{1}{\ell!}\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_{n-\ell}} \varepsilon_{i_1\cdots i_{n-\ell} j_1\cdots j_\ell} b_{j_1 1}\cdots b_{j_\ell \ell}.$$

The exterior form γ_ℓ which is built on the column vectors $\{b_{j_1 1}, \ldots, b_{j_\ell \ell}\}$ of the matrix $\mathbf{B} \in \mathbb{R}^{l\times n}$ is an element of the column space $\mathcal{R}(\mathbf{B})$. Its dual exterior form $*\gamma = \delta_{n-\ell}$, in contrast, is an element of the orthogonal complement $\mathcal{R}(\mathbf{B})^\perp$.

$$q_{i_1\cdots i_{n-\ell}} := \varepsilon_{i_1\cdots i_{n-\ell} j_1\cdots j_\ell} b_{j_1 1}\cdots b_{j_\ell \ell} \tag{3.139}$$

denote the *Grassmann coordinates* (*Plücker coordinates*) which are dual to the *Grassmann coordinates* (*Plücker coordinates*) $p_{j_1\cdots j_{n-m}}\mathbf{q} := [q_{i_1} \ldots q_{n-\ell}]$ is constituted by subdeterminants of the matrix $\mathbf{B}$, while $\mathbf{p}{:=}[p_{j_1} \ldots p_{n-m}]$ by subdeterminants of the matrix $\mathbf{A}$.

The $(\alpha, \beta, \gamma, \delta)$-diagram of *Fig. 3.9* is commutative. If $\mathcal{R}(\mathbf{B}) = \mathcal{R}(\mathbf{A})^\perp$, then $\mathcal{R}(\mathbf{B})^\perp = \mathcal{R}(\mathbf{A})$. Identify $\ell = n - m$ in order to convince yourself about the $(\alpha, \beta, \gamma, \delta)$- diagram to be commutative.

$$\alpha_m \to *\alpha_m = \beta_{n-m} = \gamma_{n-m} \to *\gamma_{n-m} = *\beta_{n-m} = **\alpha_m = (-1)^{m(n-m)}\alpha_m$$

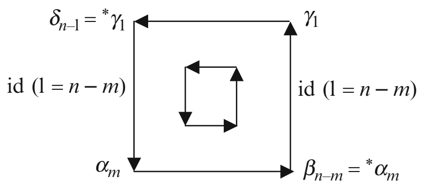


Fig. 3.9 Commutative diagram

Third, let us specialize $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{B})^{\perp}$ and $\mathcal{R}(\mathbf{A})^{\perp} = \mathcal{R}(\mathbf{B})$ by $\ell = n - m$.

$$\alpha_m \to *\alpha_m = \beta_{n-m} = \gamma_{n-m} \to *\gamma_{n-m} = *\beta_{n-m} = **\alpha_m = (-1)^{m(n-m)}\alpha_m \tag{3.140}$$

The *first* and *second example* will be our candidates for test computations of the diagram of *Fig. 3.9* to be commutative. *Box 3.26* reviews *direct* and *inverse Grassmann coordinates* (*Plücker coordinates*) for $\mathbf{A} \in \mathbb{R}^{3\times 2}$, $\mathbf{B} \in \mathbb{R}^{3\times 1}$, *Box 3.27* for $\mathbf{A} \in \mathbb{R}^{4\times 2}$, $\mathbf{B} \in \mathbb{R}^{4\times 1}$.

Box 3.25. (Direct and inverse *Grassmann coordinates* (*Plücker coordinates*))

first example

The forward computation

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 10 \end{bmatrix} \in \mathbb{R}^{3\times 2} : \mathbf{a}_1 = \sum_{i_1=1}^{n=3} \mathbf{e}_{i_1} a_{i_1 1} \ \ and \ \ a_2 = \sum_{i_2=1}^{n=3} \mathbf{e}_{i_2} a_{i_2 2}$$

$$\alpha_2 := \sum_{i_1,i_2=1}^{n=3} \frac{1}{2!}\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(\mathbb{R}^3) \subset \mathbf{\Lambda}^m(\mathbb{R}^n)$$

$$\beta_1 := *\alpha_2 := \sum_{i_1,i_2,j_1=1}^{n=3} \frac{1}{2!}\mathbf{e}_{j_1} \varepsilon_{i_1 i_2 j_1} a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(\mathbb{R}^3) \subset \mathbf{\Lambda}^m(\mathbb{R}^n)$$

Grassmann coordinates (*Plücker coordinates*)

$$\beta_1 = \frac{1}{2}\mathbf{e}_1 p_{23} + \frac{1}{2}\mathbf{e}_2 p_{31} + \frac{1}{2}\mathbf{e}_3 p_{12}$$

$$p_{23} = a_{21}a_{32} - a_{31}a_{22},\ p_{31} = a_{31}a_{12} - a_{11}a_{32},$$

$$p_{12} = a_{11}a_{22} - a_{21}a_{12},\ p_{23} = 8,\ p_{31} = -9,\ p_{12} = 1$$

The backward computation

$$\gamma_1 := \sum_{i_1,i_2,j_1=1}^{n=3} \frac{1}{1!}\mathbf{e}_{j_1} \varepsilon_{i_1 i_2 j_1} a_{i_1 1} a_{i_2 2} = \mathbf{e}_1 p_{23} + \mathbf{e}_2 p_{31} + \mathbf{e}_3 p_{12} \in \mathbf{\Lambda}^1(\mathbb{R}^3)$$

$$\delta_2 := *\gamma_1 := \frac{1}{2!} \sum_{i_1,i_2,j_1,j_2,j_3=1}^{n=3} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \varepsilon_{i_1 i_2 j_1} \varepsilon_{j_2 j_3 j_1} a_{j_2 1} a_{j_3 2} \ \Rightarrow$$

$$\delta_2 = \frac{1}{2!} \sum_{i_1,i_2,j_1,j_2,j_3=1}^{n=3} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} (\delta_{i_1 j_2}\delta_{i_2 j_3} - \delta_{i_1 j_3}\delta_{i_2 j_2}) a_{j_2 1} a_{j_3 2}$$

$$\delta_2 = \frac{1}{2} \sum_{i_1, i_2 = 1}^{n=3} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} a_{i_1 1} a_{i_2 2} = \alpha_2 \in \mathbf{\Lambda}^2(\mathbb{R}^3) \subset \mathbf{\Lambda}^m(\mathbb{R}^n)$$

inverse Grassmann coordinates
(dual *Grassmann* coordinates, dual *Plücker* coordinates)

$$\delta_2 = \alpha_2 = \frac{1}{2}\mathbf{e}_2 \wedge \mathbf{e}_3 (a_{21}a_{32} - a_{31}a_{22}) + \frac{1}{2}\mathbf{e}_3 \wedge \mathbf{e}_1 (a_{31}a_{12} - a_{11}a_{32})$$

$$+ \frac{1}{2}\mathbf{e}_1 \wedge \mathbf{e}_2 (a_{11}a_{22} - a_{21}a_{12})$$

$$\delta_2 = \alpha_2 = \frac{1}{2}\mathbf{e}_2 \wedge \mathbf{e}_3 q_{23} + \frac{1}{2}\mathbf{e}_2 \wedge \mathbf{e}_3 q_{31} + \frac{1}{2}\mathbf{e}_2 \wedge \mathbf{e}_3 q_{12} \in \mathbf{\Lambda}^2(\mathbb{R}^3).$$

Box 3.26. (Direct and inverse *Grassmann coordinates* (*Plücker coordinates*))

second example
The forward computation

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 10 \end{bmatrix} \in \mathbb{R}^{4\times 2} : \mathbf{a}_1 = \sum_{i_1=1}^{n=4} \mathbf{e}_{i_1} a_{i_1 1} \ \ and \ \ \mathbf{a}_2 = \sum_{i_2=1}^{n=4} \mathbf{e}_{i_2} a_{i_2 2}$$

$$\alpha_2 := \sum_{i_1, i_2 = 1}^{n=4} \frac{1}{2!} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(\mathbb{R}^4) \subset \mathbf{\Lambda}^m(\mathbb{R}^n)$$

$$\beta_2 := *\alpha_2 := \frac{1}{2!} \sum_{i_1, i_2, j_1, j_2 = 1}^{n=4} \frac{1}{2!} \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \varepsilon_{i_1 i_2 j_1 j_2} a_{i_1 1} a_{i_2 2} \in \mathbf{\Lambda}^2(\mathbb{R}^4) \subset \mathbf{\Lambda}^{n-m}(\mathbb{R}^n)$$

$$\beta_2 = \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_4 p_{12} + \frac{1}{4}\mathbf{e}_2 \wedge \mathbf{e}_4 p_{13} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_2 p_{14} +$$
$$+ \frac{1}{4}\mathbf{e}_4 \wedge \mathbf{e}_1 p_{23} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_1 p_{24} + \frac{1}{4}\mathbf{e}_1 \wedge \mathbf{e}_2 p_{34}$$

$$p_{12} = 1, \ p_{13} = 2, \ p_{14} = 9, \ p_{23} = 1, \ p_{34} = 7$$

The backward computation

$$\gamma_2 := \frac{1}{2!} \sum_{i_1, i_2, j_1, j_2 = 1}^{n=4} \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \varepsilon_{i_1 i_2 j_1 j_2} a_{i_1 1} a_{i_2 2} \subset \mathbf{\Lambda}^2(\mathbb{R}^4) \subset \mathbf{\Lambda}^{n-m}(\mathbb{R}^n)$$

$$\delta_2 := *\gamma_2 := \frac{1}{2!} \sum_{i_1, i_2, j_1, j_2, j_3, j_4 = 1}^{n=4} \frac{1}{2!} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \varepsilon_{i_1 i_2 j_1 j_2} \varepsilon_{j_1 j_2 j_3 j_4} a_{j_3 1} a_{j_4 2} =$$

$$= \alpha_2 \in \mathbf{\Lambda}^2(\mathbb{R}^4) \subset \mathbf{\Lambda}^m(\mathbb{R}^n)$$

$$\delta_2 = \alpha_2 = \frac{1}{4} \sum_{i_1, i_2 = 1}^{n=4} \mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} a_{i_1 1} a_{i_2 2}$$

$$\delta_2 = \alpha_2 = \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_4 q_{12} + \frac{1}{4}\mathbf{e}_2 \wedge \mathbf{e}_4 q_{13} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_2 q_{14} +$$
$$+ \frac{1}{4}\mathbf{e}_4 \wedge \mathbf{e}_1 q_{23} + \frac{1}{4}\mathbf{e}_3 \wedge \mathbf{e}_1 q_{24} + \frac{1}{4}\mathbf{e}_1 \wedge \mathbf{e}_2 q_{34}$$

$$q_{12} = p_{12},\ q_{13} = p_{13},\ q_{14} = p_{14},\ q_{23} = p_{23},\ q_{24} = p_{24},\ q_{34} = p_{34}.$$

3-35 Break Points

Throughout the analysis of *high leverage points* and *outliers* within the observational data we did assume a fixed linear model. In reality such an assumption does not apply. The functional model may change with time as *Fig.* 3.9 indicates. Indeed we have to *break-up* the linear model into *pieces*. Break points have to be introduced as those points when the linear model changes. Of course, *a hypothesis test* has to decide whether the break point exists with a certain prob-ability. Here we only highlight the notion of break points in the context of leverage points. For localizing break points we apply the *Gauss–Jacobi Combinatorial Algorithm* following *J.L. Awange and E. W. Grafarend 2005* (2002), *J. L. Awange* (2002), *A. T. Hornoch* (1950), *S. Wellisch* (1910).
Table 3.1 summarises a set of observations y_i with $n = 10$ elements. Those measurements have been taken at time instants $\{t_1, \ldots, t_{10}\}$. *Figure* 3.9 illustrates the graph of the corresponding function $y(t)$. By means of the celebrated *Gauss–Jacobi Combinatorial Algorithm* we aim at localizing break points. *First*, outlined in *Box* 3.28 we determine all the combinations of two points which allow the fit

Table 3.1 Test " break points" observations for a piecewise linear model

	y	t	
y_1	1	1	t_1
y_2	2	2	t_2
y_3	2	3	t_3
y_4	3	4	t_4
y_5	2	5	t_5
y_6	1	6	t_6
y_7	0.5	7	t_7
y_8	2	8	t_8
y_9	4	9	t_9
y_{10}	4.5	10	t_{10}

Table 3.2

combination	Parameters x_1	x_2	matrix of the metric $(\text{vech } G_x)'$
(1, 2)	0.00000	1.00000	[2, 3, 5]
(1, 3)	0.50000	0.50000	[2, 4, 10]
(1, 4)	0.33333	0.66667	[2, 5, 17]
(1, 5)	0.75000	0.25000	[2, 6, 26]
(1, 6)	1.00000	0.00000	[2, 7, 37]
(1, 7)	1.08333	-0.08333	[2, 8, 50]
(1, 8)	0.85714	0.14286	[2, 9, 65]
(1, 9)	0.62500	0.37500	[2, 10, 82]
(1,10)	0.61111	0.38889	[2, 11, 101]
(2, 3)	2.00000	0.00000	[2, 5, 13]
(2, 4)	1.00000	0.50000	[2, 6, 20]
(2, 5)	2.00000	0.00000	[2, 7, 29]
(2, 6)	2.50000	-0.25000	[2, 8, 40]
(2, 7)	2.60000	-0.30000	[2, 9, 53]
(2, 8)	2.00000	0.00000	[2, 10, 68]
(2, 9)	1.42857	0.28571	[2, 11, 85]
(2,10)	1.37500	0.31250	[2, 12, 104]
(3, 4)	-1.00000	1.00000	[2, 7, 25]
(3, 5)	2.00000	0.00000	[2, 8, 34]
(3, 6)	3.00000	-0.33333	[2, 9, 45]
(3, 7)	3.12500	-0.37500	[2, 10, 58]
(3, 8)	2.00000	0.00000	[2, 11, 73]
(3, 9)	1.00000	0.33333	[2, 12, 90]
(3,10)	0.92857	0.35714	[2, 13, 109]
(4, 5)	7.00000	-1.00000	[2, 9, 41]
(4, 6)	7.00000	-1.00000	[2, 10, 52]
(4, 7)	6.33333	-0.83333	[2, 11, 65]
(4, 8)	4.00000	-0.25000	[2, 12, 80]
(4, 9)	2.20000	0.20000	[2, 13, 97]
(4,10)	2.00000	0.25000	[2, 14, 116]
(5, 6)	7.00000	-1.00000	[2, 11, 61]
(5, 7)	5.75000	-0.75000	[2, 12, 74]
(5, 8)	2.00000	0.00000	[2, 13, 89]
(5, 9)	-0.50000	0.50000	[2, 14, 106]
(5,10)	-0.50000	0.50000	[2, 15, 125]
(6, 7)	4.00000	-0.50000	[2, 13, 85]
(6, 8)	-2.00000	0.50000	[2, 14, 100]
(6, 9)	-5.00000	1.00000	[2, 15, 117]
(6,10)	-4.25000	0.87500	[2, 16, 136]
(7, 8)	-10.00000	1.50000	[2, 15, 113]
(7, 9)	-11.75000	1.75000	[2, 16, 130]
(7,10)	-8.83333	1.33333	[2, 17, 149]
(8, 9)	-14.00000	2.00000	[2, 17, 145]
(8,10)	-8.00000	1.25000	[2, 18, 164]
(9,10)	-0.50000	0.50000	[2, 19, 181]

of a straight line without any approximation error. As a determined linear model $\mathbf{y} = \mathbf{A}\mathbf{x}$, $\mathbf{A} \in \mathbb{R}^{2\times 2}$, $rk\mathbf{A} = 2$ namely $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ we calculate (3.142) x_1 and (3.143) x_2 in a closed form. For instance, the pair of observations (y_1, y_2), in short $(1, 2)$ at $(t_1, t_2) = (1, 2)$ determines $(x_1, x_2) = (0, 1)$. Alternatively, the pair of observations (y_1, y_3), in short $(1, 3)$, at $(t_1, t_3) = (1, 3)$ leads us to $(x_1, x_2) = (0.5, 0.5)$. *Table 3.2* contains the possible 45 combinations which determine (x_1, x_2) from $(y_1, \ldots, y_{10})$. These solutions are plotted in Figure 3.10.

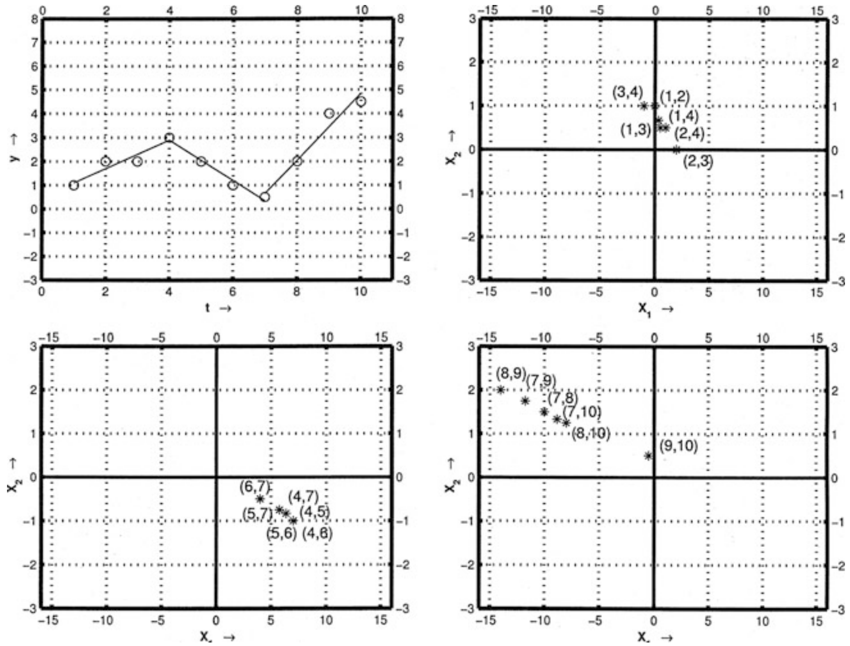


Fig. 3.10 *Top left*: Graph of the function y(t), two break points. *Top right*: Gauss–Jacobi Combinatorial Algorithm, piecewise linear model, 1st cluster: (t_i, t_j). *Bottom left*: Gauss–Jacobi Combinatorial Algorithm, 2nd cluster: (t_i, t_j). *Bottom right*: Gauss–Jacobi Combinatorial Algorithm, 3rd cluster: (t_i, t_j)

Box 3.27. (Piecewise linear model Gauss–Jacobi combinatorial algorithm)

1st step

$$\mathbf{y} = \begin{bmatrix} y(t_i) \\ y(t_j) \end{bmatrix} = \begin{bmatrix} 1 & t_i \\ 1 & t_j \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{Ax} \quad \forall i < j \in \{1, \ldots, n\} \tag{3.141}$$

$$\mathbf{y} \in \mathbb{R}^2,\ \mathbf{A} \in \mathbb{R}^{2\times 2};\ rk\mathbf{A} = 2,\ \mathbf{x} \in \mathbb{R}^2$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \Leftrightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{t_j - t_i} \begin{bmatrix} t_j & -t_i \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y(t_i) \\ y(t_j) \end{bmatrix} \tag{3.142}$$

$$x_1 = \frac{t_j y_1 - t_i y_2}{t_j - t_i} \quad \textit{and} \quad x_2 = \frac{y_j - y_i}{t_j - t_i} \tag{3.143}$$

$$\begin{aligned} & t_i = t_1 = 1,\ t_j = t_2 = 2 \\ \textit{Example:}\ & y(t_1) = y_1 = 1,\ y(t_2) = y_2 = 2 \\ & x_1 = 0,\ x_2 = 1. \end{aligned}$$

$$\textit{Example:}\quad \begin{array}{l} t_i = t_1 = 1,\ t_j = t_3 = 3 \\ y(t_1) = y_1 = 1,\ y(t_3) = y_3 = 2 \\ x_1 = 0.5 \quad \textit{and} \quad x_2 = 0.5\ . \end{array}$$

Second, we introduce the *pullback operation* $\mathbf{G_y} \to \mathbf{G_x}$. The matrix of the metric $\mathbf{G_y}$ of the observation space $\mathbb{Y}$ is pulled back to generate by (3.144) the matrix of the metric $\mathbf{G_x}$ of the parameter space $\mathbb{X}$ for the "*determined linear model*" $\mathbf{y} = \mathbf{Ax}$, $\mathbf{A} \in \mathbb{R}^{2\times 2}$, $\mathrm{rk}\mathbf{A} = 2$, namely $\mathbf{G_x} = \mathbf{A}'\mathbf{G_y}\mathbf{A}$. If the observation space $\mathbb{Y} = \mathrm{span}\{\mathbf{e}_1^y, \mathbf{e}_2^y\}$ is spanned by two orthonormal vectors $\mathbf{e}_1^y, \mathbf{e}_2^y$ relating to a pair of observations $(y_i, y_j), i < j, i, j \in \{1, \ldots, 10\}$, then the matrix of the metric $\mathbf{G_y} = \mathbf{I}_2$ of the observation space is the unit matrix. In such an experimental situation (3.145) $\mathbf{G_x} = \mathbf{A}'\mathbf{A}$ is derived. For the *first example* $(t_i, t_j) = (1, 2)$ we are led to vech $\mathbf{G_x} = [2, 3, 5]'$. "*Vech half*" shortens the matrix of the metric $\mathbf{G_x} \in \mathbb{R}^{2\times 2}$ of the *parameter space* $\mathbb{X} \ni (x_1, x_2)$ by stacking the columns of the *lower triangle* of the symmetric matrix $\mathbf{G_x}$. Similarly, for the second example $(t_i, t_j) = (1, 3)$ we produce vech $\mathbf{G_x} = [2, 4, 10]'$. For all the 45 combinations of observations (y_i, y_j).

In the last column *Table 3.2* contains the necessary information of the matrix of the metric $\mathbf{G_x}$ of the parameter space $\mathbb{X}$ formed by $(\mathrm{vech}\,\mathbf{G_x})'$. Indeed, such a notation is quite economical.

Box 3.28. (Piecewise linear model: *Gauss–Jacobi* combinatorial algorithm)

2nd step

pullback of the matrix $\mathbf{G}_X$ the metric from $\mathbf{G}_y$

$$\mathbf{G_x} = \mathbf{A}'\mathbf{G_y}\mathbf{A} \tag{3.144}$$

"if $\mathbf{G_y} = \mathbf{I}_2$, then $\mathbf{G_x} = \mathbf{A}'\mathbf{A}$"

$$\mathbf{G_x} = \mathbf{A}'\mathbf{A} = \begin{bmatrix} 2 & t_i + t_j \\ t_i + t_j & t_i^2 + t_j^2 \end{bmatrix} \quad \forall\ i < j \in \{1, \ldots, n\}. \tag{3.145}$$

Example: $t_i = t_1 = 1, t_j = t_2 = 2$

$$\mathbf{G_x} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}, \mathrm{vech}\mathbf{G_x} = [2, 3, 5]'.$$

Example: $t_i = t_1 = 1, t_j = t_3 = 3$

$$\mathbf{G_x} = \begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}, \mathrm{vech}\mathbf{G_x} = [2, 4, 10]'.$$

Third, we are left the problem to identify the break points. *C.F. Gauss* (1828) and *C.G.J. Jacobi* (1841) have proposed to take the *weighted arithmetic mean* of the combinatorial solutions $(x1, x2)^{(1,2)}, (x1, x2)^{(1,3)}$, in general $(x1, x2)^{(i,j)}, i < j$, are considered as

Pseudo-observations.

Box 3.29. (Piecewise linear model: Gauss–Jacobi combinatorial algorithm)

3rd step

pseudo-observations

Example

$$\begin{bmatrix} x_1^{(1,2)} \\ x_2^{(1,2)} \\ \hline x_1^{(1,3)} \\ x_2^{(1,3)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \hline 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \end{bmatrix} \quad \in \mathbb{R}^{4\times 1} \tag{3.146}$$

$\mathbf{G}_x$-LESS

$$\mathbf{x}_\ell := \hat{\mathbf{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \left[\mathbf{G}_x^{(1,2)} + \mathbf{G}_x^{(1,3)}\right]^{-1} \left[\mathbf{G}_x^{(1,2)}, \mathbf{G}_x^{(1,3)}\right] \begin{bmatrix} x_1^{(1,2)} \\ x_2^{(1,2)} \\ x_1^{(1,3)} \\ x_2^{(1,3)} \end{bmatrix} \quad \in \mathbb{R}^{2\times 1} \tag{3.147}$$

$$\mathrm{vech}\mathbf{G}_x^{(1,2)} = [2, 3, 5]', \mathrm{vech}\mathbf{G}_x^{(1,3)} = [2, 4, 10]'$$

$$\mathbf{G}_x^{(1,2)} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}, \mathbf{G}_x^{(1,3)} = \begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix}$$

$$\begin{bmatrix} x_1^{(1,2)} \\ x_2^{(1,2)} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\begin{bmatrix} x_1^{(1,3)} \\ x_2^{(1,3)} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$\mathbf{G}_x^{-1} = \begin{bmatrix} 4 & 7 \\ 7 & 15 \end{bmatrix}^{-1} = \frac{1}{11} \begin{bmatrix} 15 & -7 \\ -7 & 4 \end{bmatrix} = [\mathbf{G}_x^{(1,2)} + \mathbf{G}_x^{(1,3)}]^{-1}$$

$$\mathbf{x}_\ell := \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} 6 \\ 6 \end{bmatrix} \Leftrightarrow \hat{x}_1 = \hat{x}_2 = \frac{6}{11} = 0.545, 454.$$

Box 3.30 provides us with an example. For establishing the *third step* of the *Gauss–Jacobi Combinatorial Algorithm*. We outline $\mathbf{G_X}$-LESS for the set of *pseudo-observations* (3.146) $(x_1, x_2)^{(1,2)}$ and $(x_1, x_2)^{(1,3)}$ solved by (3.177) $x_\ell = (\hat{x}_1, \hat{x}_2)$.

The matrices $\mathbf{G}_x^{(1,2)}$ and $\mathbf{G}_x^{(1,3)}$ representing the metric of the parameter space $\mathbb{X}$ derived from $(x_1, x_2)^{(1,2)}$ and $(x_1, x_2)^{(1,3)}$ are additively composed and inverted, a result which is motivated by the special design matrix $\mathbf{A} = [\mathbf{I}_2, \mathbf{I}_2]'$ of "direct" *pseudo-observations*. As soon as we implement the weight matrices $\mathbf{G}_x^{(1,2)}$ and $\mathbf{G}_x^{(1,3)}$ from *Table 3.2* as well as $(x_1, x_2)^{(1,2)}$ and $(x_1, x_2)^{(1,3)}$ we are led to the weighted arithmetic mean $\hat{x}_1 = \hat{x}_2 = 6/11$. Such a result has to be compared with the componentwise median x_1 (median) = 1/4 and $x_2(median) = 3/4$.

$$\left[\begin{array}{ll} (1,2),\ (1,3) & (1,2),\ (1,3) \\ \textit{combination} & \textit{combination} \\ \mathbf{G}_x\text{-LESS} & \textit{median} \\ \hat{x}_1 = 0.545,454 & x_1(\textit{median}) = 0.250 \\ \hat{x}_2 = 0.545,454 & x_2(\textit{median}) = 0.750\,. \end{array}\right.$$

Here, the arithmetic mean of $x_1^{(1,2)}, x_1^{(1,3)}$ and $x_2^{(1,2)}, x_2^{(1,3)}$ coincides with the median neglecting the *weight* of the pseudo-observations.

Box 3.30. (Piecewise linear models and two break points: "Example")

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{t}_{n_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1}_{n_2} & \mathbf{t}_{n_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1}_{n_3} & \mathbf{t}_{n_3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} i_{y_1} \\ i_{y_2} \\ i_{y_3} \end{bmatrix} \tag{3.148}$$

I_{n_1}-*LESS*, I_{n_2}-*LESS*, I_{n_3}-*LESS*,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_\ell = \frac{\mathbf{1}}{n_1 \mathbf{t}'_{n_1}\mathbf{t}_{n_1} - (\mathbf{1}'_{n_1}\mathbf{t}_{n_1})^2} \begin{bmatrix} (\mathbf{t}'_{n_1}\mathbf{t}_{n_1})(\mathbf{1}'_{n_1}\mathbf{y}_{n_1}) - (\mathbf{1}'_{n_1}\mathbf{t}_{n_1})(\mathbf{t}'_{n_1}\mathbf{y}_{n_1}) \\ -(\mathbf{1}'_{n_1}\mathbf{t}_{n_1})(\mathbf{1}'_{n_1}\mathbf{y}_1) + n_1\mathbf{t}'_{n_1}\mathbf{y}_1 \end{bmatrix} \tag{3.149}$$

$$\begin{bmatrix} x_3 \\ x_4 \end{bmatrix}_\ell = \frac{1}{n_2 \mathbf{t}'_{n_2}\mathbf{t}_{n_2} - (\mathbf{1}'_{n_2}\mathbf{t}_{n_2})^2} \begin{bmatrix} (\mathbf{t}'_{n_2}\mathbf{t}_{n_2})(\mathbf{1}'_{n_2}\mathbf{y}_{n_2}) - (\mathbf{1}'_{n_2}\mathbf{t}_{n_2})(\mathbf{t}'_{n_2}\mathbf{y}_{n_2}) \\ -(\mathbf{1}'_{n_2}\mathbf{t}_{n_2})(\mathbf{1}'_{n_2}\mathbf{y}_2) + n_2\mathbf{t}'_{n_2}\mathbf{y}_2 \end{bmatrix} \tag{3.150}$$

$$\begin{bmatrix} x_5 \\ x_6 \end{bmatrix}_\ell = \frac{1}{n_3 \mathbf{t}'_{n_3}\mathbf{t}_{n_3} - (\mathbf{1}'_{n_3}\mathbf{t}_{n_3})^2} \begin{bmatrix} (\mathbf{t}'_{n_3}\mathbf{t}_{n_3})(\mathbf{1}'_{n_3}\mathbf{y}_{n_3}) - (\mathbf{1}'_{n_3}\mathbf{t}_{n_3})(\mathbf{t}'_{n_3}\mathbf{y}_{n_3}) \\ -(\mathbf{1}'_{n_3}\mathbf{t}_{n_3})(\mathbf{1}'_{n_3}\mathbf{y}_3) + n_3\mathbf{t}'_{n_3}\mathbf{y}_3 \end{bmatrix}. \tag{3.151}$$

Box 3.31. (Piecewise linear models and two break points: "Example")

1st interval: n = 4, m = 2, $t \in \{t_1, t_2, t_3, t_4\}$

$$\mathbf{y}_1 = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ 1 & t_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathbf{i}_{y_1} = \mathbf{1}_n x_1 + \mathbf{t}_n x_2 + \mathbf{i}_{y_1} \tag{3.152}$$

2nd interval: n = 4, m = 2, $t \in \{t_4, t_5, t_6, t_7\}$

$$\mathbf{y}_2 = \begin{bmatrix} 1 & t_4 \\ 1 & t_5 \\ 1 & t_6 \\ 1 & t_7 \end{bmatrix} \begin{bmatrix} \mathrm{x}_3 \\ x_4 \end{bmatrix} + \mathbf{i}_{y_2} = \mathbf{1}_n x_3 + \mathbf{t}_n x_4 + \mathbf{i}_{y_2} \tag{3.153}$$

3rd interval: n = 4, m = 2, $t \in \{t_7, t_8, t_9, t_{10}\}$

$$\mathbf{y}_3 = \begin{bmatrix} 1 & t_7 \\ 1 & t_8 \\ 1 & t_9 \\ 1 & t_{10} \end{bmatrix} \begin{bmatrix} x_5 \\ x_6 \end{bmatrix} + \mathbf{i}_{y_3} = \mathbf{1}_n x_5 + \mathbf{t}_n x_6 + \mathbf{i}_{y_3}. \tag{3.154}$$

Figure 3.10 have illustrated the three clusters of combinatorial solutions referring to the first, second and third straight line. Outlined in *Box 3.31* and *Box 3.31*, namely by (3.148) to (3.151), $n_1 = n_2 = n_3 = 4$, we have computed $(x_1, x_2)_\ell$ for the first segment, $(x_3, x_4)_\ell$ for the second segment and $(x_5, x_6)_\ell$ for the third segment of the least squares fit of the straight line. *Table 3.3* contains the results explicitly. Similarly, by means of the *Gauss–Jacobi Combinatorial Algorithm* of *Table 3.4* we have computed the identical solution $(x_1, x_2)_\ell$, $(x_3, x_4)_\ell$ and $(x_5, x_6)_\ell$ as "*weighted arithmetic means*" numerically presented only for the *first segment*.

Table 3.3 (I_n-LESS solutions for those segments of a straight line, *two break points*)

$$\mathbf{I}_4\text{-LESS}: \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_\ell = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}: \; y(t) = 0.5 + 0.6t$$

$$\mathbf{I}_4\text{-LESS}: \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}_\ell = \frac{1}{20}\begin{bmatrix} 126 \\ -17 \end{bmatrix}: \; y(t) = 6.30 - 0.85t$$

$$\mathbf{I}_4\text{-LESS}: \begin{bmatrix} x_5 \\ x_6 \end{bmatrix}_\ell = \frac{1}{20}\begin{bmatrix} -183 \\ 28 \end{bmatrix}: \; y(t) = -9.15 + 1.40t$$

Table 3.4 (Gauss–Jacobi Combinatorial Algorithm for the first segment of a straight line)

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \left[\mathbf{G}_x^{(1,2)} + \mathbf{G}_x^{(1,3)} + \mathbf{G}_x^{(1,4)} + \mathbf{G}_x^{(2,3)} + \mathbf{G}_x^{(2,4)} + \mathbf{G}_x^{(3,4)}\right]^{-1} *$$

$$\left[\mathbf{G}_x^{(1,2)}, \mathbf{G}_x^{(1,3)}, \mathbf{G}_x^{(1,4)}, \mathbf{G}_x^{(2,3)}, \mathbf{G}_x^{(2,4)}, \mathbf{G}_x^{(3,4)}\right] \begin{bmatrix} x_1^{(1,2)} \\ x_2^{(1,2)} \\ x_1^{(1,3)} \\ x_2^{(1,3)} \\ x_1^{(1,4)} \\ x_2^{(1,4)} \\ x_1^{(2,3)} \\ x_2^{(2,3)} \\ x_1^{(2,4)} \\ x_2^{(2,4)} \\ x_1^{(3,4)} \\ x_2^{(3,4)} \end{bmatrix} \qquad (3.155)$$

$$\left[\mathbf{G}_x^{(1,2)} + \mathbf{G}_x^{(1,3)} + \mathbf{G}_x^{(1,4)} + \mathbf{G}_x^{(2,3)} + \mathbf{G}_x^{(2,4)} + \mathbf{G}_x^{(3,4)}\right]^{-1} = \frac{1}{30}\begin{bmatrix} 15 & -5 \\ -5 & 2 \end{bmatrix}$$

$$\mathbf{G}_x^{(1,2)} \begin{bmatrix} x_1^{(1,2)} \\ x_2^{(1,2)} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$\mathbf{G}_x^{(1,3)} \begin{bmatrix} x_1^{(1,3)} \\ x_2^{(1,3)} \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 10 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$

$$\mathbf{G}_x^{(1,4)} \begin{bmatrix} x_1^{(1,4)} \\ x_2^{(1,4)} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 17 \end{bmatrix} \begin{bmatrix} 0.333 \\ 0.666 \end{bmatrix} = \begin{bmatrix} 4 \\ 13 \end{bmatrix}$$

$$\mathbf{G}_x^{(2,3)} \begin{bmatrix} x_1^{(2,3)} \\ x_2^{(2,3)} \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 10 \end{bmatrix}$$

$$\mathbf{G}_x^{(2,4)} \begin{bmatrix} x_1^{(2,4)} \\ x_2^{(2,4)} \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 5 \\ 16 \end{bmatrix}$$

$$\mathbf{G}_x^{(3,4)} \begin{bmatrix} x_1^{(3,4)} \\ x_2^{(3,4)} \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 7 & 25 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 18 \end{bmatrix}$$

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{30}\begin{bmatrix} 15 & -5 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 24 \\ 69 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.6 \end{bmatrix}.$$

3-4 Special Linear and Nonlinear Models: A Family of Means for Direct Observations

In case of direct observations, LESS of the inconsistent linear model $\mathbf{y} = \mathbf{1}_n\mathbf{x} + \mathbf{i}$ has led to

$$\mathbf{x}_\ell = \arg\{\mathbf{y} = \mathbf{1}_n\mathbf{x} + \mathbf{i} \mid ||\,\mathbf{i}\,||^2 = \min\}, \mathbf{x}_\ell = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \frac{1}{n}(y_1 + \cdots + y_n).$$

Such a mean has been the starting point of many alternatives we present to you by *Table 3.5* based upon *S.R. Wassels* (2002) review.

3-5 A Historical Note on C.F. Gauss and A.M. Legendre

The inventions of Least Squares and its generalization

The historian S.M. Stigler (1999, pp 320, 330–331) made the following comments on the history of *Least Squares*.

> "The method of *least squares* is the automobile of modern statistical analysis: despite its limitations, occasional accidents, and incidental pollution, this method and its numerous variations, extensions, and related conveyances carry the bulk of statistical analyses, and are known and valued by nearly all. But there has been some dispute, historically, as to who is the *Henry Ford* of statistics. *Adrian Marie Legendre* published the method in 1805, an American, *Robert Adrian*, published the method in late 1808 or early 1809, and *Carl Fiedrich Gauss* published the method in 1809. *Legendre* appears to have discovered the method in early 1805, and *Robert Adrain* may have "discovered" it in *Legendre's* 1805 book (Stigler 1977c, 1978c), but in 1809 *Gauss* had the temerity to claim that he had been using the method since 1795, and one of the most famous priority disputes in the history of science was off and running. It is unnecessary to repeat the details of the dispute – *R.L. Plackett* (1972) has done a masterly job of presenting and summarizing the evidence in the case.

Table 3.5 (A family of means)

Name	Formula
arithmetic mean	$x_\ell = \frac{1}{n}(y_1 + \cdots + y_n)$ $n = 2 : x_\ell = \frac{1}{2}(y_1 + y_2)$ $\mathbf{x}_\ell = (\mathbf{1}'\mathbf{G}_y\mathbf{1})^{-1}\mathbf{1}'\mathbf{G}_y\mathbf{y}$ if $\mathbf{G}_y = \mathrm{Diag}(g_1, \ldots, g_1)$
weighted arithmetic mean	$x_\ell = \dfrac{g_1 y_1 + g_n y_n}{g_1 + \cdots + g_n}$

(continued)

Table 3.5 (continued)

Name	Formula
geometric mean	$x_g = \sqrt[n]{y_1 \cdots y_n} = \left(\prod_{i=1}^{n} y_i\right)^{1/n}$ $n = 2 : x_g = \sqrt{y_1 y_2}$
logarithmic mean	$x_{\log} = \frac{1}{n}(\ln y_1 + \cdots + \ln y_n)$ $x_{\log} = \ln x_g$ $y_{(1)} < \cdots < y_{(n)}$ ordered set of observations
median	$med\ y = \left[\begin{array}{c} y_{(k+1)}\ if\ n = 2k+1 \\ \text{“}add\text{''} \\ (y_{(k)} + y_{(k+1)})/2\ if\ n = 2k \\ \text{“}even\text{''} \end{array} \right.$
Wassels family of means	$x_p = \frac{(y_1)^{p+1} + \cdots + (y_n)^{p+1}}{(y_1)^p + \cdots + (y_n)^p}$ $x_p = \sum_{i=1}^{n} (y_i)^{p+1} \Big/ \sum_{i=1}^{n} (y_i)^p$ $Case\ p = 0 : x_p = x_\ell$ $Case\ p = -1/2, n = 2 : x_p = x_\ell$
Hellenic mean	$Case\,p = -1 :$ $n = 2 :$ $H = H(y_1, y_2) = \left(\frac{y_1^{-1} + y_2^{-1}}{2}\right)^{-1} = \frac{2y_1 y_2}{y_1 + y_2}.$

Let us grant, then, that *Gauss's* later accurate were substantially accunate, and that he did device the method of least squares between 1794 and 1799, independently of *Legendre* or any other discoverer. There still remains the question, what importance did he attach to the discovery? Here the answer must be that while *Gauss* himself may have felt the method useful, he was unsuccessful in communicating its importance to other before 1805. He may indeed have mentioned the method to *Olbers, Lindemau*, or *von Zach* before 1805, but in the total lack of applications by others, despite ample opportunity, suggests the message was not understood. The fault may have been more *in the listener than in the teller*, but in this case its failure serves only to enhance our admiration for *Legendre's* 1805 success. For *Legendre's* description of the method had an immediate and widespread effect – as we have seen, it even caught the eye and under-standing of at least one of those astronomers (*Lindemau*) who had been deaf to *Gauss's message*, and perhaps it also had an influence upon the form and emphasis of *Gauss's* exposition of the method.

When Gauss did publish on least squares, *he went far beyond Legendre* in both conceptual and technical development, *linking the method to probability* and providing algorithms for the computation of estimates. His work has been discussed often, including by

> *H.L. Seal* (1967), *L. Eisenhart* (1968), *H. Goldsteine* (1977, §§ 4.9, 4.10), *D.A. Sprott* (1978), O.S. Sheynin (1979, 1993, 1994), *S.M. Stigler* (1986), *J.L. Chabert* (1989), *W.C. Waterhouse* (1990), *G.W. Stewart* (1995), and *J. Dutka* (1996). *Gauss's* development had to wait a long time before finding an appreciative audience, and much was intertwined with other's work, notably *Laplace's*. *Gauss* was the first among mathematicians of the age, but it was *Legendre* who crystallized the idea in a form that caught the mathematical public's eye. Just as the automobile was not the product of one man of genius, so too the *method of least squares* is due to many, including at least two independent *discoverers*. *Gauss* may well have been the first of these, but he was no *Henry Ford of statistics*. If these was any single scientist who first put the method within the reach of the common man, it was *Legendre*."

Indeed, these is not much to be added. *G.W. Stewart* (1995) recently translated the original Gauss text "*Theoria Combinationis Observationum Erroribus Minimis Obmaxial, Pars Prior. Pars Posterior*" from the Latin origin into English. *F. Pukelsheim* (1998) critically reviewed the sources, the reset Latin text and the quality of the translation. Since the English translation appeared in the SIAM series "*Classics in Applied Mathematics*", he concluded: "*Opera Gaussii contra SIAM defensa*".

"*Calculus probilitatis contra La Place defenses*." This is *Gauss's* famous diary entry of 17 June 1798 that he later quoted to defend priority on the *Method of Least Squares* (Werke, Band X, 1, p.533).

C.F. Gauss goes Internet

With the *Internet Address* http://gallica.bnf.fr you may reach the catalogues of digital texts of *Bibliotheque Nationale de France*. Fill the window "Auteur" by "*Carl Friedrich Gauss*" and you reach "*Types de documents*". Continue with "*Touts les documents*" and click "*Rechercher*" where you find 35 documents numbered 1 to 35. In total, 12732 "*Gauss pages*" are available. Only the *Gauss–Gerling correspondence* is missing. The origin of all texts are the resources of the Library of the Ecole Polytechnique. Meanwhile *Gauss's Werke* are also available under http://www.sub.uni-goettingen.de/. A CD-Rom is available from "*Niedersächsische Staats- und Universitätsbibliothek*."

For the early impact of the *Method of Least Squares* on *Geodesy*, namely *W. Jordan*, we refer to the documentary by *S. Nobre* and *M. Teixeira* (2000).

Chapter 4
The Second Problem of Probabilistic Regression

Special Gauss–Markov method without Datum Defect

Setup of BLUUE for the moments of first order and of BIQUUE for the central moment of second order

In probabilistic regression we study the determination of first moments as well as second central moments of type special *Gauss–Markov model without Datum Defect*. Theoretical results and geodetic applications are taken from E. Grafarend's "Variance–covariance component estimation" (Statistical and Decision, Supplement Issue No. 2, pages 401–441, 105 references, Oldenburg Verlag, München 1989). The original contributions of variance–covariance component estimation as well as geodetic examples are presented in "Gewichtsschätzung in Geodätischen Netzen" (weight estimation in geodetic networks) by E. Grafarend and A. d'Hone (Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Series A, Report 88, pages 1–43, 46 references, München 1978) namely introducing (a) quadratic unbiased estimators: QUE (b), invariant quadratic unbiased estimators: IQUE, (c) locally best, quadratic unbiased estimators, (d) locally best, invariant quadratic unbiased estimators, (e) best QUE and best IQUE, (f) quadratic unbiased estimators of minimum norm: MINQUE. The geodetic examples are reproduced here.

The references include G. Dewess (1973), H. Drygas (1972, 1977), H. Ebner (1977), J. Focke and G. Dewess (1972), E. Grafarend (1974, 1978), F. Graybill (1954), G. Hein (1977), F.R. Helmert (1907), C.R. Hendersen (1953), S.D. Horn, R.A. Horn and D.B. Dungan (1975), P.L. Hsu (1938), R. Keln (1978), J. Kleffe (1975 i, ii, 1976, 1977), J. Kleffe and R. Pincus (1974 i,ii), K.R. Koch (1976, 1977), K. Kubik (1970), L.R. Lamotte (1973), H. Neudecker (1969), R. Pincus (1974), F. Pukelsheim (1976), C.R. Rao (1970, 1971 i,ii, 1972), C.R. Rao and S.K. Mirra (1971), J.N.K. Rao (1973), R. Rudolph (1976), M. Schuler (1932), M. Schuler and H. Wolf (1954), R.S. Searle (1971), J. Seely (1970 i,ii, 1971), J. Seely and C. Zyskind (1971), E.C. Townsend and S.R. Searle (1971), W. Welsch (1977) and H. Wolf (1975).

Real problems are nonlinear. We have to divide the solutions space into three classes: (a) strong nonlinear, (b) weak nonlinear and (c) linear. Weak nonlinearity is defined by nonlinear systems which allow a *Taylor series approximation*. Linear system equations, for instance "linear best *uniformly unbiased* estimator" or "best invariant quadratic *uniformly unbiased* estimator," have the advantage that their normal equation have only one minimum solution. We call such a solution "*global*" or "*uniform.*" In contrast to linear estimation, weak nonlinear equations produce

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2_4,

© Springer-Verlag Berlin Heidelberg 2012

"local" solutions. They have the disadvantage that *many local solutions* may exist. A typical example is a geodetic network analysis by *P. Lohse* (1994). For more details read *C.R. Rao* and *J. Kleffe* (1969, pages 161–180).

For *beginners*, not for mathematicians, is our introductory example (see *Tables 4.1 and 4.2*) in *Sect. 4-1*. In Sect. 4-2 we introduce the "Best Linear Uniformly Unbiased estimator" by means of Gauss elimination. Famous is the *Gauss–Markov Equivalence Theorem* ($\mathbf{G}_y$–LESS) $\sim$ (Σ_y–BLUUE), namely *weighted least squares solution* equivalent to Σy – *best linear uniformly unbiased* estimator. For instance, F. Pukelsheim (1990, pages 20–24) is an excellence reference. For the setup of the "*best invariant quadratic uniformly unbiased*" estimation in *Sect. 4-3*, namely by BLOCK PARTITIONING, we use the E-D equivalence presented in Sect. 4-37. Alternative estimators of type IQE, "*Invariant Quadratic Uniformly Unbiased Estimation*" of variance–covariance components of type *IQUUE*, especially *F.R. Helmer's BIQUUE* which was developed as early as 1907 by two methods. He used two setups presented in detail by Sect. 4-35 "*Invariant Quadratic Uniformly Unbiased Estimators*" of variance–covariance components of *Helmert type*: *HIQUUE versus HIQE*. We give credit to *B. Schaffrin* (1983) who classified the Helmert test matrix for the first time correctly. *E. Grafarend* (1994) used variance–covariance component estimation of *Helmert type* for *condition equations* in the model $\mathbf{B}\varepsilon = \mathbf{By}-\mathbf{c}$, $\mathbf{By} \notin \mathbf{R}(\mathbf{A}) + \mathbf{c}$ (E. Grafarend 1994, pages 38–40) as well as for *condition equations with unknown parameters* in the model $\mathbf{Ax} + \mathbf{B}\varepsilon = \mathbf{By}-\mathbf{c}$, $\mathbf{By} \notin \mathbf{R}(\mathbf{A}) + \mathbf{c}$ (E. Grafarend 1994, pages 41–43), geodetic are also given.

The various methods for solving the basic problem of variance–covariance component estimation was the *final in the eighties of the twentieth century*.

Since *F.R. Helmerts invention* of the variance–covariance estimation in the early twentieth century, geodesists have used his results in many contributions listed in Sect. 4-37. The first example is a linear model $\mathbf{E}\{\mathbf{y}\} = \mathbf{Ay} = \mathbf{Ax}_1$, $\mathbf{D}\{\mathbf{y}\} = \mathbf{I}\sigma^2 = \mathbf{I}x_2$ *with one covariance component*. We estimate ($\mathbf{x}_1$ˆ, x_2ˆ) of type *linear and unbiased*. We apply again the *IPM method* in order to find x_1ˆ by (4.176), (4.177) and x2ˆ by (4.82), (4.183). The *second example* is variance *component estimation of Helmert type* resulting in two cases (a) det(H) $\neq 0$ and (b) det(H) = 0 in a two step procedure. The result is given by (4.210)–(4.212).

Special geodetic problems will be discussed in *Examples 3* and 4. E*xample 3* treats a two variance component model for a *hybrid direction – distance network*, the *Huaytapallana network* in Peru illustrated by *Fig. 4.2* and resulted in *Tables 4.1–4.3*. *Figure 4.3* illustrates the two *variance component estimation of Helmert type for the reproducing point* for this network. *Example 4* is based on a *two variance component, multivariate gyrocompass* observation model for a *first order Markov process*: special attention is payed to the *reproducing point*.

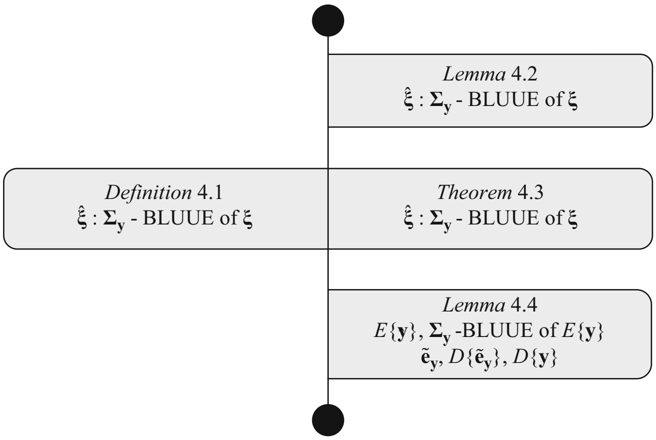
Example 5 is a special treatment of *Bayes design for first and second order statistics*, the moment estimation *using prior information* for the mean parameter vector ξ as well as the model variance σ^2. We have followed *F. Pukelsheim* (1990 pp. 270–272).

Finally we have highlighted the five *postulates* for *inhomogeneous, bilinear estimation*:

1. Inhomogeneous, multilinear
2. Invariance
3. "unbiased" or "minimum bias in the mean"
4. Second order statistics, quasi-Gauss normal distribution
5. Best estimation of the first and central second order moments

In particular, we emphasize the "*E-D-correspondance*".

Fast track reading : Read only *Theorem 4.3* and *Theorem 4.13*.



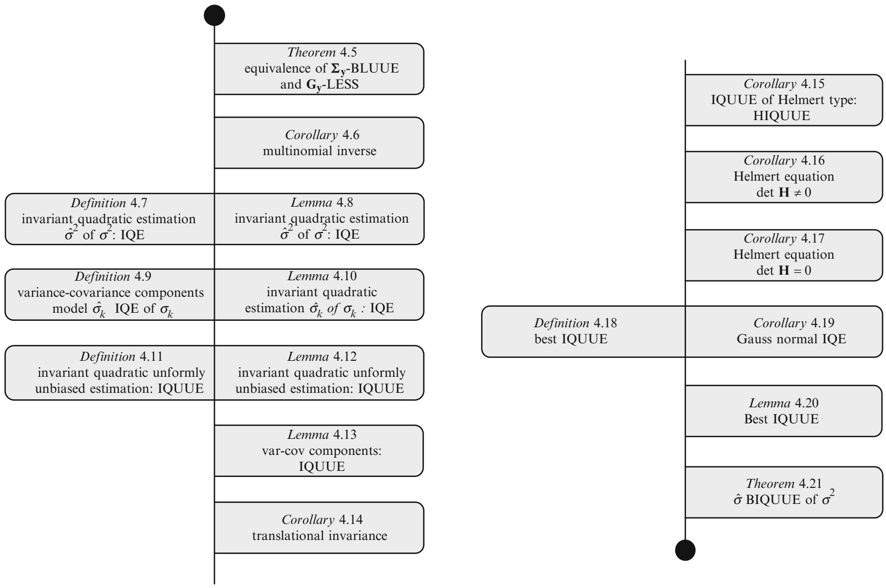
"*The first guideline of chapter four: definition, lemmas and theorem*"

In 1823, supplemented in 1828, C.F. Gauss put forward a new substantial generalization of "*least squares*" pointing out that an *integral measure of loss*, more definitely the *principle of minimum variance*, was preferable to least squares and to maximum likelihood. He abandoned both his previous postulates and set high store by the formula $\hat{\sigma}^2$ which provided an *unbiased estimate of variance* σ^2. *C. F. Gauss's* contributions to the treatment of erroneous observations, later on extended by F. *R. Helmert*, defined the state of the classical theory of errors.

To the *analyst C.F. Gauss's preference* to estimators of type BLUUE (Best Linear Uniformly Unbiased Estimator) for the moments of first order as well as of type BIQUUE (Best Invariant Quadratic Uniformly Unbiased Estimator) for the moments of second order is completely unknown. Extended by *A.A. Markov* who added *correlated observations* to the *Gauss* unbiased minimum variance estimator we present to you BLUUE of fixed effects and $\boldsymbol{\Sigma}_{\mathbf{y}}$-BIQUUE of the variance component.

"*The second guideline of chapter four:*
definitions, lemmas, corollaries and theorems"

In the third chapter we have solved a *special algebraic regression problem*, namely the inversion of a system of *inconsistent* linear equations of full column rank classified as "*overdetermined*". By means of the postulate of a least squares solution $||\mathbf{i}||^2 = ||\mathbf{y} - \mathbf{Ax}||^2 = \min$ we were able to determine m unknowns from n observations ($n > m$: *more equations n than unknowns m*). Though "LESS" generated a unique solution to the "*overdetermined*" system of linear equations with full column rank, we are unable to classify "LESS". There are two key questions we were not able to answer so far: In view of "MINOS" versus "LUMBE" we want to know whether "LESS" produces an unbiased estimation or *not*. How can we attach to an objective accuracy measure "LESS"?



The key for evaluating "LESS" is handed over to us by treating the special algebraic regression problem by means of a special *probabilistic regression problem*, namely a *special Gauss–Markov model without datum defect*. We shall prove that *uniformly unbiased estimations of the unknown parameters* of type "*fixed effects*" exist. "LESS" is replaced by "BLUUE" (**B**est **L**inear **U**niformly **U**nbiased **E**stimation). The *fixed effects* constitute the moments of first order of the under-lying probability distributions of the observations to be specified. In contrast, its central moments of second order, known as the variance-covariance matrix or dispersion matrix, open the door to associate to the estimated fixed effects an objective accuracy measure.

? What is a probabilistic problem ?

By means of certain statistical objective function, here of type

"*best linear uniformly unbiased estimation*" (BLUUE) for moments of *first order*	"*best quadratic invariant uniformly unbiased estimation*" (BIQUUE) for the central moments of *second order*

we solve the *inverse problem* of linear, lateron nonlinear equations with *fixed effects* which relates *stochastic observations to parameters*. According to the *Measurement Axiom*, observations are elements of a probability space. In terms of *second order statistics* the *observation space* $\mathbb{Y}$ of integer dimension, $dim\,\mathbb{Y} = n$, is characterized by

the first moment $E\{\mathbf{y}\}$, the expectation of (BLUUE) $\mathbf{y} \in \{\mathbb{Y}, pdf\}$	and	the central second moment $D\{\mathbf{y}\}$ the dispersion matrix or (BIQUUE) variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$.

In the case of "*fixed effects*" we consider the *parameter space* $\boldsymbol{\Xi}$, dim $\boldsymbol{\Xi} = m$, to be *metrical*. Its metric is induced from the probabilistic measure of the metric, the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$ of the observations $\mathbf{y} \in \{\mathbb{Y}, pdf\}$.

In particular, its variance-covariance matrix is *pulled-back* from the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$. In the special probabilistic regression model with unknown "*fixed effects*" $\xi \in \boldsymbol{\Xi}$ (elements of the parameter space) are estimated while the random variables like $\mathbf{y} - E\{\mathbf{y}\}$ are predicted.

4-1 Introduction

Our introduction has *four targets*. *First*, we want to introduce $\hat{\mu}$, a linear estimation of the mean value of "*direct*" observations, and $\hat{\sigma}^2$, a quadratic estimation of their variance component. For such a simple linear model we outline the postulates of *uniform unbiasedness and of minimum variance*. We shall pay special attention to the key role of the *invariant quadratic estimation* ("IQE") $\hat{\sigma}^2$ of σ^2. *Second*, we intend to analyse two data sets, the second one containing an outlier, by comparing the *arithmetic mean* and the *median* as well as the "*root mean square error*" (r.m.s.) of type BIQUUE and the "*median absolute deviation*" (m.a.d.). By proper choice of the bias term we succeed to prove identity of the weighted arithmetic mean and the median for the data set corrupted by an obvious outlier. *Third*, we discuss the competitive estimator "MALE", namely *Maximum Likelihood Estimation* which does not produce an unbiased estimation $\hat{\sigma}^2$ of σ^2, in general. *Fourth*, in order to develop the best quadratic uniformly unbiased estimation $\hat{\sigma}^2$ of σ^2, we have to highlight the need for *fourth order* statistic. "IQE" as well as "IQUUE" depend on the *central moments of fourth order* which are reduced to central moments of second order if we assume "*quasi-normal distributed*" observations.

4-11 The Front Page Example

By means of *Table 4.1* let us introduce a set of "direct" measurements $y_i, i \in 1, 2, 3, 4, 5$ of length data. We shall outline how we can compute the *arithmetic mean* 13.0 as well as the *standard deviation* of 1.6.

In contrast, *Table 4.2* presents an *augmented* observation vector: The observations six is an *outlier*. Again we have computed the new *arithmetic mean* $30.1\bar{6}$ as well as the *standard deviation* 42.1. In addition, for both examples we have calculated the *sample median* and the *sample absolute deviation* for comparison. All definitions will be given in the context as well as a careful analysis of the two data sets.

Table 4.1 "direct" observations, comparison of mean and median ($mean\ y = 13, med\ \mathbf{y} = 13,\ [n/2] = 2, [n/2]+1 = 3, med\ y = \mathbf{y}_{(3)}, mad\ y = med\ |\mathbf{y}_{(i)} - \text{med}\,\mathbf{y}| {=} 1, r.m.s.(I - BIQUUE) = 1.6$)

Number of observ.	Observ. y_i	Difference of observation and mean	Difference of observation and median	Ordered set of observations $y_{(i)}$	Ordered set of $\lvert y_{(i)} - med\ y \rvert$	Ordered set of $y_{(i)} - mean\ y$
1	15	+2	+2	11	0	+2
2	12	−1	−1	12	1	−1
3	14	+1	+1	13	1	+1
4	11	−2	−2	14	2	−1
5	13	0	0	15	2	0

Table 4.2 "direct" observations, effect of one outlier (mean $y = 30.1\bar{6}$, $med\ y = (13+14)/2 = 13.5, r.m.s.(I - BLUUE) = 42.1, med\ \left|\mathbf{y}_{(i)} - \text{med}\mathbf{y}\right| = mad y = 1.5$)

Number of observ.	Observ. y_i	Difference of observation and mean	Difference of observation and median	Ordered set of observations $y_{(i)}$	Ordered set of $\lvert y_{(i)} - med\ y \rvert$	Ordered set of $y_{(i)} - mean\ y$
1	15	$-15.1\bar{6}$	+1.5	11	0.5	$-15.1\bar{6}$
2	12	$-18.1\bar{6}$	−1.5	12	0.5	$-16.1\bar{6}$
3	14	$-16.1\bar{6}$	+0.5	13	1.5	$-17.1\bar{6}$
4	11	$-19.1\bar{6}$	−2.5	14	1.5	$-18.1\bar{6}$
5	13	$-17.1\bar{6}$	−0.5	15	2.5	$-19.1\bar{6}$
6	116	$+85.8\bar{3}$	+102.5	116	+102.5	$+85.8\bar{3}$

4-12 Estimators of Type BLUUE and BIQUUE of the Front Page Example

In terms of a *special Gauss–Markov model* our data set can be described as following. The *statistical moment of first order*, namely the expectation $E\{\mathbf{y}\} = \mathbf{1}\mu$ of the observation vector $\mathbf{y} \in \mathbb{R}^n$, here $n = 5$ and the *central statistical moment of second order*, namely the variance-covariance matrix $\mathbf{\Sigma_y}$, also called the dispersion matrix $D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2,\ D\{\mathbf{y}\} =: \mathbf{\Sigma_y} \in \mathbb{R}^{n\times n},\ rk\mathbf{\Sigma_y} = n$, of the observation vector $\mathbf{y} \in \mathbb{R}^n$, with the variance σ^2 characterize the stochastic linear model. The mean $\mu \in \mathbb{R}$ of the "direct" observations and the *variance factor* σ^2 are *unknown*. We shall estimate $(\mu,\ \sigma^2)$ by means of three postulates:

- *First postulate*: $\hat{\mu}$: *linear estimation*, $\hat{\sigma}^2$: *quadratic estimation*

$$\hat{\mu} = \sum_{p=1}^{n} l_p y_p \quad or \quad \hat{\mu} = \mathbf{l}'\mathbf{y}$$

$$\hat{\sigma}^2 = \sum_{p,q=1}^{n} m_{pq} y_p y_q \quad or \quad \begin{aligned}\hat{\sigma}^2 &= \mathbf{y}'\mathbf{M}\mathbf{y} = (\mathbf{y}' \otimes \mathbf{y}')(\mathrm{vec}\mathbf{M}) \\ &= (vec\mathbf{M})'(\mathbf{y} \otimes \mathbf{y})\end{aligned}$$

- *The second postulate: uniform unbiasedness*

$$\begin{aligned}&E\{\hat{\mu}\} = \mu \ \textit{ for all } \mu \in \mathbb{R} \\ &E\{\hat{\sigma}^2\} = \sigma^2 \ \textit{ for all } \sigma^2 \in \mathbb{R}^+\end{aligned}$$

- *The third postulate: minimum variance*

$$D\{\hat{\mu}\} = E\{[\hat{\mu} - E\{\hat{\mu}\}]^2\} = \min_{\ell} \quad and \quad D\{\hat{\sigma}^2\} = E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2\} = \min_{M}$$

$$\hat{\mu} = \arg\ \min_{\ell} D\{\hat{\mu}|\hat{\mu} = \ell'\mathbf{y},\ E\{\hat{\mu}\} = \mu\}$$

$$\hat{\sigma}^2 = \arg\ \min_{M} D\{\hat{\sigma}^2|\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y},\ E\{\hat{\sigma}^2\} = \sigma^2\}.$$

First, we begin with the postulate that the *fixed* unknown parameters $(\mu,\ \sigma^2)$ are *estimated* by means of a certain *linear* form $\hat{\mu} = \ell'\mathbf{y} + \kappa = \mathbf{y}'\ell + \kappa$ and by means of a certain *quadratic form* $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} + \mathbf{x}'\mathbf{y} + \omega = (vec\mathbf{M})'(\mathbf{y} \otimes \mathbf{y}) + \mathbf{x}'\mathbf{y} + \omega$ of the observation vector $\mathbf{y}$, subject to the symmetry condition $\mathbf{M} \in \mathrm{SYM} := \{\mathbf{M} \in \mathbb{R}^{n\times n} | \mathbf{M} = \mathbf{M}'\}$, namely the space of *symmetric matrices*. *Second* we demand $E\{\hat{\mu}\} = \mu,\ E\{\hat{\sigma}^2\} = \sigma^2$, namely unbiasedness of the estimations $(\hat{\mu},\ \hat{\sigma}^2)$. Since the estimators $(\hat{\mu},\ \hat{\sigma}^2)$ are special forms of the observation vector $\mathbf{y} \in \mathbb{R}^n$, an intuitive understanding of the *postulate of unbiasedness* is the following: If the dimension of the observation space $\mathbb{Y} \ni \mathbf{y}$, $dim\,\mathbb{Y} = n$, is going to *infinity*, we expect information about the "two values" $(\mu,\ \sigma^2)$, namely

$$\lim_{n\to\infty} \hat{\mu}(n) = \mu, \quad \lim_{n\to\infty} \hat{\sigma}^2(n) = \sigma^2.$$

Let us investigate how LUUE (*Linear Uniformly Unbiased Estimation*) of μ as well as IQUUE (*Invariant Quadratic Uniformly Unbiased Estimation*) operate.

LUUE

$$\left.\begin{array}{c} E\{\hat{\mu}\} = E\{l'\mathbf{y} + \kappa\} = l'E\{\mathbf{y}\} + \kappa \\ E\{\mathbf{y}\} = \mathbf{1}_n\mu \end{array}\right] \Rightarrow$$
$$E\{\hat{\mu}\} = l'E\{\mathbf{y}\} + \kappa = l'\mathbf{1}_n\mu + \kappa$$
$$E\{\hat{\mu}\} = \mu \Leftrightarrow \kappa = 0,\ (l'\mathbf{1}_n - 1)\mu = 0$$
$$\Leftrightarrow \kappa = 0,\ l'\mathbf{1}_n - 1 = 0 \ \textit{for all}\ \mu \in \mathbb{R}.$$

Indeed $\hat{\mu}$ is LUUE if and only if $\kappa = 0$ and $(l'\mathbf{1}_n - 1)\mu = 0\ for\,all\,\mu \in \mathbb{R}$. The zero identity $(\ell'\mathbf{1}_n - 1)\mu = 0$ is fulfilled by means of $\ell'\mathbf{1}_n - 1 = 0,\ \ell'\mathbf{1}_n = 1$, if we restrict the solution by the *quantor* "$for\,all\,\mu \in \mathbb{R}$". $\mu = 0$ is not an admissible solution. Such a situation is described as "*uniformly unbiased*". We summarize that LUUE is constrained by the zero identity

$$\ell'\mathbf{1}_n - 1 = 0.$$

Next we shall prove that $\hat{\sigma}^2$ is IQUUE if and only if

IQUUE

$$\begin{array}{ll} E\{\hat{\sigma}^2\} = E\{\mathbf{y}'\mathbf{M}\mathbf{y} + \mathbf{x}'\mathbf{y} + \omega\} = & E\{\hat{\sigma}^2\} = E\{\mathbf{y}'\mathbf{M}\mathbf{y} + \mathbf{x}'\mathbf{y} + \omega\} = \\ E\{(vec\,\mathbf{M})'(\mathbf{y} \otimes \mathbf{y}) + \mathbf{x}'\mathbf{y} + \omega\} = & E\{(\mathbf{y}' \otimes \mathbf{y}')(vec\,\mathbf{M})' + \mathbf{y}'\mathbf{x} + \omega\} = \\ (vec\,\mathbf{M})'E\{\mathbf{y} \otimes \mathbf{y}\} + \mathbf{x}'E\{\mathbf{y}\} + \omega & E\{\mathbf{y}' \otimes \mathbf{y}'\}(vec\,\mathbf{M})' + E\{\mathbf{y}'\}\mathbf{x} + \omega. \end{array}$$

$\hat{\sigma}^2$ is called *translational invariant* with respect to $\mathbf{y} \mapsto \mathbf{y} - E\{\mathbf{y}\}$ if

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} + \mathbf{x}'\mathbf{y} + \omega = (\mathbf{y} - E\{\mathbf{y}\})'\mathbf{M}(\mathbf{y} - E\{\mathbf{y}\}) + \mathbf{x}'(\mathbf{y} - E\{\mathbf{y}\}) + \omega$$

and uniformly unbiased if

$$E\{\hat{\sigma}^2\} = (vec\,\mathbf{M})'E\{\mathbf{y} \otimes \mathbf{y}\} + \mathbf{x}'E\{\mathbf{y}\} + \omega = \sigma^2 \text{ for all } \sigma^2 \in \mathbb{R}^+.$$

Finally we have to discuss the postulate of a best estimator of type BLUUE of μ and BIQUUE of σ^2. We proceed *sequentially*, *first* we determine $\hat{\mu}$ of type BLUUE and second $\hat{\sigma}^2$ of type BIQUUE. At the end we shall discuss *simultaneous estimation* of $(\hat{\mu},\ \hat{\sigma}^2)$.

The scalar $\hat{\mu} = \ell'\mathbf{y}$ is BLUUE of μ (Best Linear Uniformly Unbiased Estimation) with respect to the linear model

$$E\{\mathbf{y}\} = \mathbf{1}_n\mu, \ D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2,$$

if it is *uniformly unbiased* in the sense of and in comparison of all linear, uniformly unbiased estimations possesses the *smallest variance* in the sense of

$$\begin{aligned} D\{\hat{\mu}\} &= E\{[\hat{\mu} - E\{\hat{\mu}\}]^2\} \\ &= \sigma^2\ell'\ell = \sigma^2 \mathrm{tr}\ell'\ell = \sigma^2||\ell||^2 = \min . \end{aligned}$$

The constrained Lagrangean $\mathcal{L}(\ell, \lambda)$, namely

$$\begin{aligned} \mathcal{L}(\ell, \lambda) :&= \sigma^2\ell'\ell + 2\lambda(\ell'\mathbf{1}_n - 1) \\ &= \sigma^2\ell'\ell + 2(\mathbf{1}_n\ell - 1)\lambda = \min_{\ell,\lambda}, \end{aligned}$$

produces by means of the *first derivatives*

$$\begin{aligned} \frac{1}{2}\frac{\partial\mathcal{L}}{\partial l}(\hat{l}, \hat{\lambda}) &= \sigma^2\hat{l} + \mathbf{1}_n\hat{\lambda} = \mathbf{0} \\ \frac{1}{2}\frac{\partial\mathcal{L}}{\partial\lambda}(\hat{l}, \hat{\lambda}) &= \hat{l}'\mathbf{1}_n - 1 = \mathbf{0} \end{aligned}$$

the *normal equations* for the augmented unknown vector (ℓ, λ), also known as the *necessary conditions* for obtaining an optimum. Transpose the *first normal equation*, right multiply by $\mathbf{1}_n$, the unit column and substitute the *second normal equation* in order to solve for the *Lagrange multiplier* $\hat{\lambda}$. If we substitute the solution $\hat{\lambda}$ in the *first normal equation*, we directly find the linear operator $\hat{\ell}$.

$$\begin{aligned} \sigma^2\hat{\ell}' + \mathbf{1}'_n\hat{\lambda} &= \mathbf{0}' \\ \sigma^2\hat{\ell}'\mathbf{1}_n + \mathbf{1}'_n\mathbf{1}_n\hat{\lambda} &= \sigma^2 + \mathbf{1}'_n\mathbf{1}_n\hat{\lambda} = 0 \\ \Rightarrow \hat{\lambda} &= -\frac{\sigma^2}{\mathbf{1}'_n\mathbf{1}_n} = -\frac{\sigma^2}{n} \\ \sigma^2\hat{\ell} + \mathbf{1}_n\hat{\lambda} &= \sigma^2\hat{l} - \mathbf{1}_n\frac{\sigma^2}{n} = \mathbf{0}' \\ \Rightarrow \hat{\ell} &= \mathbf{1}_n\frac{1}{n} \ \textit{and} \ \hat{\mu} = \hat{\ell}'\mathbf{y} = \frac{1}{n}\mathbf{1}'_n\mathbf{y}. \end{aligned}$$

The *second derivatives*

$$\frac{1}{2}\frac{\partial^2\mathcal{L}}{\partial\ell\partial\ell'}(\hat{\ell}, \hat{\lambda}) = \sigma^2\mathbf{I}_n > \mathbf{0}'$$

constitute the *sufficiency condition* which is automatically satisfied. Let us briefly summarize the first result $\hat{\mu}$ BLUUE of μ. The scalar $\hat{\mu} = \ell'\mathbf{y}$ is BLUUE of μ with respect to the linear model

$$E\{\mathbf{y}\} = \mathbf{1}_n\mu, D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2,$$

if and only if

$$\hat{\ell}' = \frac{1}{n}\mathbf{1}'_n \ \ and \ \ \hat{\mu} = \frac{1}{n}\mathbf{1}'_n\mathbf{y}$$

is the *arithmetic mean*. The observation space $\mathbf{y} \in \{\mathbb{Y}, pdf\}$ is decomposed into

$$\mathbf{y}(\text{BLUUE}) := \mathbf{1}_n\hat{\mu} \ \ versus \ \ \mathbf{e_y}(\text{BLUUE}) := \mathbf{y} - \mathbf{y}(\text{BLUUE}),$$

$$\mathbf{y}(\text{BLUUE}) = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\mathbf{y} \ \ versus \ \ \mathbf{e_y}(\text{BLUUE}) = [\mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)]\mathbf{y},$$

which are orthogonal in the sense of

$$\left\langle \mathbf{e_y}(\text{BLUUE}) \mid \mathbf{y}(\text{BLUUE}) \right\rangle = 0 \ \ or \ \ \left[\mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)\right]\frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n) = \mathbf{0}.$$

Before we continue with the setup of the *Lagrangean* which guarantees BIQUUE, we study beforehand $\mathbf{e_y} := \mathbf{y} - E\{\mathbf{y}\}$ and $\mathbf{e_y}(BLUUE) := \mathbf{y} - \mathbf{y}(BLUUE)$. Indeed the *residual vector* $\mathbf{e_y}(BLUUE)$ is a linear form of *residual vector* $\mathbf{e_y}$.

$$\mathbf{e_y}(BLUUE) = \left[\mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)\right]\mathbf{e_y}.$$

For the proof we depart from

$$\begin{aligned}
\mathbf{e_y}(BLUUE) : &= \mathbf{y} - \mathbf{1}_n\hat{\mu} = \left[\mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)\right]\mathbf{y} \\
&= \left[\mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)\right](\mathbf{y} - E\{\mathbf{y}\}) \\
&= \mathbf{I}_n - \frac{1}{n}(\mathbf{1}_n\mathbf{1}'_n),
\end{aligned}$$

where we have used the *invariance property* $\mathbf{y} \mapsto \mathbf{y} - E\{\mathbf{y}\}$ based upon the idempotence of the matrix $[\mathbf{I}_n - (\mathbf{1}_n\mathbf{1}'_n)/n]$.

Based upon the fundamental relation $\mathbf{e_y}(\text{BLUUE}) = \mathbf{D}\mathbf{e_y}$, where $\mathbf{D} := \mathbf{I}_n - (\mathbf{1}_n\mathbf{1}'_n)/n$ is a projection operator onto the normal space $\mathcal{R}(\mathbf{1}_n)^{\perp}$, we are able to derive an unbiased estimation of the variance component σ^2. Just compute

$$\begin{aligned} E\{\mathbf{e}'_{\mathbf{y}}(BLUUE)\mathbf{e}_{\mathbf{y}}(BLUUE)\} &= \mathrm{tr}E\{\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE})\mathbf{e}'_{\mathbf{y}}(\mathrm{BLUUE})\} \\ &= \mathrm{tr}\mathbf{D}\,E\{\mathbf{e}_{\mathbf{y}}\mathbf{e}'_{\mathbf{y}}\}\mathbf{D}' = \sigma^2\mathrm{tr}\mathbf{D}\,\mathbf{D}' = \sigma^2\mathrm{tr}\mathbf{D} \\ tr\mathbf{D} &= \mathrm{tr}\,(\mathbf{I}_n) - \mathrm{tr}\tfrac{1}{n}(\mathbf{1}_n\mathbf{1}'_n) = n-1 \\ E\{\mathbf{e}'_{\mathbf{y}}(\mathrm{BLUUE})\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE})\} &= \sigma^2(n-1). \end{aligned}$$

Let us define the *quadratic estimator* $\hat{\sigma}^2$ of σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n-1}\mathbf{e}'_{\mathbf{y}}(\mathrm{BLUUE})\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE}),$$

which is unbiased according to

$$E\{\hat{\sigma}^2\} = \frac{1}{n-1}E\{\mathbf{e}'_{\mathbf{y}}(\mathrm{BLUUE})\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE})\} = \sigma^2.$$

Let us briefly summarize the first result $\hat{\sigma}^2$ IQUUE of σ^2. The scalar

$$\hat{\sigma}^2 = \mathbf{e}'_{\mathbf{y}}(\mathrm{BLUUE})\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE})/(n-1)$$

is IQUUE of σ^2 based upon the BLUUE-residual vector

$$\mathbf{e}_{\mathbf{y}}(\mathrm{BLUUE}) = \left[\mathbf{I}_n - \tfrac{1}{n}(\mathbf{1}_n\mathbf{1}'_n)\right]\mathbf{y}.$$

Let us highlight $\hat{\sigma}^2$ BIQUUE of σ^2.

A scalar $\hat{\sigma}^2$ is BIQUUE of σ^2 (Best Invariant Quadratic Uniformly Unbiased Estimation) with respect to the linear model

$$E\{\mathbf{y}\} = \mathbf{1}_n\mu,\ D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2,$$

if it is

(i) *Uniformly unbiased* in the sense of $E\{\hat{\sigma}^2\} = \sigma^2$ for all $\sigma^2 \in \mathbb{R}^+$
(ii) *Quadratic* in the sense of $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y}$ for all $\mathbf{M} = \mathbf{M}'$
(iii) *Translational invariant* in the sense of $\mathbf{y}'\mathbf{M}\mathbf{y} = (\mathbf{y} - E\{\mathbf{y}\})'\mathbf{M}(\mathbf{y} - E\{\mathbf{y}\}) = (\mathbf{y} - \mathbf{1}_n\mu)'\mathbf{M}(\mathbf{y} - \mathbf{1}_n\mu)$
(vii) *Best* if it possesses the *smallest variance* in the sense of $D\{\hat{\sigma}^2\} = E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2\} = \min_{\mathbf{M}}$.

First, let us consider the most influential postulate of *translational invariance* of the *quadratic estimation*

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\mathrm{vec}\mathbf{M})'(\mathbf{y}\otimes\mathbf{y}) = (\mathbf{y}'\otimes\mathbf{y}')(\mathrm{vec}\mathbf{M})$$

to comply with

$$\hat{\sigma}^2 = \mathbf{e'}_\mathbf{y}\mathbf{M}\mathbf{e}_\mathbf{y} = (\mathrm{vec}\mathbf{M})'(\mathbf{e}_\mathbf{y} \otimes \mathbf{e}_\mathbf{y}) = (\mathbf{e'}_\mathbf{y} \otimes \mathbf{e'}_\mathbf{y})(\mathrm{vec}\mathbf{M})$$

subject to

$$\mathbf{M} \in SYM\!: = \{\mathbf{M} \in \mathbb{R}^{n \times n} | \mathbf{M} = \mathbf{M}'\}.$$

Translational invariance is understood as the action of transformation group

$$\mathbf{y} = E\{\mathbf{y}\} + \mathbf{e}_\mathbf{y} = \mathbf{1}_n \mu + \mathbf{e}_\mathbf{y}$$

with respect to the linear model of "direct" observations. Under the action of such a transformation group the quadratic estimation $\hat{\sigma}^2$ of σ^2 is specialized to

$$\begin{aligned}
\hat{\sigma}^2 &= \mathbf{y}'\mathbf{M}\mathbf{y} = \left[E\{\mathbf{y}\} + \mathbf{e}_\mathbf{y}\right]' \mathbf{M} \left[E\{\mathbf{y}\} + \mathbf{e}_\mathbf{y}\right] = (\mathbf{1'}_n \mu + \mathbf{e'}_\mathbf{y})\mathbf{M}(\mathbf{1}_n \mu + \mathbf{e}_\mathbf{y}) \\
\hat{\sigma}^2 &= \mu^2 \mathbf{1'}_n \mathbf{M} \mathbf{1}_n + \mu \mathbf{1'}_n \mathbf{M} \mathbf{e}_\mathbf{y} + \mu \mathbf{e'}_\mathbf{y} \mathbf{M} \mathbf{1}_n + \mathbf{e'}_\mathbf{y} \mathbf{M} \mathbf{e}_\mathbf{y} \\
\mathbf{y}'\mathbf{M}\mathbf{y} &= \mathbf{e'}_\mathbf{y}\mathbf{M}\mathbf{e}_\mathbf{y} \Leftrightarrow \mathbf{1'}_n \mathbf{M} = \mathbf{0}' \, and \, \mathbf{1'}_n \mathbf{M}' = \mathbf{0}'.
\end{aligned}$$

IQE, namely $\mathbf{1'}_n\mathbf{M} = \mathbf{0}'$ and $\mathbf{1'}_n\mathbf{M}' = \mathbf{0}'$ has a definite consequence. It is independent of μ, the *first moment* of the *probability distribution* ("*pdf*"). Indeed, the estimation procedure of the *central second moment* σ^2 is *decoupled* from the estimation of the *first moment* μ. Here we find the key role of the invariance principle. Another aspect is the general solution of the homogeneous equation $\mathbf{1'}_n\mathbf{M} = \mathbf{0}'$ subject to the symmetry postulate $\mathbf{M} = \mathbf{M}'$.

$$\mathbf{1}'\mathbf{M} = \mathbf{0}' \Leftrightarrow \left[\begin{array}{l} \mathbf{M} = \left[\mathbf{I}_n - \mathbf{1'}_n(\mathbf{1'}_n\mathbf{1}_n)^{-1}\mathbf{1'}_\mathbf{n}\right]\mathbf{Z} \\ \mathbf{M} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1'}_n)\mathbf{Z}\,, \end{array}\right.$$

where $\mathbf{Z}$ is an arbitrary matrix. The general solution of the homogeneous matrix equation contains the *left inverse* (*generalized inverse* $(\mathbf{1'}_n\mathbf{1}_n)^{-1}\mathbf{1'}_n = \mathbf{1}_n^-$) which takes an exceptionally simple form, here. The general form of the matrix $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is in no agreement with the symmetry postulate $\mathbf{M} = \mathbf{M}'$.

$$\begin{array}{ll} \mathbf{1'}_n\mathbf{M} & = \mathbf{0}' \\ \mathbf{M} & = \mathbf{M}' \end{array} \Leftrightarrow \mathbf{M} = \alpha(\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1'}_n).$$

Indeed, we made the choice $\mathbf{Z} = \alpha\mathbf{I}_n$ which reduces the unknown parameter space to one-dimension. Now by means of the postulate "*best*" under the constraint generated by "*uniform unbiasedness*" $\hat{\sigma}^2$ of σ^2 we shall determine the parameter $\alpha = 1/(n-1)$.

The postulate IQUUE is materialized by

$$E\{\hat{\sigma}^2 | \mathbf{\Sigma}_\mathbf{y} = \mathbf{I}_n\sigma^2\} = \sigma^2 \Leftrightarrow \mathrm{tr}\mathbf{M} = 1 \Leftrightarrow \mathrm{tr}\mathbf{M} - 1 = 0.$$

For the simple case of "i.i.d." observations, namely $\mathbf{\Sigma}_\mathbf{y} = \mathbf{I}_n\sigma^2$, $E\{\hat{\sigma}^2\} = \sigma^2$ for an IQE, IQUUE is equivalent to $\mathrm{tr}\mathbf{M} = 1$ or $(\mathrm{tr}\mathbf{M}) - 1 = 0$ as a *condition equation*.

$$\mathrm{tr}\mathbf{M} = 1 \Leftrightarrow \begin{array}{c} \alpha\mathrm{tr}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n) = \alpha(n-1) = 1 \\ \alpha \qquad\qquad = \dfrac{1}{n-1}. \end{array}$$

IQUUE of the simple case

$$\left.\begin{array}{r} \mathit{invariance} : (i)\mathbf{1}'\mathbf{M} = \mathbf{0}' \, and \, \mathbf{M} = \mathbf{M}' \\ \text{QUUE} : (ii)\mathrm{tr}\mathbf{M} - 1 = 0 \end{array}\right] \Rightarrow \mathbf{M} = \frac{1}{n-1}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n\right)$$

has already solved our problem of generating the symmetric matrix $\mathbf{M}$.

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = \frac{1}{n-1}\mathbf{y}'(\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}'_n)\mathbf{y} \in \text{IQUUE}$$

1.1 Is there still a need to apply "best" as an optimability condition for BIQUUE ?

Yes, there is! The general solution of the homogeneous equations $\mathbf{1}'_n\mathbf{M} = \mathbf{0}'$ and $\mathbf{M}'\mathbf{1}_n = \mathbf{0}$ generated by the postulate of *translational invariance* of IQE did not produce a symmetric matrix. Here we present the *simple symmetrization*. An alternative approach worked depart from

$$\frac{1}{2}(\mathbf{M} + \mathbf{M}') = \frac{1}{2}\{[\mathbf{I}_n - \mathbf{1}_n(\mathbf{1}'_n\mathbf{1}_n)^{-1}\mathbf{1}'_n]\mathbf{Z} + \mathbf{Z}'[\mathbf{I}_n - \mathbf{1}_n(\mathbf{1}'_n\mathbf{1}_n)^{-1}\mathbf{1}'_n]\},$$

leaving the general matrix $\mathbf{Z}$ as an unknown to be determined. Let us therefore develop BIQUUE for the linear model

$$E\{\mathbf{y}\} = \mathbf{1}_n\mu, D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$$

$$D\{\hat{\sigma}^2\} = E\{(\hat{\sigma}^2 - E\{\hat{\sigma}^2\})^2\} = E\{\hat{\sigma}^4\} - E\{\hat{\sigma}^2\}^2.$$

Apply the *summation convention* over repeated indices $i, j, k, \ell \in \{1, \ldots, n\}$.

$$1st : E\{\hat{\sigma}^2\}^2$$

$$E\{\hat{\sigma}^2\}^2 = m_{ij}E\{e^y_i e^y_j\}m_{k\ell}E\{e^y_k e^y_\ell\} = m_{ij}m_{kl}\pi_{ij}\pi_{k\ell}$$

subject to

$$\pi_{ij} := E\{e^y_i e^y_j\} = \sigma^2\delta_{ij} and \pi_{k\ell} := E\{e^y_k e^y_\ell\} = \sigma^2\delta_{k\ell}$$

$$E\{\hat{\sigma}^2\}^2 = \sigma^4 m_{ij}\delta_{ij}m_{k\ell}\delta_{k\ell} = \sigma^4(tr\mathbf{M})^2$$

$$2nd : E\{\hat{\sigma}^4\}$$

$$E\{\hat{\sigma}^4\} = m_{ij}m_{k\ell}E\{e^y_i e^y_j e^y_k e^y_\ell\} = m_{ij}m_{k\ell}\pi_{ijk\ell}$$

subject to

$$\pi_{ijk\ell} := E\{e_i^y e_j^y e_k^y e_\ell^y\} \forall i, j, k, \ell \in \{1, \ldots, n\}.$$

For a normal pdf, the *fourth order moment* $\pi_{ijk\ell}$ can be reduced to *second order moments*. For a more detailed presentation of "*normal models*" we refer to Appendix A-7

$$\begin{aligned}
\pi_{ijk\ell} &= \pi_{ij}\pi_{k\ell} + \pi_{ik}\pi_{j\ell} + \pi_{i\ell}\pi_{jk} = \sigma^4(\delta_{ij}\delta_{k\ell} + \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) \\
E\{\hat{\sigma}^4\} &= \sigma^4 m_{ij} m_{k\ell}(\delta_{ij}\delta_{k\ell} + \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) \\
E\{\hat{\sigma}^4\} &= \sigma^4[(\mathrm{tr}\mathbf{M})^2 + 2\mathrm{tr}\mathbf{M}'\mathbf{M}].
\end{aligned}$$

Let us briefly summarize the representation of the variance $D\{\hat{\sigma}^2\} = E\{(\hat{\sigma}^2 - E\{\hat{\sigma}^2\})^2\}$ for normal models.

Let the linear model of i.i.d. direct observations be defined by

$$E\{\mathbf{y}|pdf\} = \mathbf{1}_n\mu, D\{\mathbf{y}|pdf\} = \mathbf{I}_n\sigma^2.$$

The variance of a normal IQE can be represented by

$$\begin{aligned}
D\{\hat{\sigma}^2\} := E\{(\hat{\sigma}^2 - E\{\hat{\sigma}^2\})^2\} = \\
= 2\sigma^4[(\mathrm{tr}\mathbf{M})^2 + \mathrm{tr}(\mathbf{M^2})].
\end{aligned}$$

In order to construct BIQUUE, we shall define a *constrained Lagrangean* which takes into account the conditions of translational invariance, uniform unbiased-ness and symmetry.

$$\mathcal{L}(\mathbf{M}, \lambda_0, \lambda_1, \lambda_2) := 2\mathrm{tr}\mathbf{M}'\mathbf{M} + 2\lambda_0(\mathrm{tr}\mathbf{M} - 1) + 2\lambda_1\mathbf{1}'_n\mathbf{M}\mathbf{1}_n + 2\lambda_2\mathbf{1}'_n\mathbf{M}'\mathbf{1}_n = \min_{\mathbf{M}, \lambda_0, \lambda_1, \lambda_2}.$$

Here we used the condition of translational invariance in the special form

$$\mathbf{1}'_n\frac{1}{2}(\mathbf{M} + \mathbf{M}')\mathbf{1}_n = 0 \Leftrightarrow \mathbf{1}'_n\mathbf{M}\mathbf{1}_n = 0 \; and \; \mathbf{1}'_n\mathbf{M}'\mathbf{1}_n = 0,$$

which accounts for the symmetry of the unknown matrix. We here conclude with the normal equations for BIQUUE generated from

$$\frac{\partial\mathcal{L}}{\partial(\mathrm{vec}\mathbf{M})} = 0, \frac{\partial\mathcal{L}}{\partial\lambda_0} = 0, \frac{\partial\mathcal{L}}{\partial\lambda_1} = 0, \frac{\partial\mathcal{L}}{\partial\lambda_2} = 0.$$

$$\begin{bmatrix} 2(\mathbf{I}_n \otimes \mathbf{I}_n) & \mathrm{vec}\mathbf{I}_n & \mathbf{I}_n \otimes \mathbf{1}_n & \mathbf{1}_n \otimes \mathbf{I}_n \\ (\mathrm{vec}\mathbf{I}_n)' & 0 & 0 & 0 \\ \mathbf{I}_n \otimes \mathbf{1}'_n & 0 & 0 & 0 \\ \mathbf{1}'_n \otimes \mathbf{I}_n & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{vec}\mathbf{M} \\ \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

These normal equations will be solved lateron. Indeed $\mathbf{M} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n)/(n-1)$ is a solution.

$$BIQUUE: \quad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-1}\mathbf{y}'(\mathbf{I_n} - \tfrac{1}{n}\mathbf{1_n 1'_n})\mathbf{y} \\ D\{\hat{\sigma}^2\} &= \frac{2}{n-1}\sigma^4 \end{aligned}$$

Such a result is *based upon*

$$(\mathrm{tr}\mathbf{M})^2(\mathrm{BIQUUE}) = \frac{1}{n-1}, \quad (\mathrm{tr}\mathbf{M}^2)(\mathrm{BIQUUE}) = \frac{1}{n-1},$$

$$D\{\hat{\sigma}^2|\mathrm{BIQUUE}\} = D\{\hat{\sigma}^2\} = 2\sigma^4[(\mathrm{tr}\mathbf{M})^2 + (\mathrm{tr}\mathbf{M}^2)](\mathrm{BIQUUE}),$$

$$D\{\hat{\sigma}^2\} = \frac{2}{n-1}\sigma^4.$$

Finally, we are going to outline the simultaneous estimation of $\{\mu, \sigma^2\}$ for the *linear model of direct observations*.

- *First postulate*: inhomogeneous, multilinear (bilinear) estimation

$$\hat{\mu} = \kappa_1 + \ell'_1\mathbf{y} + \mathbf{m}'_1(\mathbf{y} \otimes \mathbf{y})$$
$$\hat{\sigma}^2 = \kappa_2 + \ell'_2\mathbf{y} + (\mathrm{vec}\mathbf{M_2})'(\mathbf{y} \otimes \mathbf{y})$$
$$\begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \begin{bmatrix} \ell'_1 & \mathbf{m}'_1 \\ \ell'_2 & (\mathrm{vec}\mathbf{M_2})' \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{y} \otimes \mathbf{y} \end{bmatrix}$$
$$\mathbf{x} = \mathbf{XY} \Leftrightarrow \mathbf{x} := \begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} \kappa_1 & \ell'_1 & \mathbf{m}'_1 \\ \kappa_2 & \ell'_2 & (\mathrm{vec}\mathbf{M_2})' \end{bmatrix}$$
$$\mathbf{Y} := \begin{bmatrix} 1 \\ \mathbf{y} \\ \mathbf{y} \otimes \mathbf{y} \end{bmatrix}$$

- *Second postulate*: uniform unbiasedness

$$E\{\mathbf{x}\} = E\left\{\begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix}\right\} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$$

- *Third postulate*: minimum variance

$$D\{\mathbf{x}\} := \mathrm{tr}E\{[\mathbf{x} - E\{\mathbf{x}\}][\mathbf{x} - E\{\mathbf{x}\}]'\} = \min.$$

4-13 BLUUE and BIQUUE of the Front Page Example, Sample Median, Median Absolute Deviation

According to *Tables 4.1* and *4.2* we presented you with two sets of observations $y_i \in \mathbb{Y}, \mathtt{dim}\mathbb{Y} = n, i \in \{1, \ldots, n\}$ the second one qualifies to certain "*one outlier*". We aim at a definition of the *median* and of the *median absolute deviation* which is compared to the definition of the *mean* (weighted mean) and of the *root-mean-square error*. *First* we *order* the observations according to $y_{(1)} < y_{(2)} < \ldots < y_{(n-1)} < y_{(n)}$ by means of the *permutation matrix*

$$\begin{bmatrix} y_{(1)} \\ y_{(2)} \\ \ldots \\ y_{(n-1)} \\ y_{(n)} \end{bmatrix} = \mathbf{P} \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_{n-1} \\ y_n \end{bmatrix},$$

namely

data set one *data set two*

$$\begin{bmatrix} 11 \\ 12 \\ 13 \\ 14 \\ 15 \end{bmatrix} = \begin{bmatrix} 0\,0\,0\,1\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,1\,0\,0 \\ 1\,0\,0\,0\,0 \end{bmatrix} \begin{bmatrix} 15 \\ 12 \\ 14 \\ 11 \\ 13 \end{bmatrix} \quad versus \quad \begin{bmatrix} 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 116 \end{bmatrix} = \begin{bmatrix} 0\,0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1 \end{bmatrix} \begin{bmatrix} 15 \\ 12 \\ 14 \\ 11 \\ 13 \\ 116 \end{bmatrix}$$

$$\mathbf{P}_5 = \begin{bmatrix} 0\,0\,0\,1\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,1 \\ 0\,0\,1\,0\,0 \\ 1\,0\,0\,0\,0 \end{bmatrix} \quad versus \quad \mathbf{P}_6 = \begin{bmatrix} 0\,0\,0\,1\,0\,0 \\ 0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0 \\ 0\,0\,1\,0\,0\,0 \\ 1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1 \end{bmatrix}.$$

Note $\mathbf{PP}' = \mathbf{I}$, $\mathbf{P}^{-1} = \mathbf{P}'$. *Second*, we define the *sample median* med $\mathbf{y}$ as well as the median absolute deviation mad $\mathbf{y}$ of $\mathbf{y} \in \mathbb{Y}$ by means of

$$\mathrm{med}\mathbf{y} := \left[\begin{array}{c} y_{([n/2]+1)} \text{ if } n \text{ is an } odd \text{ number} \\ \frac{1}{2}(y_{(n/2)} + y_{(n/2+1)}) \text{ if } n \text{ is an } even \text{ number} \end{array} \right.$$

$$\mathrm{mad}\mathbf{y} := \mathrm{med}|y_{(i)} - \mathrm{med}\mathbf{y}|,$$

where $[n/2]$ denotes the largest integer ("natural number") $\leq n/2$.

Third, we compute I-LESS, namely *mean* $\mathbf{y} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{y} = \frac{1}{n}\mathbf{1}'\mathbf{y}$ listed in *Table 4.3*. Obviously for the second observational data set the *Euclidean metric* of the observation space $\mathbb{Y}$ is not isotropic. Indeed let us compute $\mathbf{G_y}$-LESS, namely the weighted mean $\mathbf{y} = (\mathbf{1}'\mathbf{G_y}\mathbf{1})^{-1}\mathbf{1}'\mathbf{G_y}\mathbf{y}$. A particular choice of the matrix of the metric, also called "*weight matrix*", is $\mathbf{G_y} = \mathrm{Diag}(1, 1, 1, 1, 1, x)$ such that

$$weighted\ mean\,\mathbf{y} = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 x}{5 + x},$$

where x is the *unknown weight* of the *extreme value* ("outlier") y_6. A special robust design of the *weighted mean* $\mathbf{y}$ is the *median* $\mathbf{y}$, namely

$$weighted\ mean\,\mathbf{y} = med\ \mathbf{y}$$

such that

$$x = \frac{y_1 + y_2 + y_3 + y_4 + y_5 - 5\,med\ \mathbf{y}}{med\ \mathbf{y} - y_6}$$

here

$$x = 0.024, 390, 243 \sim \frac{24}{1000}.$$

Indeed the weighted mean with respect to the weight matrix $\mathbf{G_y} = \mathrm{Diag}(1, 1, 1, 1, 1, 24/1000)$ *reproduces* the median of the second data set. The extreme value has been down-weighted by a weight $24/1000$ approximately. *Four*, with respect to the

Table 4.3 "direct" observations, comparison two data sets by means of med y, mad y (I-LESS, $\mathbf{G_y}$-LESS), r.m.s. (I-BIQUUE)

data set one	*data set two*
$n = 5$ ("odd")	$n = 6$ ("even")
$n/2 = 2.5, [n/2] = 2$	$n/2 = 3, n/2 + 1 = 4$
$[n/2] + 1 = 3$	
$med\mathbf{y} = y_{(3)} = 13$	$med\mathbf{y} = 13.5$
$mad\mathbf{y} = 1$	$mad\mathbf{y} = 1.5$
$mean\ \mathbf{y}(I\text{-LESS}) = 13$	$mean\ \mathbf{y}(I\text{-LESS}) = 30.1\bar{6}$
	weighted mean y
	$(\mathbf{G_y}\text{-LESS}) = 13.5$
	$\mathbf{G_y} = \mathrm{Diag}(1, 1, 1, 1, 1, \frac{24}{1000})$
$\hat{\mu}(\mathbf{I}\text{-}BLUUE) = 13$	$\hat{\mu}(\mathbf{I}\text{-}BLUUE) = 30.1\bar{6}$
$\hat{\sigma}^2(\mathbf{I}\text{-}BIQUUE) = 2.5$	$\hat{\sigma}^2(\mathbf{I}\text{-}BIQUUE) = 1770.1$
r.m.s. $(\mathbf{I}\text{-}BIQUUE) =$	r.m.s. $(\mathbf{I}\text{-}BIQUUE) =$
$\hat{\sigma}(\mathbf{I}\text{-}BIQUUE) = 1.6$	$\hat{\sigma}(\mathbf{I}\text{-}BIQUUE) = 42.1$

simple linear model $E\{\mathbf{y}\} = \mathbf{1}\mu$, $D\{\mathbf{y}\} = \mathbf{I}\sigma^2$ we compute I-BLUUE of μ and I-BIQUUE of σ^2, namely

$$\hat{\mu} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \tfrac{1}{n}\mathbf{1}'\mathbf{y}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\mathbf{y}'\left[\mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\right]\mathbf{y} = \frac{1}{n-1}\mathbf{y}'\left[\mathbf{I} - \tfrac{1}{n}\mathbf{1}\mathbf{1}'\right]\mathbf{y} = \frac{1}{n-1}(\mathbf{y} - \mathbf{1}\hat{\mu})'(\mathbf{y} - \mathbf{1}\hat{\mu}).$$

Obviously the extreme value y_6 in the second data set has spoiled the specification of the *simple linear model*. The r.m.s. (I-BLUUE) = 1.6 of the first data set is increased to the r.m.s. (I-BIQUUE) = 42.1 of the second data set.

Five, we setup the alternative linear model for the second data set, namely

$$E\left\{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ \hline y_6 \end{bmatrix}\right\} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \hline \mu_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \hline \mu + \nu \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\mu + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\nu$$

$$E\{\mathbf{y}\} = \mathbf{A}\xi : \begin{bmatrix} \mathbf{A} := [\mathbf{1}, \mathbf{a}] \in^{5\times 2}, \mathbf{1} := [1,1,1,1,1,1]' \in \mathbb{R}^{6\times 1} \\ \xi := [\mu, \nu]' \in^{2\times 1}, \mathbf{a} := [0,0,0,0,0,1]' \in \mathbb{R}^{6\times 1} \end{bmatrix}$$

$$D\{\mathbf{y}\} = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{bmatrix}\sigma^2 \in \mathbb{R}^{6\times 6}$$

$$D\{\mathbf{y}\} = \mathbf{I_6}\sigma^2, \sigma^2 \in \mathbb{R}^+,$$

adding to the observation y_6 the bias term ν. Still we assume the variance-covariance matrix $D\{\mathbf{y}\}$ of the observation vector $y \in \mathbb{R}^{6\times 6}$ to be isotropic with one variance component as an unknown. $(\hat{\mu}, \hat{\nu})$ is $\mathbf{I_6}$-BLUUE if

$$\begin{bmatrix} \hat{\mu} \\ \hat{\nu} \end{bmatrix} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{\nu} \end{bmatrix} = \begin{bmatrix} 13 \\ 103 \end{bmatrix}$$

$$\hat{\mu} = 13, \hat{\nu} = 103, \mu_1 = \hat{\mu} = 13, \hat{y}_2 = 116$$

$$D\left\{\begin{bmatrix}\hat{\mu}\\ \hat{\nu}\end{bmatrix}\right\} = (\mathbf{A}'\mathbf{A})^{-1}\sigma^2$$

$$D\left\{\begin{bmatrix}\hat{\mu}\\ \hat{\nu}\end{bmatrix}\right\} = \frac{\sigma^2}{5}\begin{bmatrix}1 & -1\\ -1 & 6\end{bmatrix}$$

$$\sigma^2_{\hat{\mu}} = \frac{\sigma^2}{5}, \sigma^2_{\hat{\nu}} = \frac{6}{5}\sigma^2, \sigma_{\hat{\mu}\hat{\nu}} = -\frac{1}{5}\sigma^2$$

$$\hat{\sigma}^2$$

is $\mathbf{I_6}$-BIQUUE if

$$\hat{\sigma}^2 = \frac{1}{n - \mathrm{rk}\mathbf{A}}\mathbf{y}'\left[\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{y}$$

$$\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{5}\begin{bmatrix}4 & -1 & -1 & -1 & -1 & 0\\ -1 & 4 & -1 & -1 & -1 & 0\\ -1 & -1 & 4 & -1 & -1 & 0\\ -1 & -1 & -1 & 4 & -1 & 0\\ -1 & -1 & -1 & -1 & 4 & 0\\ 0 & 0 & 0 & 0 & 0 & 5\end{bmatrix}$$

$$\mathbf{r}_i := \left[\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]_{ii} = \left(\frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, \frac{4}{5}, 1\right) \forall i \in \{1, \ldots, 6\}$$

are the *redundancy numbers*.

$$\mathbf{y}'(\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} = 13{,}466$$

$$\hat{\sigma}^2 = \frac{1}{4}13{,}466 = 3{,}366.5, \hat{\sigma} = 58.02$$

$$\sigma^2_{\hat{\mu}}(\hat{\sigma}^2) = \frac{3{,}366.5}{5} = 673.3, \sigma_{\hat{\mu}}(\hat{\sigma}) = 26$$

$$\sigma^2_{\hat{\nu}}(\hat{\sigma}^2) = \frac{6}{5}3{,}366.5 = 4{,}039.8, \sigma_{\hat{\nu}}(\hat{\sigma}) = 63.6.$$

Indeed the r.m.s. value of the partial mean $\hat{\mu}$ as well as of the estimated bias $\hat{\nu}$ have changed the results remarkably, namely from r.m.s. (simple linear model) 42.1 to r.m.s. (linear model) 26. A r.m.s. value of the bias $\hat{\nu}$ in the order of 63.6 has been documented. Finally let us compute the empirical "error vector" $\tilde{l}$ and is variance-covariance matrix by means of

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \left[\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\mathbf{y},$$

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \left[\mathbf{I_6} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right]\sigma^2,$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \left[2 \; -1 \; 1 \; -2 \; 0 \; 116\right]'$$

$$D\left\{\tilde{l}\right\} = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 & 0 \\ -1 & 4 & -1 & -1 & -1 & 0 \\ -1 & -1 & 4 & -1 & -1 & 0 \\ -1 & -1 & -1 & 4 & -1 & 0 \\ -1 & -1 & -1 & -1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix} \frac{\sigma^2}{5}$$

$$D\{\tilde{l}|\hat{\sigma}^2\} = 673.3 * \begin{bmatrix} 4 & -1 & -1 & -1 & -1 & 0 \\ -1 & 4 & -1 & -1 & -1 & 0 \\ -1 & -1 & 4 & -1 & -1 & 0 \\ -1 & -1 & -1 & 4 & -1 & 0 \\ -1 & -1 & -1 & -1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

4-14 Alternative Estimation Maximum Likelihood (MALE)

Maximum Likelihood Estimation ("MALE") is a competitor to BLUUE of the first moments $E\{\mathbf{y}\}$ and to the BIQUUE of the second central moments $D\{\mathbf{y}\}$ of a random variable $y \in \{\mathbb{Y}, pdf\}$, which we like to present at least by an example.

Maximum Likelihood Estimation
:linear model:
$E\{\mathbf{y}\} = \mathbf{1}_n\mu, D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$
"independent, identically normal distributed observations"
$[y_1, \ldots, y_n]'$
"direct observations"
unknown parameter: $\{\mu, \sigma\} \in \{\mathbb{R}, \mathbb{R}^+\} =: \mathbb{X}$
"simultaneous estimations of $\{\mu, \sigma^2\}$".

Given the above linear model of independent, identically, normal distributed observations $[\mathbf{y_1}, \ldots, \mathbf{y}_n]' = y \in \{\mathbb{R}^n, \; pdf\}$. The first moment μ as well as the central second moment σ^2 constitute the *unknown parameters* $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$ where $\mathbb{R} \times \mathbb{R}^+$ is the *admissible parameter space*. The estimation of the unknown parameters (μ, σ^2) is based on the following *optimization problem*

Maximize the log-likelihood function

$$\ln f(y_1, \; \ldots, \; y_n \left|\mu, \sigma^2\right.) = \ln \prod_{i=1}^{n} f(y_i \left|\mu, \sigma^2\right.)$$

$$= \ln \left\{ \frac{1}{(2\pi)^{\frac{n}{2}\sigma^n}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right) \right\} =$$

$$= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 = \max_{\mu, \sigma^2}$$

of the independent, identically normal distributed random variables $\{y_1, \ldots, y_n\}$. The log-likelihood function is *simple* if we introduce the first sample moment m_1 and the second sample moment m_2, namely

$$m_1 := \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \mathbf{1}' y, m_2 := \frac{1}{n} \sum_{i=1}^{n} y_i^2 = \frac{1}{n} \mathbf{y}'\mathbf{y}$$

$$\ln f\left(y_1, \ldots, y_n \left|\mu, \sigma^2\right.\right) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} \left(m_2 - 2m_1\mu + \mu^2\right),$$

Now we are able to define the optimization problem

$$\ell\left(\mu, \sigma^2\right) := \ln f\left(y_1, \ldots, y_n \left|\mu, \sigma^2\right.\right) = \max_{\mu,\, \sigma^2}$$

more precisely.

Definition. (Maximum Likelihood Estimation, linear model $E\{\mathbf{y}\} = \mathbf{1}_n\mu$, $D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$, independent, identically normal distributed observations $\{y_1, \ldots, y_n\}$):

A 2x1 vector$[\mu_\ell, \sigma_\ell^2]'$ is called MALE of $[\mu, \sigma^2]'$, (**M**aximum **L**ikelihood **E**stimation) with respect to the linear model 0.1 if its log-likelihood function

$$\begin{aligned} \ell(\mu, \sigma^2) &:= \ln f(y_1, \ldots, y_n \left|\mu, \sigma^2\right.) = \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{n}{2\sigma^2} (m_2 - 2m_1 + \mu^2) \end{aligned}$$

is minimal. The simultaneous estimation of $(\mu,\ \sigma^2)$ of type MALE can be characterized as following.

Corollary. (MALE with respect to the linear model $E\{\mathbf{y}\} = \mathbf{1}_n\mu, D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$, independent identically normal distributed observations $\{y_1, \ldots, y_n\}$):

The log-likelihood function $\ell\left(\mu, \sigma^2\right)$ with respect to the linear model $E\{\mathbf{y}\} = 1_n\mu, D\{\mathbf{y}\} = I_n\sigma^2$,$(\mu, \sigma^2 \in \mathbb{R} \times \mathbb{R}^+)$, of independent, identically normal distributed observations $\{y_1, \ldots, y_n\}$ is maximal if

$$\mu_\ell = m_1 = \frac{1}{n} 1' y, \sigma_\ell^2 = m_2 - m_1^2 = \frac{1}{n} (y - y_\ell)'(y - y_\ell)$$

is a simultaneous estimation of the mean volume (first moment) μ_ℓ and of the variance (second moment) σ_ℓ^2.

Proof. The *Lagrange function*

$$L(\mu,\sigma^2) := -\frac{n}{2}\ln\sigma^2 - \frac{n}{2\sigma^2}(m_2 - 2m_1\mu + \mu^2) = \max_{\mu,\sigma^2}$$

leads to the *necessary conditions*

$$\begin{aligned}
\frac{\partial L}{\partial\mu}(\mu,\sigma^2) &= \frac{nm_1}{\sigma^2} = \frac{n\mu}{\sigma^2} = 0\\
\frac{\partial L}{\partial\sigma^2}(\mu,\sigma^2) &= -\frac{n}{2\sigma^2} + \frac{n}{2\sigma^4}(m_2 - 2\mu m_1 + \mu^2) = 0,
\end{aligned}$$

also called the *likelihood normal equations*. Their solution is

$$\begin{bmatrix}\mu_1\\ \sigma_\ell^2\end{bmatrix} = \begin{bmatrix}m_1\\ m_2 - m_1^2\end{bmatrix} = \frac{1}{n}\begin{bmatrix}\mathbf{1}'\mathbf{y}\\ \mathbf{y}'\mathbf{y} - (\mathbf{1}'\mathbf{y})^2\end{bmatrix}.$$

The matrix of *second derivates* constitutes as a negative matrix the *sufficiency conditions*.

$$\frac{\partial^2 L}{\partial(\mu,\sigma^2)\partial(\mu,\sigma^2)'}(\mu_\ell,\sigma_\ell) = -n\begin{bmatrix}\frac{1}{\sigma_\ell^2} & 0\\ 0 & \frac{1}{\sigma_\ell^4}\end{bmatrix} > 0.$$

Finally we can immediately check that $\ell(\mu,\sigma^2) \to -\infty$ as (μ,σ^2) approaches the *boundary of the parameter space*. If the log-likelihood function is sufficiently regular, we can expand it as

$$\begin{aligned}
\ell(\mu,\ \sigma^2) = \ell(\mu_\ell,\ \sigma_\ell^2) + D\ell(\mu_\ell,\ \sigma_\ell^2)\begin{bmatrix}\mu-\mu_\ell\\ \sigma^2-\sigma_\ell^2\end{bmatrix} +\\
+\frac{1}{2}D^2\ell(\mu_\ell,\ \sigma_\ell^2)\begin{bmatrix}\mu-\mu_\ell\\ \sigma^2-\sigma_\ell^2\end{bmatrix}\otimes\begin{bmatrix}\mu-\mu_\ell\\ \sigma^2-\sigma_\ell^2\end{bmatrix} + O_3.
\end{aligned}$$

Due to the likelihood normal likelihood equations $D\ell(\mu_\ell,\sigma_\ell^2)$ vanishes. Therefore the behavior of $\ell(\mu,\ \sigma^2)$ near$(\mu_\ell,\ \sigma_\ell^2)$ is largely determined by $D^2\ell(\mu_\ell,\ \sigma_\ell^2) \in \mathbb{R}\times\mathbb{R}^+$, which is a measure of the local curvature the log-likelihood function $\ell(\mu,\ \sigma^2)$. The *non-negative Hesse matrix of second derivatives*

$$I(\mu_\ell,\ \sigma_\ell^2) = -\frac{\partial^2\ell}{\partial(\mu,\ \sigma^2)\partial(\mu,\ \sigma^2)'}(\mu_\ell,\ \sigma_\ell^2) > 0$$

is called *observed Fischer information*. It can be regarded as an *index* of the steepness of the log-likelihood function moving away from (μ,σ^2), and as an

Table 4.4 $(\mu_\ell, \sigma_\ell^2)$ MALE of $(\mu, \sigma^2) \in \{\mathbb{R}, \mathbb{R}^+\}$: the front page examples

	μ_ℓ	σ_ℓ^2	$\lvert\sigma_\ell\rvert$
1st example (n = 5)	13	2	$\sqrt{2}$
2nd example (n = 6)	30.16	1474.65	36.40

indicator of the *strength of preference* for the MLE point with respect to the other points of the parameter space. Finally, compare by means of *Table 4.4* $(\mu_\ell, \sigma_\ell^2)$ MALE of (μ, σ^2) for the front page example of *Tables 4.1* and *4.2*

4-2 Setup of the Best Linear Uniformly Unbiased Estimator

of type BLUUE for the moments of first order

Let us introduce the *special Gauss–Markov model* $\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$ specified in *Box 4.1*, which is given for the *first order moments* in the form of a *inconsistent system of linear equations relating the first non-stochastic* ("fixed"), *real-valued vector* ξ of unknowns to the expectation $E\{\mathbf{y}\}$ of the *stochastic*, real-valued vector $\mathbf{y}$ of observations, $\mathbf{A}\xi = E\{\mathbf{y}\}$, since $E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A})$ is an element of the column space $\mathcal{R}(\mathbf{A})$ of the real-valued, *non-stochastic* ("fixed") "*first order design matrix*" $\mathbf{A} \in \mathbb{R}^{n \times m}$. The rank of the fixed matrix $\mathbf{A}$, rk$\mathbf{A}$, equals the number m of unknowns, $\xi \in \mathbb{R}^m$. In addition, the *second order central moments*, the *regular* variance-covariance matrix $\mathbf{\Sigma_y}$, also called *dispersion matrix* $D\{\mathbf{y}\}$ constitute the second matrix $\mathbf{\Sigma_y} \in \mathbb{R}^{n \times m}$ of *unknowns* to be specified as a linear model further on.

Box 4.1. (Special Gauss–Markov model $\mathbf{y} = \mathbf{A}\xi + \mathbf{e}$)

1st moments

$$\mathbf{A}\xi = E\{\mathbf{y}\},\ \mathbf{A} \in \mathbb{R}^{n \times m},\ E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A}),\ \text{rk}\mathbf{A} = m \tag{4.1}$$

2nd moments

$$\mathbf{\Sigma_y} = D\{\mathbf{y}\} \in \mathbb{R}^{n \times n},\ \mathbf{\Sigma_y}\ \text{positive definite},\ \text{rk}\mathbf{\Sigma_y} = n \tag{4.2}$$

$$\mathbf{\Xi},\ E\{\mathbf{y}\},\ \mathbf{y} - E\{\mathbf{y}\} = \mathbf{e}\ unknown$$
$$\mathbf{\Sigma_y}\ unknown\,.$$

4-21 The Best Linear Uniformly Unbiased Estimation $\hat{\xi}$ of ξ : $\Sigma_{\mathbf{y}}$-BLUUE

Since we are dealing with a linear model, it is "a *natural choice*" to *setup a linear form to estimate the parameters* ξ of fixed effects, namely

$$\hat{\xi} = \mathbf{Ly} + \kappa, \tag{4.3}$$

where $\{\mathbf{L} \in \mathbb{R}^{m \times n}, \ \kappa \in \mathbb{R}^{m}\}$ are fixed unknowns. In order to determine the real-valued $m \times n$ matrix $\mathbf{L}$ and the real-valued $m \times 1$ vector κ, independent of the variance-covariance matrix $\Sigma_{\mathbf{y}}$, the *inhomogeneous linear estimation* $\hat{\xi}$ of the vector ξ of fixed effects has to fulfil certain *optimality conditions*.

(1st) $\hat{\xi}$ is an inhomogeneous linear unbiased estimation of ξ

$$E\{\hat{\xi}\} = E\{\mathbf{Ly} + \kappa\} = \xi \textit{ for all } \xi \in \mathbb{R}^{m}, \tag{4.4}$$

and (2nd) in comparison to all other linear uniformly unbiased estimations $\hat{\xi}$ has minimum variance

$$\begin{aligned} \operatorname{tr} D\{\hat{\xi}\} &:= E\{(\hat{\xi} - \xi)'(\hat{\xi} - \xi)\} = \\ &= \operatorname{tr} \mathbf{L}\, \Sigma_{\mathbf{y}}\, \mathbf{L}' = ||\mathbf{L}'||_{\Sigma} = \min_{\mathbf{L}}. \end{aligned} \tag{4.5}$$

First the condition of a linear uniformly unbiased estimation $E\{\hat{\xi}\} = E\{\mathbf{Ly}+\kappa\} = \xi$ with respect to the *Special Gauss–Markov model* (4.1), (4.2) has to be considered in more detail. As soon as we substitute the linear model (4.1) into the postulate of uniformly unbiasedness (4.4) we are led to

$$E\{\hat{\xi}\} = E\{\mathbf{Ly} + \kappa\} = \mathbf{L}E\{\mathbf{y}\} + \kappa = \xi \textit{ for all } \xi \in \mathbb{R}^{m} \tag{4.6}$$

and

$$\mathbf{LA}\xi + \kappa = \xi \textit{ for all } \xi \in \mathbb{R}^{m}. \tag{4.7}$$

Beside $\kappa = 0$ the postulate of linear uniformly unbiased estimation with respect to the *special Gauss–Markov model* (4.1), (4.2) leaves us with one condition, namely

$$\mathbf{LA} - \mathbf{I}_m \xi = 0 \textit{ for all } \xi \in \mathbb{R}^{m} \tag{4.8}$$

or

$$\mathbf{LA} - \mathbf{I}_m = 0. \tag{4.9}$$

Note that there are *locally unbiased estimations* such that $(\mathbf{LA} - \mathbf{I}_m)\xi_0 = 0$ for $\mathbf{LA} - \mathbf{I}_m \neq 0$. Alternatively, *B. Schaffrin* (2000) has *softened the constraint of unbiasedness* (4.9) by replacing it by the *stochastic matrix constraint* $\mathbf{A}'\mathbf{L}' = \mathbf{I}_m +$

$\mathbf{E}_0$ subject to $E\{\text{vec}\,\mathbf{E}_0\} = 0$, $D\{\text{vec}\,\mathbf{E}_0\} = (\mathbf{I}_m \otimes \mathbf{\Sigma}_0)$, $\mathbf{\Sigma}_0$ a *positive definite* matrix. For $\mathbf{\Sigma}_0 \to 0$, uniform unbiasedness is restored. Estimators which fulfill the stochastic matrix constraint $\mathbf{A}'\mathbf{L}' = \mathbf{I}_m + \mathbf{E}_0$ for finite $\mathbf{\Sigma}_0$ are called "*softly unbiased*" or "*unbiased in the mean*".

Second, the choice of norm for "*best*" of type *minimum variance* has to be discussed more specifically. Under the condition of a *linear uniformly unbiased estimation* let us derive the specific representation of the *weighted Frobenius matrix norm* of $\mathbf{L}'$. Indeed let us define the *dispersion matrix*

$$\begin{aligned} D\{\hat{\xi}\} &:= E\{(\hat{\xi} - E\{\hat{\xi}\})(\hat{\xi} - E\{\hat{\xi}\})'\} = \\ &= E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\}\,, \end{aligned} \tag{4.10}$$

which by means of the inhomogeneous linear form $\hat{\xi} = \mathbf{L}\mathbf{y} + \kappa$ is specified to

$$D\{\hat{\xi}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' \tag{4.11}$$

and

Definition 4.1. ($\hat{\xi}$ $\mathbf{\Sigma}_\mathbf{y}$-BLUUE of ξ):

An $m1$ vector $\hat{\xi} = \mathbf{L}\mathbf{y} + \kappa$ is called $\mathbf{\Sigma}_\mathbf{y}$-BLUUE of ξ (*Best Linear Uniformly Unbiased Estimation*) with respect to the $\mathbf{\Sigma}_\mathbf{y}$-norm in (4.1) if

(1st) $\hat{\xi}$ is *uniformly unbiased* in the sense of

$$\text{tr}\,D\{\hat{\xi}\} := \text{tr}\mathbf{L}\,D\{\mathbf{y}\}\,\mathbf{L}' = ||\mathbf{L}'||_{\mathbf{\Sigma}_\mathbf{y}}\,. \tag{4.12}$$

Now we are prepared for

Lemma 4.2. ($\hat{\xi}$ $\mathbf{\Sigma}_\mathbf{y}$-BLUUE of ξ):

An $m1$ vector $\hat{\xi} = \mathbf{L}\mathbf{y} + \kappa$ is $\mathbf{\Sigma}_\mathbf{y}$-BLUUE of ξ in (4.1) if and only if

$$\kappa = 0 \tag{4.13}$$

holds and the matrix L fulfils the system of normal equations

$$\begin{bmatrix} \Sigma_\mathbf{y} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{L}' \\ \mathbf{\Lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_m \end{bmatrix} \tag{4.14}$$

or

$$\Sigma_\mathbf{y}\mathbf{L}' + \mathbf{A}\Lambda = \mathbf{0} \tag{4.15}$$

and

$$\mathbf{A}'\mathbf{L}' = \mathbf{I}_m \tag{4.16}$$

with the $m \times m$ matrix of "*Lagrange multipliers*".

Proof. Due to the postulate of an *inhomogeneous linear uniformly unbiased estimation* with respect to the parameters $\xi \in \mathbb{R}^m$ of the *special Gauss–Markov model* we were led to $\kappa = 0$ and one *conditional constraint* which makes it plausible to minimize the *constraint Lagrangean*

$$\mathcal{L}(\mathbf{L}, \boldsymbol{\Lambda}) := \operatorname{tr} \mathbf{L}\boldsymbol{\Sigma}_\mathbf{y}\mathbf{L}' + 2\operatorname{tr} \boldsymbol{\Lambda}(\mathbf{A}'\mathbf{L}' - \mathbf{I}_m) = \min_{\mathbf{L},\boldsymbol{\Lambda}} \tag{4.17}$$

for $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE. The *necessary conditions* for the minimum of the *quadratic constraint Lagrangean* $\mathcal{L}(\mathbf{L}, \boldsymbol{\Lambda})$ are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}, \hat{\boldsymbol{\Lambda}}) := 2(\boldsymbol{\Sigma}_\mathbf{y}\hat{\mathbf{L}}' + \mathbf{A}\hat{\boldsymbol{\Lambda}})' = 0 \tag{4.18}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Lambda}}(\hat{\mathbf{L}}, \hat{\boldsymbol{\Lambda}}) := 2(\mathbf{A}'\hat{\mathbf{L}}' - \mathbf{I}_m) = 0 \tag{4.19}$$

which agree to the *normal equations* (4.14).

The second derivatives

$$\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\mathbf{L})\partial(\operatorname{vec}\mathbf{L})'}(\hat{\mathbf{L}}, \hat{\boldsymbol{\Lambda}}) = 2(\boldsymbol{\Sigma}_\mathbf{y} \otimes \mathbf{I}_m) > 0 \tag{4.20}$$

constitute the *sufficiency conditions* due to the *positive-definiteness* of the matrix $\boldsymbol{\Sigma}$ for $\mathcal{L}(\mathbf{L}, \boldsymbol{\Lambda}) = \min$ ♣.

Obviously, a *homogeneous linear form* $\hat{\xi} = \mathbf{Ly}$ is sufficient to generate $\boldsymbol{\Sigma}$-BLUUE for the *special Gauss–Markov model* (4.1), (4.2). Explicit representations of $\boldsymbol{\Sigma}$-BLUUE of type $\hat{\xi}$ as well as of its dispersion matrix $D\{\hat{\xi}|\hat{\xi}\ \boldsymbol{\Sigma}_\mathbf{y}$-BLUUE generated by solving the normal equations (4.14) are collected in

Theorem 4.3. ($\hat{\xi}$ $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ):

Let $\hat{\xi} = \mathbf{Ly}$ be $\boldsymbol{\Sigma}$-BLUUE of ξ, in the *special linear Gauss–Markov model* (4.1), (4.2). Then

$$\hat{\xi} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y} \tag{4.21}$$

$$\hat{\xi} = \boldsymbol{\Sigma}_{\hat{\xi}}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y} \tag{4.22}$$

are equivalent to the representation of the solution of the normal equations (4.14) subjected to the related dispersion matrix

$$D\{\hat{\xi}\} := \boldsymbol{\Sigma}_{\hat{\xi}} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}.$$

Proof. We shall present two proofs of the above theorem: The first one is based upon *Gauss elimination* in solving the normal equations (4.14), the second one uses the power of the IPM method (Inverse Partitioned Matrix, *C.R. Rao's* Pandora Box).

(i) *forward step (Gauss elimination)*:

Multiply the first normal equation by $\mathbf{\Sigma}_\mathbf{y}^{-1}$, multiply the reduced equation by $\mathbf{A}'$ and subtract the result from the second normal equation. Solve for $\mathbf{\Lambda}$

$$\left.\begin{array}{l} \Sigma_\mathbf{y}\hat{\mathbf{L}}' + \mathbf{A}\mathbf{\Lambda} = \mathbf{0} \text{ (first equation:} \\ \text{multiply by} -\mathbf{A}'\Sigma_\mathbf{y}^{-1}) \\ \quad \mathbf{A}\hat{\mathbf{L}}' = \mathbf{I}_m \text{ (second equation)} \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \left.\begin{array}{c} -\mathbf{A}'\hat{\mathbf{L}}' - \mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A}\hat{\mathbf{\Lambda}} = \mathbf{0} \\ \mathbf{A}'\hat{\mathbf{L}}' = \mathbf{I}_m \end{array}\right] \Leftrightarrow$$

$$\Leftrightarrow \quad -\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A}\hat{\mathbf{\Lambda}} = \mathbf{I} \quad \Rightarrow$$

$$\Rightarrow \quad \hat{\mathbf{\Lambda}} = -(\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1} \tag{4.23}$$

(ii) *backward step (Gauss elimination)*:

Substitute $\hat{\mathbf{\Lambda}}$ in the modified first normal equation and solve for $\hat{\mathbf{L}}$.

$$\Leftrightarrow \left.\begin{array}{c} \hat{\mathbf{L}}' + \Sigma_\mathbf{y}^{-1}\mathbf{A}\hat{\mathbf{\Lambda}} = \mathbf{0} \Leftrightarrow \hat{\mathbf{L}} = \hat{\mathbf{\Lambda}}'\boldsymbol{A}'\Sigma_\mathbf{y}^{-1} \\ \hat{\mathbf{\Lambda}} = -(\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1} \end{array}\right] \Leftrightarrow$$

$$\Rightarrow \quad \hat{\mathbf{L}} = (\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_\mathbf{y}^{-1}. \tag{4.24}$$

(iii) *IPM (Inverse Partitioned Matrix)*:

Let us partition the symmetric matrix of the normal equations (4.14)

$$\begin{bmatrix} \Sigma_\mathbf{y} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{0} \end{bmatrix}.$$

$$\begin{bmatrix} \Sigma_\mathbf{y} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{bmatrix}$$

$$\begin{aligned} \mathbf{B}_{11} &= \mathbf{I}_m - \Sigma_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_\mathbf{y}^{-1} \\ \mathbf{B}'_{12} &= (\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_\mathbf{y}^{-1} \\ \mathbf{B}_{22} &= -(\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A})^{-1}. \end{aligned}$$

The normal equations are now solved by

$$\begin{bmatrix} \hat{\mathbf{L}}' \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_m \end{bmatrix}$$

$$\begin{aligned} \hat{\mathbf{L}} &= \mathbf{B}'_{12} = (\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1} \\ \hat{\boldsymbol{\Lambda}} &= \mathbf{B}_{22} = -(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}. \end{aligned} \tag{4.25}$$

(iv) dispersion matrix The related dispersion matrix is computed by means of the "*Error Propagation Law*".

$$D\{\hat{\xi}\} = D\{\mathbf{L}\mathbf{y} \mid \hat{\mathbf{L}} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\} = \hat{\mathbf{L}}D\{\mathbf{y}\}\hat{\mathbf{L}}'$$

$$D\{\hat{\xi}\} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}$$

$$D\{\hat{\xi}\} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}. \tag{4.26}$$

Here is our proof's end.

By means of *Theorem 4.3* we succeeded to produce $\hat{\xi}$-BLUUE of ξ. In consequence, we have to estimate $\widehat{E\{\mathbf{y}\}}$ as $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of $E\{\mathbf{y}\}$ as well as the "*error vector*"

$$\mathbf{e}_{\mathbf{y}} := \mathbf{y} - E\{\mathbf{y}\} \tag{4.27}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} := \mathbf{y} - \widehat{E\{\mathbf{y}\}} = \mathbf{y} - \mathbf{A}\hat{\xi} = (\mathbf{I}_n - \mathbf{A}\mathbf{L})\mathbf{y} \tag{4.28}$$

out of

Lemma 4.4. ($\widehat{E\{\mathbf{y}\}}$ $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of $E\{\mathbf{y}\}$, $\tilde{\mathbf{e}}_{\mathbf{y}}$, $D\{\tilde{\mathbf{e}}_{\mathbf{y}}\}$, $D\{\mathbf{y}\}$):

(i) Let $\widehat{E\{\mathbf{y}\}}$ be $\boldsymbol{\Sigma}$-BLUUE of $E\{\mathbf{y}\} = \mathbf{A}\xi$ with respect to the *special Gauss–Markov model* (4.1), (4.2), then

$$\widehat{E\{\mathbf{y}\}} = \mathbf{A}\hat{\xi} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y} \tag{4.29}$$

leads to the singular variance-covariance matrix (dispersion matrix)

$$D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'. \tag{4.30}$$

(ii) If the error vector $\mathbf{e}$ is empirically determined, we receive for

$$\tilde{\mathbf{e}}_{\mathbf{y}} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}]\mathbf{y} \tag{4.31}$$

and its singular variance-covariance matrix (dispersion matrix)

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \boldsymbol{\Sigma}_{\mathbf{y}} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}', \quad \mathrm{rk}\mathbf{D}\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = n - m \tag{4.32}$$

(iii) The dispersion matrices of the *special Gauss–Markov model* (4.1), (4.2) are related by

$$D\{\mathbf{y}\} = D\{\mathbf{A}\hat{\xi} + \tilde{\mathbf{e}}_\mathbf{y}\} = D\{\mathbf{A}\hat{\xi}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}$$
$$= D\{\tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}, \tag{4.33}$$

$$C\{\tilde{\mathbf{e}}_\mathbf{y},\ \mathbf{A}\hat{\xi}\} = 0,\ C\{\tilde{\mathbf{e}}_\mathbf{y},\ \tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} = 0. \tag{4.34}$$

$\tilde{\mathbf{e}}_\mathbf{y}$ and $\mathbf{A}\hat{\xi}$ are *uncorrelated.*

Proof.

$$(i)\ \widehat{E\{\mathbf{y}\}} = \mathbf{A}\hat{\xi} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y}$$

As soon as we implement $\hat{\xi}$ $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ, namely (4.21), into $\mathbf{A}\hat{\xi}$ we are directly led to the desired result.

$$(ii)\ \ D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'$$

$\hat{\xi}$ $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ, namely (4.21), implemented in

$$D\{\mathbf{A}\hat{\xi}\} :\, = E\{\mathbf{A}(\hat{\xi} - E\{\hat{\xi}\})(\hat{\xi} - E\{\hat{\xi}\})'\mathbf{A}'\}$$
$$D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}E\{(\hat{\xi} - E\{\hat{\xi}\})(\hat{\xi} - E\{\hat{\xi}\})'\}\mathbf{A}'$$
$$D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}E\{(\mathbf{y} - E\{\mathbf{y}\})(\mathbf{y} - E\{\mathbf{y}\})'\}\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'$$
$$D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'$$
$$D\{\mathbf{A}\hat{\xi}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'$$

leads to the proclaimed result.

$$(iii)\ \ \tilde{\mathbf{e}}_\mathbf{y} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]\mathbf{y}.$$

Similarly if we substitute $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ, namely (4.21), in

$$\tilde{\mathbf{e}}_\mathbf{y} = \mathbf{y} - \widehat{E\{\mathbf{y}\}} = \mathbf{y} - \mathbf{A}\hat{\xi} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]\mathbf{y}$$

we gain what we wanted!

$$(iv)\ \ D\{\hat{\mathbf{e}}_\mathbf{y}\} = \boldsymbol{\Sigma} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} := E\{(\tilde{\mathbf{e}}_\mathbf{y} - E\{\tilde{\mathbf{e}}_\mathbf{y}\})(\tilde{\mathbf{e}}_\mathbf{y} - E\{\tilde{\mathbf{e}}_\mathbf{y}\})'\}.$$

As soon as we substitute

$$E\{\tilde{\mathbf{e}}_\mathbf{y}\} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]E\{\mathbf{y}\}$$

in the definition of the dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$, we are led to

$$\begin{aligned}
D\{\tilde{\mathbf{e}}_\mathbf{y}\} :&= [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]\,\boldsymbol{\Sigma}\,[\mathbf{I}_n - \boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'],\\
D\{\tilde{\mathbf{e}}_\mathbf{y}\} &= [\boldsymbol{\Sigma}_\mathbf{y} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}']\,[\mathbf{I}_n - \boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}']\\
&= \boldsymbol{\Sigma}_\mathbf{y} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}' - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}' + \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\\
&= \boldsymbol{\Sigma}_\mathbf{y} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'.\\
\mathrm{rk}D\{\tilde{\mathbf{e}}_\mathbf{y}\} &= \mathrm{rk}D\{\mathbf{y}\} - \mathrm{rk}\mathbf{A}(\mathbf{A}\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}' = n - m.\\
(v)\ \ D\{\mathbf{y}\} &= D\{\mathbf{A}\hat{\xi} + \tilde{\mathbf{e}}_\mathbf{y}\} = D\{\mathbf{A}\hat{\xi}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\} = D\{\tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}\\
\mathbf{y} - E\{\mathbf{y}\} &= \mathbf{y} - \mathbf{A}\xi = \mathbf{y} - \mathbf{A}\hat{\xi} + \mathbf{A}(\hat{\xi} - \xi)\\
\mathbf{y} - E\{\mathbf{y}\} &= \mathbf{A}(\hat{\xi} - \xi) + \tilde{\mathbf{e}}_\mathbf{y}.
\end{aligned}$$

The additive decomposition of the residual vector $\mathbf{y} - E\mathbf{y}$ left us with two terms, namely the predicted residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and the term which is a linear functional of $\hat{\xi} - \xi$. The corresponding *product decomposition*

$$\begin{aligned}
[\mathbf{y} - E\{\mathbf{y}\}][\mathbf{y} - E\{\mathbf{y}\}]' &= \mathbf{A}(\hat{\xi} - \xi)(\hat{\xi} - \xi)' + \mathbf{A}(\hat{\xi} - \xi)\tilde{\mathbf{e}}_\mathbf{y}'\\
&\quad + \tilde{\mathbf{e}}_\mathbf{y}(\hat{\xi} - \xi)'\mathbf{A}' + \tilde{\mathbf{e}}_\mathbf{y}\tilde{\mathbf{e}}_\mathbf{y}'
\end{aligned}$$

for $\hat{\xi}$ $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ, in particular $E\{\hat{\xi}\} = \xi$, *and*

$$\begin{aligned}
[\mathbf{y} - E\{\mathbf{y}\}][\mathbf{y} - E\{\mathbf{y}\}]' &= \mathbf{A}(\hat{\xi} - E\{\hat{\xi}\})(\hat{\xi} - E\{\hat{\xi}\})' + \mathbf{A}(\hat{\xi} - E\{\hat{\xi}\})\tilde{\mathbf{e}}_\mathbf{y}'\\
&\quad + \tilde{\mathbf{e}}_\mathbf{y}(\hat{\xi} - E\{\hat{\xi}\})'\mathbf{A}' + \tilde{\mathbf{e}}_\mathbf{y}\tilde{\mathbf{e}}_\mathbf{y}'\\
D\{\mathbf{y}\} &= E\{[\mathbf{y} - E\{\mathbf{y}\}][\mathbf{y} - E\{\mathbf{y}\}]'\} = D\{\mathbf{A}\hat{\xi}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}\\
&= D\{\tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}
\end{aligned}$$

due to

$$\begin{aligned}
E\{\mathbf{A}(\hat{\xi} - E\{\hat{\xi}\})\tilde{\mathbf{e}}_\mathbf{y}'\} &= E\{\mathbf{A}(\hat{\xi} - E\{\hat{\xi}\})(\mathbf{y} - \mathbf{A}\hat{\xi})'\} = 0\\
E\{\tilde{\mathbf{e}}_\mathbf{y}(\hat{\xi} - E\{\hat{\xi}\})'\mathbf{A}'\} &= E\{\tilde{\mathbf{e}}_\mathbf{y}(\hat{\xi} - E\{\hat{\xi}\})'\mathbf{A}'\} = 0
\end{aligned}$$

or

$$C\{\mathbf{A}\hat{\xi},\ \tilde{\mathbf{e}}_\mathbf{y}\} = 0,\ \ C\{\tilde{\mathbf{e}}_\mathbf{y},\ \mathbf{A}\hat{\xi}\} = 0.$$

These covariance identities will be proven next.

$$\begin{aligned}
C\{\mathbf{A}\hat{\xi}, \tilde{\mathbf{e}}_\mathbf{y}\} &= \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}^-)^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}E\{(\mathbf{y} - E\{\mathbf{y}\})(\mathbf{y} - E\{\mathbf{y}\})'\}\\
&\quad \times [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]'
\end{aligned}$$

$$C\{\mathbf{A}\hat{\xi}, \tilde{\mathbf{e}}_\mathbf{y}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\boldsymbol{\Sigma}_\mathbf{y}[\mathbf{I}_n - \boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}']$$
$$C\{\mathbf{A}\hat{\xi}, \tilde{\mathbf{e}}_\mathbf{y}\} = \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}' - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}' = 0.$$

Here is our proof's end.

We recommend to consider the exercises as follows.

Exercise 4.1. (translation invariance: $\mathbf{y} \mapsto \mathbf{y} - E\{\mathbf{y}\}$):

Prove that the error prediction of type $\hat{\xi}$ $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE of ξ , namely

$$\tilde{\mathbf{e}}_\mathbf{y} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]\mathbf{y}$$

is translation invariant in the sense of $\mathbf{y} \mapsto \mathbf{y} - E\{\mathbf{y}\}$ that is

$$\tilde{\mathbf{e}}_\mathbf{y} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}]\mathbf{e}_\mathbf{y}$$

subject to $\mathbf{e}_\mathbf{y} := \mathbf{y} - E\{\mathbf{y}\}$.

Exercise 4.2. (idempotence):

Is the matrix $\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ *idempotent*?

Exercise 4.3. (projection matrices):

Are the matrices $\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ *and* $\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ *projection matrices*?

4-22 The Equivalence Theorem of $\mathbf{G}_\mathbf{y}$-LESS and $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE

We have included the *fourth chapter* on $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE in order to interpret $\mathbf{G}_\mathbf{y}$-LESS of the *third chapter*. The key question is open:

?When are $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE and $\mathbf{G}_\mathbf{y}$-LESS equivalent?

The answer will be given by

Theorem 4.5. (equivalence of $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE and $\mathbf{G}_\mathbf{y}$-LESS):

With respect to the *special linear Gauss–Markov model of full column rank* (4.1), (4.2) $\hat{\xi} = \mathbf{Ly}$ is $\boldsymbol{\Sigma}_\mathbf{y}$-BLUUE, if $\xi_\ell = \mathbf{Ly}$ is $\mathbf{G}_\mathbf{y}$-LESS of (3.1) for

$$\mathbf{G}_\mathbf{y} = \boldsymbol{\Sigma}_\mathbf{y}^{-1} \sim \mathbf{G}_\mathbf{y}^{-1} = \boldsymbol{\Sigma}_\mathbf{y}. \tag{4.35}$$

In such a case, $\hat{\xi} = \xi_\ell$ is the unique solution of the system of normal equations

$$(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})\hat{\xi} = \mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y} \tag{4.36}$$

attached with the regular dispersion matrix

$$D\{\hat{\xi}\} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}. \tag{4.37}$$

The proof is straight forward if we compare the solution (3.11) of $\mathbf{G_y}$-LESS and (4.21) of $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE. Obviously the *inverse dispersion matrix* $D\{\mathbf{y}\}$, $\mathbf{y} \in \{\mathbb{Y}, \; pdf\}$ is equivalent to the *matrix of the metric* $\mathbf{G_y}$ of the *observation space* $\mathbb{Y}$. Or conversely the inverse matrix of the metric of the *observation space* $\mathbb{Y}$ determines the variance-covariance matrix $D\{\mathbf{y}\} \sim \boldsymbol{\Sigma}_{\mathbf{y}}$ of the random variable $\mathbf{y} \in \{\mathbb{Y}, \; pdf\}$.

4-3 Setup of the Best Invariant Quadratic Uniformly Unbiased Estimator

of type BIQUUE for the central moments of second order

The subject of variance-covariance component estimation within Mathematical Statistics has been one of the central research topics in the nineteen eighties. In a remarkable bibliography up-to-date to the year 1977 *H. Sahai* listed more than 1,000 papers on variance-covariance component estimations, where his basic source was "*Statistical Theory and Method*" abstracts (published for the International Statistical Institute by Longman Groups Limited), "*Mathematical Reviews*" and "*Abstract Service of Quality Control and Applied Statistics*". Excellent review papers and books exist on the topic of variance-covariance estimation such as *C.R. Rao and J. Kleffe* (1988), *R.S. Rao* (1977) *S. B. Searle* (1978), *L.R. Verdooren* (1980), *J. Kleffe* (1980), and *R. Thompson* (1980). The PhD Thesis of B. *Schaffrin* (1983) offers a critical review of state-of-the-art of variance-covariance component estimation.

In *Geodetic Sciences* variance components estimation originates from *F.R. Helmert* (1924) who used least squares residuals to estimate heterogeneous variance components. *R. Kelm* (1974) and *E. Grafarend, A. Kleusberg and B. Schaffrin* (1980) proved the relation of $\boldsymbol{\Sigma}_0$ Helmert type IQUUE balled $\boldsymbol{\Sigma}$-HIQUUE to BIQUUE and MINQUUE invented by *C.R. Rao*. Most notable is the Ph.D. Thesis of *M. Serbetci* (1968) whose gravimetric measurements were analyzed by $\boldsymbol{\Sigma}_0$-HIQUUE Geodetic extensions of the *Helmert method* to compete variance components originate from *H. Ebner* (1972, 1977), *W. Förstner* (1979, 1980), *W. Welsch* (1977, 1978, 1979, 1980), *K. R. Koch* (1978, 1981), C. G. Persson (1981), *L. Sjoeberg* (1978), *E. Grafarend and A. d'Hone* (1978), *E. Grafarend* (1984) *B. Schaffrin* (1979, 1980, 1981). *W. Förstner* (1979), *H. Fröhlich* (1980), and *K.R. Koch* (1981) used the estimation of variance components for the adjustment of geodetic networks and the estimation of a length dependent variance of distances. A special field of geodetic application has been *oscillation analysis* based upon a fundamental paper by *H. Wolf* (1975), namely *M. Junasevic* (1977) for the estimation of signal-to-noise ratio in *gyroscopic azimuth observations*. The

Helmert method of variance component estimation was used by *E. Grafarend and A. Kleusberg* (1980) and *A. Kleusberg and E. Grafarend* (1981) to estimate variances of signal and noise in *gyrocompass observations*. Alternatively *K. Kubik* (1967a, b, c, 1970) pioneered the method of *Maximum Likelihood* (MALE) for estimating weight ratios in a hybrid distance–direction network. "MALE" and "FEMALE" extensions were proposed by *B. Schaffrin* (1983), *K. R. Koch* (1986), and *Z. C. Yu* (1996).

A typical problem with $\mathbf{\Sigma}_0$-*Helmert type* IQUUE is that it does *not* produce *positive variances* in general. The problem of generating a *positive-definite variance-covariance* matrix from variance-covariance component estimation has already been highlighted by *J. R. Brook* and *T. Moore* (1980), *K.G. Brown* (1977, 1978), *O. Bemk and H. Wandl* (1980), *V. Chew* (1970), *Han Chien-Pai* (1978), *R. R. Corbeil and S. R. Searle* (REML, 1976), *F. J. H. Don and J. R. Magnus* (1980), *H. Drygas* (1980), *S. Gnot, W. Klonecki and R. Zmyslony* (1977). *H. O. Hartley and J. N. K. Rao* (ML, 1967), in particular *J. Hartung* (1979, 1980), *J. L. Hess* (1979), *S. D. Horn and R. A. Horn* (1975), S. D. Horn, R. A. Horn and D. B. Duncan (1975), *C. G. Khatri* (1979), *J. Kleffe* (1978, 1980),), J. Kleffe and *J. Zöllner* (1978), in particular L. R. Lamotte (1973, 1980), *S. K. Mitra* (1971), *R. Pincus* (1977), in particular *F. Pukelsheim* (1976, 1977, 1979, 1981 a, b), *F. Pukelsheim and G. P. Styan* (1979), *C. R. Rao* (1970, 1978), *S. R. Searle* (1979), S. *R. Searle and H. V. Henderson* (1979), *J. S. Seely* (1972, 1977), in particular *W. A. Thompson* (1962, 1980), *L. R. Verdooren* (1979), and *H. White* (1980).

In view of available textbooks, review papers and basic contributions in scientific journals we are only able to give a short introduction. *First*, we outline the general model of variance-covariance components leading to a linear structure for the central second order moment, known as the variance-covariance matrix. *Second*, for the example of one variance component we discuss the key role of the postulate's (i) symmetry, (ii) invariance, (iii) uniform unbiasedness, and (iv) minimum variance. *Third*, we review variance-covariance component estimations of Helmert type.

4-31 Block Partitioning of the Dispersion Matrix and Linear Space Generated by Variance-Covariance Components

The *variance-covariance component* model is defined by the block partitioning (4.33) of a *variance-covariance matrix* $\mathbf{\Sigma}_{\mathbf{y}}$, also called dispersion matrix $D\{\mathbf{y}\}$, which follows from a corresponding *rank partitioning* of the observation vector $\mathbf{y} = [y_1', \ldots, y_\ell']'$. The integer number l is the number of blocks. For instance, the variance-covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ in (4.41) is partitioned into $l = 2$ blocks. The various blocks consequently factorized by *variance* σ_j^2 and by *covariances* $\sigma_{jk} = \rho_{jk}\sigma_j\sigma_k$. $\rho_{jk} \in [-1, +1]$ denotes the correlation coefficient between the blocks. For *instance*, $D\{\mathbf{y}_1\} = \mathbf{V}_{11}\sigma_1^2$ is a *variance factorization*, while $D\{\mathbf{y}_1, \mathbf{y}_2\} = \mathbf{V}_{12}\sigma_{12} = \mathbf{V}_{12}\rho_{12}\sigma_1\sigma_2$ is a *covariance factorization*. The matrix blocks $\mathbf{V}_{jj}$ are built into the matrix $\mathbf{C}_{jj}$, while the off-diagonal blocks $\mathbf{V}_{jk}, \mathbf{V}'_{jk}$ into the matrix $\mathbf{C}_{jk}$ of the same dimensions.

$$\dim \boldsymbol{\Sigma} = \dim \mathbf{C}_{jj} = \dim \mathbf{C}_{jk} = n \times n.$$

The *collective matrices* $\mathbf{C}_{jj}$ and $\mathbf{C}_{jk}$ enable us to develop an additive decomposition (4.36), (4.43) of the block partitioning variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$. As soon as we collect all variance-covariance components in an peculiar true order, namely

$$\boldsymbol{\sigma} := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \ldots, \sigma_{\ell-1\ell,}, \sigma_\ell^2]',$$

we are led to a linear form of the dispersion matrix (4.37), (4.43) as well as of the dispersion vector (4.39), (4.44). Indeed the dispersion vector $\mathrm{d}(\mathbf{y}) = \mathbf{X}\boldsymbol{\sigma}$ builds up a linear form where the second order design matrix X, namely

$$\mathbf{X} := [\mathrm{vec}\mathbf{C}_1, \ldots, \mathrm{vec}\mathbf{C}_{\ell(\ell+1)}] \in \mathbb{R}^{n^2 \times \ell(\ell+1)/2},$$

reflects the block structure. There are $l(l+1)/2$ matrices $\mathbf{C}_j, j \in \{1, \ldots, \ell(\ell+1)/2\}$. For instance, for $l = 2$ we are left with 3 block matrices $\{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\}$.

Before we analyze the *variance-covariance component* model in more detail, we briefly mention the multinominal inverse $\boldsymbol{\Sigma}^{-1}$ of the block partitioned matrix $\boldsymbol{\Sigma}$. For instance by "JPM" and "SCHUR" we gain the block partitioned inverse matrix $\boldsymbol{\Sigma}^{-1}$ with elements $\{\mathbf{U}_{11}, \mathbf{U}_{12}, \mathbf{U}_{22}\}$ (4.51)–(4.54) derived from the block partitioned matrix $\boldsymbol{\Sigma}$ with elements $\{\mathbf{V}_{11}, \mathbf{V}_{12}, \mathbf{V}_{22}\}$ (4.47). "*Sequential JPM*" solves the block inverse problems for any block partitioned matrix. With reference to *Box 4.2* and *Box 4.3*

$$\boldsymbol{\Sigma} = \mathbf{C}_1\sigma_1 + \mathbf{C}_2\sigma_2 + \mathbf{C}_3\sigma_3 \Rightarrow \boldsymbol{\Sigma}^{-1} = E_1(\sigma) + E_2(\sigma) + E_3(\sigma)$$

is an example.

Box 4.2. (Partitioning of variance-covariance matrix):

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{V}_{11}\sigma_1^2 & \mathbf{V}_{12}\sigma_{12} & \cdots & \mathbf{V}_{1\ell-1}\sigma_{1\ell-1} & \mathbf{V}_{1\ell}\sigma_{1\ell} \\ \mathbf{V}'_{12}\sigma_{12} & \mathbf{V}_{22}\sigma_2^2 & \cdots & \mathbf{V}_{2\ell-1}\sigma_{2\ell-1} & \mathbf{V}_{2\ell}\sigma_{2\ell} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{V}'_{1\ell-1}\sigma_{1\ell-1} & \mathbf{V}'_{2\ell-1}\sigma_{2\ell-1} & \cdots & \mathbf{V}_{\ell-1\ell-1}\sigma_{\ell-1}^2 & \mathbf{V}_{\ell-1\ell}\sigma_{\ell-1\ell} \\ \mathbf{V}'_{1\ell}\sigma_{1\ell} & \mathbf{V}'_{2\ell}\sigma_{2\ell} & \cdots & \mathbf{V}'_{\ell-1\ell}\sigma_{\ell-1\ell} & \mathbf{V}_{\ell\ell}\sigma_\ell^2 \end{bmatrix} > 0 \qquad (4.38)$$

l second moments σ^2 of type *variance*, $\ell(\ell-1)/2$ second moment σ_{jk} of type *covariance*

matrix blocks of second order design

$$\mathbf{C}_{jj} := \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \mathbf{V}_{jj} & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \quad \forall j \in \{1, \ldots, \ell\} \tag{4.39}$$

$$\mathbf{C}_{jk} := \begin{bmatrix} 0 & & & 0 \\ \cdots & 0 & \mathbf{V}_{jk} & \cdots \\ & \mathbf{V}_{kj} & & \\ \cdots & & & \cdots \\ 0 & & & 0 \end{bmatrix} \quad \left[\begin{array}{l} \textit{subject to } j < k \\ \textit{and } j, k \in \{1, \ldots, \ell\} \end{array} \right. \tag{4.40}$$

$$\mathbf{\Sigma} = \sum_{j=1}^{\ell} \mathbf{C}_{jj} \sigma_j^2 + \sum_{j=1, k=2, j<k}^{\ell-1, \ell} \mathbf{C}_{jk} \sigma_{jk} \tag{4.41}$$

$$\mathbf{\Sigma} = \sum_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j \sigma_j \in \mathbb{R}^{n \times m} \tag{4.42}$$

$$[\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \ldots, \sigma_{\ell-1\ell,}, \sigma_\ell^2]' =: \sigma \tag{4.43}$$

"*dispersion vector*"

$$D\{\mathbf{y}\} := \mathbf{\Sigma}_\mathbf{y} \Leftrightarrow d\{\mathbf{y}\} = \mathrm{vec} D\{\mathbf{y}\} = \mathrm{vec} \mathbf{\Sigma}$$

$$d(\mathbf{y}) = \sum_{j=1}^{\ell(\ell+1)/2} (\mathrm{vec} \mathbf{C}_j) \sigma_j = \mathbf{X} \sigma \tag{4.44}$$

"**X** *is called second order design matrix*"

$$\mathbf{X} := [\mathrm{vec} \mathbf{C}_1, \ldots, \mathrm{vec} \mathbf{C}_{\ell(\ell+1)/2}] \tag{4.45}$$

"*dimension identities*."

$$d(\mathbf{y}) \in \mathbb{R}^{n^2 \times 1}, \sigma \in \mathbb{R}, \mathbf{X} \in \mathbb{R}^{n^2 \times \ell(\ell+1)/2}.$$

Box 4.3. (Multinomial inverse):

$$: Input :$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}'_{12} & \mathbf{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{11} \sigma_1^2 & \mathbf{V}_{12} \sigma_{12} \\ \mathbf{V}'_{12} \sigma_{12} & \mathbf{V}_{22} \sigma_{22} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{V}_{11} & 0 \\ 0 & 0 \end{bmatrix} \sigma_1^2 + \begin{bmatrix} 0 & \mathbf{V}_{12} \\ \mathbf{V}'_{12} & 0 \end{bmatrix} \sigma_{12} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{V}_{22} \end{bmatrix} \sigma_2^2 \in \mathbb{R}^{n \times m} \tag{4.46}$$

$$\mathbf{C}_{11} := \mathbf{C}_1 := \begin{bmatrix} \mathbf{V}_{11} & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{C}_{12} := \mathbf{C}_2 := \begin{bmatrix} 0 & \mathbf{V}_{12} \\ \mathbf{V}'_{12} & 0 \end{bmatrix}, \mathbf{C}_{22} := \mathbf{C}_3 := \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{V}_{22} \end{bmatrix} \tag{4.47}$$

$$\mathbf{\Sigma} = \mathbf{C}_{11}\sigma_1^2 + \mathbf{C}_{12}\sigma_{12} + \mathbf{C}_{22}\sigma_2^2 = \mathbf{C}_1\sigma_1 + \mathbf{C}_2\sigma_2 + \mathbf{C}_3\sigma_3 = \sum_{j=1}^{3} \mathbf{C}_j \sigma_j \tag{4.48}$$

$$\operatorname{vec}\varSigma = \sum_{j=1}^{3} (\operatorname{vec}\mathbf{C}_j)\sigma_j = [\operatorname{vec}\mathbf{C}_1, \operatorname{vec}\mathbf{C}_2, \operatorname{vec}\mathbf{C}_3] \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \mathbf{X}\sigma \tag{4.49}$$

$$\operatorname{vec}\mathbf{C}_j \in \mathbb{R}^{n^2 \times 1} \quad \forall j \in \{1, \ldots, \ell(\ell+1)/2\}$$

"**X** *is called second order design matrix*"

$$\mathbf{X} := [\operatorname{vec}\mathbf{C}_1, \ldots, \operatorname{vec}\mathbf{C}_{\ell(\ell+1)/2}] \in \mathbb{R}^{n^2 \times \ell(\ell+1)/2}$$

here: $l = 2$:output:

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \mathbf{U}_{11} & 0 \\ 0 & 0 \end{bmatrix} \sigma_1^{-2} + \begin{bmatrix} 0 & \mathbf{U}_{12} \\ \mathbf{U}'_{12} & 0 \end{bmatrix} \sigma_{12}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{U}_{22} \end{bmatrix} \sigma_2^{-2} \tag{4.50}$$

subject to

$$\mathbf{U}_{11} = \mathbf{V}_{11}^{-1} + q\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{S}\mathbf{V}'_{12}\mathbf{V}_{11}^{-1} \tag{4.51}$$

$$\mathbf{U}_{12} = \mathbf{U}'_{21} = -q\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{S} \tag{4.52}$$

$$\mathbf{U}_{22} = \mathbf{S} = (\mathbf{V}_{22} - q\mathbf{V}'_{12}\mathbf{V}_{11}^{-1}\mathbf{V}_{12})^{-1} \tag{4.53}$$

$$q := \frac{\sigma_{12}^2}{\sigma_1^2\sigma_2^2} \tag{4.54}$$

$$\mathbf{\Sigma}^{-1} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 = \sum_{j=1}^{\ell(\ell+1)/2=3} \mathbf{E}_j \tag{4.55}$$

$$\mathbf{E}_1(\sigma) := \begin{bmatrix} \mathbf{U}_{11} & 0 \\ 0 & 0 \end{bmatrix} \sigma_1^{-2}, \mathbf{E}_2(\sigma) := \begin{bmatrix} 0 & \mathbf{U}_{12} \\ \mathbf{U}'_{12} & 0 \end{bmatrix} \sigma_{12}^{-1}, \mathbf{E}_3(\sigma) := \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{U}_{22} \end{bmatrix} \sigma_2^{-2} \tag{4.56}$$

The general result that inversion of a block partitioned symmetric matrix conserves the block structure is presented in

Corollary 4.6. Corollary 4.6 *(multinomial inverse):*

$$\mathbf{\Sigma} = \sum_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j \sigma_j \Leftrightarrow \mathbf{\Sigma}^{-1} = \sum_{j=1}^{\ell(\ell+1)/2} \mathrm{E}_j(\sigma) \tag{4.57}$$

We shall take advantage of the block structured multinominal inverse when we are reviewing HIQUUE or variance-covariance estimations of Helmert type.

The variance component model as well as the variance-covariance model are defined next. A variance component model is a linear model of type

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{V}_{11}\sigma_1^2 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{V}_{22}\sigma_2^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{V}_{\ell-1\ell-1}\sigma_{\ell-1}^2 & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{V}_{\ell\ell}\sigma_\ell^2 \end{bmatrix} \tag{4.58}$$

$$d\{\mathbf{y}\} = \mathrm{vec}\boldsymbol{\Sigma} = [\mathrm{vec}\mathbf{C}_{11}, \ldots, \mathrm{vec}\mathbf{C}_{jj}] \begin{bmatrix} \sigma_1^2 \\ \cdots \\ \sigma_\ell^2 \end{bmatrix} \tag{4.59}$$

$$d\{\mathbf{y}\} = \mathbf{X}\sigma \quad \forall\, \sigma \in \mathbb{R}^+. \tag{4.60}$$

In contrast, the general model (4.49) is the variance-covariance model with a linear structure of type

$$d\{\mathbf{y}\} = \mathrm{vec}\boldsymbol{\Sigma} = [\mathrm{vec}\mathbf{C}_{11}, \mathrm{vec}\mathbf{C}_{12}, \mathrm{vec}\mathbf{C}_{12}, \ldots, \mathrm{vec}\mathbf{C}_{\ell\ell}] \begin{bmatrix} \sigma_1^2 \\ \sigma_{12} \\ \sigma_2^2 \\ \cdots \\ \sigma_\ell^2 \end{bmatrix} \tag{4.61}$$

$$d\{\mathbf{y}\} = \mathbf{X}\boldsymbol{\sigma} \quad \forall\, \sigma_j^2 \in R^+,\ \ \boldsymbol{\Sigma}\ \textit{positive definite}. \tag{4.62}$$

The most popular cases of variance-covariance components are collected in the examples.

Example 4.1. (one variance components, i.i.d. observations)

$$D\{\mathbf{y}\} : \boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{I}_n\sigma^2 \textit{ subject to } \boldsymbol{\Sigma}_{\mathbf{y}} \in SYM(\mathbb{R}^{n\times n}),\ \sigma^2 \in \mathbb{R}^+.$$

Example 4.2. (one variance component, correlated observations)

$$D\{\mathbf{y}\} : \boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{V}_n\sigma^2 \textit{ subject to } \boldsymbol{\Sigma}_{\mathbf{y}} \in SYM(\mathbb{R}^{n\times n}),\ \sigma^2 \in \mathbb{R}^+.$$

Example 4.3. (two variance components, two sets of totally uncorrected observations “heterogeneous observations”)

$$D\{\mathbf{y}\} : \boldsymbol{\Sigma}_{\mathbf{y}} = \begin{bmatrix} \mathbf{I}_{n_1}\sigma_1^2 & 0 \\ 0 & \mathbf{I}_{n_2}\sigma_2^2 \end{bmatrix} \textit{ subject to } \begin{bmatrix} n = n_1 + n_2 \\ \boldsymbol{\Sigma}_{\mathbf{y}} = SYM(\mathbb{R}^{n\times n}) \\ \sigma_1^2 \in^+, \sigma_2^2 \in \mathbb{R}^+. \end{bmatrix} \tag{4.63}$$

Example 4.4. (two variance components, one covariance components, two sets of correlated observations “heterogeneous observations”)

$$D\{\mathbf{y}\} : \mathbf{\Sigma_y} = \begin{bmatrix} \mathbf{V}_{11}\sigma_1^2 & \mathbf{V}_{12}\sigma_{12} \\ \mathbf{V'}_{12}\sigma_{12} & \mathbf{V}_{11}\sigma_2^2 \end{bmatrix} \textsf{ subject to } \left[\begin{array}{c} n = n_1 + n_2 \\ \mathbf{V}_{11} \in \mathbb{R}^{n_1 \times n_1}, \mathbf{V}_{22} \in \mathbb{R}^{n_2 \times n_2} \\ \mathbf{V}_{12} \in \mathbb{R}^{n_1 \times n_2} \end{array} \right.$$
$$\mathbf{\Sigma_y} \in SYM(\mathbb{R}^{n\times n}),\ \sigma_1^2 \in \mathbb{R}^+,\ \sigma_2^2 \in \mathbb{R}^+,\ \mathbf{\Sigma_y} \textsf{ positive definite.} \tag{4.64}$$

$$\textit{Special case}: \mathbf{V}_{11} = \mathbf{I}_{n_1},\ \mathbf{V}_{22} = \mathbf{I}_{n_2}.$$

Example 4.5. (elementary error model, random effect model)

$$\mathbf{e_y} = \mathbf{y} - E\{\mathbf{y}\,|\mathbf{z}\} = \sum_{j=1}^{\ell} \mathbf{A}_j(\mathbf{z}_j - E\{\mathbf{z}_j\}) = \sum_{j=1}^{\ell} \mathbf{A}_j \mathbf{e}_{\mathbf{z}}^j \tag{4.65}$$

$$E\{\mathbf{e}_{\mathbf{z}}^j\} = \mathbf{0}, \quad E\{\mathbf{e}_{\mathbf{z}}^j, \mathbf{e'}_{\mathbf{z}}^k\} = \delta_{jk}\mathbf{I}_q \tag{4.66}$$

$$D\{\mathbf{y}\} : \mathbf{\Sigma_y} = \sum_{j=1}^{\ell} \mathbf{A}_j \mathbf{A'}_j \sigma_j^2 + \sum_{j,k=1, j<k}^{\ell} (\mathbf{A}_j \mathbf{A'}_k + \mathbf{A}_k \mathbf{A'}_j)\sigma_{jk} \tag{4.67}$$

At this point, we should emphasize that a linear space of variance-covariance components can be build up *independently of the block partitioning* of the dispersion matrix $D\{\mathbf{y}\}$. For future details and explicit examples let us refer to *B. Schaffrin* (1983).

4-32 *Invariant Quadratic Estimation of Variance-Covariance Components of Type IQE*

By means of *Definition 4.2* (one variance component) and *Definition 4.9* (variance-covariance components) we introduce

$$\hat{\sigma}^2\ IQE\ of\ \sigma^2\ and\ \hat{\sigma}_k\ IQE\ of\ \sigma_k.$$

Those conditions of IQE, represented in *Lemmas 4.7* and *4.9* enable us to separate the estimation process of first moments ξ_j (like BLUUE) from the estimation process of central second moments σ_k (like BIQUUE). Finally we provide you with the general solution (4.74) of the in homogeneous matrix equations $\mathbf{M}_k^{1/2}\mathbf{A} = 0$ (orthogonality conditions) for all k∈1,···,l(l+1)/2 where $l(l+1)/2$ is the number of variance-covariance components, restricted to the *special Gauss–Markov model* $E\,\{\mathbf{y}\} = \mathbf{A}\xi$, $d\,\{\mathbf{y}\} = \mathbf{X}\sigma$ of "*full column rank*", $\mathbf{A} \in \mathbb{R}^{n\times m}$, rk$\mathbf{A} = m$.

Definition 4.2. (invariant quadratic estimation $\hat{\sigma}^2$ of σ^2 : IQE):

The scalar $\hat{\sigma}^2$ is called IQE (*Invariant Quadratic Estimation*) of $\sigma^2 \in \mathbb{R}^+$ with respect to the *special Gauss–Markov model of full column rank.*

$$\begin{aligned} &E\{\mathbf{y}\} = \mathbf{A}\xi,\ \mathbf{A} \in \mathbb{R}^{n \times m},\ \text{rk}\mathbf{A} = m \\ &D\{\mathbf{y}\} = \mathbf{V}\sigma^2,\ \mathbf{V} \in \mathbb{R}^{n \times n},\ \text{rk}\mathbf{V} = n,\ \sigma^2 \in \mathbb{R}^+, \end{aligned} \tag{4.68}$$

if the "variance component σ^2 is σ"
(i) a quadratic estimation

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\text{vec}\mathbf{M})'(\mathbf{y} \otimes \mathbf{y}) = (\mathbf{y}' \otimes \mathbf{y}')(\text{vec}\mathbf{M}) \tag{4.69}$$

subject to

$$\mathbf{M} \in SYM := \{\mathbf{M} \in \mathbb{R}^{n \times n} | \mathbf{M}' = \mathbf{M}\} \tag{4.70}$$

(ii) transformational invariant: $\mathbf{y} \to \mathbf{y} - E\{\mathbf{y}\} =: \mathbf{e_y}$ *in the sense of*

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{e_y}'\mathbf{M}\mathbf{e_y} \tag{4.71}$$

or

$$\hat{\sigma}^2 = (\text{vec}\mathbf{M})'(\mathbf{y} \otimes \mathbf{y}) = (\text{vec}\mathbf{M})'(\mathbf{e_y} \otimes \mathbf{e_y}) \tag{4.72}$$

or

$$\hat{\sigma}^2 = \text{tr}(\mathbf{M}\mathbf{y}\mathbf{y}') = \text{tr}(\mathbf{M}\mathbf{e_y}\mathbf{e_y}') \tag{4.73}$$

Already in the introductory paragraph we emphasized the key of "IQE". Indeed by the postulate "IQE" the estimation of the first moments $E\{\mathbf{y}\} = \mathbf{A}\xi$ is supported by the estimation of the central second moments $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$ or $d\{\mathbf{y}\} = \mathbf{X}\sigma$. Let us present to you the fundamental result of "$\hat{\sigma}^2$ IQE OF σ^2".

Lemma 4.8. (invariant quadratic estimation $\hat{\sigma}^2$ of $\hat{\sigma}^2$:IQE):

Let $\mathbf{M} = (\mathbf{M}^{1/2})'\mathbf{M}^{1/2}$ be a multiplicative decomposition of the symmetric matrix $\mathbf{M}$. The scalar $\hat{\sigma}^2$ is IQE of σ^2, if and only if

$$\mathbf{M}^{1/2} = 0 \, or \, \mathbf{A}'(\mathbf{M}^{1/2})' = 0 \tag{4.74}$$

for all $\mathbf{M}^{1/2} \in \mathbb{R}^{n \times n}$.

Proof. First, we substitute the transformation $\mathbf{y} = E\{\mathbf{y}\} + \mathbf{e_y}$ subject to expectation identity $E\{\mathbf{y}\} = \mathbf{A}\xi,\ \mathbf{A} \in \mathbb{R}^{n \times m}$, rk$\mathbf{A} = m$, into $\mathbf{y}'\mathbf{M}\mathbf{y}$.

$$\mathbf{y}'\mathbf{M}\mathbf{y} = \xi'\mathbf{A}'\mathbf{M}\mathbf{A}\xi + \xi'\mathrm{A}'\mathrm{M}\mathrm{e}_\mathbf{y} + \mathbf{e}_\mathbf{y}'\mathbf{M}\mathbf{A}\xi + \mathbf{e}_\mathbf{y}'\mathbf{M}\mathbf{e}_\mathbf{y}.$$

Second, we take advantage of the multiplicative decomposition of the matrix **M**, namely

$$\mathbf{M} = (\mathbf{M}^{1/2})'\mathbf{M}^{1/2}, \tag{4.75}$$

which generates the symmetry of the matrix

$$\mathbf{M} \in SYM{:} = \{\mathbf{M} \in \mathbb{R}^{m\times n}|\mathbf{M}' = \mathbf{M}\}$$

$$\mathbf{y}'\mathbf{M}\mathbf{y} = \xi'\mathbf{A}'(\mathbf{M}^{1/2})'\mathbf{M}^{1/2}\mathbf{A}\xi + \xi'A'(\mathbf{M}^{1/2})'\mathbf{M}^{1/2}\mathbf{e}_\mathbf{y}$$
$$+\mathbf{e}_\mathbf{y}'(\mathbf{M}^{1/2})'\mathbf{M}^{1/2}\mathbf{A}\xi + \mathbf{e}_\mathbf{y}'\mathbf{M}\mathbf{e}_\mathbf{y}.$$

Third, we postulate "IQE".

$$\mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{e}_\mathbf{y}'\mathbf{M}\mathbf{e}_\mathbf{y} \Leftrightarrow \mathbf{M}^{1/2}\mathbf{A} = 0 \Leftrightarrow \mathbf{A}'(\mathbf{M}^{1/2})' = 0.$$

For the proof, here is our journey's end.

Let us extend "IQE" from a "one variance component model" to a "variance-covariance components model". *First*, we define "IQE" (4.81) for variance-covariance components, *second* we give necessary and sufficient conditions identifying "IQE".

Definition 4.9. (variance-covariance components model $\hat{\sigma}_k$ IQE of σ_k):

The dispersion vector $\hat{d}(\mathbf{y})$ is called IQE ("*Invariant Quadratic Estimation*") with respect to the *special Gauss–Markov model of full column rank.*

$$\left[\begin{array}{l} E\{\mathbf{y}\} = \mathbf{A}\xi,\ \mathbf{A} \in \{\mathbb{R}^{n\times m}\};\ \mathrm{rk}\mathbf{A} = m \\ d\{\mathbf{y}\} = \mathbf{X}\boldsymbol{\sigma},\ D\{\mathbf{y}\} \sim \boldsymbol{\Sigma}_\mathbf{y}\, positive\, definite,\ \mathrm{rk}\boldsymbol{\Sigma}_\mathbf{y} = n, \end{array}\right. \tag{4.76}$$

if the variance-covariance components

$$\sigma := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \ldots, \sigma_\ell^2]' \tag{4.77}$$

(i) bilinear estimations

$$\hat{\sigma}_k = \mathbf{y}'\mathbf{M}\mathbf{y} = (\mathrm{vec}\mathbf{M})'(\mathbf{y}\otimes\mathbf{y}) = \mathrm{tr}\mathbf{M}_k\mathbf{y}\mathbf{y}' \quad \forall\mathbf{M}_k \in \mathbb{R}^{n\times n\times\ell(\ell+1)/2} \tag{4.78}$$

$$subject to$$

$$\mathbf{M}_k \in SYM := \{\mathbf{M}_k \in \mathbb{R}^{n\times n\times\ell(\ell+1)/2}|\mathbf{M}_k = \mathbf{M}_k'\}, \tag{4.79}$$

(ii) translational invariant

$$\mathbf{y} \to \mathbf{y} - E\{\mathbf{y}\} =: \mathbf{e}_\mathbf{y}$$

$$\hat{\sigma}_k = \mathbf{y}'\mathbf{M}_k\mathbf{y} = \mathbf{e}_\mathbf{y}'\mathbf{M}_k\mathbf{e}_\mathbf{y} \tag{4.80}$$

$$\hat{\sigma}_k = (\text{vec}\mathbf{M}_k)'(\mathbf{y} \otimes \mathbf{y}) = (\text{vec}\mathbf{M}_k)'(\mathbf{e}_\mathbf{y} \otimes \mathbf{e}_\mathbf{y}). \tag{4.81}$$

Note the fundamental lemma "$\hat{\sigma}_k$ IQE of σ_k" whose proof follows the same line as the proof of *Lemma 4.7*.

Lemma 4.10. (invariant quadratic estimation $\hat{\sigma}_k$ of σ_k : IQE):

Let $\mathbf{M}_k = (\mathbf{M}_k^{1/2})'\mathbf{M}_k^{1/2}$ be a multiplicative decomposition of the symmetric matrix $\mathbf{M}_k$. The dispersion vector $\hat{\sigma}_k$ is IQE of σ_k, if and only if

$$\mathbf{M}_k^{1/2}\mathbf{A} = 0 \, or \, \mathbf{A}'(\mathbf{M}_k^{1/2})' = 0 \tag{4.82}$$

for all $\mathbf{M}_k^{1/2} \in R^{n \times n \times \ell(\ell+1)/2}$.

? How can we characterize "$\hat{\sigma}^2$ IQE of σ^2" or "$\hat{\sigma}_k$ IQE of σ_k" ?

The problem is left with the orthogonality conditions (4.74) and (4.82). *Box 4.4* reviews the general solutions of the *homogeneous equations* (4.83) and (4.85) for our "*full column rank linear model*".

Box 4.4. (General solutions of homogeneous matrix equations):

$$\mathbf{M}_k^{1/2} = 0 \quad \Leftrightarrow \quad \mathbf{M}_k = \mathbf{Z}_k(\mathbf{I}_n - \mathbf{A}\mathbf{A}^-) \tag{4.83}$$

$$\text{“}\textit{for all } \mathbf{A}^- \in \mathbf{G} := \{\mathbf{A}^- \in \mathbb{R}^{n \times m} | \mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}\}''$$

$$: \text{rk}\mathbf{A} = m$$

$$\mathbf{A}^- = \mathbf{A}_L^- = (\mathbf{A}'\mathbf{G}_\mathbf{y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}_\mathbf{y} \tag{4.84}$$

$$\text{“}\textit{for all left inverses } \mathbf{A}_L^- \in \{\mathbf{A}^- \in R^{m \times n} | (\mathbf{A}^-\mathbf{A})' = \mathbf{A}^-\mathbf{A}\}''$$

$$\left.\begin{matrix} \mathbf{M}_k^{1/2} = 0 \\ \text{rk}\mathbf{A} \leq m \end{matrix}\right] \Rightarrow \mathbf{M}_k^{1/2} = \mathbf{Z}_k[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G}_y\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}_y] \tag{4.85}$$

"*unknown matrices* : $\mathbf{Z}_k$ *and* $\mathbf{G}_\mathbf{y}$."

First, (4.83) is a representation of the general solutions of the inhomogeneous matrix equations (4.82) where $\mathbf{Z}_k, k \in \{1, \ldots, \ell(\ell+1)/2\}$, are arbitrary matrices. Note that $k = 1, \mathbf{M}_1$ describes the "*one variance component model*", otherwise the general variance-covariance components model. Here we are dealing with a *special Gauss–Markov model* of "*full column rank*", rk$\mathbf{A} = m$. In this case, the generalized inverse $\mathbf{A}^-$ is specified as the "*weighted left inverse*" $\mathbf{A}_L^-$ of type (4.70) whose weight $\mathbf{G}_\mathbf{y}$ is unknown. In summarizing, representations of two matrices $\mathbf{Z}_k$ and $\mathbf{G}_\mathbf{y}$ to be unknown, given $\mathbf{H}_k^{1/2}$, $\mathbf{M}_k$ is computed by

$$\mathbf{M}_k = (\mathbf{M}_k^{1/2})'\mathbf{M}_k^{1/2} = [\mathbf{I}_n - \mathbf{G_y A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}']\mathbf{Z}'_k\mathbf{Z}_k[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_y A})^{-1}\mathbf{A}'\mathbf{G_y}] \tag{4.86}$$

definitely as a *symmetric matrix*.

4-33 Invariant Quadratic Uniformly Unbiased Estimations of Variance-Covariance Components of Type IQUUE

Unbiased estimations have already been introduced for the first moments $E\{\mathbf{y}\} = \mathbf{A}\xi, \mathbf{A} \in \mathbb{R}^{n\times m}, \text{rk}\mathbf{A} = m$. Similarly we like to develop the theory of the one variance component σ^2 and the variance-covariance unbiased estimations for the central second moments, namely components σ_k, $k \in \{1,\dots,\ell(\ell+1)/2\}$, where l is the number of blocks. *Definition 4.11* tells us when we use the terminology "invariant quadratic uniformly unbiased estimation" $\hat{\sigma}^2$ of σ^2 or $\hat{\sigma}_k$ of σ_k, in short "IQUUE" *Lemma 4.12* identifies $\hat{\sigma}^2$IQUUE of σ^2 by the additional $\text{tr}\mathbf{VM} = 1$. In contrast, *Lemma 4.12* focuses on $\hat{\sigma}_k$IQUUE of σ_k by means of the additional conditions $\text{tr}\mathbf{C}_j\mathbf{M}_k = \boldsymbol{\delta}_{jk}$. Examples are given in the following paragraphs.

Definition 4.11. (invariant quadratic uniformly unbiased estimation $\hat{\sigma}^2$ of σ^2 and $\hat{\sigma}_k$ of σ_k : IQUUE):

The vector of variance-covariance components $\hat{\sigma}_k$ is called IQUUE (*Invariant Quadratic Uniformly Unbiased Estimation*) of σ_k with respect to the *special Gauss–Markov model of full column rank.*

$$\left[\begin{array}{l} E\{\mathbf{y}\} = \mathbf{A}\xi, \mathbf{A} \in \mathbb{R}^{n\times m}, \text{rk}\mathbf{A} = m \\ d\{\mathbf{y}\} = \mathbf{X}\boldsymbol{\sigma}, \mathbf{X} \in \mathbb{R}^{n^2\times \ell(\ell+1)/2}, D\{\mathbf{y}\} \sim \boldsymbol{\Sigma}_\mathbf{y}\, \textit{positive definite}, \text{rk}\boldsymbol{\Sigma}_\mathbf{y} \\ \text{rk}\Sigma_\mathbf{y} = n,\ \text{vech}D\{\mathbf{y}\} = d\{\mathbf{y}\}, \end{array}\right. \tag{4.87}$$

if the variance-covariance components

$$\boldsymbol{\sigma} := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \dots, \sigma_\ell^2] \tag{4.88}$$

are

(i) a bilinear estimation

$$\begin{gathered} \hat{\sigma}_k = \mathbf{y}'\mathbf{M}_k\mathbf{y} = (\text{vec}\mathbf{M}_k)'(\mathbf{y}\otimes\mathbf{y}) = \text{tr}\mathbf{M}_k\mathbf{y}\mathbf{y}' \\ \forall \mathbf{M}_k \in R^{n\times n\times \ell(\ell+1)/2} \end{gathered} \tag{4.89}$$

subject to

$$\mathbf{M}_k = (\mathbf{M}_k^{1/2})'(\mathbf{M}_k^{1/2}) \in SYM := \{\mathbf{M}_k \in R^{n\times m\times \ell(\ell+1)/2} | \mathbf{M}_k = \mathbf{M}'_k\} \tag{4.90}$$

(ii) translational invariant in the sense of

$$\mathbf{y} \to \mathbf{y} - E\{\mathbf{y}\} =: \mathbf{e_y}$$

$$\hat{\sigma}_k = \mathbf{y}'\mathbf{M}_k\mathbf{y} = \mathbf{e}'_\mathbf{y}\mathbf{M}_k\mathbf{e_y} \tag{4.91}$$

or

$$\hat{\sigma}_k = (\mathrm{vec}\mathbf{M}_k)'(\mathbf{y} \otimes \mathbf{y}) = (\mathrm{vec}\mathbf{M}_k)'(\mathbf{e}_y \otimes \mathbf{e}_y) \tag{4.92}$$

or

$$\hat{\sigma}_k = \mathrm{tr}\mathrm{M}_k\mathbf{y}\mathbf{y}' = \mathrm{tr}\mathbf{M}_k\mathbf{e_y}\mathbf{e}'_\mathbf{y}, \tag{4.93}$$

(iii) uniformly unbiased in the sense of

$k = 1$ (one variance component):

$$E\{\hat{\sigma}^2\} = \sigma^2, \forall\sigma^2 \in \mathbb{R}^+, \tag{4.94}$$

$k \geq 1$ (variance-covariance components):

$$E\{\hat{\sigma}_k\} = \sigma_k, \forall\sigma_k \in \{\mathbb{R}^{\ell(\ell+1)/2} | \mathbf{\Sigma_y}\, \textit{positive definite}\}, \tag{4.95}$$

with l variance components and $l(l-1)/2$ covariance components.

Note the quantor "for all $\sigma^2 \in \mathbb{R}^+$" within the definition of *uniform unbiasedness* (4.79) for one variance component. Indeed, *weakly unbiased estimators exist without the quantor* (*B. Schaffrin* 2000). A similar comment applies to the quantor "for all $\sigma_k \in \{\mathbb{R}^{\ell(\ell+1)/2} | \mathbf{\Sigma_y}\, \textit{positive definite}$" within the definition of *uniform unbiasedness* (4.80) for variance-covariance components. Let us characterize "$\hat{\sigma}^2$IQUUE *of* σ^2".

Lemma 4.12. ($\hat{\sigma}^2$ IQUUE of σ^2):

The scalar $\hat{\sigma}^2$ is IQUUE of σ^2 with respect to the *special Gauss–Markov model of full column rank*.

$$\left[\begin{array}{l} \text{“\textit{first moment}''} : E\{\mathbf{y}\} = \mathbf{A}\xi, \mathbf{A} \in \mathbb{R}^{n\times m}, \mathrm{rk}\mathbf{A} = m \\ \text{“\textit{central second moment}''} : D\{\mathbf{y}\} = \sigma^2, \mathbf{V} \in \mathbb{R}^{n\times n}, \mathrm{rk}\mathbf{V} = n,\ \sigma^2 \in \mathbb{R}^+, \end{array}\right.$$

if and only if

$$(i)\, \mathbf{M}^{1/2}\mathbf{A} = 0 \tag{4.96}$$

and

$$(ii)\, \mathrm{tr}\mathbf{V}\mathbf{M} = 1. \tag{4.97}$$

:*Proof*:

First, we compute $E\{\hat{\sigma}^2\}$.

$$\hat{\sigma}^2 = \text{tr}\mathbf{M}\mathbf{e}_\mathbf{y}\mathbf{e}'_\mathbf{y} \Rightarrow E\{\hat{\sigma}^2\} = \text{tr}\mathrm{M}\Sigma_\mathbf{y} = \text{tr}\boldsymbol{\Sigma}_\mathbf{y}\mathbf{M}.$$

Second, we substitute the "*one variance component model*" $\boldsymbol{\Sigma}_\mathbf{y} = \mathbf{V}\sigma^2$.

$$E\{\hat{\sigma}^2\} := \sigma^2 \;\; \forall\sigma^2 \in \mathbb{R}^- \quad \Leftrightarrow \text{tr}\mathbf{VM} = 1.$$

Third, we adopt the first condition of type "IQE".

The conditions for "$\hat{\sigma}_k$ IQUUE *of* σ_k" are only slightly more complicated.

Lemma 4.13. ($\hat{\sigma}_k$ IQUUE of σ^2):

The vector $\hat{\sigma}_k, k \in \{1, \ldots, \ell(\ell+1)/2\}$ is IQUUE of σ_k with respect to the *block partitioned special Gauss–Markov model of full column rank.*

"*first moment*"

$$E\left\{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{\ell-1} \\ \mathbf{y}_\ell \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_{\ell-1} \\ \mathbf{A}_\ell \end{bmatrix} \xi = \mathbf{A}\xi, \mathbf{A} \in \mathbb{R}^{n_1 n_2 \cdots n_{\ell-1}, n_\ell \times m}, \text{rk}\mathbf{A} = m \tag{4.98}$$

$$n_1 + n_2 + \cdots + n_{\ell-1} + n_\ell = n$$

"*central second moment*"

$$D\left\{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_{\ell-1} \\ \mathbf{y}_\ell \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{V}_{11}\sigma_1^2 & \mathbf{V}_{12}\sigma_{12} & \cdots & \mathbf{V}_{1\ell-1}\sigma_{1\ell-1} & \mathbf{V}_{1\ell}\sigma_{1l} \\ \mathbf{V}_{12}\sigma_{12} & \mathbf{V}_{22}\sigma_2^2 & \cdots & \mathbf{V}_{2\ell-1}\sigma_{2\ell-1} & \mathbf{V}_{2\ell}\sigma_{2\ell} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{V}_{1\ell-1}\sigma_{1\ell-1} & \mathbf{V}_{2\ell-1}\sigma_{2\ell-1} & \cdots & \mathbf{V}_{\ell-1,\ell-1}\sigma_{\ell-1}^2 & \mathbf{V}_{\ell-1,\ell}\sigma_{\ell-1,\ell} \\ \mathbf{V}_{1\ell}\sigma_{1l} & \mathbf{V}_{1\ell}\sigma_{1l} & \cdots & \mathbf{V}_{\ell-1,\ell}\sigma_{\ell-1,\ell} & \mathbf{V}_{\ell,\ell}\sigma_\ell^2 \end{bmatrix} \tag{4.99}$$

$$D\{\mathbf{y}\} = \sum_{j=1}^{\ell} \mathbf{C}_{jj}\sigma_j^2 + \sum_{\substack{j,k=1 \\ j<k}}^{\ell(\ell+1)/2} \mathbf{C}_{jk}\sigma_{jk} \tag{4.100}$$

$$D\{\mathbf{y}\} = \sum_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j\sigma_j \tag{4.101}$$

$$\boldsymbol{\sigma} := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \ldots, \sigma_\ell^2] \in \mathbb{R}^{\ell(\ell+1)/2+1} \tag{4.102}$$

$$\mathbf{C}_j \in \mathbb{R}^{n \times n \times \ell(\ell+1)/2} \; (3d \; array) \tag{4.103}$$

$$\mathbf{V}_{11} \in \mathbb{R}^{n_1 \times n_1}, \mathbf{V}_{12} \in \mathbb{R}^{n_1 \times n_2}, \ldots, \mathbf{V}_{\ell-1,\ell} \in \mathbb{R}^{n_{\ell-1} \times n_l}, \mathbf{V}_{\ell\ell} \in \mathbb{R}^{n_\ell \times n_\ell} \tag{4.104}$$

$$D\{\mathbf{y}\} \sim \mathbf{\Sigma_y} \in \mathbb{R}^{n\times n} = \mathbb{R}^{(n_1+\ldots+n_l)\times(n_1+\ldots+n_l)} \tag{4.105}$$

$$\mathrm{rk}\mathbf{\Sigma}_y = n,\ \ \mathbf{\Sigma}_y\, positive\ \ definite \tag{4.106}$$

if and only if

$$(i)\, \mathbf{M}_k^{1/2}\mathbf{A} = 0 \tag{4.107}$$

and

$$(ii)\, \mathrm{tr}\mathbf{C}_j\mathbf{M}_k = \delta_{jk}. \tag{4.108}$$

Before we continue with the proof we have to comment on our setup of the variance-covariance components model. For a more easy access of an *analyst* we have demonstrated the blocks partitioning of

the observation vector and the variance-covariance

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_\ell \end{bmatrix} \begin{matrix} ,\ \dim \mathbf{y}_1 = n_1 \\ \vdots \\ ,\ \dim \mathbf{y}_\ell = n_\ell \end{matrix}, \qquad \mathbf{\Sigma_y} = \begin{bmatrix} \mathbf{V}_{11}\sigma_1^2 & \cdots & \mathbf{V}_{1\ell}\sigma_{1\ell} \\ \vdots & & \vdots \\ \mathbf{V}_{1\ell}\sigma_{1\ell} & \cdots & \mathbf{V}_{\ell\ell}\sigma_\ell^2 \end{bmatrix}.$$

n_1 observations build up the observation vector y_1 as well as the variance factor V11. Similarly, n_2 observations build up the *variance factor* V_{22}. Both observations collected in the observations vectors y_1 and y_2, constitute the *covariance factor* V_{12}.

This scheme is to be continued for the other observations and their corresponding variance and covariance factors. The matrices $\mathbf{C}_{jj}$ and $\mathbf{C}_{jk}$ which map variance components $\sigma_{jk}(k > j)$ to the variance-covariance matrix $\mathbf{\Sigma_y}$ contain the variance factors $\mathbf{V}_{jj}$ at colj, rowj while the covariance factors contain$\{\mathbf{V}'_{jk}, \mathbf{V}_{jk}\}$ at col_k, row_j and col_j, row_k, respectively. The following proof of *Lemma 4.12* is based upon the linear structure (4.85).

:Proof:

First, we compute $E\{\hat{\sigma}_k\}$.

$$E\{\hat{\sigma}_k\} = \mathrm{tr}\mathbf{M}_k\mathbf{\Sigma_y} = \mathrm{tr}\mathbf{\Sigma_y}\mathbf{M}_k.$$

Second, we substitute the *block partitioning* of the variance-covariance matrix $\mathbf{\Sigma_y}$.

$$\left.\begin{matrix} \mathbf{\Sigma_y} = \sum\limits_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j\sigma_j \\ E\{\hat{\sigma}_k\} = \mathrm{tr}\mathbf{\Sigma_y}\mathbf{M}_k \end{matrix}\right] \Rightarrow \mathrm{tr}\mathbf{\Sigma_y}\mathbf{M}_k = \mathrm{tr} \sum_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j\mathbf{M}_k\sigma_j \tag{4.109}$$

$$\mathrm{tr}\mathbf{C}_j\mathbf{M}_k - \delta_{jk} = 0.$$

Third, we adopt the first conditions of the type “IQE”.

4-34 Invariant Quadratic Uniformly Unbiased Estimations of One Variance Component (IQUUE) from $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE: HIQUUE

Here is our first example of "how to use IQUUE". Let us adopt the residual vector $\tilde{\mathbf{e}}_{\mathbf{y}}$ as *predicted* by $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE for a "one variance component" dispersion model, namely $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$, $\text{rk}\mathbf{V} = m$. *First*, we prove that $\mathbf{M}^{1/2}$ generated by $\mathbf{V}$-BLUUE fulfils both the conditions of IQUUE namely $\mathbf{M}^{1/2}\mathbf{A} = 0$ and $\text{tr}\mathbf{V}\mathbf{M} = \text{tr}\mathbf{V}(\mathbf{M}^{1/2})'\mathbf{M}^{1/2} = 1$. As outlined in *Box 4.5*, the one condition of uniform unbiasedness leads to the solutions for one unknown α within the "*ansatz*" $Z'Z = \alpha\mathbf{V}^{-1}$, namely the number $n-m$ of "*degrees of freedom*" or the "*surjectivity defect*". *Second*, we follow "*Helmert's*" ansatz to setup IQUUE of *Helmert type*, in Short "HIQUUE".

Box 4.5. (IQUUE : one variance component)

1st variations

$$\{E\{\mathbf{y}\} = \mathbf{A}\mathbf{x},\ \mathbf{A} \in \mathbb{R}^{n\times m},\ \text{rk}\mathbf{A} = m,\ D\{\mathbf{y}\} = \mathbf{V}\sigma^2,\ \text{rk}\mathbf{V} = m,\ \sigma^2 \in \mathbb{R}^+\}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}(4.31)$$

1st test: IQE

$$\mathbf{M}^{1/2}\mathbf{A} = 0$$

$$\text{“}if\ \mathbf{M}^{1/2} = \mathbf{Z}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}],\ ,then\,\mathbf{M}^{1/2}\mathbf{A} = 0''$$

2nd test : IQUUE "if tr$\mathbf{V}\mathbf{M} = 1$, then

$$\text{tr}\{\mathbf{V}[\mathbf{I}_n - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}']\mathbf{Z}'\mathbf{Z}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\} = 1$$

$$ansatz : \mathbf{Z}'\mathbf{Z} = \alpha\mathbf{V}^{-1} \tag{4.110}$$

$$\text{tr}\mathbf{V}\mathbf{M} = \alpha\text{tr}\{\mathbf{V}[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}][\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\} = 1$$

$$\text{tr}\mathbf{V}\mathbf{M} = \alpha\text{tr}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})\mathbf{A}'\mathbf{V}^{-1}] = 1$$

$$\left.\begin{array}{l}\text{tr}\mathbf{I}_n = 0\\ \text{tr}[\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}] = \text{tr}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}\mathbf{A} = \text{tr}\mathbf{I}_m = m\end{array}\right] \Rightarrow$$

$$\text{tr}\mathbf{V}\mathbf{M} = \alpha(n-m) = 1 \Rightarrow \alpha = \frac{1}{n-m} \tag{4.111}$$

Let us make a statement about the *translational invariance* of $\tilde{e}_y$ predicted by $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE and specified by the "*one variance component*" model $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{V}\sigma^2$.

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \mathbf{e}_{\mathbf{y}}(\boldsymbol{\Sigma}_{\mathbf{y}}\text{-BLUUE}) = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}]\mathbf{y}. \tag{4.112}$$

Corollary 4.14. *(translational invariance):*

$$\tilde{e}_y = [\mathbf{I} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}]e_y = \mathbf{P}e_y \tag{4.113}$$

subject to

$$P := \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_y^{-1}. \tag{4.114}$$

The proof is "*a nice exercise*": Use $\tilde{\mathbf{e}}_y = \mathbf{P}\mathbf{y}$ and replace $\mathbf{y} = E\{\mathbf{y}\} + \mathbf{e}_y = \mathbf{A}\xi + \mathbf{e}_y$. The result is our statement, which is based upon the "*orthogonality condition*" $\mathbf{PA} = \mathbf{0}$. Note that $\mathbf{P}$ is idempotent in the sense of $\mathbf{P} = \mathbf{P}^2$. In order to generate "$\hat{\sigma}^2$ IQUUE of σ^2" we start from "*Helmert's ansatz*".

Box 4.6. (Helmert's ansatz)

one variance component

$$\tilde{\mathbf{e}}_{\mathbf{y}}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} = \mathbf{e}'_{\mathbf{y}}\mathbf{P}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{P}\mathbf{e}_{\mathbf{y}} = \operatorname{tr}\mathbf{P}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{P}\mathbf{e}_{\mathbf{y}}\mathbf{e}'_{\mathbf{y}} \tag{4.115}$$

$$E\{\tilde{\mathbf{e}}_{\mathbf{y}}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}\} = \operatorname{tr}(\mathbf{P}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{P}\,E\{\mathbf{e}_{\mathbf{y}}\mathbf{e}'_{\mathbf{y}}\}) = \operatorname{tr}(\mathbf{P}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{P}\boldsymbol{\Sigma}_{\mathbf{y}}) \tag{4.116}$$

"*one variance component*"

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{V}\sigma^2 = C_1\sigma^2$$

$$E\{\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}\} = (\operatorname{tr}\mathbf{P}'\mathbf{V}^{-1}\mathbf{P}\mathbf{V})\sigma^2 \quad \forall\sigma^2 \in \mathbb{R}^2 \tag{4.117}$$

$$tr\,\mathbf{P}'\mathbf{V}^{-1}\mathbf{P}\mathbf{V} = \operatorname{tr}[\mathbf{I}_n - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})\mathbf{A}'] = n - m \tag{4.118}$$

$$E\{\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}\} = (n-m)\sigma^2 \tag{4.119}$$

$$\hat{\sigma}^2 := \frac{1}{n-m}\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} \quad\Rightarrow\quad E\{\hat{\sigma}^2\} = \sigma^2. \tag{4.120}$$

Let us finally collect the result of "Helmert's ansatz" in

Corollary 4.15. *($\hat{\sigma}^2$ of HIQUUE of σ^2):*

Helmert's ansatz

$$\hat{\sigma}^2 = \frac{1}{n-m}\mathbf{e}_y'\mathbf{V}^{-1}\mathbf{e}_y \tag{4.121}$$

is IQUUE, also called HIQUUE.

4-35 Invariant Quadratic Uniformly Unbiased Estimators of Variance Covariance Components of Helmert Type: HIQUUE Versus HIQE

In the previous paragraphs we succeeded to prove that *first* $M^{1/2}$ generated by $\tilde{\mathbf{e}}_y = \mathbf{e}_y(\boldsymbol{\Sigma}_\mathbf{y}\text{-BLUUE})$ with respect to "*one variance component*" leads to IQUUE and *second Helmert's ansatz* generated "$\hat{\sigma}^2$IQUUE of σ^2". Here we reverse the order. *First*, we prove that *Helmert's ansatz* for estimating variance-covariance components may lead (or may, in general, not) lead to

$$\text{“}\hat{\sigma}_k \textit{ IQUUE of } \sigma_k\text{''}.$$

Second, we discuss the proper choice of $\mathbf{M}_k^{1/2}$ and test whether (i) $\mathbf{M}_k^{1/2}\mathbf{A} = \mathbf{0}$ and (ii) $\operatorname{tr}\varepsilon_j\mathbf{M}_k = \delta_{jk}$ is fulfilled by HIQUUE of whether $\mathbf{M}_k^{1/2}\mathbf{A} = \mathbf{0}$ is fulfilled by HIQE.

Box 4.7. (*Helmert's ansatz variance-covariance components*):

step one: make a sub order device of variance-covariance components:

$$\sigma_0 := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{12}, \ldots, \sigma_\ell^2]_0'$$

$$\textit{step two}: \textit{compute}\ \boldsymbol{\Sigma}_0 := (\boldsymbol{\Sigma}_y)_0 = \boldsymbol{\Sigma} \sum_{j=1}^{\ell(\ell+1)/2} \mathbf{C}_j\sigma_j(\sigma_0) \tag{4.122}$$

step three: compute $\tilde{\mathbf{e}}_\mathbf{y} = \mathbf{e}_\mathbf{y}(\boldsymbol{\Sigma}_0\text{–BLUUE})$, namely

$$\mathbf{P}(\sigma_0) := (\mathbf{I} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_0^{-1} \tag{4.123}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \mathbf{P}_0\mathbf{y} = \mathbf{P}_0\mathbf{e}_\mathbf{y} \tag{4.124}$$

step four: *Helmert's* ansatz

$$\tilde{\mathbf{e}}_\mathbf{y}'\boldsymbol{\Sigma}_0^{-1}\tilde{\mathbf{e}}_\mathbf{y} = \mathbf{e}'_\mathbf{y}\mathbf{P}'_0\boldsymbol{\Sigma}_0^{-1}\mathbf{P}_\mathbf{0}\mathbf{e}_\mathbf{y} = \operatorname{tr}(\mathbf{P}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{P}_0'\mathbf{e}_\mathbf{y}\mathbf{e}'_\mathbf{y}) \tag{4.125}$$

$$\mathbf{E}\{\hat{\mathbf{e}}_\mathbf{y}'\boldsymbol{\Sigma}_0^{-1}\tilde{\mathbf{e}}_\mathbf{y}\} = tr(\mathbf{P}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{P}_0'\boldsymbol{\Sigma}) \tag{4.126}$$

"*variance-covariance components*"

$$\boldsymbol{\Sigma}_\mathbf{y} = \boldsymbol{\Sigma} \sum_{k=1}^{\ell(\ell+1)/2} \mathbf{C}_k\sigma_k \tag{4.127}$$

$$E\{\tilde{\mathbf{e}}_\mathbf{y}'\boldsymbol{\Sigma}_0^{-1}\tilde{\mathbf{e}}_\mathbf{y}'\} = \operatorname{tr}(\mathbf{P}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{P}_0'\mathbf{C}_k)\sigma_k \tag{4.128}$$

step five: multinomial inverse

$$\mathbf{\Sigma} = \sum_{k=1}^{\ell(\ell+1)/2} \mathbf{C}_k \sigma_k \Leftrightarrow \mathbf{\Sigma}^{-1} = \sum_{k=1}^{\ell(\ell+1)/2} \mathbf{E}_k(\sigma_j) \tag{4.129}$$

input:σ_0, $\mathbf{\Sigma}_0$, *output*:$\mathbf{E}_k(\sigma_0)$.
step six: *Helmert's* equation

$$\begin{gathered} i, j \in \{1, \dots, \ell(\ell+1)/2\} \\ E\{\tilde{\mathbf{e}}'_{\mathbf{y}} \mathbf{E}_i(\sigma_0) \tilde{\mathbf{e}}_{\mathbf{y}}\} = \sum_{k=1}^{\ell(\ell+1)/2} (\mathrm{tr}\mathbf{P}(\sigma_0)\mathbf{E}_i(\sigma_0)\mathbf{P}'(\sigma_0)\mathbf{C}_j)\sigma_j \end{gathered} \tag{4.130}$$

"*Helmert's choice*"

$$\mathbf{e'}_{\mathbf{y}} \mathbf{E}_i\ (\sigma_0)\ \mathbf{e}_{\mathbf{y}} = \sum_{j=1}^{\ell(\ell+1)/2} (\mathrm{tr}\mathbf{P}(\sigma_0)\mathbf{E}_i(\sigma_0)\mathbf{P}'(\sigma_0)\mathbf{C}_j)\sigma_j \tag{4.131}$$

$$q = \mathbf{H}\hat{\boldsymbol{\sigma}} \left[\begin{array}{l} q := \tilde{\mathbf{e}}\mathbf{y}'\mathbf{E}_i(\sigma_0)\tilde{\mathbf{e}}\mathbf{y} \\ \mathbf{H} := \mathrm{tr}\mathbf{P}(\sigma_0)\mathbf{E}_i(\sigma_0)\mathbf{P}'(\sigma_0)\mathbf{C}_j \quad (\textit{Helmert's process}) \\ \hat{\boldsymbol{\sigma}} := [\hat{\sigma}_1^2, \hat{\sigma}_{12}, \hat{\sigma}_2^2, \hat{\sigma}_{13}, \hat{\sigma}_{23}, \hat{\sigma}_3^2, \dots, \hat{\sigma}_\ell^2] \,. \end{array} \right. \tag{4.132}$$

Box 4.7 summarizes the essential steps which lead to "$\hat{\sigma}_k$ HIQUUE of σ_k" if $\det \mathbf{H} = 0$, where $\mathbf{H}$ is the *Helmert matrix*. For the *first step*, we use some prior information $\sigma_0 = \hat{\sigma}_0$ for the unknown variance-covariance components. For instance, $(\mathbf{\Sigma}_y)_0 = \mathbf{\Sigma}_0 = \mathrm{Diag}[(\sigma_1^2)_0, \dots, (\sigma_\ell^2)_0]$ may be the available information on *variance components*, but leaving the *covariance components with zero*. *Step two* enforces the *block partitioning* of the variance-covariance matrix generating the linear space of variance-covariance components. $\tilde{\mathbf{e}}_y = \mathbf{D}_0\mathbf{e}_y$ in *step three* is the local generator of the *Helmert ansatz* in *step four*. Here we derive the key equation $\mathbf{E}\{\tilde{\mathbf{e}}'_{\mathbf{y}} \mathbf{\Sigma}_{\mathbf{0}}^{-\mathbf{1}} \tilde{\mathbf{e}}_{\mathbf{y}}\} = tr(\mathbf{D}_0\mathbf{\Sigma}_0^{-1}\mathbf{D}'_0\mathbf{\Sigma})\,\sigma_k$. *Step five* focuses on the multinormal inverse of the block partitioned matrix $\mathbf{\Sigma}$, also called "*multiple IPM*". *Step six* is taken if we replace Σ_0^{-1} by the block partitioned inverse matrix, on the "*Helmert's ansatz*". The fundamental expectation equation which maps the variance-covariance components σ_j by means of the "*Helmert traces*" $\mathbf{H}$ to the quadratic terms $q(\sigma_0)$. Shipping the expectation operator on the left side, we replace σ_j by their estimates $\hat{\sigma}_j$. As a result we have found the aborted *Helmert equation* $\mathbf{q} = H\hat{\sigma}$ which has to be inverted. Note $E\{\mathbf{q}\} = \mathbf{H}\sigma$ reproducing unbiasedness.

Let us classify the solution of the *Helmert equation* $\mathbf{q} = \mathbf{H}\sigma$ with respect to bias. *First* let us assume that the *Helmert matrix* is of full rank, vk$\mathbf{H} = \ell(\ell+1)/2$ the number of unknown variance-covariance components. The inverse solution, *Box 4.8*, produces an update $\hat{\sigma}_1 = \mathbf{H}^{-1}(\hat{\sigma}_0)'\mathbf{q}(\hat{\sigma}_0)$ out of the zero order information $\hat{\sigma}_0$ we have implemented. For the next step, we iterate $\hat{\sigma}_2 = \mathbf{H}^{-1}(\hat{\sigma}_1)\mathbf{q}(\hat{\sigma}_1)$ up to the *reproducing point* $\hat{\sigma}_w = \hat{\sigma}_{w-1}$ with in computer arithmetic when iteration ends. Indeed, we assume "*Helmert is contracting*".

Box 4.8. (Solving Helmert's equation):

the fast case : $\mathrm{rk}\mathbf{H} = \ell\,(\ell+1)\,/2, \det\mathbf{H} \neq 0$

$$\text{"}\textit{iterated Helmert equation}\text{"}:$$
$$\hat{\sigma}_1 = \mathbf{H}^{-1}(\hat{\sigma}_0)q(\hat{\sigma}_0), \ldots, \hat{\sigma}_\omega = \mathbf{H}_\omega^{-1}(\hat{\sigma}_{\omega-1})q(\hat{\sigma}_{\omega-1}) \tag{4.133}$$
$$\text{"}\textit{reproducing point}\text{"}$$
$$\textit{start}: \quad \sigma_0 = \hat{\sigma}_0 \Rightarrow \hat{\sigma}_1 = \mathbf{H}_0^{-1}q_0 \Rightarrow \hat{\sigma}_2 = \mathbf{H}_1^{-1}q_1$$
$$\textit{subject to}\,\mathbf{H_1} := \mathbf{H}(\hat{\sigma}_1), q_1 := q(\hat{\sigma}_1)$$
$$\Rightarrow \cdots \Rightarrow \hat{\sigma}_\omega = \hat{\sigma}_{\omega-1}\,(\textit{computer arithmetic}):\ \textit{end}.$$

?Is the special Helmert variance-covariance estimator

Corollary 4.16 gives a positive answer.

Corollary 4.16. *(Helmert equation, det H≠ 0);*

In case the *Helmert matrix* $\mathbf{H}$ is a full rank matrix, namely $\mathrm{rk}\mathbf{H} = \ell\,(\ell+1)\,/2$

$$\hat{\sigma} = \mathbf{H}^{-1}\mathbf{q} \tag{4.134}$$

is $\mathbf{\Sigma}_\infty$-HIQUUE at reproducing point.

Proof.

$$\mathbf{q} := \tilde{\mathbf{e}}_y'\mathbf{E}_i\tilde{\mathbf{e}}_y$$

$E\{\hat{\sigma}\} = \mathbf{H}^{-1}E\{\mathbf{q}\} = \mathbf{H}^{-1}\mathbf{H}\sigma = \sigma.$ ♣

For the second case of our classification, let us assume that *Helmert matrix* is no longer of full rank, $\mathrm{rk}\,\mathbf{H} < \ell(\ell+1)/2$, $\det\mathbf{H} = 0$. Now we are left with the central question.

? Is the special Helmert variance-covariance estimator $\sigma = \mathbf{H}_{l_n-1}^{-}q = \mathbf{H}^{+}q$ of type "MINOLESS" "IQUUE"?

Unfortunately, the MINOLESS of the rank factorized Helmert equation $\mathbf{q} = \mathbf{JK}\hat{\sigma}$ outlined in *Box 4.9* by the *weighted Moore–Penrose solution*, indicates a negative answer. Instead, *Corollary 4* proves $\hat{\sigma}$ is only HIQE, but resumes also in *establishing estimable variance-covariance components* as "*Helmert linear combinations*" of them.

Box 4.9. (Solving Helmerts equation the second case):

rk $\mathbf{H} < \ell(\ell+1)/2$, det $\mathbf{H} = 0$ *"rank factorization" "MINOLESS"*

$$\mathbf{H} = \mathbf{JK},\ \text{rk}\mathbf{H} = \text{rk}\mathbf{F} = \text{rk}\mathbf{G} =: \mathbf{v} \tag{4.135}$$

"dimension identities″

$$\mathbf{H} \in \mathbb{R}^{\ell(\ell+1)/2\times\ell(\ell+1)/2},\ \mathbf{J} \in \mathbb{R}^{\ell(\ell+1)/2\times v},\ \mathbf{G} \in \mathbb{R}^{v\times\ell(\ell+1)/2}$$

$$\mathbf{H}_{lm}^{-} = \mathbf{H}^{+}(weighted) = \mathbf{K}_R^{-}(weighted) = \mathbf{J}_L^{-}(weighted) \tag{4.136}$$

$$\hat{\sigma}_{lm} = \mathbf{G}_\sigma^{-1}\mathbf{K}'(\mathbf{K}\mathbf{G}_\sigma^{-1}\mathbf{K}^{-1})(\mathbf{J}'\mathbf{G}_\mathbf{q}\mathbf{J})^{-1}\mathbf{G}_q q = \mathbf{H}_{\sigma,q}^{+} q. \tag{4.137}$$

In case "det $\mathbf{H} = 0$" Helmerts variance-covariance components estimation is *no longer unbiased*, but estimable functions like $H\hat{\sigma}$ exist:

Corollary 4.17. *(Helmert equation, det H=0):*

In case the Helmert matrix H, rkH¡$\ell(\ell+1)/2$, det $\mathbf{H} = 0$, is rank deficient, the *Helmert equation no longer generates an unbiased* IQE. An estimable parameter set is H$\hat{\sigma}$:

$$(i)\ H\hat{\sigma} = \mathbf{H}\mathbf{H}^{+}\mathbf{q}\ \text{is}\boldsymbol{\Sigma}_0 - \text{HIQUUE} \tag{4.138}$$

$$(ii)\ \hat{\sigma}\ is\ IQE.$$

Proof.

$$(i)\ E\{\hat{\sigma}\} = \mathbf{H}^{+}E\{\mathbf{q}\} = \mathbf{H}^{+}\mathbf{H}\sigma \neq \sigma,\ \hat{\sigma} \in \text{IQE}$$

$$(ii)\ E\{H\hat{\sigma}\} = \mathbf{H}\mathbf{H}^{+}E\{\mathbf{q}\} = \mathbf{H}\mathbf{H}^{+}\mathbf{H}\sigma = \mathbf{H}\sigma,\ H\hat{\sigma} \in \text{HIQUUE}.$$

♣

In summary, we lost a bit of our illusion that $\tilde{\sigma}_\mathbf{y}(\boldsymbol{\Sigma}_\mathbf{y}-\text{BLUUE})$ now always produces IQUUE.

"The illusion of progress is short, but exciting"
"Solving the Helmert equations"

Figure 4.1 illustrates the result of *Corollary 4* and *Corollary 5*. Another drawback is that we have no guarantee that

HIQE or HIQUUE

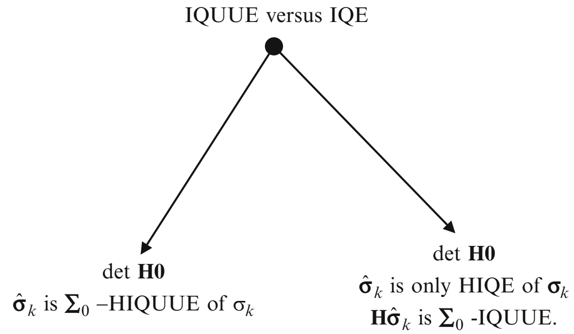


Fig. 4.1 Solving the Helmert equation for estimating variance-covariance-components

generates a positive definite variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$. Such a postulate can be enforced by means of an *inequality constraint* on the *Helmert equation* $H\hat{\sigma} = \mathbf{q}$ of type "$\hat{\sigma} > 0$" or "$\hat{\sigma} > \sigma$" in symbolic writing. Then consult the text books on "positive variance-covariance component estimation". At this end, we have to give credit to *B. Schaffrin* (1983, p. 62) who classified Helmerts variance-covariance components estimation for the first time correctly.

4-36 Best Quadratic Uniformly Unbiased Estimations of One Variance Component: BIQUUE

First, we give a definition of "best" $\hat{\sigma}^2$IQUUE of σ^2 within *Definition 4.18* namely for a *Gauss normal random variable* $\mathbf{y} \in \mathbf{Y} = \{\mathbb{R}^n, pdf\}$. *Definition 4.19* presents a basic result representing "Gauss normal" BIQUUE. In particular we outline the *reduction* of fourth order moments to second order moments if the random variable y is Gauss normal or, more generally, quasi-normal. At same length we discuss the suitable choice of the proper constrained Lagrangean generating $\hat{\sigma}^2$ BIQUUE of σ^2. The highlighted is *Lemma 4* where we resume the normal equations typical for BIQUUE and *Theorem 4* with explicit representations of $\hat{\sigma}^2$, $D\{\hat{\sigma}^2\}$ and $\hat{D}\{\hat{\sigma}^2\}$ of type BIQUUE with respect to the *special Gauss–Markov model with full column rank*.

? What is the "best" $\hat{\sigma}^2$ IQUUE of σ^2?

First, let us define what is "best" IQUUE.

Definition 4.18. ($\hat{\sigma}^2$ best invariant quadratic uniformly unbiased estimation of σ^2 : BIQUUE)

Let $\mathbf{y} \in \{\mathbb{R}^n, pdf\}$ be a *Gauss normal random variable* representing the stochastic observation vector. Its central moments up to order four

$$E\{e_i^{\mathbf{y}}\} = 0,$$

$$E\{e_i^{\mathbf{y}} e_j^{\mathbf{y}}\} = \pi_{ij} = v_{ij}\sigma^2 \tag{4.139}$$

$$E\{e_i^{\mathbf{y}} e_j^{\mathbf{y}} e_k^{\mathbf{y}}\} = \pi_{ijk} = 0, \text{ (obliquity)} \tag{4.140}$$

$$E\{e_i^{\mathbf{y}} e_j^{\mathbf{y}} e_k^{\mathbf{y}} e_l^{\mathbf{y}}\} = \pi_{ijkl} = \pi_{ij}\pi_{kl} + \pi_{ik}\pi_{jl} + \pi_{il}\pi_{jk}$$
$$= (v_{ij}v_{kl} + v_{ik}v_{jl} + v_{il}v_{jk})\sigma^4 \tag{4.141}$$

relate to the "centralized random variable"

$$\mathbf{e_y} := \mathbf{y} - E\{\mathbf{y}\} = [e_i^{\mathbf{y}}]. \tag{4.142}$$

The moment arrays are taken over the index set $i, j, k, l \in 1{,}000, n$ when the natural number n is identified as the number of observations. n is the dimension of the observation space $\mathbf{y} \in \{\mathbb{R}^n, pdf\}$.

The scalar $\hat{\sigma}^2$ is called BIQUUE of σ^2(Best Invariant Quadratic Uniformly Unbiased Estimation) of the *special Gauss–Markov model of full column rank.*

"first moments$''$:

$$E\{\mathbf{y}\} = \mathbf{A}\xi, \ \mathbf{A} \in \mathbb{R}^{n\times m}, \text{¸} \in \mathbb{R}^m, \ \text{rk}\mathbf{A} = m \tag{4.143}$$

"central second moments$''$:

$$D\{\mathbf{y}\} \sim \mathbf{a_y} = \mathbf{V}\sigma^2, \mathbf{V} \in \mathbb{R}^{n\times m}, \sigma^2 \in \mathbb{R}^+, \text{rk}\mathbf{V} = n \tag{4.144}$$

where ¸ $\in \mathbb{R}^m$ is the first unknown vector and $\sigma^2 \in \mathbb{R}^+$ the second un-known "one variance component", if it is.

(i) *a quadratic estimation (IQE)*:

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\text{vec}\mathbf{M})'\mathbf{y} \otimes \mathbf{y} = \text{tr}\mathbf{M}\mathbf{y}\mathbf{y}' \tag{4.145}$$

subject to

$$\mathbf{M} = (\mathbf{M}^{\frac{1}{2}})'\mathbf{M}^{\frac{1}{2}} \in SYM := \{\mathbf{M} \in \mathbb{R}^{n\times m} \mid \mathbf{M} = \mathbf{M}'\} \tag{4.146}$$

(ii) *translational invariant, in the sense of*

$$\mathbf{y} \to \mathbf{y} - E\{\mathbf{y}\} =: \mathbf{e_y} \tag{4.147}$$

$$\hat{\sigma}^2 = y'\mathbf{M}y = e_y'\mathbf{M}e_y \tag{4.148}$$

or equivalently

$$\hat{\sigma}^2 = (\text{vec}\mathbf{M})'y \otimes y = (\text{vec}\mathbf{M})'e_y \otimes e_y \tag{4.149}$$

$$\hat{\sigma}^2 = \text{tr}\mathbf{M}yy' = \text{tr}\mathbf{M}\mathbf{e}_y\mathbf{e}'_y \tag{4.150}$$

(iii) *uniformly unbiased in the sense of*

$$E\{\hat{\sigma}^2\} = \sigma^2, \forall \sigma^2 \in \mathbb{R}^+ \tag{4.151}$$

and

(iv) *of minimal variance in the sense*

$$D\{\hat{\sigma}^2\} := E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2\} = \min_{\mathbf{M}}. \tag{4.152}$$

In order to produce "best" IQUUE we have to analyze the variance $E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^{-1}\}$ of the invariant quadratic estimation $\hat{\sigma}^2$ the "*one variance component*", of σ^2. In short, we present to you the result in

Corollary 4.19. *(the variance of $\hat{\sigma}$ with respect to a Gauss normal IQE):*

If $\hat{\sigma}^2$is IQE of σ^2, then for a Gauss normal observation space $\mathbf{Y} = \{\mathbb{R}^n, pdf\}$ the variance of σ^2 of type IQE is represented by

$$E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2\} = 2\mathrm{tr}\mathbf{M}'\mathbf{V}\mathbf{M}\mathbf{V}. \tag{4.153}$$

Proof. ansatz : IQE

$$\begin{aligned}
\hat{\sigma}^2 &= \mathrm{tr}\mathbf{M}\mathbf{e_y}\mathbf{e}'_\mathbf{y} \;\Rightarrow\; E\{\hat{\sigma}^2\} = (\mathrm{tr}\mathbf{M}\mathbf{V})\sigma^2 \\
E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]\} &= E\{[\mathrm{tr}\mathbf{M}\mathbf{e_y}\mathbf{e}'_\mathbf{y} - (\mathrm{tr}\mathbf{M}\mathbf{V})\sigma^2][\mathrm{tr}\mathbf{M}\mathbf{e_y}\mathbf{e}'_\mathbf{y} - (\mathrm{tr}\mathbf{M}\mathbf{V})\sigma^2]\} \\
&= E\{(\mathrm{tr}\mathbf{M}\mathbf{e_y}\mathbf{e}'_\mathbf{y})(\mathrm{tr}\mathbf{M}\mathbf{e_y}\mathbf{e}'_\mathbf{y})\} - (\mathrm{tr}\mathbf{M}\mathbf{V})^2\sigma^4 \;\Rightarrow\; (4.153).
\end{aligned}$$

♣

With the "ansatz" $\hat{\sigma}^2$ IQE of σ^2 we have achieved the first decomposition of var$\{\hat{\sigma}^2\}$. The second decomposition of the first term will lead us to central moments of fourth order which will be decomposed into central moments of second order for a *Gauss normal random variable y*. The computation is easiest in "Ricci calculus". An alternative computation of the reduction "fourth moments to second moments" in "*Cayley calculus*" which is a bit more advanced, is gives in Appendix A-7.

$$\begin{aligned}
E\{(\mathrm{tr}\mathbf{M}\mathbf{e}_y\mathbf{e}'_y)(\mathrm{tr}\mathbf{M}\mathbf{e}_y\mathbf{e}'_y)\} &= \sum_{i,j,k,l=1}^{n} \mathbf{m}_{ij}\mathbf{m}_{kl} E\{\mathbf{e}_i^y\mathbf{e}_j^y\mathbf{e}_k^y\mathbf{e}_l^y\} = \sum_{i,j,k,l=1}^{n} \mathbf{m}_{ij}\mathbf{m}_{kl}\boldsymbol{\beta}_{ijkl} \\
&= \sum_{i,j,k,l=1}^{n} \mathbf{m}_{ij}\mathbf{m}_{kl}(\boldsymbol{\beta}_{ij}\boldsymbol{\beta}_{kl} + \boldsymbol{\beta}_{ik}\boldsymbol{\beta}_{jl} + \boldsymbol{\beta}_{il}\boldsymbol{\beta}_{jk}) \\
&= \sum_{i,j,k,l=1}^{n} \mathbf{m}_{ij}\mathbf{m}_{kl}(\mathbf{v}_{ij}\mathbf{v}_{kl} + \mathbf{v}_{ik}\mathbf{v}_{jl} + \mathbf{v}_{il}\mathbf{v}_{jk})\sigma^4
\end{aligned}$$

$$\mathrm{E}\{(\mathrm{tr}\mathbf{M}\mathbf{e}_y\mathbf{e}'_y)(\mathrm{tr}\mathrm{M}\mathbf{e}_y\mathbf{e}'_y)\} = \sigma^4(\mathrm{tr}\mathbf{M}\mathbf{V})^2 + 2\sigma^4\mathrm{tr}(\mathbf{M}\mathbf{V})^2. \tag{4.154}$$

A combination of the first and second decomposition leads to the final result.

$$
\begin{aligned}
E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]\} &= E\{(\mathrm{tr}\mathbf{M}\mathbf{e}_{\mathbf{y}}\mathbf{e}'_{\mathbf{y}})(\mathrm{tr}\mathbf{M}\mathbf{e}_{\mathbf{y}}\mathbf{e}'_{\mathbf{y}})\} - \sigma^4(\mathrm{tr}\mathbf{M}\mathbf{V}) \\
&= 2\sigma^4(\mathrm{tr}\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{V}).
\end{aligned}
\qquad \clubsuit
$$

A *first choice of a constrained Lagrangean* for the optimization problems "BIQUUE", namely (4.158) of *Box 4.10*, is based upon the variance

$$E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}] \mid \mathrm{IQE}\}$$

constrained to "IQE" *and*

(i) the condition of uniform unbiasedness

$$(\mathrm{tr}\mathbf{V}\mathbf{M}) - 1 = 0$$

as well as

(ii) the condition of the invariant quadratic estimation $\mathbf{A}'(\mathbf{M}^{1/2}) = 0$.

A *second choice of a constrained Lagrangean* generating $\hat{\sigma}^2$ BIQUUE of σ^2, namely (4.160) of *Box 4.10*, takes advantage of the general solution of the homogeneous matrix equation $\mathbf{M}^{1/2}\mathbf{A} = 0$ which we already obtained for "IQE". Equation (4.72) is the matrix container for $\mathbf{M}$. In consequence, building into the Lagrangean the structure of the matrix $\mathbf{M}$, desired by the condition of the *invariance quadratic estimation* $\hat{\sigma}^2$ IQE ofσ^2 reduces the first Lagrangean by the second condition. Accordingly, the *second choice of the Lagrangean* (4.160) includes only one condition, in particular the condition for an *uniformly unbiased* estimation $(\mathrm{tr}\mathbf{V}\mathbf{M}) - 1 = 0$. Still we are left with the problem to make a proper choice for the matrices $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{G}_{\mathbf{y}}$. The first "*ansatz*" $\mathbf{Z}'\mathbf{Z} = \alpha \mathrm{G}_y$ produces a specific matrix $\mathbf{M}$, while the second "*ansatz*" $\mathbf{G}_{\mathbf{y}} = \mathbf{V}^{-1}$ couples the matrix of the metric of the observation space to the inverse variance factor $\mathbf{V}^{-1}$. Those "*natural specifications*" reduce the second Lagrangean to a specific form (4.161), a third Lagrangean which only depends on two unknowns, α and λ_0. Now we are prepared to present the basic result for $\hat{\sigma}^2$ BIQUUE of σ^2.

Box 4.10. (Choices of constrained Lagrangeans generating $\hat{\sigma}^2$ BIQUUE of σ^2)

"*a first choice*"

$$L(\mathbf{M}^{1/2}, \lambda_0, \mathbf{A}_1) := 2\mathrm{tr}(\mathbf{M}\mathbf{V}\mathbf{M}\mathbf{V}) + 2\lambda_0[(\mathrm{tr}\mathbf{V}\mathbf{M}) - 1] + 2\mathrm{tr}\mathbf{A}_1\mathbf{A}'(\mathbf{M}^{1/2})' \tag{4.155}$$

"*a second choice*"

$$\mathbf{M} = (\mathbf{M}^{1/2})'\mathbf{M}^{1/2} = [\mathbf{I}_n - \mathbf{G}_{\mathbf{y}}\mathbf{A}(\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{A})^{-1}\mathbf{A}']\mathbf{Z}'\mathbf{Z}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}_{\mathbf{y}}] \tag{4.156}$$

ansatz : $\mathbf{Z}'\mathbf{Z} = \alpha G_{\mathbf{y}}$

$$\mathbf{M} = \alpha \mathrm{G}_{\mathbf{y}}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G}_{\mathbf{y}}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G}_{\mathbf{y}}] \tag{4.157}$$

$$\mathbf{VM} = \alpha \mathrm{V} G_{\mathbf{y}}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{G_yA})^{-1}\mathbf{A}'\mathbf{G_y}] \tag{4.158}$$

$$\textit{ansatz} : \mathbf{G_y} = \mathbf{V}^{-1}$$

$$\mathbf{VM} = \alpha[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}] \tag{4.159}$$

$$L(\alpha, \lambda_0) = \mathrm{tr}\mathbf{MVMV} + 2\lambda_0[(\mathbf{VM} - 1)] \tag{4.160}$$

$$\mathrm{tr}\mathbf{MVMV} = \alpha^2 \mathrm{tr}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}] = \alpha^2(n-m)$$

$$\mathrm{tr}\mathbf{VM} = \alpha \mathrm{tr}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}] = \alpha(n-m)$$

$$L(\alpha, \lambda_0) = \alpha^2(n-m) + 2\lambda_0[\alpha(n-m) - 1] = \min_{\alpha,\lambda_0}. \tag{4.161}$$

Lemma 4.20. ($\hat{\sigma}^2$ BIQUUE of σ^2):

The scalar $\hat{\sigma}^2 = \mathbf{y}'\mathbf{My}$ is BIQUUE of σ^2 with respect to *special Gauss–Markov model of full column rank*, if and only if the matrix α together with the "*Lagrange multiplier*" fulfills the system of normal equations

$$\begin{bmatrix} 1 & 1 \\ n-m & 0 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda}_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{4.162}$$

solved by

$$\hat{\alpha} = \frac{1}{n-m}, \quad \lambda_0 = -\frac{1}{n-m}. \tag{4.163}$$

: Proof:

Minimizing the *constrained Lagrangean*

$$L(\alpha, \lambda_0) = \alpha^2(n-m) + 2\lambda_0[\alpha(n-m) - 1] = \min_{\alpha,\lambda_0}$$

leads us to the *necessary conditions*

$$\frac{1}{2}\frac{\partial L}{\partial \alpha}(\hat{\alpha}, \hat{\lambda}_0) = \hat{\alpha}(n-m) + \hat{\lambda}_0(n-m) = 0$$

$$\frac{1}{2}\frac{\partial L}{\partial \lambda_0}(\hat{\alpha}, \hat{\lambda}_0) = \hat{\alpha}(n-m) - 1 = 0$$

or

$$\begin{bmatrix} 1 & 1 \\ n-m & 0 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda}_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

solved by $\hat{\alpha} = -\hat{\lambda}_0 = \frac{1}{n-m}$.

$$\frac{1}{2}\frac{\partial^2 L}{\partial \alpha^2}(\hat{\alpha}, \hat{\lambda}_0) = n - m \times 0$$

constitutes the necessary condition, automatically fulfilled. Such a solution for the parameter α leads us to the "BIQUUE" representation of the matrix **M**.

$$\mathbf{M} = \frac{1}{n-m}\mathbf{V}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}] \tag{4.164}$$

Explicit representations $\hat{\sigma}^2$ BIQUUE ofσ^2, of the variance $D\{\hat{\sigma}^2\}$ and its estimate $D\{\hat{\sigma}^2\}$ are highlighted by

Theorem 4.21. ($\hat{\sigma}$ BIQUUE of σ^2):

Let $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\text{vec}\mathbf{M})'(\mathbf{y} \otimes \mathbf{y}) = \text{tr}\mathbf{M}\mathbf{y}\mathbf{y}'$ be BIQUUE of σ^2 with reseat to the *special Gauss–Markov model of full column rank*.

(i) $\hat{\sigma}^2$ BIQUUE of σ^2
Explicit representations of $\hat{\sigma}^2$ BIQUUE of σ^2

$$\hat{\sigma}^2 = (n-m)^{-1}\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y} \tag{4.165}$$

$$\hat{\sigma}^2 = (n-m)^{-1}\tilde{\mathbf{e}}'\mathbf{V}^{-1}\tilde{\mathbf{e}} \tag{4.166}$$

$$subjectto\tilde{\mathbf{e}} = \tilde{\mathbf{e}}$$

(BLUUE).

(ii) D$\hat{\sigma}^2$ | BIQUUE
BIQUUEs variance is explicitly represented by

$$D\{\hat{\sigma}^2|BIQUUE\} = E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2 \mid \text{BIQUUE}\} = 2(n-m)^{-1}(\sigma^2)^2. \tag{4.167}$$

(iii) $\tilde{D}\{\hat{\sigma}^2\}$
An estimate of BIQUUEs variance is

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-m)^{-1}(\hat{\sigma}^2) \tag{4.168}$$

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-m)^{-3}(\tilde{\mathbf{e}}'\mathbf{V}^{-1}\tilde{\mathbf{e}})^2. \tag{4.169}$$

: Proof:

We have already prepared the proof for (i). Therefore we continue to prove (ii) and (iii)

$$(ii)\, D\{\hat{\sigma}^2 \mid \text{BIQUUE}\}$$

$$D\{\hat{\sigma}^2\} = E\{[\hat{\sigma}^2 - E\{\hat{\sigma}^2\}]^2\} = 2\sigma^2\text{tr}\mathbf{MVMV},$$

$$\mathbf{MV} = \frac{1}{n-m}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}],$$

$$\mathbf{MVMV} = \frac{1}{(n-m)^2}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}],$$

$$\mathrm{tr}\mathbf{MVMV} = \frac{1}{n-m} \Rightarrow$$

$$\Rightarrow \quad D\{\hat{\sigma}^2\} = 2(n-m)^{-1}(\sigma^2)^2.$$

$$(iii)\ D\{\hat{\sigma}^2\}$$

Just replace within $D\{\hat{\sigma}^2\}$ the variance σ^2 by the estimate $\hat{\sigma}^2$.

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-m)^{-1}(\hat{\sigma}^2)^2. \qquad \clubsuit$$

Upon writing the chapter on variance-covariance component estimation we learnt about the untimely death of *J.F. Seely*, Professor of Statistics at Oregon State University, on *23 February 2002*. *J.F. Seely*, born on *11 February 1941* in the small town of *Mt. Pleasant, Utah*, who made various influential contributions to the theory of *Gauss–Markov linear model*, namely the quadratic statistics for estimation of variance components. His Ph.D. adviser *G. Zyskind* had elegantly characterized the situation where *ordinary least squares approximation* of fixed effects remains optimal for mixed models: the regression space should be invariant under multiplication by the *variance-covariance* matrix. *J.F. Seely* extended this idea to *variance-covariance component estimation*, introducing the notion of *invariant quadratic subspaces* and their relation to completeness. By characterizing the class of *admissible unbiased estimators* of variance-covariance components. In particular, the usual ANOVA estimator in 2-variance component models is inadmissible. Among other contributions to the theory of *mixed models*, he succeeded in generalizing and improving on several existing procedures for *tests* and *confidence intervals on variance-covariance components*.

Additional Reading

Seely. J. and *Lee, Y.* (confidence interval for a variance: 1994), *Azzam, A., Birkes, A.D. and Seely, J.* (admissibility in linear models, polyhydral covariance structure: 1988), *Seely, J. and Rady, E.* (random effects–fixed effects, linear hypothesis: 1988), *Seely, J. and Hogg, R.V.* (unbiased estimation in linear models: 1982), *Seely, J.* (confidence intervals for positive linear combinations of variance components, 1980), *Seely, J.* (minimal sufficient statistics and completeness, 1977), *Olsen, A., Seely, J. and Birkes, D.* (invariant quadratic unbiased estimators for two variance components, 1975), *Seely, J.* (quadratic sub-spaces and completeness, 1971) and *Seely, J.* (linear spaces and unbiased estimation, 1970).

4-37 *Simultaneous Determination of First Moment and the Second Central Moment, Inhomogeneous Multilinear Estimation, the $E-D$ Correspondence, Bayes Design with Moment Estimations*

The general linear model with linear dispersion structure is reviewed with respect to (i) *unbiasedness*, (ii) *invariance* and (iii) *nonnegativeness* of the variance-covariance component estimation. The $E-D$ *correspondence* is outlined as well as the importance of quadratic dispersion component estimations and their *reproducing property*. We illustrate the theoretical investigations by two examples: (i) one variance component estimation and (ii) variance-covariance component estimation of *Helmert type*. These results are extended by three other examples: (iii) hybrid direction-distance network, (iv) gyroscopic azimuth observations and (v) *Bayes design with moment estimation*.

The subject of *variance-covariance component estimation* within an adjustment process was one of the central research topics in the last decade, both in geodesy and mathematical statistics. In a remarkable bibliography which up-to-date to the year *1977 H. Sahai* listed more than 1,000 papers on variance component estimation where the basic source was *"Statistical Theory and Method Abstracts"*, published for the International Statistical Institute by Longman Group Limited, "Mathematical Reviews" and "Abstract Service of Quality Control and Applied Statistics". Let us refer also to his bibliography (1979) as well as *H. Sahai, A.J. Khuri and C.H. Kapadia* (1985), *C.R. Rao* and *J. Kleffe* (1988), *R.D. Anderson* (1979), *C.R. Henderson* (1953), *J. Kleffe* (1975, 1977), *H. Drygas* (1977,1981), *S. Gnot and J. Kleffe* (1983), *S. Gnot, J. Kleffe and R. Zymslong* (1985), *P.S.R.S. Rao* (1977), *S.R. Searle*(1978, 1979), *L.R. Verdooren* (1979,1980) and *R. Thompson* (1962, 1980). The PhD thesis of *B. Schaffrin* (1983) offers a critical review of the *state-of-the-art*.

In *Geodetic Sciences* variance component estimation originates from *F.R. Helmert* (1907) who used least squares residuals to estimate *heterogeneous variance components*. Instead *K. Kubik* (1967 i, ii, iii, 1970) used early the method on *Maximum Likelihood* for estimating *weight ratios* in a hybrid distance-direction network. In this context we refer also to *H.O Hartley and J.N.K Rao*, *H.D. Patterson and R. Thompson* 1971) and *K. R. Koch* (1986). *R. Kelm* (1978) and E. Grafarend (1978 i, ii, 1982) used *Helmert's variance component estimation* for geodetic examples. *Monte-Carlo algorithms* were used for variance components estimation by *J. Kusche* (2003). In his PhD thesis *M. Serbetci* analyzed *gravimetric measurements* by the *Helmert estimation techniques*. Numerical extensions of his estimation method originate from *H. Ebner* (1972, 1977), *W. Foerstner* (1979, 1982) and *W. Welsch* (1977, 1978, 1979, 1980). MINQUE estimators were used by *R. Kelm* (1978), with respect to the *Gauss–Markov* model various geodetic applications have proven the power of variance component estimation: *H. Fröhlich* (1982, 1983, 1984), *H. Fröhlich and D. Duddek* (1983), *K. R. Koch* (1981) and *L. Sjöberg* (1980) used variance component estimation for the analysis of the *constant and distance dependent variance factor* in electro-optical distance measurements.

The *analysis of deforming networks* was based on variance component estimation by *Y. Chen* (1983), *K. R. Koch*(1981), *B. Schaffrin* (1981, 1983) and *P. Schwintzer* (1984). *H. Pelzer* (1982) presented a special variance component model for *eccentric direction measurements* . *J. D. Bossler and R.H. Hanson* (1980) used *Bayesian variance component estimation* for *densification networks*. A first application of variance component estimation of *Helmert type* has been in heterogeneously observed networks, for instance *direction and distance measurements* characterized by an unknown variance component. Examples are given by *E. Grafarend and A. d'Hone* (1978), *E. Grafarend, A. Kleusberg and B. Schaffrin* (1980), *K. Kubik* (1967), *O. Remmer* (1971) and *W. Welsch* (1978, 1979, 1980). A special field of geodetic application has been oscillation analysis based on a basic contribution of *H. Wolf* (1975) which was applied by *M. Junasevic* (1977) for estimating the signal-to-noise ratios in *gyroscopic azimuth observations* and has been extended by *E. Grafarend and A. Kleusberg* (1980) and *A. Kleusberg and E. Grafarend* (1981). Geodetic examples in *Condition Adjustment* and *Condition Adjustment with Unknowns (Gauss–Helmert model)* including variance-covariance component estimation, namely in collocation problems, have been given by *E. Grafarend* (1984), *C. G. Persson* (1981, 1982) and *L. Sjöberg* (1983 i, ii, 1983, 1984 i, ii, 1985, 1995).

A. Amiri-Simkooej (2007) applied variance component estimation to GPS (*Global Positioning System*). *W. Lindlohr* (1983) used variance component estimation for the analysis of a special *auto regression process in photogrammetric networks*. *P. Xu* et al. (2007) discussed the estimating problem of variance-covariance components and proved conditions with respect to the *rank* of the variance-covariance matrix to be estimate. Of course, *not* all components of a full variance and covariance matrix are estimated, only $r(r-1)|2$ when we assume a variance–covariance matrix of rank $r(r-1)$. In a definition in *P. Xu et al* (2006), they presented *a zero order Thykhonov regularization* method for estimating variance-covariance components. A basic result which they found is the simultaneous estimation of the regularization parameter and related variance component estimation *is advantages*. *P.J.G. Teunissen and A. R. Amiri-Simkooej* (2008) studied least-squares variance component estimation: they treated an elliptical contoured distribution, best linear uniformly unbiased estimator (*BLUE*), best invariant quadratic unbiased estimator (*BIQUE*), minimum norm quadratic unbiased estimator (*MINQUE*) and restricted maximum likelihood estimator (*REML*). *N. Crocetto, M. Gatti and P. Russo* (2000) presented simplified formulae for the *BIQUE* estimation observation groups. *K. R. Koch* (1986, 1987) compares *Maximum Likelihood Estimation* and *Bayesian Inference* for variance components. They also studied in K. R. Koch and J. Kusche (2002) the regularization of potential determination from satellite data by variance components. J. Kusche (2003) analyzed noise variance estimation and optimal weight determination for GOCE gravity recovery. *Z. Ou* (1989, 1991) discussed various methods for variance-covariance component estimation and applied it to *approximation Bayes estimation*.

A special procedure for estimating the variance function of linear models for GPS carrier-phase observations has been developed by *W. Bischoff et al* (2006).

C. C. J. M. Tiberius and F. Kenselaar applied variance components estimation for *precise GPS positioning*. Z. Yu (1992, 1996) tried to generalizes variance-covariance component estimation and applied it to *Maximum Likelihood Estimation*. Finally we have to refer to the most important geodetic contributions of *K. R. Koch* (1980, 1990, 1999, 2000) and *K.R. Koch and M. Schmidt* (1994): They estimated *a set-up of a covariance* and designed *a test about the identity* with respect to a prescribed design matrix in *K. R. Koch* (1976). Various estimators of variance components were analyzed in *K. R. Koch* (1978, 1979i (mixed models), 1979ii(*Gauss–Helmert*)) of variance-covariance components in K. R. Koch (1981, "*Gauge unknown in distance measurements*", 1983 "*Multivariate and incomplete multivariate models*"), *Bayes* estimates of variance and *Bayes Inference* (1994, 1988i, ii) as well as maximum likelihood estimates of variance components in *K. R. Koch and A. J. Pope* (1986). Notable is also the contribution "*Comments on P. Xu et al.*" in *K. R. Koch and J. Kusche* (2006) relating to *variance component estimation in linear inverse ill-posed models*.

The basic problem with the analysis of first moments and central second moments with second order statistics is their problematic structure design of their components. Have the *first order moments* a linear or nonlinear structure? *Is the rank detectable*? With respect to *second order control moments* it is very difficult to design. How many variances and covariances are needed for a reasonable structure? In case of variances it is very difficult to find out what is the rank of the *variance matrix*. A more question is the structure of covariances or of a *variance-covariance matrix*. What is the rank of an *estimate set* of the *variances and of the covariances*. *Who decides* upon a fixing those variances and covariances?

One key problem, in addition, is the one how to produce realistic, *non-negative variances* or *non-negative variance-covariance components*. Most of the analysis programs work well to determine variances and covariances when there is no bias among the observations and the redundancy of the observations is "reasonable high". How good are the *initial values* when variance-covariance values in an iteration process are used? How can we guarantee non-negative estimates for a *variance-covariance matrix without any physical meaning*? *L.R. La Motte* (1973) presented "a" non-negative quadratic unbiased estimation procedure for variance components. An alternative *non-negative estimation procedure* for variance which is "*almost*" unbiased non-negative iterative estimator was designed by S.D. Horn (1975). *W. Foerstner* (1979) came up with another non-negative estimator of variances. It is well-known that MINQUE ("*Minimum Norm Quadratic Unbiased Estimator*") and the *Foerstner's estimator* provide the same results in an iterative process.

If the a priori weights are correct, no bias exist in the model as was proven by *C. G. Persson* (1980). It is also well known that MINQUE produces the same result as

Best Quadratic Unbiased Estimator (BQUE)

for instance shown by *L. Sjöberg* (1984i). Based on results by *F. Pukelsheim* (1981) and *L. R. la Motte* (1973), *Best Quadratic Unbiased Non-negative Estimator (BQUNE)*

were defined by *L. Sjöberg* (1984i) for the linear *Gauss–Markov model* and generalized to the *Gauss–Helmert model*. A special contribution of *L. Sjöberg* (1984i) is appreciated which is taken care of the problem when *negative* variance components appear. *However*, the two constraints of *unbiasedness* and *non-negative* of variance components *cannot necessarily be satisfied* simultaneously, an effect which led *L. Sjöberg* (1984) to consider.

Best Quadratic Minimum Bias Non-negative Estimator (BQMBNE)

presented in *M. Eshagh and L. Sjöberg* (2008) partially. They focussed mainly on a *Modified Best-Quadratic unbiased Non-negative Estimator.*

Example 1 (one variance components)

Assume a linear model which will be characterized by first and second moments on the $n \times 1$ vector of observations

$$E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{x_1} \qquad D\{\mathbf{y}\} = \mathbf{I}\sigma^2 = \mathbf{I}x_2,$$

where $\mathbf{A}$ is the $n \times m$ first order design matrix, $\mathbf{I}$ the $n \times n$ second order design matrix (here: a unit matrix). $\boldsymbol{\mu} = \mathbf{x_1}$ is the $n \times 1$ vector of unknown parameters, $\sigma^2 = x_2$ the 1×1 vector (scalar) of unknown variance components.

We intend to estimate the unknown vector $\boldsymbol{\mu}$ or $\mathbf{x_1}$ linearly from the vector $\mathbf{y}$ of observations, the unknown σ^2 or x_2 quadratically from the observation vector $\mathbf{y}$.

1st postulate

$$\tilde{\mu_i} = \sum_{p=1}^{n} \mathbf{L}_{ip}\mathbf{y}_p \quad or \quad \tilde{\boldsymbol{\mu}} = \mathbf{L}\mathbf{y} \tag{4.170}$$

$$\tilde{\sigma}^2 = \sum_{p,q=1}^{n} \mathbf{M}_{pq}\mathbf{y}_p\mathbf{y}_q \quad or \quad \tilde{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} \tag{4.171}$$

In the classical procedure we at first estimate $\mathbf{x_1}$: (i) Estimation of $\mathbf{x_1}$

2nd postulate

We postulate unbiased estimation of $\mathbf{x_1}$:

$$E\{\tilde{\mathbf{x}}_1\} = \mathbf{x_1} \qquad \forall \mathbf{x_1}$$

Corollary 1.

$$\left.\begin{array}{r} E\{\mathbf{y}\} = \mathbf{A}\mathbf{x_1} \\ E\{\mathbf{x_1}\} = \mathbf{x_1} \\ \tilde{\mathbf{x_1}} = \mathbf{L}\mathbf{y} \end{array}\right] \Rightarrow \mathbf{L}\mathbf{A} - \mathbf{I} = 0$$

PROOF:

$$\boxed{\begin{array}{c} E\{\tilde{\mathbf{x}}_1\} = E\{\mathbf{Ly}\} = \mathbf{L}E\{\mathbf{y}\} = \mathbf{LAx_1} = \mathbf{x_1} \\ \forall \mathbf{x_1} \Longleftrightarrow (\mathbf{LA} - \mathbf{I})\mathbf{x_1} = 0 \qquad \forall \mathbf{x_1} \Longleftrightarrow \mathbf{LA} - \mathbf{I} = 0 \end{array}}$$

Corollary 2.

$$\mathbf{LA} - \mathbf{I} = 0 \quad \text{has a solution} \quad \Longleftrightarrow rk\mathbf{A} = m$$

$\mathbf{x_1}$ is unbiased estimable if and only if the rank $rk\mathbf{A} = r$ of the first order design matrix is identical with the number m of unknown parameters: $r = m$.

3rd postulate

$$\begin{aligned} \text{var}\, \tilde{\mathbf{x}}_1 &= \min \qquad \text{or} \\ E\{[\tilde{\mathbf{x}}_1 - E\{\tilde{\mathbf{x}}_1\}]'[\tilde{\mathbf{x}}_1 - E\{\tilde{\mathbf{x}}_1\}]\} &= \min_{\mathbf{L}} \end{aligned}$$

Corollary 3.

$$\text{var}\, \tilde{\mathbf{x}}_1 = \text{tr}\sigma^2\mathbf{LL}'$$

Proof:

$$\begin{aligned} \text{var}\, \tilde{\mathbf{x}}_1 &= E\{[\mathbf{y} - E\{\mathbf{y}\}]'\mathbf{L}'\mathbf{L}[\mathbf{y} - E\{\mathbf{y}\}]\} \\ &= E\{\boldsymbol{\varepsilon}'\mathbf{L}'\mathbf{L}\boldsymbol{\varepsilon}\} = \text{tr}\mathbf{L}E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\}\mathbf{L}' = \text{tr}\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' = \text{tr}\sigma^2\mathbf{LL}' \end{aligned}$$

In order to assure BLUE (best linear unbiased estimations) of the vector $\mathbf{x_1}$ of unknowns we minimize

$$\text{tr}\sigma^2\mathbf{LL}' + 2\text{tr}(\mathbf{LA} - \mathbf{I})\Lambda = \min_{\mathbf{L}, \mathbf{A}}$$

where the postulate of $\mathbf{x_1}$ – unbiasedness is added by the matrix Λ of the *Lagrange multipliers*. As a result we find

$$\begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{L}}' \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{I} \end{bmatrix}$$

$$\det \begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & 0 \end{bmatrix} = \det(\sigma^2\mathbf{I})\det(-\mathbf{A}\sigma^{-2}\mathbf{A}) \neq 0, \tag{4.172}$$

$$\text{if} \quad rk\mathbf{A} = m$$

IPM-method (‘Pandora Box’ of *C.R. Rao*, “inverse partitioned matrix”) :

$$\begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & 0 \end{bmatrix}\begin{bmatrix} \sigma^{-2}\mathbf{D} & \mathbf{C} \\ \mathbf{C}' & \mathbf{E} \end{bmatrix} = \begin{bmatrix} \sigma^{-2}\mathbf{D} & \mathbf{C} \\ \mathbf{C}' & \mathbf{E} \end{bmatrix}\begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{A} \\ \mathbf{A}' & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & \mathbf{I}_m \end{bmatrix} \tag{4.173}$$

$$\left.\begin{array}{l} \mathbf{ID} + \mathbf{AC}' = \mathbf{DI} + \mathbf{CA}' = \mathbf{I}_n \\ \mathbf{A}'\mathbf{C} = \mathbf{C}'\mathbf{A} = \mathbf{I}_m \\ \sigma^2\mathbf{IC} + \mathbf{AE} = \sigma^{-2}\mathbf{DA} = 0 \\ \mathbf{A}'\mathbf{D}\sigma^{-2} = \mathbf{C}'\sigma^2\mathbf{I} \ \# \ \mathbf{EA}' = 0 \end{array}\right\} \Rightarrow$$

$$\begin{aligned} &\mathbf{D} = \mathbf{I} - \sigma^{-2}\mathbf{A}(\mathbf{A}'\sigma^{-2}\mathbf{A})^{-1}\mathbf{A}' = \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \\ &\mathbf{C} = \sigma^{-2}\mathbf{A}(\mathbf{A}'\sigma^{-2}\mathbf{A})^{-1} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1} \\ &\mathbf{E} = -(\mathbf{A}'\sigma^{-2}\mathbf{A})^{-1} \end{aligned} \tag{4.174}$$

Result:

$$\hat{\mathbf{L}} = \mathbf{C}' = (\mathbf{A}'\sigma^{-2}\mathbf{A})^{-1}\mathbf{A}'\sigma^{-2} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \tag{4.175}$$

Corollary 4.

$$\mathbf{x_1} - \text{BLUE}$$

$$\hat{\mathbf{x}}_1 = \hat{\mathbf{L}}\mathbf{y} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \tag{4.176}$$

$$D\{\hat{\mathbf{x}}_1\} = \sigma^2(\mathbf{A}'\mathbf{A})^{-1} = -\mathbf{E} \tag{4.177}$$

Next in the classical procedure we shall estimate $\mathbf{x_2}$: (ii) Estimation of $\mathbf{x_2}$ The residual vector $\hat{\boldsymbol{\varepsilon}}$ is defined by

$$\hat{\boldsymbol{\varepsilon}} := \mathbf{y} - \widehat{E(\mathbf{y})} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\mu}}$$

Corollary 5.

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{D}\boldsymbol{\varepsilon} \tag{4.178}$$

Proof:

$$\begin{aligned} (a)\ &\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\boldsymbol{\mu}} = \mathbf{y} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}\mathbf{y} = (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} = \mathbf{D}\mathbf{y} \\ &\mathbf{D} := (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}') \quad \text{“projection operator, idempotent matrix”} \end{aligned} \tag{4.179}$$

$$(b)\ \mathbf{Dy} = \mathbf{D}(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}) = \mathbf{D}\boldsymbol{\varepsilon} \quad \text{“invariance property } \mathbf{y} \mapsto \mathbf{y} - \mathbf{A}\boldsymbol{\mu}\text{”} \tag{4.180}$$

since

$$\begin{aligned}(\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}) \\ = (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} - (-\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A})\boldsymbol{\mu} \\ = (\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y}\end{aligned}$$

Corollary 6.

$$E\{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\} = \sigma^2(n - m) \tag{4.181}$$

Proof:

$$\begin{aligned}E\{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\} &= \mathrm{tr}E\{\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'\} = \mathrm{tr}\sigma^2\mathbf{DD}' = \sigma^2\mathrm{tr}\mathbf{D} \\ &= \sigma^2\mathrm{tr}(\mathbf{I}_n) - \sigma^2\mathrm{tr}(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}') \\ &= \sigma^2\mathrm{tr}(\mathbf{I}_n) - \sigma^2\mathrm{tr}((\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}) \\ &= \sigma^2\mathrm{tr}(\mathbf{I}_n) - \sigma^2\mathrm{tr}(\mathbf{I}_m) = \sigma^2(n - m)\end{aligned}$$

2nd postulate

We postulate unbiased estimation of x_2:

$$E\{\tilde{x}_2\} = x_2 \qquad \forall x_2$$

Corollary 7.

If

$$\sigma^2 = (n - m)^{-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \tag{4.182}$$

then

$$E\{\hat{\sigma}^2\} = \sigma^2 \quad \text{or} \quad E\{\hat{x}_2\} = x_2 \tag{4.183}$$

It can also be shown that $\hat{x}_2 = \hat{\sigma}^2$ is BIQUE (best invariant quadratic unbiased estimation). The construction of an unbiased estimation of $\bar{\sigma}^2$ described above is the guiding principle for the estimation of variance components according to F.R. Helmert.

Example 2 (variance component estimation of Helmert type):

Assume a linear model which will be characterized by first and second moments of the $n \times 1$ vector of observations

$$E\{\mathbf{y}\} = \sum_{i=1}^{m} \mathbf{a}_i \boldsymbol{\mu}_i \tag{4.184}$$

$$D\{\mathbf{y}\} = \sum_{j=1}^{l} \mathbf{C}_{jj}\sigma_j^2 + \sum_{j=1}^{l-1}\sum_{k=2}^{l} \mathbf{C}_{jk}\sigma_{jk} = \tag{4.185}$$

$$= \begin{bmatrix} \sigma_1^2\ \mathbf{Q}_{11} & \sigma_{12}\ \mathbf{Q}_{12} & \dots & \sigma_{1l}\ \mathbf{Q}_{1l} \\ \sigma_{12}^2\ \mathbf{Q}_{12} & \sigma_2\ \mathbf{Q}_{22} & \dots & \sigma_{2l}\ \mathbf{Q}_{2l} \\ \cdot & \cdot & \cdot & \\ \sigma_{1l}^2\ \mathbf{Q}'_{11} & \sigma_{2l}\ \mathbf{Q}'_{2l} & \dots & \sigma_l^2\ \mathbf{Q}_{ll} \end{bmatrix} \tag{4.186}$$

where the dispersion or variance-covariance matrix $\mathbf{D}\{\mathbf{y}\}$ is assumed to be *positive definite*.

$$\mathbf{C}_{jj} = \begin{array}{cc} \begin{bmatrix} 0 & \dots & 0 \\ \dots & \mathbf{Q}_{jj} & \dots \\ 0 & \dots & 0 \end{bmatrix} & \text{row j} \\ \text{column j} & \end{array} \tag{4.187}$$

$$j = 1, \dots, l$$

$$\mathbf{C}_{jk} = \begin{array}{cc} \text{column k} & \\ \begin{bmatrix} 0 & & & 0 \\ & 0 & \mathbf{Q}_{jk} & \\ \dots & & & \dots \\ & \mathbf{Q}'_{jk} & 0 & \\ 0 & & & 0 \end{bmatrix} & \begin{array}{l} \\ \text{row j} \\ \\ \text{row k} \\ \\ \end{array} \\ \text{column j} & \end{array} \tag{4.188}$$

$\mathbf{A} := [\mathbf{a}_1, \dots, \mathbf{a}_m]$ is the n $\times$ m *first order design* matrix, $\mathbf{C}_{jj}$, $\mathbf{C}_{jk}$ are n x n *second order design* matrix blocks. m first moments (regression parameter), l second moments σ_j^2 of type *variance* and $l(l-1)/2$ second moments σ_{jk}^2 of type *covariance* are unknown. The above linear dispersion structure may be derived from the *linear structure of the residual vector*

$$\left.\begin{array}{l} \boldsymbol{\varepsilon} := \mathbf{y} - E\{\mathbf{y}\} = \mathbf{y} - \mathbf{A}\mu = \sum_{j=1}^{l} \mathbf{A}_j \boldsymbol{\varepsilon}_j \\ E\{\boldsymbol{\varepsilon}_j\} = 0, E\{\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}'_k\} = \boldsymbol{\delta}_{jk} I \end{array}\right] \Rightarrow \tag{4.189}$$

$$\mathbf{D}\{\mathbf{y}\} = \sum_{j=1}^{l} \mathbf{A}_j \mathbf{A}'_j \sigma_j^2 + \sum_{\substack{j,k \\ j<k}}^{l} (\mathbf{A}_j \mathbf{A}'_k + \mathbf{A}_k \mathbf{A}'_j \sigma_{jk}^2 \tag{4.190}$$

The variance-covariance model of *Helmert type* may be referred to as a linear dispersion structure of heterogeneous, but independent observations with respect to each other.

Let us rewrite the first two moments in vector form.

$$E\{\mathbf{y}\} = \sum_{i=1}^{m} \mathbf{a}_i \mu_i \quad \text{or} \quad E\{\mathbf{y}\} = \mathbf{A}\mu = \mathbf{A}\mathbf{x}_1 \tag{4.191}$$

$$D\{\mathbf{y}\} = \sum_{j=1}^{l(l+1)/2} \mathbf{C}_j \sigma_j \quad \text{or} \quad d\{\mathbf{y}\} := \mathrm{vec} D\{\mathbf{y}\} = \mathbf{B}\sigma = \mathbf{B}\mathbf{x}_2 \tag{4.192}$$

$$\mathbf{x}_1 := [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m]' = \boldsymbol{\mu}, o(\boldsymbol{\mu}) = m \times 1$$

$$\mathbf{x}_2 := [\sigma_i^2, \sigma_{12}, \sigma_2^2, \sigma_{13}, \sigma_{23}, \sigma_3^2, \ldots, \sigma_{l-1,l}, \sigma_l^2]' =: \sigma$$

$$o(\boldsymbol{\mu}) = l(l+1)/2$$

$\boxed{\text{The first step}}$: estimation of σ_k (reverse of the classical procedure)

$$\tilde{\sigma_k} = \sum_{p,q=1}^{n} \mathbf{M}_{kpq} \mathbf{y}_p \mathbf{y}_q \tag{4.193}$$

(0)
Choose a priori a vector σ_o of variance-covariance components such that

$$\mathbf{D}_o\{\mathbf{y}\} = \boldsymbol{\Sigma}(\sigma_o) = \boldsymbol{\Sigma}_o$$

is positive definite. Construct the residual vector $\hat{\boldsymbol{\varepsilon}}$ which belongs to the $\boldsymbol{\Sigma}_o$ – *least square solution*

(1)

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}} &:= \mathbf{y} - E\hat{\{\mathbf{y}\}} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\mu}} = (\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_o^{-1})\mathbf{y} \\ &= \mathbf{D}(\sigma_o)\mathbf{y} = \mathbf{D}(\sigma_o)(\mathbf{y} - \mathbf{A}\mu) = \mathbf{D}(\sigma_o)\boldsymbol{\varepsilon}\end{aligned} \tag{4.194}$$

Thus the estimator $\hat{\boldsymbol{\varepsilon}}$ is *invariant with respect to* $\mathbf{y} \mapsto \mathbf{y} - \mathbf{A}\boldsymbol{\mu}$

(2)

$$\hat{\boldsymbol{\varepsilon}}'\boldsymbol{\Sigma}_o^{-1}\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'\mathbf{D}'(\sigma_o)\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\sigma_o)\boldsymbol{\varepsilon} = \mathrm{tr}\mathbf{D}'(\sigma_o)\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\sigma_o)\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \tag{4.195}$$

(3)

$$E\{\hat{\boldsymbol{\varepsilon}}'\boldsymbol{\Sigma}_o^{-1}\hat{\boldsymbol{\varepsilon}}\} = \mathrm{tr}\mathbf{D}'(\sigma_o)\boldsymbol{\Sigma}_o^{-1}\mathbf{D}(\sigma_o)\boldsymbol{\Sigma} \tag{4.196}$$

Corollary. (multinomial inverse):

$$\boldsymbol{\Sigma} = \sum_{k=1}^{l(l+1)/2} \mathbf{C}_k \sigma_k \Rightarrow \boldsymbol{\Sigma}^{-1} = \sum_{k=1}^{l(l+1)/2} \mathbf{E}_k \tag{4.197}$$

e.g.

$$\begin{aligned}\boldsymbol{\Sigma} &= \sigma_1^2\begin{bmatrix}\mathbf{Q}_{11} & 0\\ 0 & 0\end{bmatrix} + \sigma_{12}\begin{bmatrix}0 & \mathbf{Q}_{12}\\ \mathbf{Q}'_{12} & 0\end{bmatrix} + \sigma_2^2\begin{bmatrix}0 & 0\\ 0 & \mathbf{Q}_{22}\end{bmatrix} = \\ &= \sigma_1^2\mathbf{C}_{11} + \sigma_{12}\mathbf{C}_{12} + \sigma_2^2\mathbf{C}_{22}\end{aligned} \tag{4.198}$$

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} &= \sigma_1^{-2}\begin{bmatrix}\mathbf{P}_{11} & 0\\ 0 & 0\end{bmatrix} + \sigma_{12}^{-1}\begin{bmatrix}0 & \mathbf{P}_{12}\\ \mathbf{P}'_{12} & 0\end{bmatrix} + \sigma_2^{-2}\begin{bmatrix}0 & 0\\ 0 & \mathbf{P}_{22}\end{bmatrix} = \\ &= \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3\end{aligned} \tag{4.199}$$

$$\mathbf{P}_{11} = \mathbf{Q}_{11}^{-1} + \mathbf{k}\mathbf{K}\mathbf{P}_{22}\mathbf{K}' \tag{4.200}$$

$$\mathbf{P}_{12} = \mathbf{P}'_{12} = -\mathbf{k}\mathbf{K}\mathbf{P}_{22} \tag{4.201}$$

$$\mathbf{P}_{22} = (\mathbf{Q}_{22} - \mathbf{k}\mathbf{K}'\mathbf{Q}_{11}\mathbf{K})^{-1} \tag{4.202}$$

$$k := \frac{\sigma_{12}^2}{\sigma_1^2\sigma_2^2}, \qquad \mathbf{K} := \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12} \tag{4.203}$$

Once we apply the above corollary for the multinomial inverse we arrive at

$$E\{\hat{\boldsymbol{\varepsilon}}'\mathbf{E}_i(\sigma_o)\hat{\boldsymbol{\varepsilon}}\} = \sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{D}'(\sigma_o)\mathbf{E}_i(\sigma_o)\mathbf{D}(\sigma_o)\mathbf{C}_j\sigma_j \tag{4.204}$$

The *Helmert equation* for variance-covariance component estimation is now gained once we write for all $i = 1, \ldots, l(l+1)/2$

$$\boxed{\hat{\boldsymbol{\varepsilon}}'\mathbf{E}_i(\sigma_o)\hat{\boldsymbol{\varepsilon}} = \sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{D}'(\sigma_o)\mathbf{E}_i(\sigma_o)\mathbf{D}(\sigma_o)\mathbf{C}_j\hat{\sigma_j}} \tag{4.205}$$

or in a more suitable form

$$\boxed{\mathbf{q} = \mathbf{H}\hat{\sigma}} \tag{4.206}$$

$$\mathbf{q} \sim \mathbf{q}_i := \hat{\boldsymbol{\varepsilon}}'\mathbf{E}_i(\sigma_o)\hat{\boldsymbol{\varepsilon}} = \mathbf{D}'(\sigma_o)\mathbf{E}_i(\sigma_o)\mathbf{D}(\sigma_o)\mathbf{y} \tag{4.207}$$

$$\mathbf{H} \sim \mathbf{H}_{ij} := \mathrm{tr}\mathbf{D}'(\sigma_o)\mathbf{E}_i(\sigma_o)\mathbf{D}(\sigma_o)\mathbf{C}_j \tag{4.208}$$

$$\hat{\sigma} = \mathbf{H}^{-}\mathbf{q} \tag{4.209}$$

is the g-inverse solution of the *Helmert equation*. The inversion of the *Helmert equation* is supposed to give the factors which latter on guarantee an unbiased estimation of variance-covariance estimation. Two cases will be reported.

Case 1:

$$\det \mathbf{H} \neq 0 \Rightarrow \operatorname{rk}\mathbf{H} = l(l+1)/2 \Rightarrow$$

$$\begin{aligned} \hat{\sigma} \quad \text{is} \quad & \boldsymbol{\Sigma}_o - \text{HIQUE}(\boldsymbol{\Sigma}_o - \text{Helmert type IQUE}) \\ & \neq \boldsymbol{\Sigma}_o - \text{HIQUE} \\ & \neq \boldsymbol{\Sigma}_o - \text{MINIQUE}(\boldsymbol{\Sigma}_o - \text{minimum IQUE}) \end{aligned}$$

Case 2:

$$\det \mathbf{H} = 0 \Rightarrow \operatorname{rk}\mathbf{H} < l(l+1)/2 \Rightarrow$$

$$\hat{\sigma} \quad \text{is} \quad \boldsymbol{\Sigma}_o - \text{HIQUE}(\boldsymbol{\Sigma}_o - \text{Helmert type invariant quadratic estimation})$$

Note that in case 2 there is no word about unbiasedness. *Still there remains the problem of positive variance components* $\widehat{\sigma_1^2}, \widehat{\sigma_2^2}, \ldots, \widehat{\sigma_l^2}$.

Actually the estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ may differ significantly from the assumed dispersion matrix $\boldsymbol{\Sigma}_o$. But the special case $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_o$ is called of *reproducing type*. We also say a priori and a posteriori variance-covariance components coincide.

The second step: estimation of $\boldsymbol{\mu}_i$ (reverse of the classical procedure)

$$\tilde{\boldsymbol{\mu}}_i = \sum_{p=1}^{m} \mathbf{L}_{ip}\mathbf{y}_p \tag{4.210}$$

If $\tilde{\boldsymbol{\mu}}$ solves the consistent system of equations

$$\mathbf{A}'\boldsymbol{\Sigma}^{-1}(\hat{\sigma})\mathbf{A}\hat{\boldsymbol{\mu}} = \mathbf{A}'\boldsymbol{\Sigma}^{-1}(\hat{\sigma})\mathbf{y} \tag{4.211}$$

then $\hat{\boldsymbol{\mu}}$ is called $\hat{\sigma}$ – least squares solution where

$$\mathbf{L}(\hat{\sigma}) = (\mathbf{A}'\boldsymbol{\Sigma}^{-1}(\hat{\sigma})\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}^{-1}(\hat{\sigma}) \tag{4.212}$$

Many problems in regression parameter estimation, e.g. deformation analysis, are characterized by a more general linear dispersion structure that the Helmert one. A lot of study has been done for the *linear model with linear dispersion structure*, namely:

$$\boxed{\begin{aligned} E\{\mathbf{y}\} &= \mathbf{A}\boldsymbol{\mu} = \mathbf{A}\mathbf{x}_1 \\ \operatorname{vec}\ D\{\mathbf{y}\} &= \mathbf{B}\sigma = \mathbf{B}\mathbf{x}_2 \end{aligned}}$$

Solution for the unknown variance-covariance components are called *admissible*, if the dispersion matrix $D\{\mathbf{y}\}$ is nonnegative definite and the variance components $\widehat{\sigma_j^2}$ are nonnegative

The procedure of the *simultaneous estimation* of the unknowns $\mathbf{x_1}$ and $\mathbf{x_2}$ of first and second moment type is the proper estimation concept pioneered by J. Kleffe (1978), S. Gnot et al. (1977).

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$$

$$\tilde{\mathbf{x}}_1 = \mathbf{K}_1 + \mathbf{L}_1\mathbf{y} + \mathbf{M}_1(\mathbf{y} \otimes \mathbf{y})$$

$$\tilde{\mathbf{x}}_2 = \mathbf{K}_2 + \mathbf{L}_2\mathbf{y} + \mathbf{M}_2(\mathbf{y} \otimes \mathbf{y})$$

$$\begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{L}_1 & \mathbf{M}_1 \\ \mathbf{L}_2 & \mathbf{M}_2 \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{y} \otimes \mathbf{y} \end{bmatrix}$$

"$\otimes$" indicates the "*Kronecker-Zehfuß product*".

1st postulate

"inhomogeneous, multilinear (bilinear) estimation"

$$\tilde{\mathbf{x}} = \mathbf{X}\mathbf{Y}$$

$$\mathbf{X} := \begin{bmatrix} \mathbf{K}_1 & \mathbf{L}_1 & \mathbf{M}_1 \\ \mathbf{K}_2 & \mathbf{L}_2 & \mathbf{M}_2 \end{bmatrix}, \mathbf{Y} := \begin{bmatrix} 1 \\ \mathbf{y} \\ \mathbf{y} \otimes \mathbf{y} \end{bmatrix}$$

2nd postulate

"invariance"

$$\mathbf{y} \mapsto \mathbf{y} - \mathbf{A}\boldsymbol{\mu} : \hat{\sigma} \quad \text{invariant}$$

3rd postulate

"unbiasedness" or "minimum bias"
Let us define the bias vector **b** and the bias matrix **S** by

$$\mathbf{b} \ : \ E\{\tilde{\mathbf{x}}\} - \mathbf{x} = \mathbf{S}\mathbf{x}, \quad \mathbf{S} := \mathbf{X} \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} - \mathbf{I}$$

$$\|\mathbf{S}\| = \sqrt{\mathrm{tr}\mathbf{S}'\mathbf{S}} = \min\|\hat{\mathbf{S}}\| = \min\{\|\mathbf{S}\| \ \ |\mathbf{S} = \mathbf{X} \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} - \mathbf{I}\}$$

The resulting condition equations have to be added via Lagrange multipliers to the postulate $\|\mathbf{D}\{\tilde{\mathbf{x}}\}\| = \min$ of best estimation. A special case will be noted:

$$\mathbf{K}_1 = 0, \quad \mathbf{M}_1 = 0 : \mathbf{b}_1 = (\mathbf{L}\mathbf{A} - \mathbf{I})\boldsymbol{\mu}$$

$$\mathbf{K}_2 = 0, \quad \mathbf{M}_2 = 0 : \mathbf{b}_2 = \mathbf{M}_2(\mathbf{A}\boldsymbol{\mu} \otimes \mathbf{A}\boldsymbol{\mu}) + \mathbf{M}_2 \ \ \mathrm{vec} \ \ \boldsymbol{\Sigma} - \sigma$$

4th postulate

Note that the dispersion matrix $D\{\tilde{\mathbf{x}}\}$ contains beside second moments also third and fourth moments of the observation vector **y**. In order to restrict the multilinear estimation process to *second order statistics*, *normally (or quasinormally) distributed* *observations* are introduced which guarantee the reduction, for instance,

$$\begin{aligned}\sigma_{ijk} &= E\{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_k\} = 0 \\ \sigma_{ijkl} &= E\{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_l\} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}\end{aligned}$$

or

$$\begin{aligned}\sigma_{ijk} &= = 0 \\ \sigma_{ijkl} &= \mathbf{f}_{ijklpqrs}\sigma_{pq}\sigma rs\end{aligned}$$

5th postulate

<u>"best estimation"</u>

$$\begin{aligned}D\{\tilde{\mathbf{x}}\} &:= E\{[\tilde{\mathbf{x}} - E\{\tilde{\mathbf{x}}\}][\tilde{\mathbf{x}} - E\{\tilde{\mathbf{x}}\}]'\} \\ \|D\{\tilde{\mathbf{x}}\}\| &:= \sqrt{\mathrm{tr} D\{\tilde{\mathbf{x}}\}} = \underset{\mathbf{X}}{\min}\end{aligned}$$

Finally within the context of the five postulates important results will be mentioned.

(1) For the linear model $E\{\mathbf{y}\} = \mathbf{A}\mu$, $\mathrm{vec}D\{\mathbf{y}\} = \mathbf{B}\sigma$, $D\{\mathbf{y}\}$ positive definite, there always exist linear *unbiased estimations* $\hat{\mathbf{z}} = \mathbf{Z}\mathbf{Y}$ of block-diagonal structure

$$\begin{aligned}&\mathbf{Z} := \begin{bmatrix} \mathbf{L} & 0 \\ 0 & \mathbf{M} \end{bmatrix}, \mathbf{Y} := \begin{bmatrix} \mathbf{y} \\ \mathbf{y} \otimes \mathbf{y} \end{bmatrix}, \mathbf{o}(\mathbf{Y}) = n(n+1) \times 1 \\ &\mathbf{z} := \mathrm{vec}[\mathbf{F}\boldsymbol{\mu}, \mathbf{G}\sigma]\end{aligned}$$

of certain optimality properties.

(2) For the linear model $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\mu}$, $\mathrm{vec}D\{\mathbf{y}\} = \mathbf{B}\sigma$, $D\{\mathbf{y}\}$ positive definite, there exist *invariant quadratic estimations* of arbitrary vectors $G\sigma$ which are independent of the special estimations of vectors $\mathbf{A}\boldsymbol{\mu}$ by $\mathbf{L}_y = \mathbf{A}\mathbf{A}^{-}\mathbf{y}$. This result motivates *the two steps estimation process: At first* the vector $\mathbf{G}\sigma$ is invariant estimated, *secondly* the vector $\mathbf{F}\boldsymbol{\mu}$ is estimated dependent of $\mathbf{G}\hat{\sigma}$.

(3) "E-D-correspondence"
Once we construct invariant quadratic estimation of vectors $\mathbf{G}\sigma$, namely linear in $\mathbf{Py} \otimes \mathbf{Py}$ where $\mathbf{P} := \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$, then the whole theory of regression parameter estimation with respect to a *singular Gauß–Markov model* can be transferred to the estimation of dispersion components, if the observation vector y is at least quasi-normally distributed. This idea has been pioneered by *S.K. Mitra* (1971), *F.Pukelsheim* (1976,1977,1979,1981) and *J. Seely* (1971,1972,1977).

(4) We have briefly mentioned that the dispersion component estimation based on the a priori dispersion matrix $\boldsymbol{\Sigma}_o$ leads to the concept of reproducing estimations in the sense of $\boldsymbol{\Sigma}_o = \hat{\boldsymbol{\Sigma}}$. Actually an initial estimate $\boldsymbol{\Sigma}_o$ does not lead to the value $\hat{\boldsymbol{\Sigma}}$. We have to iterate, e.g. the *Helmert equation*, up to the reproducing by iteration. *B. Schaffrin* (1983) has described quite a number of iteration procedures. For instance, he proved that the diagonalization technique of the *Helmert equation* pioneered by *W. Förstner* (1979) leads to repro-BIQUE.

(5) With respect to the *Helmert equation* we have mentioned that unbiased dispersion component estimations might not exist. Then we would prefer *minimum bias estimation*. In addition, variance component might be *negative*, a well-known result from practical work. In this context we mention the *nonnegative quadratic minimum bias estimator*, pioneered by *J. Hartung* (1979,1980). *B. Schaffrin* (1981ii) has efficiently used the linear complementarity problem (LCP algorithm), e.g. to solve the *Helmert equation* under the constraints of *nonnegative variance components*.

Let us discuss the intuitive *Helmert estimation* of variance-covariance components with respect to the linear model with linear dispersion structure. We take advantage of the quadratic set-up of variance-covariance components estimation of invariant type, namely

$$\hat{\sigma} = \mathbf{X}\hat{\varepsilon} \otimes \hat{\boldsymbol{\varepsilon}} = \mathbf{X}\text{vec} \;\; \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'$$
$$\hat{\boldsymbol{\varepsilon}} = \mathbf{D}_o\boldsymbol{\varepsilon} = \mathbf{D}_o\mathbf{y}$$

where the projection matrix $\mathbf{D}_o := \mathbf{I} - \mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}$ refers to $\boldsymbol{\Sigma}_o - \boldsymbol{\mu}$ by 4.194. Combine (4.213) in order to find

$$\hat{\sigma} = \mathbf{X}\text{vec} \;\; \mathbf{D}_o\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{D}_o' = \mathbf{X}\mathbf{D}_o \otimes \mathbf{D}_o\text{vec} \;\; \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \tag{4.213}$$

The postulate of *unbiased* variance-covariance component estimation and the linear dispersion structure lead to

$$E\{\hat{\sigma}\} = \mathbf{X}\mathbf{D}_o \otimes \mathbf{D}_o\text{vec}\boldsymbol{\Sigma} = \mathbf{X}\mathbf{D}_o \otimes \mathbf{D}_o\mathbf{B}\sigma = \sigma \quad \text{for all} \quad \sigma \tag{4.214}$$

or

$$\mathbf{X}\mathbf{D}_o \otimes \mathbf{D}_o\mathbf{B} = \mathbf{I} \tag{4.215}$$

Obviously

$$\mathbf{X} = [(\mathbf{D}_o \otimes \mathbf{D}_o)\mathbf{B}]_l^- \tag{4.216}$$

is the *left generalized inverse* which solve (4.215) such that

$$\boxed{\hat{\sigma} = [(\mathbf{D}_o \otimes \mathbf{D}_o)\mathbf{B}]_l^- \mathbf{D}_o \otimes \mathbf{D}_o \;\; \text{vec} \;\; \mathbf{y}\mathbf{y}'} \tag{4.217}$$

We mention two general representations of the left generalized inverse which are equivalent, namely

$$\mathbf{X} = \{\mathbf{B}'(\mathbf{D}_o \otimes \mathbf{D}_o)'\mathbf{R}(\mathbf{D}_o \otimes \mathbf{D}_o)\mathbf{B}\}^{-1}\mathbf{B}'(\mathbf{D}_o \otimes \mathbf{D}_o)'\mathbf{R} \tag{4.218}$$

and

$$\mathbf{X} = \{\mathbf{S}'(\mathbf{D}_o \otimes \mathbf{D}_o)\mathbf{B}\}^{-1}\mathbf{S}' \tag{4.219}$$

$$\mathbf{R} = \mathbf{S}\{\mathbf{B}'(\mathbf{D}_o \otimes \mathbf{D}_o)'\mathbf{S}\}^{-1}\{\mathbf{S}'(\mathbf{D}_o \otimes \mathbf{D}_o)\mathbf{B}\}^{-1}\mathbf{S}' \tag{4.220}$$

The arbitrary matrices **R** and **S** have to be chosen in such a way that the regular inverse of the matrices { } exist.The equivalence of the various representations can be shown once we use $\mathbf{R} = \mathbf{X}'\mathbf{X}$, for instance. The Helmert estimator is obtained once we choose

$$\begin{aligned}\mathbf{S}_H = \{[\boldsymbol{\Sigma}_o^{-1} - \boldsymbol{\Sigma}_o^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}]\otimes \\ \boldsymbol{\Sigma}_o^{-1} - \boldsymbol{\Sigma}_o^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}\mathbf{A})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_o^{-1}\}\{\boldsymbol{\Sigma}_o \otimes \boldsymbol{\Sigma}_o\}\mathbf{B}_o\end{aligned} \tag{4.221}$$

where we have introduced the vector of a multinomial inverse of $\boldsymbol{\Sigma}_o$ according to (4.197) namely

$$\text{vec}\ \ \boldsymbol{\Sigma}_o^{-1} = \mathbf{E}_o\mathbf{s} \tag{4.222}$$

and **s** the "summation vector".

Finally we present two characteristic geodetic numerical example:

Example 3. (two variance components, hybrid direction-distance network): An example for a combined distance-direction two-dimensional network has been taken from W. Welsch. *Figure* 4.2 is the network graph, *Table* 4.1 the set of approximate (x,y)-coordinates, *Table* 4.2, the set of 40 distance and 99 direction observations. There are 12 points, 21 coordinate correction unknowns and 23

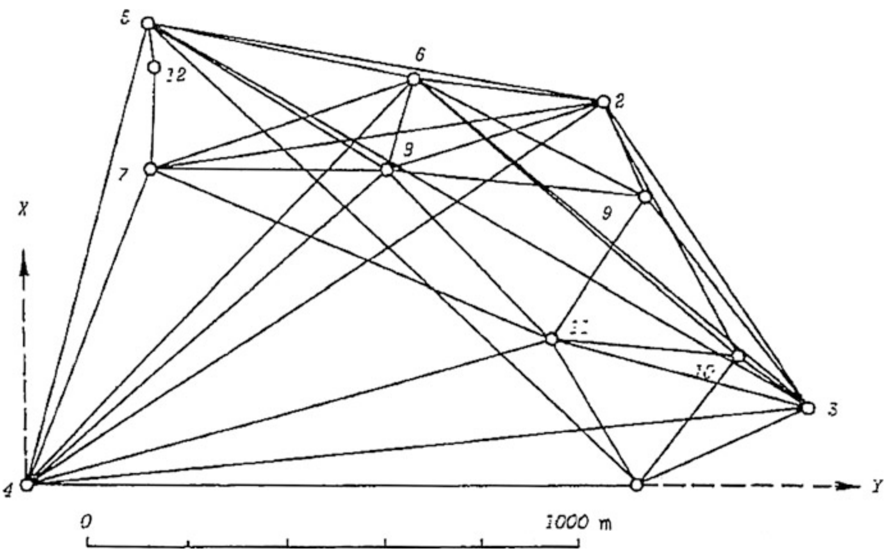


Fig. 4.2 Huaytapallana network, Peru

Table 4.1 Approximate coordinates, Huaytapallana network, Peru

	y	x
2	2 175.874 m	1 737.156 m
3	2 577.368 m	1 148.348 m
5	1 252.818 m	1 893.092 m
6	1 774.584 m	1 778.635 m
7	1 262.644 m	1 596.958 m
8	1 725.782 m	1 596.606 m
9	2 253.863 m	1 554.891 m
10	2 424.436 m	1 237.101 m
11	2 053.959 m	1 282.128 m
12	1 252.980 m	1 871.469 m
1	2 232.194 m	1 000.000 m
4	1 000.000 m	1 000.000 m

Table 4.2 Distance observations, Huaytapallana network, Peru

1	5	1325.4268 m	4	6	1098.2724 m
1	4	1232.1958 m	4	8	939.4676 m
1	3	375.7020 m	4	11	1091.0600 m
1	11	333.6888 m	5	6	534.1382 m
1	10	305.2402 m	5	8	558.1884 m
2	3	712.6460 m	5	12	21.6076 m
2	10	558.3987 m	6	7	543.2001 m
2	8	471.5552 m	6	9	528.9682 m
2	6	403.4410 m	6	8	188.4908 m
2	4	1387.8132 m	6	10	845.9200 m
2	9	198.2577 m	7	8	463.1096 m
2	7	923.9211 m	7	11	851.6534 m
2	5	936.1232 m	7	12	274.6834 m
3	11	540.2310 m	8	9	529.7660 m
3	5	1519.5508 m	8	11	454.5350 m
3	4	1584.3316 m	8	10	785.7282 m
3	6	1020.6574 m	9	3	519.5172 m
3	10	176.8214 m	9	11	338.1710 m
4	7	652.1740 m	9	10	360.6380 m
4	5	928.1742 m	10	11	373.1988 m

orientation unknowns. The coordinate system is geodetically oriented (y right, x top) $[\Delta x_1, \Delta x_4, \Delta y_4]' = 0$ has been chosen for the network datum. Units of the computations are meter for distances and rad for directions. Figure 6.183 is the plot of a posteriori variance ratio $\hat{\sigma_1^2}/\hat{\sigma_2^2}$ as a function of the a priori variance ration σ_1^2/σ_2^2 in the interval $10^2 \leq \sigma_1^2/\sigma_2^2 \leq 10^5$ with the reproducing point $\hat{\sigma_1^2}/\hat{\sigma_2^2} = \sigma_1^2/\sigma_2^2 = 2.2 \times 10^4$ or $\hat{\sigma}_1 = 0.0022\,\text{m}$ for distances and $\hat{\sigma}_2 = 9, 4^{cc}$ for directions. Example 3 has been taken from *E. Grafarend, A Kleusberg and B. Schaffrin* (1980). Note that the abscissa in Fig. 4.3 contains logarithmic scale while its ordinate linear scale which explains the dashed curve which in linear-linear scale is a straight line.

Table 4.3 Direction observations, Huaytapallana network, Peru

1	11	0	11	4	0
	10	79.24383 g		8	65.28951 g
	3	110.03086 g		10	224.34846 g
	4	335.87130 g		3	232.58056 g
				1	280.78056 g
2	3	0			
	10	8.73148 g	2	8	0
	9	12.35741 g		7	9.57346 g
	4	102.45062 g		6	25.82932 g
	8	118.82870 g		5	29.92500 g
	6	144.65833 g			
			3	10	59.31235 g
3	1	0		6	68.20741 g
	4	19.87037 g		2	87.73488 g
	11	41.76821 g			
	5	58.44753 g	4	5	0
	2	87.73951 g		8	38.63519 g
				11	65.78457 g
4	5	0		3	76.46451 g
	7	8.82377 g		1	82.43395 g
	6	32.26790 g			
	8	38.63395 g	5	3	0
	2	46.78519 g		8	3.04444 g
				1	14.45988 g
5	2	0		12	66.93951 g
	6	3.09259 g		4	84.95370 g
	8	24.99630 g			
	4	106.90741 g	6	3	0
				10	1.85370 g
6	2	0		4	107.46080 g
	9	21.24722 g		5	171.37500 g
	8	110.11574 g			
	4	143.27346 g	7	12	0
	7	171.73210 g		8	102.29383 g
	5	207.18827 g		11	126.34753 g
				4	228.62654 g
7	6	0			
	2	12.01327 g	8	10	0
	8	21.76358 g		11	18.38642 g
	4	148.09691 g		4	125.94784 g
				5	205.39877 g
8	2	0			
	9	24.28580 g	10	11	0
	4	175.47037 g		8	22.55494 g
	7	219.32438 g		6	36.53086 g
	5	254.92191 g		9	60.94383 g
	6	335.94846 g	11	4	0
				7	40.75679 g

(Continued)

Table 4.3 (Continued)

9	3	0		9	156.91512 g
	10	11.42994 g		3	232.57685 g
	11	83.05864 g			
	8	147.80833 g	1	4	0
	6	170.59784 g		5	47.06883 g
	2	217.04722 g	1	4	0
				10	143.37284 g
10	1	0			
	11	64.32253 g	12	5	0
	6	100.85370 g		7	198.20988 g
	2	127.25864 g			
	3	290.10370 g			

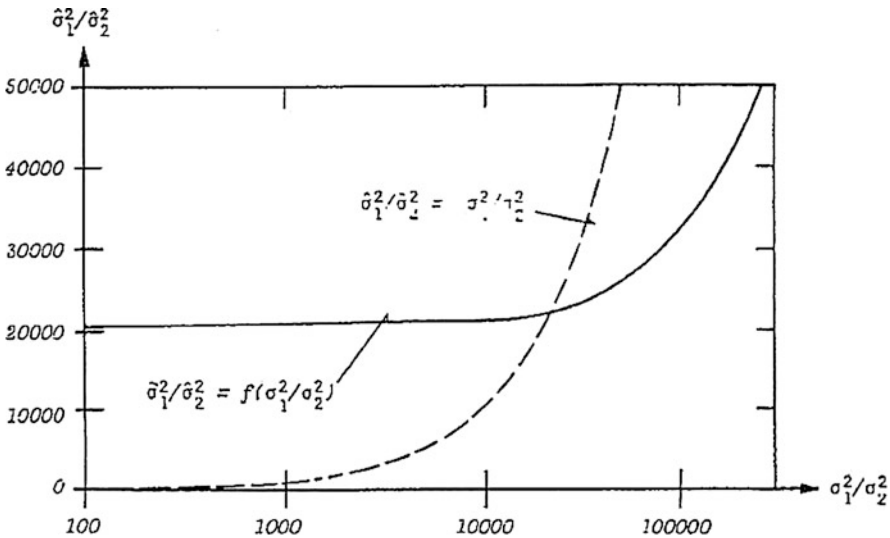


Fig. 4.3 Two variance component estimations of Helmert type: $\hat{\sigma}_1^2/\hat{\sigma}_2^2 = f(\sigma_1^2/\sigma_2^2)$, Huaytapallana network, Peru, $\hat{\sigma_1} = 0.0022$m, $\hat{\sigma_2} = 9, 4^{cc}$ at reproducing point

Example 4. (two variance components, multivariate gyrocompass observations):

An experimental data set of gyrocompass reversal point observations has been taken from *M. Junasevic* (1977) and are listed in *Table* 4.4: They include 18 reversal points with our TK3 instrument. The data analysis was based on a frequency oscillation model with linear damping and disturbances of *first order Markov type*, namely $\boldsymbol{\varepsilon}_i = \boldsymbol{\xi}_i + \rho\boldsymbol{\varepsilon}_{i-1}$. The variance components within $E(\boldsymbol{\xi}_i\boldsymbol{\xi}_j) = \sigma_1^2\boldsymbol{\delta}_{ij}$, $E(\boldsymbol{\varepsilon}_0^2) = \sigma_2^2$ have been estimated according to the *Helmert method*. Figure 4.4 is a plot of the a posteriori variance ratios as a function of a priori variance ratios $\hat{\sigma_1^2}/\hat{\sigma_2^2} = f(\sigma_1^2/\sigma_2^2)$ and of different values ρ we could find positive variances. For $\rho = 0.6$ we arrived at $\sigma_1 = 7$", $\sigma_2 = 97$" and for $\rho = 0.6$ $\sigma_1 = 9$", $\sigma_2 = 75$" at the reproducing point.

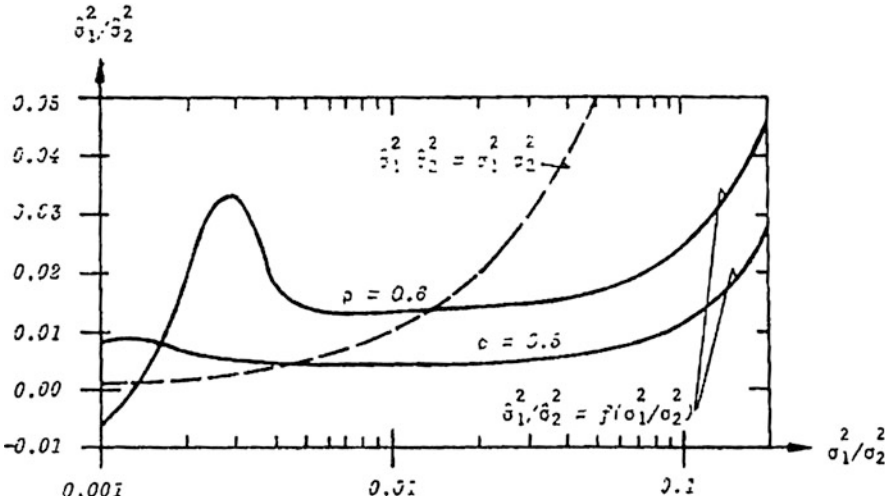


Fig. 4.4 Two variance component estimations of Helmert type: $\hat{\sigma}_1^2/\hat{\sigma}_2^2 = f(\sigma_1^2/\sigma_2^2)$, linear damping, first order Markov model $\rho = 0.5$, $\hat{\sigma}_1 = 7$", $\hat{\sigma}_2 = 97$" at reproducing point, $\rho = 0.6$, $\hat{\sigma}_1 = 9$", $\hat{\sigma}_2 = 75$" at the righthand true reproducing point

Table 4.4 Data set of 18 reversal points, gyrocompass TK3

$68.^{g}080$	$70.^{g}114$	$68.^{g}072$	$70.^{g}108$	$68.^{g}072$	$70.^{g}104$
68.080	70.094	68.085	70.092	68.090	70.083
68.110	70.071	68.113	70.070	68.118	70.064

Inputs of $\rho \leq 0.4$ or $\rho \leq 0.7$ led to instabilities. The examples has been chosen since it demonstrates that, in general, *there might exist more than one reproducing point*, e.g. at $\sigma_1^2/\sigma_2^2 = 0,0005$ and $\sigma_1^2/\sigma_2^2 = 0,015$ for $\rho = 0.6$. Only in the latter case the *single* variance components have been reproducing. Thus in using numerical algorithms to find the reproducing point by iteration, *there is the danger that alternative reproducing points are overlooked*. *Example* 4 has been taken from *E. Grafarend and A. Kleusberg* (1980). Again we note the logarithmic-linear scale of Fig. 4.4 similar to Fig. 4.3.

Example 5. (Bayes design with moment estimation)

A Bayes linear model with moment assumption is using, for instance, *prior information* of type mean parameter vector $\boldsymbol{\xi}$ and the model variance σ^2: This is taking into account by using $\boldsymbol{\xi}$ and σ^2 follow a distribution, turned the *prior distribution* which is to be special by the *experiment at* the beginning of the modelling phase. Thus the parameters $\boldsymbol{\xi}$ and σ^2 become random variables, in addition the response vector $\mathbf{y}$. The model lying distribution now determines the *joint distribution* of $\mathbf{y}$, $\boldsymbol{\xi}$ and σ^2. The expecting vector and the dispersion matrix of the response vector $\mathbf{y}$ conditionally with $\boldsymbol{\xi}$ and σ^2 given, are given by 4.223. The expectation vector and the dispersion matrix is determined by the *prior mean vector* $\boldsymbol{\xi}_o$, a *prior dispersion matrix* $\mathbf{R}_o$, non-negative definite and a *prior sample* *size* $\boldsymbol{\mu}_o \geq 0$, given by 4.224. Finally we take advantage of *prior model variance* $\sigma^2 > 0$ being the expected value of the model variance σ^2 of type 4.223. The assumption (i) 4.223 (ii) 4.224, and (iii) 4.225 are called the Bayes linear model with moment assumptions.

The assumption (ii) 4.224 are the *critical ones*: The prior estimation for $\boldsymbol{\xi}$, before any sampling evidence is available, is $\boldsymbol{\xi}_o$ and has uncertainty $(\sigma^2/\boldsymbol{\mu}_o)\mathbf{R}_o$. In case of a full rank m, the *Gauss–Markov theorem yields* the sampling estimate $\hat{\boldsymbol{\xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$, with dispersion matrix $(\sigma^2/\boldsymbol{\mu})\mathbf{M}^{-1}$, where $\mathbf{M} := \mathbf{A}'\mathbf{A}/\boldsymbol{\mu}$. The *uncertainty of the prior estimate* and the *variability of the sampling estimate* are made comparable to the experimentor through specifying $\boldsymbol{\xi}_o$, assesses the weight of the prior information *on the same observation basis* that applies to the sampling information. The *scaling issue* is of central importance because *it touches on the essence of the Bayes philosophy*, namely to *combine prior information and sampling information* being insured in comparable units. Accordingly the optimal estimator is easily understood. Otherwise, looks bad.

Bayes design with moment estimation

$$\boxed{\text{Assumption of type one}}$$

$$\mathbf{E}_p\{\mathbf{y}|\boldsymbol{\xi},\sigma^2\} = \mathbf{A}\boldsymbol{\xi}, \mathbf{D}_p\{\mathbf{y}|\boldsymbol{\xi},\sigma^2\} = \mathbf{I}_n\sigma^2 \tag{4.223}$$

$$\boxed{\text{Assumption of type two}}$$

$$\mathbf{E}_p\{\boldsymbol{\xi}|\sigma^2\} = \boldsymbol{\xi}_o, \mathbf{D}_p\{\boldsymbol{\xi}|\sigma^2\} = \frac{\sigma^2}{\boldsymbol{\mu}_o}\mathbf{R}_o \tag{4.224}$$

$$\boxed{\text{Assumption of type three}}$$

$$\mathbf{E}_p\{\sigma^2\} = \sigma_o^2 \tag{4.225}$$

$$\boxed{\text{Matrix-valued risk function (mean squared error matrix)}}$$

$$\mathbf{MSE}\{\mathbf{T}(\mathbf{y}); \mathbf{K}'\boldsymbol{\xi}\} = \mathbf{E}_p[(\mathbf{T}(\mathbf{y}) - \mathbf{K}'\boldsymbol{\xi})(\mathbf{T}(\mathbf{y}) - \mathbf{K}'\boldsymbol{\xi})'] \tag{4.226}$$

Let us assume a full column rank coefficient matrix $\mathbf{K}$ in order to *transform the parameter system* $\mathbf{K}'\boldsymbol{\xi}$ into estimated quantities $\mathbf{T}(\mathbf{y})$. A proper choice of a *matrix-valued risk function* called *mean squared error* of type 4.227 is used to minimize the *mean squared error* matrix among all *affine estimators* $\mathbf{By} + \mathbf{b}$ where the matrix $\mathbf{B} \in \mathbb{R}^{l \times n}$ and the *shift* $\mathbf{B} \in \mathbb{R}^l$. The shift b is needed since prior to sampling there is a *bias*.

$$\mathbf{E}_p\{\mathbf{K}'\boldsymbol{\xi}\} = E_p[E_p(\mathbf{K}'\boldsymbol{\xi}\|\sigma^2)] = E_p(\mathbf{K}'\boldsymbol{\xi}_o) = \mathbf{K}'\boldsymbol{\xi}_o \tag{4.227}$$

Unless $\xi_o = 0$, Any affine estimator that guarantees the minimum mean squared error matrix is called

$$\underline{\text{a Bayes estimator for } \mathbf{K}'\boldsymbol{\xi}}$$

In the prescribed *Bayes linear model* with moment assumptions, let the *prior dispersion matrix* $\mathbf{R}_o$ be *positive definite*, Then the unique Bayes estimator for $\mathbf{K}'\boldsymbol{\xi}$ is $\mathbf{K}'\boldsymbol{\xi}$ given by 4.228. The mean squared error matrix is accordingly given by 4.229. For the proof we leave it to *F. Pukelheim* (1990), pp. 270–272. Unique Bayes estimator for $\mathbf{K}'\boldsymbol{\xi}$

$$\tilde{\boldsymbol{\xi}} = (\boldsymbol{\mu}_o \mathbf{R}_o^{-1} + \mathbf{A}'\mathbf{A})^{-1}(\boldsymbol{\mu}_o \mathbf{R}_o^{-1}\boldsymbol{\xi}_o + \mathbf{A}'\mathbf{y}) \tag{4.228}$$

Minimum mean squared error matrix

$$\mathbf{MSE}_p(\mathbf{K}'\tilde{\xi}; \mathbf{K}'\boldsymbol{\xi}) = \sigma_o^2 \mathbf{K}'(\boldsymbol{\mu}_o \mathbf{R}_o^{-1} + \mathbf{A}'\mathbf{A})^{-1}\mathbf{K} \tag{4.229}$$

Chapter 5
The Third Problem of Algebraic Regression

Inconsistent system of linear observational equations with datum defect: overdetermined – undetermined system of linear equations: $\{\mathbf{Ax} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m},\ \mathbf{y} \notin \mathcal{R}(\mathbf{A}) \sim \mathrm{rk}\mathbf{A} < \min\{m, n\}\}$

The optimisation problem which appears *in general linear systems* and *weakly nonlinear systems* is a standard topic in many textbooks on optimization. Here we again define nearly nonlinear systems of equations as a problem which allows a Taylor expansion. We start in the *first section* with a front example, an inconsistent system of linear equations with datum defect. We illustrate (a) *algebraic partitioning* (b) *geometric partitioning* and (c) *set-theoretic partitioning*. This setup is written not for mathematicians, but for the analyst with remote control of the notions herein. We construct the *minimum norm-least squares solution* (MINOLESS) leads to the solution by *additive or multiplicative rank partitioning* (5.55). The *second section* is based on a review of *Minimum Norm - Least Squares Solution*, both weighted norms, based on *the four axioms of the generalized inverse*, again called *Rao's Pandora box* (C.R. Rao and S.K. Mitra (1971, pages 50–55, "three basic g-inverses, Sect. 5.3.3: minimum norm least squares solution")). A special highlight is the *eigenvalue decomposition* of $(\mathbf{G}_x, \mathbf{G}_y)$-MINOLESS and its canonical representation: "left" and "right" eigenspace analysis versus synthesis. Please, pay attention to our notes. In *Sect. 5-3* we shortly review *Total Least Squares*, namely the hybrid approximation solution of type "α-HAPS" and the celebrated *Tykhonov-Phillips regularization*, also called "Ridge Estimation" or "α-HAPS": (5.108–5.110).

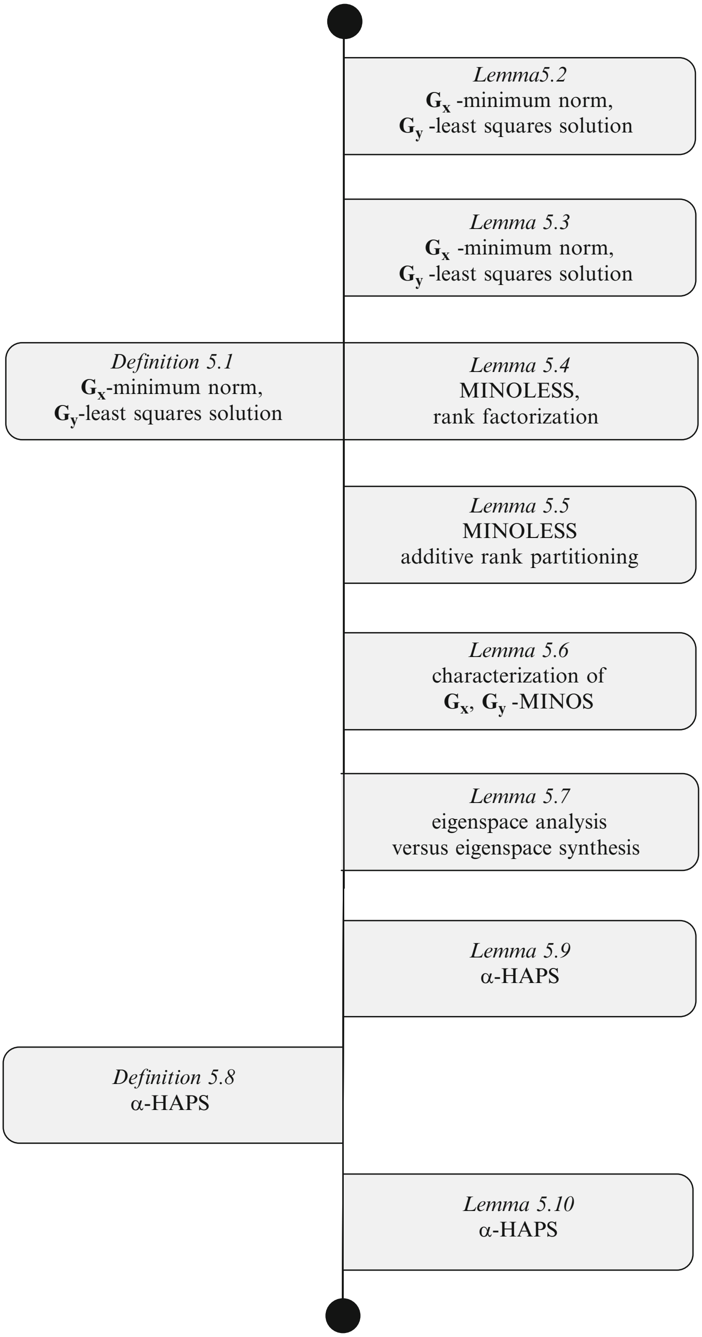
Read only *Lemma 5.6* (MINOS) and *Lemma 5.9* (HAPS).

We shall outline three aspects of the general inverse problem given in discrete form (i) set-theoretic (*fibering*), (ii) algebraic (*rank partitioning*; "IPM", the *Implicit Function Theorem*) and (iii) geometrical (*slicing*). Here we treat the third problem of algebraic regression, also called the general linear inverse problem: An inconsistent system of linear observational equations

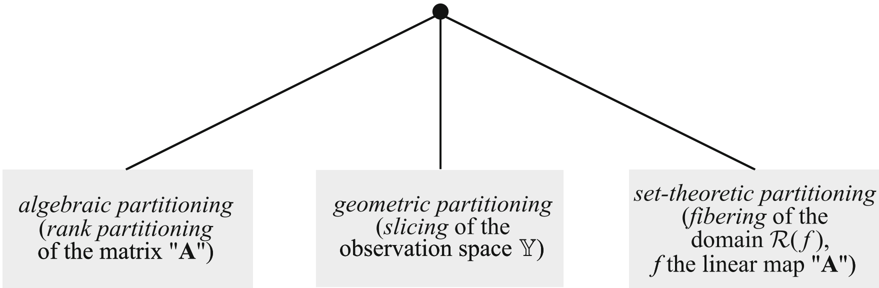
$$\{\mathbf{Ax} + \mathbf{i} = \mathbf{y}\ |\mathbf{A} \in \mathbb{R}^{n \times m}, \mathrm{rk}\,\mathbf{A} < \min\{n,\ m\}\}$$

also called "under determined – over determined system of linear equations" is solved by means of an optimization problem. The *introduction* presents us with the *front page example* of inhomogeneous equations with unknowns. In terms of *boxes and figures* we review the minimum norm, least squares solution ("MINOLESS") of such an inconsistent, rank deficient system of linear equations which is based upon the trinity.

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2_5,
© Springer-Verlag Berlin Heidelberg 2012



Lemma5.2
$\mathbf{G_x}$ -minimum norm,
$\mathbf{G_y}$ -least squares solution
Lemma 5.3
$\mathbf{G_x}$ -minimum norm,
$\mathbf{G_y}$ -least squares solution
Definition 5.1
$\mathbf{G_x}$-minimum norm,
$\mathbf{G_y}$-least squares solution
Lemma 5.4
MINOLESS,
rank factorization
Lemma 5.5
MINOLESS
additive rank partitioning
Lemma 5.6
characterization of
$\mathbf{G_x}$, $\mathbf{G_y}$ -MINOS
Lemma 5.7
eigenspace analysis
versus eigenspace synthesis
Lemma 5.9
α-HAPS
Definition 5.8
α-HAPS
Lemma 5.10
α-HAPS



5-1 Introduction

With the introductory paragraph we explain the *fundamental concepts* and *basic notions* of this section. For you, *the analyst*, who has the difficult task to deal with measurements, observational data, modeling and modeling equations we present *numerical examples* and *graphical illustrations* of all *abstract notions*. The elementary introduction is written *not for a mathematician*, but for you, *the analyst*, with limited remote control of the notions given hereafter. *May we gain your interest*?

Assume an n-dimensional *observation space*, here a *linear space* parameterized by n observations (finite, discrete) as coordinates $\mathbf{y} = [y_1, \ldots, y_n]' \in \mathbb{R}^n$ in which an m-dimensional *model manifold* is embedded (immersed). The model manifold is described as the *range* of a *linear operator* f from an m-dimensional *parameter space* $\mathbb{X}$ into the *observation space* $\mathbb{Y}$. As a mapping f is established by the mathematical equations which relate all observables to the unknown parameters. Here the *parameter space* $\mathbb{X}$, the domain of the *linear operator* f, will be also restricted to a *linear space* which is parameterized by coordinates $\mathbf{x} = [x_1, \ldots, x_m]' \in \mathbb{R}^m$. In this way the linear operator f can be understood as a *coordinate mapping* $\mathbf{A} : \mathbf{x} \mapsto \mathbf{y} = \mathbf{A}\mathbf{x}$. The linear mapping $f : \mathbb{X} \to \mathbb{Y}$ is geometrically characterized by its *range* $\mathcal{R}(f)$, namely $\mathcal{R}(A)$, defined by $\mathcal{R}(f) := \{\mathbf{y} \in \mathbb{R}^n | \mathbf{y} = \mathrm{f}(\mathbf{x})\ \text{forsome}\ \mathbf{x} \in \mathbb{X}\}$ which *in general* is a linear subspace of $\mathbb{Y}$ and its *kernel* $\mathcal{N}(f)$, namely $\mathcal{N}(\mathbf{A})$, defined by $\mathcal{N}(f) := \{\mathbf{x} \in | f(\mathbf{x}) = 0\}$. Here the range $\mathcal{R}(f)$, namely the *range space* $\mathcal{R}(\mathbf{A})$, *does not coincide* with the n-dimensional *observation space* $\mathbb{Y}$ such that $\mathbf{y} \notin \mathcal{R}(f)$, namely $\mathbf{y} \notin \mathcal{R}(\mathbf{A})$. In addition, we shall assume here that the *kernel* $\mathcal{N}(f)$, namely *null space* $\mathcal{N}(\mathbf{A})$ is not trivial: Or we may write $\mathcal{N}(f) \neq 0$.

First, Example 1.3 confronts us with an inconsistent system of linear equations with a datum defect. *Second*, such a system of equations is formulated as a special linear model in terms of *matrix algebra*. In particular we are aiming at an explanation of the terms "*inconsistent*" and "*datum defect*". The *rank* of the matrix $\mathbf{A}$ is introduced as the *index* of the *linear operator* $\mathbf{A}$. The *left complementary index* $n - \operatorname{rk} \mathbf{A}$ is responsible for surjectivity defect, which its *right complementary index* $m - \operatorname{rk} \mathbf{A}$ for the injectivity (datum defect). As a linear mapping f is neither "onto", nor "*one-to-one*" or neither *surjective*, nor *injective*. *Third*, we are going to open the *toolbox of partitioning*. By means of additive rank partitioning (horizontal and vertical *rank partitioning*) we construct the *minimum norm – least squares solution* (*MINOLESS*) of the inconsistent system of linear equations with datum defect $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, $\mathrm{rk}\mathbf{A} \leq \min\{n, m\}$. *Box 5.3* is an explicit solution of the MINOLESS of our front page example.

Fourth, we present an alternative solution of type "MINOLESS" of the front page example by *multiplicative rank partitioning*. *Fifth*, we succeed to identify the *range space* $\mathcal{R}(\mathbf{A})$ and the *null space* $\mathcal{N}(\mathbf{A})$ using the door opener "*rank partitioning*".

5-11 The Front Page Example

Example 5.1. (inconsistent system of linear equations with datum defect: $\mathbf{Ax} + \mathbf{i} = \mathbf{y},\ \mathbf{x} \in \mathbb{X} = \mathbb{R}^m,\ \mathbf{y} \in \mathbb{Y} \in \mathbb{R}^n\ \mathbf{A} \in \mathbb{R}^{n \times m},\ \ r = \mathrm{rk}\mathbf{A} \leq \min\{n, m\}$):

Firstly, the *introductory example* solves the front page inconsistent system of linear equations with datum defect,

$$\begin{array}{rcl} -x_1 + x_2 \doteq 1 & & -x_1 + x_2 + i_1 = 1 \\ -x_2 + x_3 \doteq 1 & or & -x_2 + x_3 + i_2 = 1 \\ +x_1 - x_3 \doteq -3 & & +x_1 - x_3 + i_3 = -3 \end{array}$$

obviously in general dealing with the linear space $\mathbb{X} = \mathbb{R}^m \ni \mathbf{x}$, $dim\,\mathbb{X} = m$, here $m = 3$, called the *parameter space*, and the linear space $\mathbb{Y} = \mathbb{R}^n \ni \mathbf{y}$, $dim\mathbb{Y} = n$, here $n = 3$, called the *observation space*.

5-12 The Front Page Example in Matrix Algebra

Secondly, by means of *Box 5.1* and according to *A. Cayley's doctrine* let us specify the inconsistent system of linear equations with datum defect in terms of *matrix algebra*.

Box 5.1. (Special linear model):

three observations, three unknowns, $\mathrm{rk}\,\mathbf{A} = 2$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} \Leftrightarrow$$

$$\Leftrightarrow \quad \mathbf{y} = \mathbf{Ax} + \mathbf{i} : \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} \quad \Leftrightarrow$$

$$\Leftrightarrow \ \mathbf{x}' = [x_1, \ x_2, \ x_3], \ \mathbf{y}' = [y_1, \ y_2, \ y_3] = [1, \ 1, \ -3], \ \mathbf{i}' = [i_1, \ i_2, \ i_3]$$

$$\mathbf{x} \in \mathbb{R}^{3\times 1}, \ \mathbf{y} \in \mathbb{Z}^{3\times 1} \subset \mathbb{R}^{3\times 1}$$

$$\mathbf{A} := \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \in \mathbb{Z}^{3\times 3} \subset \mathbb{R}^{3\times 3}$$

$$\boxed{r = \mathrm{rk}\mathbf{A} = 2.}$$

The matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, *here* $\mathbf{A} \in \mathbb{R}^{3\times 3}$, is an element of $\mathbb{R}^{n\times m}$ generating a linear mapping $f : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. A mapping f is called linear if $f(\lambda x_1 + x_2) = \lambda f(x_1) + f(x_2)$ holds. The *range* $\mathcal{R}(f)$, in geometry called "the *range space* $\mathcal{R}(\mathbf{A})$", and the *kernel* $\mathcal{N}(f)$, in geometry called "the *null space* $\mathcal{N}(\mathbf{A})$" characterized the linear mapping as we shall see.

? Why is the front page system of linear equations called inconsistent ?

For instance, let us solve the first two equations, namely $-x_1 + x_3 = 2\, or\, x_1 - x_3 = -2$, in order to solve for x_1 and x_3. As soon as we compare this result to the third equation we are led to the inconsistency $-2 = -3$. Obviously such a system of linear equations needs general inconsistency para-meters $(i_1, \ i_2, \ i_3)$ in order to avoid contradiction. Since the right-hand side of the equations, namely the inhomogeneity of the system of linear equations, has been *measured* as well as the linear model (the model equations) has been fixed, we have no alternative but inconsistency.

Within matrix algebra the index of the *linear operator* $\mathbf{A}$ is the *rank* $r = \mathrm{rk}\,\mathbf{A}$, here $r = 2$, which coincides *neither* with $dim\,\mathbb{X} = m$, ("*parameter space*") *nor* with $dim\,\mathbb{Y} = n$ ("*observation space*"). Indeed $r = \mathrm{rk}\,\mathbf{A} < min n, m$, here $r = \mathrm{rk}\,\mathbf{A} < min 3, 3$. In the terminology of the linear mapping f, f is neither onto ("*surjective*"), nor one-to-one ("*injective*"). The *left complementary index* of the linear mapping f, namely the linear operator $\mathbf{A}$, which accounts for the *surjectivity defect*, is given by $d_s = n - \mathrm{rk}\mathbf{A}$, also called "*degree of freedom*" (here $d_s = n - \mathrm{rk}\mathbf{A} = 1$). In contrast, the *right complementary index* of the linear mapping f, namely the linear operator $\mathbf{A}$, which accounts for the injectivity defect is given by $d = m - \mathrm{rk}\mathbf{A}$ (here $d = m - \mathrm{rk}\mathbf{A} = 1$). While "*surjectivity*" relates to the range $\mathcal{R}(f)$ or "the range space $\mathcal{R}(A)$" and "injectivity" to the *kernel* $\mathcal{N}(f)$ or "the *null space* $\mathcal{N}(\mathbf{A})$" we shall constructively introduce the notion of

range$\mathcal{R}(f)$	versus	kernel$\mathcal{N}(f)$
range space$\mathcal{R}(\mathbf{A})$		*nullspace*$\mathcal{N}(f)$

by consequently solving the inconsistent system of linear equations with datum defect. But beforehand let us ask:

? Why is the inconsistent system of linear equations called deficient with respect to the datum ?

At this point we have to go back to the *measurement process*. Our front page numerical example has been generated from measurements with a *leveling instrument*: Three *height differences* $(y_{\alpha\beta},\ y_{\beta\gamma},\ y_{\gamma\alpha})$ in a triangular network have been *observed*. They are related to absolute height $x_1 = h_\alpha,\ x_2 = h_\beta,\ x_3 = h_\gamma$ by means of $h_{\alpha\beta} = h_\beta - h_\alpha,\ h_{\beta\gamma} = h_\gamma - h_\beta,\ h_{\gamma\alpha} = h_\alpha - h_\gamma$ at points $\{P_\alpha,\ P_\beta,\ P_\gamma\}$, outlined in more detail in *Box 5.1*.

Box 5.2. (The measurement process of leveling and its relation to the linear model):

$$\begin{aligned} y_1 &= y_{\alpha\beta} = h_{\alpha\beta} + i_{\alpha\beta} = -h_\alpha + h_\beta + i_{\alpha\beta} \\ y_2 &= y_{\beta\gamma} = h_{\beta\gamma} + i_{\beta\gamma} = -h_\beta + h_\gamma + i_{\beta\gamma} \\ y_3 &= y_{\gamma\alpha} = h_{\gamma\alpha} + i_{\gamma\alpha} = -h_\gamma + h_\alpha + i_{\gamma\alpha} \end{aligned}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -h_\alpha + h_\beta + i_{\alpha\beta} \\ -h_\beta + h_\gamma + i_{\beta\gamma} \\ -h_\gamma + h_\alpha + i_{\gamma\alpha} \end{bmatrix} = \begin{bmatrix} -x_1 + x_2 + i_1 \\ -x_2 + x_3 + i_2 \\ -x_3 + x_1 + i_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix}.$$

Thirdly, let us begin with a more detailed analysis of the linear mapping $f : \mathbf{Ax} \doteq \mathbf{y}$ or $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$, namely of the *linear operator* $\mathbf{A} \in \mathbb{R}^{n \times m}$, $r = \operatorname{rk} \mathbf{A} \leq min\{n, m\}$. We shall pay special attention to the *three fundamental partitioning*, namely

(i) Algebraic partitioning called *additive* and *multiplicative rank partitioning* of the matrix $\mathbf{A}$
(ii) Geometric partitioning called slicing of the linear space $\mathbb{X}$ (*parameter space*) as well as of the linear space $\mathbb{Y}$ (*observation space*)
(iii) Set-theoretical partitioning called *fibering* of the set $\mathbb{X}$ of parameter and the set $\mathbb{Y}$ of observations.

5-13 Minimum Norm: Least Squares Solution of the Front Page Example by Means of Additive Rank Partitioning

Box 5.3 is a setup of the *minimum norm – least squares solution* of the inconsistent system of inhomogeneous linear equations with datum defect following the first principle "*additive rank partitioning*"). The term "*additive*" is taken from the additive decomposition $\mathbf{y}_1 = \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2$ *and* $\mathbf{y}_2 = \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2$ of the observational equations subject to $\mathbf{A}_{11} \in \mathbb{R}^{n \times m}$, $\operatorname{rk} \mathbf{A}_{11} \leq min\{n, m\}$.

Box 5.3. (Minimum norm-least squares solution of the inconsistent system of inhomogeneous linear equations with datum defect, "*additive rank partitioning*".

The solution of the hierarchical optimization problem (1st) $||\mathbf{i}||^2_{\mathrm{I}} = \min_{\mathbf{x}}$:

$$\mathbf{x}_l = \arg\{||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathrm{I}} = \min |\mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{y},\ \mathbf{A} \in \mathbb{R}^{n\times m}, rk\mathbf{A} \leq \min\{n,\ m\}\}$$

$$(2nd)\ ||\mathbf{x}_l||^2_{\mathrm{I}} = \min_{\mathbf{x}_l} :$$

$$\mathbf{x}_{lm} = \arg\{||\mathbf{x}_l||^2_{\mathrm{I}} = \min |\mathbf{A}'\mathbf{A}\mathbf{x}_l = \mathbf{A}'\mathbf{y},\ \mathbf{A}'\mathbf{A} \in \mathbb{R}^{m\times m},\ rk\mathbf{A}'\mathbf{A} \leq m\}$$

is based upon the simultaneous *horizontal and vertical rank partitioning* of the matrix $\mathbf{A}$, namely

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},\ \mathbf{A}_{11} \in \mathbb{R}^{r\times r},\ \mathrm{rk}\mathbf{A}_{11} = \mathrm{rk}\mathbf{A} =: r$$

with respect to the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$$

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix}, \left| \begin{array}{l} \mathbf{y}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{x}_1 \in \mathbb{R}^{r\times 1} \\ \mathbf{y}_2 \in \mathbb{R}^{(n-r)\times 1},\ \mathbf{x}_2 \in \mathbb{R}^{(m-r)\times 1}. \end{array} \right.$$

First, as shown before, we compute the least-squares solution $||\mathbf{i}||^2_{\mathrm{I}} = \min_{\mathbf{x}}$ or$||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathrm{I}} = \min_{\mathbf{x}}$ which generates standard normal equations

$$\mathbf{A}'\mathbf{A}\mathbf{x}_l = \mathbf{A}'\mathbf{y}$$

or

$$\begin{bmatrix} \mathbf{A}'_{11}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{A}_{21} & \mathbf{A}'_{11}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{A}_{22} \\ \mathbf{A}'_{12}\mathbf{A}_{11} + \mathbf{A}'_{22}\mathbf{A}_{21} & \mathbf{A}'_{12}\mathbf{A}_{12} + \mathbf{A}'_{22}\mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{21} \\ \mathbf{A}'_{12} & \mathbf{A}'_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

or

$$\begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1l} \\ \mathbf{x}_{2l} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}$$

subject to

$$\begin{array}{lll} \mathbf{N}_{11} := \mathbf{A}'_{11}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{A}_{21}, & \mathbf{N}_{12} := \mathbf{A}'_{11}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{A}_{22}, & \mathbf{m}_1 = \mathbf{A}'_{11}\mathbf{y}_1 + \mathbf{A}'_{21}\mathbf{y}_2 \\ \mathbf{N}_{21} := \mathbf{A}'_{12}\mathbf{A}_{11} + \mathbf{A}'_{22}\mathbf{A}_{21}, & \mathbf{N}_{22} := \mathbf{A}'_{12}\mathbf{A}_{12} + \mathbf{A}'_{22}\mathbf{A}_{22}, & \mathbf{m}_2 = \mathbf{A}'_{12}\mathbf{y}_1 + \mathbf{A}'_{22}\mathbf{y}_2, \end{array}$$

which are consistent linear equations with an (injectivity) defect $d = m - \mathrm{rk}\mathbf{A}$. The front page example leads us to

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \left[\begin{array}{c|c} \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \hline \begin{bmatrix} 1 & 0 \end{bmatrix} & \text{-1} \end{array}\right]$$

or

$$\mathbf{A}_{11} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$\mathbf{A}_{21} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad \mathbf{A}_{22} = -1$$

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

$$\mathbf{N}_{11} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{N}_{12} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad |\mathbf{N}_{11}| = 3 \neq 0,$$
$$\mathbf{N}_{21} = \begin{bmatrix} -1 & -1 \end{bmatrix}, \quad \mathbf{N}_{22} = 2$$

$$\mathbf{y}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{m}_1 = \mathbf{A}'_{11}\mathbf{y}_1 + \mathbf{A}'_{21}\mathbf{y}_2 = \begin{bmatrix} -4 \\ 0 \end{bmatrix}$$
$$\mathbf{y}_2 = -3, \quad \mathbf{m}_2 = \mathbf{A}'_{12}\mathbf{y}_1 + \mathbf{A}'_{22}\mathbf{y}_2 = 4.$$

Second, we compute as shown before the minimum norm solution $||\mathbf{x}_l||^2_{\mathrm{I}} = \min_{\mathbf{x}_l}$ or $\mathbf{x}'_1\mathbf{x}_1 + \mathbf{x}'_2\mathbf{x}_2$ which generates the standard normal equations in the following way.

$$\begin{aligned} \mathcal{L}(\mathbf{x}_1,\ \mathbf{x}_2) &= \mathbf{x}'_1\mathbf{x}_1 + \mathbf{x}'_2\mathbf{x}_2 \\ &= (\mathbf{x}'_{2l}\mathbf{N}'_{12}\mathbf{N}_{11}^{-1} - \mathbf{m}'_1\mathbf{N}_{11}^{-1})(\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{x}_{2l} - \mathbf{N}_{11}^{-1}\mathbf{m}_1) + \mathbf{x}'_{2l}\mathbf{x}_{2l} = \min_{\mathbf{x}_2} \end{aligned}$$

"*additive decomposition of the Lagrangean*"

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2$$
$$\mathcal{L}_0 := \mathbf{m}'_1\mathbf{N}_{11}^{-2}\mathbf{m}_1, \quad \mathcal{L}_1 := -2\mathbf{x}'_{2l}\mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{m}_1$$
$$\mathcal{L}_2 := \mathbf{x}'_{2l}\mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12}\mathbf{x}_{2l} + \mathbf{x}'_{2l}\mathbf{x}_{2l}$$

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial \mathbf{x}_2}(\mathbf{x}_{2lm}) = 0 \Leftrightarrow \frac{1}{2}\frac{\partial \mathcal{L}_1}{\partial \mathbf{x}_2}(\mathbf{x}_{2lm}) + \frac{1}{2}\frac{\partial \mathcal{L}_2}{\partial \mathbf{x}_2}(\mathbf{x}_{2lm}) = 0 \\ &\Leftrightarrow -\mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{m}_1 + (\mathbf{I} + \mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12})\mathbf{x}_{2lm} = 0 \Leftrightarrow \\ &\Leftrightarrow \quad \mathbf{x}_{2lm} = (\mathbf{I} + \mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12})^{-1}\mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{m}_1, \end{aligned}$$

which constitute the *necessary conditions*. Using Cayley inverse sum of two matrices formula, namely $(\mathbf{I} + \mathbf{B}\mathbf{C}^{-1}\mathbf{A}')^{-1}\mathbf{B}\mathbf{C}^{-1} = \mathbf{B}(\mathbf{A}\mathbf{B} + \mathbf{C})^{-1}$ for appropriate dimensions of the involved matrices, such that the identities holds

$$(\mathbf{I} + \mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12})^{-1}\mathbf{N}'_{12}\mathbf{N}_{11}^{-2} = \mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}^{2})^{-1}$$

we finally find

$$\mathbf{x}_{2lm} = \mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}^{2})^{-1}\mathbf{m}_1.$$

The *second derivatives*

$$\frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}_2 \partial \mathbf{x}'_2}(\mathbf{x}_{2lm}) = (\mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12} + \mathbf{I}) > 0$$

due to positive-definiteness of the matrix $\mathbf{I} + \mathbf{N}'_{12}\mathbf{N}_{11}^{-2}\mathbf{N}_{12}$ generate the *sufficiency condition* for obtaining the minimum of the *unconstrained Lagrangean*. Finally let us backward transform

$$\mathbf{x}_{2l} \mapsto \mathbf{x}_{1m} = -\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{x}_{2l} + \mathbf{N}_{11}^{-1}\mathbf{m}_1,$$

$$\mathbf{x}_{1lm} = -\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}^{2})^{-1}\mathbf{m}_1 + \mathbf{N}_{11}^{-1}\mathbf{m}_1.$$

Let us right multiply the identity

$$\mathbf{N}_{12}\mathbf{N}'_{12} = -\mathbf{N}_{11}\mathbf{N}'_{11} + \mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11}$$

by $(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1}$ such that

$$\mathbf{N}_{12}\mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1} = -\mathbf{N}_{11}\mathbf{N}'_{11}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1} + \mathbf{I}$$

holds, and *left* multiply by $\mathbf{N}_{11}^{-1}$, namely

$$\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1} = -\mathbf{N}'_{11}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1} + \mathbf{N}_{11}^{-1}.$$

Obviously we have generated the *linear form*

$$\boxed{\begin{gathered}
\left[\begin{aligned} \mathbf{x}_{1lm} &= \mathbf{N}'_{11}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1}\mathbf{m}_1 \\ \mathbf{x}_{2lm} &= \mathbf{N}'_{12}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1}\mathbf{m}_1 \end{aligned}\right. \\
or \\
\mathbf{x}_{lm} = \begin{bmatrix} \mathbf{x}_{1lm} \\ \mathbf{x}_{2lm} \end{bmatrix} = \begin{bmatrix} \mathbf{N}'_{11} \\ \mathbf{N}'_{12} \end{bmatrix}(\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11})^{-1}\mathbf{m}_1 \\
or \\
\begin{aligned}
\mathbf{x}_{lm} = &\begin{bmatrix} \mathbf{A}'_{11}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{A}_{21} \\ \mathbf{A}'_{12}\mathbf{A}_{11} + \mathbf{A}'_{22}\mathbf{A}_{21} \end{bmatrix} \\
&*[(\mathbf{A}'_{11}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{A}_{22})(\mathbf{A}'_{12}\mathbf{A}_{11} + \mathbf{A}'_{22}\mathbf{A}_{21}) + (\mathbf{A}'_{11}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{A}_{21})^2]^{-1} \\
&*[(\mathbf{A}'_{11}\mathbf{y}_1 + \mathbf{A}'_{21}\mathbf{y}_2].
\end{aligned}
\end{gathered}}$$

Let us compute numerically $\mathbf{x}_{lm}$ for the front page example.

$$\mathbf{N}_{11}\mathbf{N}'_{11} = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}, \ \mathbf{N}_{12}\mathbf{N}'_{12} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
$$\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11} = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}, \ [\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11}]^{-1} = \frac{1}{27}\begin{bmatrix} 6 & 3 \\ 3 & 6 \end{bmatrix}$$
$$\mathbf{m}_1 = \begin{bmatrix} -4 \\ 0 \end{bmatrix} \ \Rightarrow \ \mathbf{x}_{1lm} = \frac{1}{3}\begin{bmatrix} -4 \\ 0 \end{bmatrix}, \ \mathbf{x}_{2lm} = \frac{4}{3}, \ ||\mathbf{x}_{lm}||_{\mathrm{I}}^2 = \frac{4}{3}\sqrt{2}$$

$$x_{1lm} = \hat{h}_\alpha = -\frac{4}{3}, \ x_{2lm} = \hat{h}_\beta = 0, \ x_{3lm} = \hat{h}_\gamma = \frac{4}{3}$$
$$||\mathbf{x}_{lm}||_{\mathrm{I}}^2 = \frac{4}{3}\sqrt{2}$$
$$x_{1lm} + x_{2lm} + x_{3lm} = 0 \ \sim \ \hat{h}_\alpha + \hat{h}_\beta + \hat{h}_\gamma = 0.$$

The vector $\mathbf{i}_{lm}$ of inconsistencies has to be finally computed by means of

$$\mathbf{i}_{lm} = \mathbf{y} - \mathbf{A}\mathbf{x}_{lm}$$

$$\mathbf{i}_{lm} = -\frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{A}'\mathbf{i}_l = 0, \ ||\mathbf{i}_{lm}||_{\mathrm{I}}^2 = \frac{1}{3}\sqrt{3}.$$

The technique of horizontal and vertical rank partitioning has been pioneered by *H. Wolf* (1972,1973). ♣

5-14 Minimum Norm: Least Squares Solution of the Front Page Example by Means of Multiplicative Rank Partitioning

Box 5.4 is a setup of the *minimum norm-least squares solution* of the inconsistent system of inhomogeneous linear equations with datum defect following the first principle "*multiplicative rank partitioning*". The term "*multiplicative*" is taken from the multiplicative decomposition $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{D}\mathbf{E}\mathbf{y} + \mathbf{i}$ of the observational equations subject to

$$\mathbf{A} = \mathbf{D}\mathbf{E}, \ D \in R^{n \times r}, E \in R^{r \times m}, \ rk\mathbf{A} = rk\mathbf{D} = rk\mathbf{E} \leq \min\{n, m\}.$$

Box 5.4. (Minimum norm-least squares solution of the inconsistent system of inhomogeneous linear equations with datum defect *multiplicative rank partitioning*)

The solution of the hierarchical optimization problem

$$(1\text{st})||\mathbf{i}||_{\mathrm{I}}^2 = \min_{\mathbf{x}}$$
$$\mathbf{x}_l = \arg\{||\mathbf{y} - \mathbf{A}\mathbf{x}||_{\mathrm{I}}^2 = \min|\mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{y},\ \mathbf{A} \in \mathbb{R}^{n\times m}, rk\mathbf{A} \le \min\{n,\ m\}$$

$$(2\text{nd})||\mathbf{x}_l||_{\mathrm{I}}^2 = \min_{\mathbf{x}_l}$$
$$\mathbf{x}_{lm} = \arg\{||\mathbf{x}_l||_{\mathrm{I}}^2 = \min|\mathbf{A}'\mathbf{A}\mathbf{x}_l = \mathbf{A}'\mathbf{y},\ \mathbf{A}'\mathbf{A} \in \mathbb{R}^{m\times m}, rk\mathbf{A}'\mathbf{A} \le m\}$$

is based upon the rank factorization $\mathbf{A} = \mathbf{DE}$ of the matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ subject to simultaneous *horizontal* and *vertical rank partitioning* of the matrix $\mathbf{A}$, namely

$$\mathbf{A} = \mathbf{DE} = \begin{bmatrix} \mathbf{D} \in \mathbb{R}^{n\times r}, & \mathrm{rk}\mathbf{D} = \mathrm{rk}\mathbf{A} =: r \le \min\{n,\ m\} \\ \mathbf{E} \in \mathbb{R}^{r\times m}, & \mathrm{rk}\mathbf{E} = \mathrm{rk}\mathbf{A} =: r \le \min\{n,\ m\} \end{bmatrix}$$

with respect to the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$$
$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{DE}\mathbf{x} + \mathbf{i} \quad \left[\begin{array}{l} \mathbf{E}\mathbf{x} =: \mathbf{z} \\ \mathbf{DE}\mathbf{x} = \mathbf{D}\mathbf{z} \end{array}\right. \quad \Rightarrow \mathbf{y} = \mathbf{D}\mathbf{z} + \mathbf{i}.$$

First, as shown before, we compute the least-squares solution $||\mathbf{i}||_{\mathrm{I}}^2 = \min_{\mathbf{x}}$ or$||\mathbf{y} - \mathbf{A}\mathbf{x}||_{\mathrm{I}}^2 = \min_{\mathbf{x}}$ which generates *standard normal equations*

$$\mathbf{D}'\mathbf{D}\mathbf{z}_l = \mathbf{D}'\mathbf{y} \quad \Rightarrow \quad \mathbf{z}_l = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} = \mathbf{D}'_l\mathbf{y},$$

which are consistent linear equations of rank $\mathrm{rk}\mathbf{D} = \mathrm{rk}\mathbf{D}'\mathbf{D} = \mathrm{rk}\mathbf{A} = r$. The front page example leads us to

$$\mathbf{A} = \mathbf{DE} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \left[\begin{array}{c|c} \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \hline \begin{bmatrix} 1 & 0 \end{bmatrix} & -1 \end{array}\right], \ \mathbf{D} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

or

$$\mathbf{D}'\mathbf{DE} = \mathbf{D}'\mathbf{A} \quad \Rightarrow \quad \mathbf{E} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}$$
$$\mathbf{D}'\mathbf{D} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad (\mathbf{D}'\mathbf{D})^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \Rightarrow \mathbf{E} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$
$$\mathbf{z}_l = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}\mathbf{y}$$

$$\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \Rightarrow \mathbf{z}_l = -\frac{4}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\mathbf{z}_l = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}\mathbf{y}$$

$$\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \Rightarrow \mathbf{z}_l = -\frac{4}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Second, as shown before, we compute the minimum norm solution $||\mathbf{x}_\ell||^2_{\mathbf{I}} = \min_{\mathbf{x}_\ell}$ of the consistent system of linear equations with datum defect, namely

$$\mathbf{x}_{lm} = \arg\{||\mathbf{x}_l||^2_{\mathbf{I}} = \min_{\mathbf{x}_l} \mid \mathbf{E}\mathbf{x}_l = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}\}.$$

As outlined in *Box 1.3* the minimum norm solution of consistent equations with datum defect namely $\mathbf{E}\mathbf{x}_l = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}$, $rk\mathbf{E} = rk\mathbf{A} = r$ is

$$\boxed{\begin{aligned} \mathbf{x}_{lm} &= \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} \\ \mathbf{x}_{lm} &= \mathbf{E}^-_m\mathbf{D}^-_l\mathbf{y} = \mathbf{A}^-_{lm}\mathbf{y} = \mathbf{A}^+\mathbf{y}, \end{aligned}}$$

which is limit on the minimum norm generalized inverse. In summary, the minimum norm-least squares solution generalized inverse (MINOLESS g-inverse) also called pseudo-inverse $\mathbf{A}^+$ or *Moore–Penrose inverse* is the product of the MINOS **g**-inverse $\mathbf{E}^-_m$ (*right inverse*) and the LESS **g**-inverse $\mathbf{D}^-_l$ (*left inverse*). For the front page example we are led to compute

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{E}\mathbf{E}' = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$(\mathbf{E}\mathbf{E}')^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{x}_{lm} = \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} = \frac{1}{3}\begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}\mathbf{y}$$

$$\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} \Rightarrow \mathbf{x}_{lm} = \frac{1}{3}\begin{bmatrix} -4 \\ 0 \\ +4 \end{bmatrix} = \frac{4}{3}\begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix}, \quad ||\mathbf{x}_{lm}|| = \frac{4}{3}\sqrt{2}$$

$$\boxed{\begin{gathered}\mathbf{x}_{lm} = \begin{bmatrix} x_{1lm} \\ x_{2lm} \\ x_{3lm} \end{bmatrix} = \begin{bmatrix} \hat{h}_\alpha \\ \hat{h}_\beta \\ \hat{h}_\gamma \end{bmatrix} = \frac{4}{3}\begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix} \\ ||\mathbf{x}_{lm}|| = \frac{4}{3}\sqrt{2} \\ x_{1lm} + x_{2lm} + x_{3lm} = 0 \quad \sim \quad \hat{h}_\alpha + \hat{h}_\beta + \hat{h}_\gamma = 0. \end{gathered}}$$

The vector $\mathbf{i}_{lm}$ of inconsistencies has to be finally computed by means of

$$\mathbf{i}_{lm} := \mathbf{y} - \mathbf{A}\mathbf{x}_{lm} = [\mathbf{I}_n - \mathbf{A}\mathbf{A}^-_{lm}]\mathbf{y},$$
$$\mathbf{i}_{lm} = [\mathbf{I}_n - \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{y};$$

$$\boxed{\mathbf{i}_{lm} = -\frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{A}'\mathbf{i}_l = 0, \quad ||\mathbf{i}_{lm}|| = \frac{1}{3}\sqrt{3}}$$

♣

Box 5.5 summarizes the algorithmic steps for the diagnosis of the simultaneous horizontal and vertical rank partitioning to generate $(F_{m_1} G_{\mathbf{y}})$-MINOS.

Box 5.5. (algorithm):

The diagnostic *algorithm* for solving a general rank deficient system of linear equations

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \ \ \mathbf{A} \in \mathbb{R}^{n\times m}, \ \ \text{rk}\mathbf{A} < \min\{n, m\}$$

by means of simultaneous *horizontal and vertical rank partioning*

$$\boxed{\begin{gathered}\textit{Determine} \\ \texttt{the rank of the matrix}\mathbf{A}\ \text{rk}\mathbf{A} < \min\{n, m\}.\end{gathered}}$$

$$\downarrow$$

$$\boxed{\begin{gathered}\textit{Compute} \\ \text{“}\texttt{the simultaneous horizontal and vertical rank partioning}\text{''} \\ \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \begin{matrix} \mathbf{A}_{11} \in \mathbb{R}^{r\times r}, \mathbf{A}_{12} \in \mathbb{R}^{r\times(m-r)} \\ \mathbf{A}_{21} \in \mathbb{R}^{(n-r)\times r}, \mathbf{A}_{22} \in \mathbb{R}^{(n-r)\times(m-r)} \end{matrix} \\ \text{“}n-r\ \texttt{is called the left complementary index,} \\ m-r\ \texttt{the right complementary index}\text{''} \\ \text{“}\mathbf{A}\ \texttt{as a linear operator is neither injective}\ (m-r \neq 0), \\ \texttt{nor surjective}\ (n-r=0)\,.\text{''}\end{gathered}}$$

$$\downarrow$$

$$\boxed{\begin{array}{c} \textit{Compute} \\ \texttt{the range space}\,\mathcal{R}(\mathbf{A})\,\texttt{and the null space}\,\mathcal{N}(\mathbf{A}) \\ \texttt{of the linear operator}\,\mathbf{A} \\ \mathcal{R}(\mathbf{A}) = \texttt{span}\,\{wl_1(\mathbf{A}),\ldots,\,wl_r(\mathbf{A})\} \\ \mathcal{N}(\mathbf{A}) = \{\mathbf{x}\in\mathbb{R}^n \,|\, \mathbf{N}_{11}\mathbf{x}_{1\ell} + \mathbf{N}_{12}\mathbf{x}_{2\ell} = 0\} \\ or \\ \mathbf{x}_{1\ell} = -\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{x}_{2\ell}. \end{array}}$$

$$\downarrow$$

$$\boxed{\begin{array}{c} \textit{Compute} \\ (T_m,\,G_{\mathbf{y}}) - MINOS \\ \mathbf{x}_{\ell m} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} = \begin{bmatrix}\mathbf{N}'_{11}\\ \mathbf{N}'_{12}\end{bmatrix} = [\mathbf{N}_{12}\mathbf{N}'_{12} + \mathbf{N}_{11}\mathbf{N}'_{11}]^{-1}[\mathbf{A}'_{11}\mathbf{G}^y_{11}, \mathbf{A}'_{21}\mathbf{G}^y_{12}]\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} \\ \mathbf{N}_{11} := \mathbf{A}'_{11}\mathbf{G}^y_{11}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{G}^y_{22}\mathbf{A}_{21},\ \ \mathbf{N}_{12} := \mathbf{A}'_{11}\mathbf{G}^y_{11}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{G}^y_{22}\mathbf{A}_{22} \\ \mathbf{N}_{21} := \mathbf{N}'_{12},\ \ \mathbf{N}_{22} := \mathbf{A}'_{12}\mathbf{G}^y_{11}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{G}^y_{22}\mathbf{A}_{22}. \end{array}}$$

5-15 The Range R(f) and the Kernel N(f) Interpretation of "MINOLESS" by Three Partitionings

(i) Algebraic (rank partitioning)
(ii) Geometric (slicing)
(iii) Set-theoretical (fibering)

Here we will outline by means of *Box 5.6* the range space as well as the null space of the *general inconsistent system* of linear equations.

Box 5.6. (The range space and the null space of the general inconsistent system of linear equations):

$$\mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{y},\ \ \mathbf{A}\in\mathbb{R}^{n\times m},\ \ \mathrm{rk}\mathbf{A} \le \min\{n,\ m\}$$

"*additive rank partitioning*".

The matrix **A** is called a simultaneous *horizontal and vertical rank partitioning*, if

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix},\ \ \mathbf{A}_{11} = \mathbb{R}^{r\times r},\ \ \mathrm{rk}\mathbf{A}_{11} = \mathrm{rk}\mathbf{A} =: r$$

with respect to the linear model

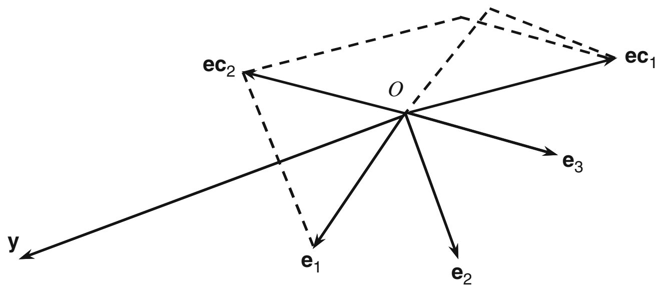


Fig. 5.1 Range $\mathcal{R}(f)$ range space $\mathcal{R}(\mathbf{A})$, $(\mathbf{y} \notin \mathcal{R}(\mathbf{A}))$

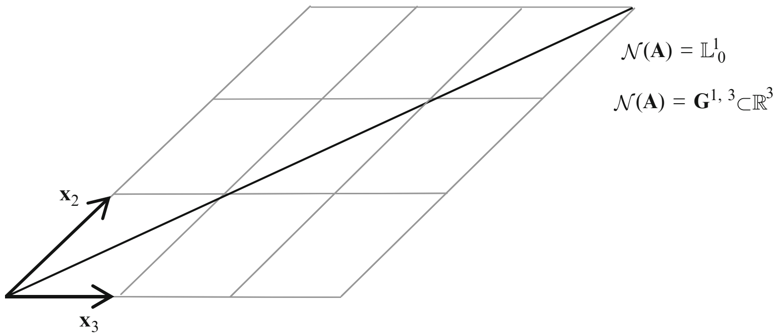


Fig. 5.2 Kernel $\mathcal{N}(f)$, null space $\mathcal{N}(\mathbf{A})$, "the null space $\mathcal{N}(\mathbf{A})$ as the linear manifold (Grassmann space $\mathbf{G}^{1,3}$) slices the parameter space $\mathbb{X} = \mathbb{R}^3$", $\mathbf{x}_3$ is not displayed

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i},\ \mathbf{A} \in \mathbb{R}^{n\times m},\ \mathrm{rk}\mathbf{A} \leq \min\{n,\ m\}$$

$$\boxed{\textit{identification of the range space}}$$

$$\mathcal{R}(\mathbf{A}) = span\,\{\sum_{i=1}^{n} \mathbf{e}_i a_{ij} \mid j \in \{1, \ldots, r\}\}$$

"*front page example*"

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{-3} \end{bmatrix},\ \mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ \hline 1 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 3},\ \mathrm{rk}\mathbf{A} =:\ r = 2$$

$$\boxed{\mathcal{R}(\mathbf{A}) = span\,\{\mathbf{e}_1 a_{11} + \mathbf{e}_2 a_{21} + \mathbf{e}_3 a_{31}, \mathbf{e}_1 a_{12} + \mathbf{e}_2 a_{22} + \mathbf{e}_3 a_{32}\} \subset \mathbb{R}^3}$$

or

$$\mathcal{R}(\mathbf{A}) = span\,\{-\mathbf{e}_1 + \mathbf{e}_3, \mathbf{e}_1 - \mathbf{e}_2\} \subset \mathbb{R}^3 = \mathbb{Y}$$

$$\mathbf{c}_1 = [-1,\ 0,\ 1],\ \mathbf{c}_2 = [1,\ -1,\ 0],\ \mathbb{R}^3 = \mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$$

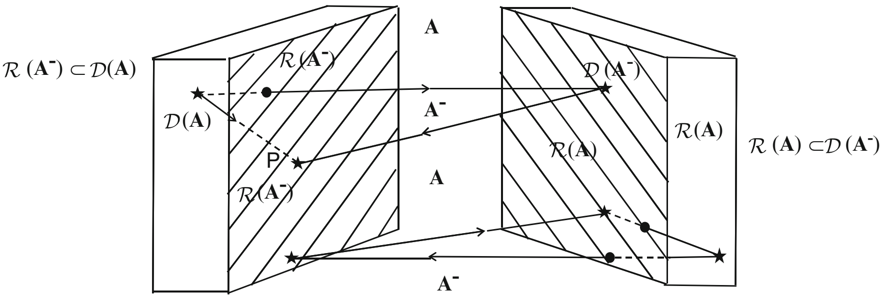


Fig. 5.3 Least squares, minimum norm generalized inverse $\mathbf{A}^-_{\ell m}(\mathbf{A}^{1,2,3,4}\, or\, \mathbf{A}^+)$, the *Moore–Penrose-inverse (Tseng inverse)*

$$\boxed{\textit{identification of the null space}}$$

$$\begin{aligned}\mathbf{N}_{11}\mathbf{x}_{1\ell} + \mathbf{N}_{12}\mathbf{x}_{2\ell} &= \mathbf{A}'_{11}\mathbf{y}_1 + \mathbf{A}'_{21}\mathbf{y}_2\\ \mathbf{N}_{12}\mathbf{x}_{1\ell} + \mathbf{N}_{22}\mathbf{x}_{2\ell} &= \mathbf{A}'_{12}\mathbf{y}_1 + \mathbf{A}'_{22}\mathbf{y}_2\end{aligned}$$

$$\mathcal{N}(\mathbf{A}) := \{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} | \mathbf{N_{11}x_{1\ell}} + \mathbf{N_{12}x_{2\ell}} = \mathbf{0}\}$$

or

$$\mathbf{N}_{11}\mathbf{x}_{1\ell} + \mathbf{N}_{12}\mathbf{x}_{2\ell} = 0 \;\Leftrightarrow\; \mathbf{x}_{1\ell} = -\mathbf{N}_{11}^{-1}\mathbf{N}_{12}\mathbf{x}_{2\ell}$$

"*front page example*"

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_3\end{bmatrix}_\ell = -\frac{1}{3}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}-1\\ -1\end{bmatrix}\mathbf{x}_{3\ell} = \begin{bmatrix}\mathbf{x}_3\\ \mathbf{x}_3\end{bmatrix}_\ell$$

$$\mathbf{N}_{11} = \begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix},\ \mathbf{N}_{11}^{-1} = \frac{1}{3}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix},\ \mathbf{N}_{12} = \begin{bmatrix}-1\\ -1\end{bmatrix}$$

$$\mathbf{x}_{1\ell} = u,\ \mathbf{x}_{2\ell} = u,\ \mathbf{x}_{3\ell} = u$$

$$\mathcal{N}(A) = H_0^1 = G^{1,3}.$$

Box 5.7 is a summary of MINOLESS of a *general inconsistent system of linear equations* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$. Based on the notion of the rank $r = \text{rk}\mathbf{A} < \min\{n, m\}$, we designed the generalized inverse of

$$\begin{gathered}\textit{MINOS type}\\ \textit{or}\\ \mathbf{A}^-_{\ell m}\ or\ \mathbf{A}^{1,2,3,4}\end{gathered}$$

Box 5.7. (MINOLESS of a general inconsistent system of linear equations):

$$f : \mathbf{x} \to \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i},\ \mathbf{x} \in \mathbb{X} = \mathbb{R}^m\ (parameter\,space),$$
$$\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n\ (observation\,space)$$
$$r := \mathrm{rk}\mathbf{A} < \min\{n, m\}$$
$$\mathbf{A} - \texttt{generalized inverse of MINOS type}:$$
$$\mathbf{A}^{1,2,3,4}\ or\ \mathbf{A}^-_{\ell m}$$

$\boxed{\texttt{Condition\#1}}$	$\boxed{\texttt{Condition\#1}}$
$f(\mathbf{x}) = f(g(\mathbf{y}))$	$\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x}$
$\Leftrightarrow$	$\Leftrightarrow$
$\mathrm{f} = f \circ g \circ f$	$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$
$\boxed{\texttt{Condition\#2}}$	$\boxed{\texttt{Condition\#2}}$
$g(\mathbf{y}) = g(f(\mathbf{x}))$	$\mathbf{A}^-\mathbf{y} = \mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{y}$
$\Leftrightarrow$	$\Leftrightarrow$
$g = g \circ f \circ g$	$\mathbf{A}^-\mathbf{A}\mathbf{A}- = \mathbf{A}^-$
$\boxed{\texttt{Condition\#3}}$	$\boxed{\texttt{Condition\#3}}$
$f(g(y)) = yR(A)$	$\mathbf{A}^-\mathbf{A}\mathbf{y} = \mathbf{y}_{R(\mathbf{A})}$
$\Leftrightarrow$	$\Leftrightarrow$
$f \circ g = P_{\mathcal{R}(\mathbf{A})}$	$\mathbf{A}^-\mathbf{A} = P_{\mathcal{R}(\mathbf{A}-)}$
$\boxed{\texttt{Condition\#4}}$	$\boxed{\texttt{Condition\#4}}$
$g(f(x)) = \mathbf{y}_{\mathcal{R}(\mathbf{A})^\perp}$	$\mathbf{A}\mathbf{A}^- = P_{\mathcal{R}(\mathbf{A})}$

$$g \circ f = P_{\mathcal{R}(g)}$$
$$\mathbf{A}^-\mathbf{A} = P_{\mathcal{R}(\mathbf{A}-)}.$$

$$\mathbf{A}\mathbf{A}^- = P_{\mathcal{R}(\mathbf{A})}$$
$$f \circ g = P_{\mathcal{R}(f)}$$

A similar construction of the generalized inverse of a matrix applies to the diagrams of the mappings:

(1) *Under the mapping* $\mathbf{A}$:

$$\mathcal{D}(\mathbf{A}) \to \mathcal{R}(\mathbf{A})$$
$$\mathbf{A}\mathbf{A}^- = P_{\mathcal{R}(\mathbf{A})}$$
$$f \circ g = P_{\mathcal{R}(f)}$$

(2) *Under the mapping* $\mathbf{A}-$:

$$\mathcal{R}(\mathbf{A}) \to P_{\mathcal{R}(\mathbf{A}-)}$$
$$\mathbf{A}^-\mathbf{A} = P_{\mathcal{R}(\mathbf{A}-)}$$
$$g \circ f = P_{\mathcal{R}(g)}.$$

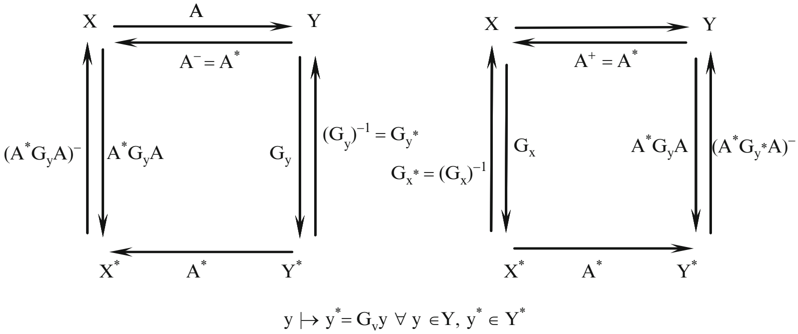


Fig. 5.4 Orthogonal inverses and adjoints in reflexive dual Hilbert spaces

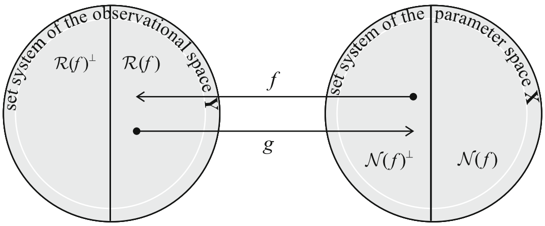


Fig. 5.5 Venn diagram, trivial fibering of the domain $\mathcal{D}(f)$: Trivial fibers $\mathcal{N}(f)^{\perp}$, trivial fibering of the range $\mathcal{R}(f)$: trivial fibers $\mathcal{R}(f)$ *and*$\mathcal{R}(f)^{\perp}$, $f : \mathbb{R}^m = \mathbb{X} \to \mathbb{Y} = \mathbb{R}^n$, $\mathbb{X}$ set system of the parameter space, $\mathbb{Y}$ set system of the observation space

In addition, we follow *Figs. 5.4* and *5.5* for the *characteristic diagrams* describing:
(i) Orthogonal inverses and adjoints in reflexive dual Hilbert spaces
(ii) Venn diagrams, trivial fiberings

In particularly, if **Gy** is rank defect we proceed as follows.

$$\mathbf{G_y} \qquad \mathbf{G}_\mathbf{y}^* = \begin{bmatrix} \Lambda_\mathbf{y} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\text{synthesis} \qquad \text{analysis}$$

$$\begin{aligned} \mathbf{G_y} &= \mathbf{U}\mathbf{G}_\mathbf{y}^*\mathbf{U}' \\ &= \mathbf{U}_1\boldsymbol{\Lambda}_\mathbf{y}\mathbf{U}'_1 \end{aligned} \qquad \begin{aligned} \mathbf{G}_\mathbf{y}^* = \mathbf{U}'\mathbf{G_y}\mathbf{U} &= \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix} \mathbf{G_y}[\mathbf{U}_1, \mathbf{U}_2] \\ &= \begin{bmatrix} \mathbf{U}'_1\mathbf{G_y}\mathbf{U}_1 & \mathbf{U}'_1\mathbf{G_y}\mathbf{U}_2 \\ \mathbf{U}'_2\mathbf{G_y}\mathbf{U}'_1 & \mathbf{U}'_2\mathbf{G_y}\mathbf{U}_2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} &\boldsymbol{\Lambda}_\mathbf{y} = \mathbf{U}'_1\mathbf{G_y}\mathbf{U}_1 \Rightarrow \mathbf{U}_1\boldsymbol{\Lambda}_\mathbf{y} = \mathbf{G_y}\mathbf{U}_1 \\ &0 = \mathbf{G_y}\mathbf{U}'_2 \ \ and \ \ \mathbf{U}'_1\mathbf{G_y}\mathbf{U}_2 = 0 \end{aligned}$$

$$\boxed{\left.\begin{array}{c} ||\mathbf{y}-\mathbf{A}\mathbf{x}||^2_{\mathbf{G}_\mathbf{y}} = \ ||\mathbf{i}||^2 = \mathbf{i}'\mathbf{G}_\mathbf{y}\mathbf{i} \\ \mathbf{G}_\mathbf{y} = \mathbf{U}'_1\mathbf{\Lambda}_\mathbf{y}\mathbf{U}_1 \end{array}\right] \Rightarrow \\ (\mathbf{y}-\mathbf{A}\mathbf{x})'\mathbf{U}'_1\mathbf{\Lambda}_\mathbf{y}\mathbf{U}_1(\mathbf{y}-\mathbf{A}\mathbf{x}') = \min_\mathbf{x} \Rightarrow}$$

$$\boxed{\Rightarrow \mathbf{U}_1(\mathbf{y}-\mathbf{A}\mathbf{x}) = \mathbf{U}_1\mathbf{i} = \mathbf{i}^*}$$

If we use *simultaneous horizontal and vertical rank partitioning*

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix}$$

subject to special dimension identities

$$\begin{array}{l} \mathbf{y}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{y}_2 \in \mathbb{R}^{(n-r)\times 1} \\ \mathbf{A}_{11} \in \mathbb{R}^{r\times r},\ \mathbf{A}_{12} \in \mathbb{R}^{r\times(m-r)} \\ \mathbf{A}_{21} \in \mathbb{R}^{(n-r)\times r},\ \mathbf{A}_{22} \in \mathbb{R}^{(n-r)\times(m-r)}, \end{array}$$

we arrive at *Lemma 5.0.*

Lemma 5.1. ($\mathbf{G}_\mathbf{x}, \mathbf{G}_\mathbf{y}$)-MINOLESS, simultaneous horizontal and vertical rank partitioning):

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1 \\ \mathbf{i}_2 \end{bmatrix} \tag{5.1}$$

subject to the dimension identities

$$\begin{array}{l} \mathbf{y}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{y}_2 \in \mathbb{R}^{(n-r)\times 1},\ \mathbf{x}_1 \in \mathbb{R}^{r\times 1},\ \mathbf{x}_2 \in \mathbb{R}^{(m-r)\times 1} \\ \mathbf{A}_{11} \in \mathbb{R}^{r\times r},\ \mathbf{A}_{12} \in \mathbb{R}^{r\times(m-r)} \\ \mathbf{A}_{21} \in \mathbb{R}^{(n-r)\times r},\ \mathbf{A}_{22} \in \mathbb{R}^{(n-r)\times(m-r)} \end{array}$$

is a *simultaneous horizontal and vertical rank partitioning* of the linear model (5.1)

$$\{\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i},\ \mathbf{A} \in^{n\times m},\ r := \mathrm{rk}\mathbf{A} < \min\{n, m\}\} \tag{5.2}$$

r is the index of the linear operator $\mathbf{A}$,
$n-r$ is the left complementary index and
$m-r$ is the right complementary index.
$\mathbf{x}_\ell$ is $\mathbf{G}_y$-LESS if it fulfils the *rank* $\mathbf{x}_{\ell m}$ is MINOS of $\mathbf{A}'\mathbf{G}_y\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{G}_y\mathbf{y}$, if

$$
\begin{aligned}
(\mathbf{x}_1)_{\ell m} &= -\mathbf{N}_{11}^{-1}\mathbf{N}_{12}[\mathbf{N}'_{12}\mathbf{N}_{11}^{-1}\mathbf{G}_{11}^{\mathbf{x}} - 2\mathbf{G}_{21}^{\mathbf{x}}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} + \mathbf{G}_{22}^{\mathbf{x}}]^{-1} \\
&\quad *(\mathbf{N}'_{12}\mathbf{N}_{11}^{-1}\mathbf{G}_{11}^{\mathbf{x}}\mathbf{N}_{11}^{-1} - 2\mathbf{G}_{21}^{\mathbf{x}}\mathbf{N}_{11}^{-1})m_1 + \mathbf{N}_{11}^{-1}m_1
\end{aligned} \tag{5.3}
$$

$$
\begin{aligned}
(\mathbf{x}_2)_{\ell m} &= [\mathbf{N}'_{12}\mathbf{N}_{11}^{-1}\mathbf{G}_{11}^{\mathbf{x}}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} - 2\mathbf{G}_{21}^{\mathbf{x}}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} + \mathbf{G}_{22}^{\mathbf{x}}]^{-1} \\
&\quad *(\mathbf{N}'_{12}\mathbf{N}_{11}^{-1}\mathbf{G}_{11}^{\mathbf{x}}\mathbf{N}_{11}^{-1} - 2\mathbf{G}_{21}^{\mathbf{x}}\mathbf{N}_{11}^{-1})m_1.
\end{aligned} \tag{5.4}
$$

The symmetric matrices $(\mathbf{G_x}, \mathbf{G_y})$ of the metric of the *parameter space* $\mathbb{X}$ as well as of the *observation space* $\mathbb{Y}$ are consequently partitioned as

$$
\mathbf{G_y} = \begin{bmatrix} \mathbf{G}_{11}^{\mathbf{y}} & \mathbf{G}_{12}^{\mathbf{y}} \\ \mathbf{G}_{21}^{\mathbf{y}} & \mathbf{G}_{22}^{\mathbf{y}} \end{bmatrix} \qquad \textit{and} \qquad \begin{bmatrix} \mathbf{G}_{11}^{\mathbf{x}} & \mathbf{G}_{12}^{\mathbf{x}} \\ \mathbf{G}_{21}^{\mathbf{x}} & \mathbf{G}_{22}^{\mathbf{x}} \end{bmatrix} = \mathbf{G_x} \tag{5.5}
$$

subject to the dimension identities

$$
\begin{matrix} \mathbf{G}_{11}^{\mathbf{y}} \in \mathbb{R}^{r\times r},\ \mathbf{G}_{12}^{\mathbf{y}} \in \mathbb{R}^{r\times(n-r)} \\ \mathbf{G}_{21}^{\mathbf{y}} \in \mathbb{R}^{(n-r)\times r},\ \mathbf{G}_{22}^{\mathbf{y}} \in \mathbb{R}^{(n-r)\times(n-r)} \end{matrix} \quad \textit{versus} \quad \begin{matrix} \mathbf{G}_{11}^{\mathbf{x}} \in \mathbb{R}^{r\times r},\ \mathbf{G}_{12}^{\mathbf{x}} \in \mathbb{R}^{r\times(m-r)} \\ \mathbf{G}_{21}^{\mathbf{x}} \in \mathbb{R}^{(m-r)\times r},\ \mathbf{G}_{22}^{\mathbf{x}} \in \mathbb{R}^{(m-r)\times(m-r)} \end{matrix}
$$

deficient normal equations

$$
\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{5.6}
$$

or

$$
\begin{bmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}_\ell = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} \tag{5.7}
$$

subject to

$$
\mathbf{N}_{11} := \mathbf{A}'_{11}\mathbf{G}_{11}^{\mathbf{y}}\mathbf{A}_{11} + \mathbf{A}'_{21}\mathbf{G}_{21}^{\mathbf{y}}\mathbf{A}_{11} + \mathbf{A}'_{11}\mathbf{G}_{12}^{\mathbf{y}}\mathbf{A}_{21} + \mathbf{A}'_{21}\mathbf{G}_{22}^{\mathbf{y}}\mathbf{A}_{21} \tag{5.8}
$$

$$
\mathbf{N}_{12} := \mathbf{A}'_{11}\mathbf{G}_{11}^{\mathbf{y}}\mathbf{A}_{12} + \mathbf{A}'_{21}\mathbf{G}_{21}^{\mathbf{y}}\mathbf{A}_{12} + \mathbf{A}'_{11}\mathbf{G}_{12}^{\mathbf{y}}\mathbf{A}_{22} + \mathbf{A}'_{21}\mathbf{G}_{22}^{\mathbf{y}}\mathbf{A}_{22}, \tag{5.9}
$$

$$
\mathbf{N}_{21} = \mathbf{N}'_{12}, \tag{5.10}
$$

$$
\mathbf{N}_{22} := \mathbf{A}'_{12}\mathbf{G}_{11}^{\mathbf{y}}\mathbf{A}_{12} + \mathbf{A}'_{22}\mathbf{G}_{21}^{\mathbf{y}}\mathbf{A}_{12} + \mathbf{A}'_{12}\mathbf{G}_{12}^{\mathbf{y}}\mathbf{A}_{22} + \mathbf{A}'_{22}\mathbf{G}_{22}^{\mathbf{y}}\mathbf{A}_{22}, \tag{5.11}
$$

$$
\mathbf{M}_{11} := \mathbf{A}'_{11}\mathbf{G}_{11}^{\mathbf{y}} + \mathbf{A}'_{21}\mathbf{G}_{21}^{\mathbf{y}},\ \mathbf{M}_{12} := \mathbf{A}'_{11}\mathbf{G}_{12}^{\mathbf{y}} + \mathbf{A}'_{21}\mathbf{G}_{22}^{\mathbf{y}}, \tag{5.12}
$$

$$
\mathbf{M}_{21} := \mathbf{A}'_{12}\mathbf{G}_{11}^{\mathbf{y}} + \mathbf{A}'_{22}\mathbf{G}_{21}^{\mathbf{y}},\ \mathbf{M}_{22} := \mathbf{A}'_{12}\mathbf{G}_{12}^{\mathbf{y}} + \mathbf{A}'_{22}\mathbf{G}_{22}^{\mathbf{y}}, \tag{5.13}
$$

$$
\mathbf{m}_1 := \mathbf{M}_{11}\mathbf{y}_1 + \mathbf{M}_{12}\mathbf{y}_2,\ \mathbf{m}_2 := \mathbf{M}_{21}\mathbf{y}_1 + \mathbf{M}_{22}\mathbf{y}_2. \tag{5.14}
$$

5-2 MINOLESS and Related Solutions Like Weighted Minimum Norm-Weighted Least Squares Solutions

5-21 The Minimum Norm-Least Squares Solution: "MINOLESS"

The system of the inconsistent, rank deficient linear equations $\mathbf{Ax} + \mathbf{i} = \mathbf{y}$ subject to $\mathbf{A} \in \mathbb{R}^{n \times m},\ \text{rk}\mathbf{A} < \min\{n,\ m\}$ allows certain solutions which we introduce by means of *Definition* 5.1 as a solution of a certain hierarchical optimization problem. *Lemma* 5.2 contains the normal equations of the hierarchical optimization problems. The solution of such a system of the normal equations is presented in *Lemma* 5.3 for the special case (i) $|\mathbf{G_x}| \neq 0$ and case (ii) $|\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A}| \neq 0$, *but* $|\mathbf{G_x}| = 0$. For the analyst:

Lemma 5.4 presents the toolbox of MINOLESS for multiplicative rank partitioning, known as rank factorization.	*and*	*Lemma* 5.5 presents the toolbox of MINOLESS for additive rank partitioning.

Definition 5.1. ($\mathbf{G_x}$-minimum norm-$\mathbf{G_y}$-least squares solution):

A vector $\mathbf{x}_{\ell m} \in \mathbf{X} = \mathbb{R}^m$ is called $\mathbf{G_x}$, $\mathbf{G_y}$-MINOLESS (MInimum NOrm with respect to the $\mathbf{G_x}$-seminorm-Least Squares Solution with respect to the $\mathbf{G_y}$-seminorm) of the inconsistent system of linear equations with datum defect

$$\mathbf{Ax} + \mathbf{i} = \mathbf{y} \left[\begin{array}{c} \text{rk}\mathbf{A} \leq \min\limits_{\mathbf{A} \in \mathbb{R}^{n \times m}} \{n,\ m\} \\ \mathbf{y} \notin \mathcal{R}(\mathbf{A}),\ \mathcal{N}(\mathbf{A}) \neq \{0\} \\ \mathbf{x}_{\ell m} \in \mathbb{X} = \mathbb{R}^m,\ \mathbf{y} \in \mathbf{Y} = \mathbb{R}^n, \end{array} \right. \tag{5.15}$$

if and only if

first

$$\mathbf{x}_\ell = \arg\{||\mathbf{i}||_{\mathbf{G_y}} = \min_{\mathbf{x}} |\mathbf{Ax} + \mathbf{i} = \mathbf{y},\ \text{rk}\mathbf{A} \leq \min\{n,\ m\}\}, \tag{5.16}$$

second

$$\mathbf{x}_{\ell m} = \arg\{||\mathbf{x}||_{\mathbf{G_x}} = \min_{\mathbf{x}} |\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{G_y}\mathbf{y}\} \tag{5.17}$$

is $\mathbf{G_y}$-MINOS of the system of normal equations $\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{G_y}\mathbf{y}$ which are $\mathbf{G_x}$-LESS. The solutions of type $\mathbf{G_x}$, $\mathbf{G_y}$-MINOLESS can be characterized as following.

Lemma 5.2. ($\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$ least squares solution):

A vector $\mathbf{x}_{\ell m} \in \mathbb{X} = \mathbb{R}^m$ is called $\mathbf{G_x}$, $\mathbf{G_y}$-MINOLESS of (5.1), if and only if the system of normal equations

$$\begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\ell m} \\ \lambda_{\ell m} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{A'G_yy} \end{bmatrix} \tag{5.18}$$

with respect to the vector $\lambda_{\ell m}$ of "*Lagrange multipliers*" is fulfilled. $\mathbf{x}_{\ell m}$ always exists and is uniquely determined, if the augmented matrix $[\mathbf{G_x}, \mathbf{A'G_yA}]$ agrees to the *rank identity*

$$\mathrm{rk}[\mathbf{G_x}, \mathbf{A'G_yA}] = m \tag{5.19}$$

or, equivalently, if the matrix $\mathbf{G_x} + \mathbf{A'G_yA}$ is *regular*.

Proof. $\mathbf{G_y}$-MINOS of the system of normal equations $\mathbf{A'G_yAx}_\ell = \mathbf{A'G_y}$ is constructed by means of the *constrained Lagrangean*

$$\mathcal{L}(\mathbf{x}_\ell, \lambda_\ell) := \mathbf{x'}_\ell \mathbf{G_x x}_\ell + 2\lambda'_\ell(\mathbf{A'G_yAx}_\ell - \mathbf{A'G_yy}) = \min_{\mathbf{x},\lambda},$$

such that the *first derivatives*

$$\left.\begin{aligned} \frac{1}{2}\frac{\partial \mathcal{L}}{\partial \mathbf{x}}(\mathbf{x}_{\ell m}, \lambda_{\ell m}) &= \mathbf{G_x x}_{\ell m} + \mathbf{A'G_yA}\lambda_{\ell m} = 0 \\ \frac{1}{2}\frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{x}_{\ell m}, \lambda_{\ell m}) &= \mathbf{A'G_yAx}_{\ell m} - \mathbf{A'G_yy} = 0 \end{aligned}\right] \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\ell m} \\ \lambda_{\ell m} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{A'G_yy} \end{bmatrix}$$

constitute the *necessary conditions*. The second derivatives

$$\frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}\partial \mathbf{x'}}(\mathbf{x}_{\ell m}, \lambda_{\ell m}) = \mathbf{G_x} \geq 0 \tag{5.20}$$

due to the positive semidefiniteness of the matrix $\mathbf{G_x}$ generate the *sufficiency condition* for obtaining the minimum of the *constrained Lagrangean*. Due to the assumption$\mathbf{A'G_yy} \in \mathcal{R}(\mathbf{A'G_xA})$ the *existence* of $\mathbf{G_y}$-MINOS $\mathbf{x}_{\ell m}$ is granted. In order to prove *uniqueness* of $\mathbf{G_y}$-MINOS $\mathbf{x}_{\ell m}$ we have to consider

case (i)	and	case (ii)
$\mathbf{G_x}$ positive definite		$\mathbf{G_x}$ positive semidefinite.

case (i): $\mathbf{G_x}$ positive definite

$$\begin{vmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & 0 \end{vmatrix} = |\mathbf{G_x}| \left| -\mathbf{A'G_yAG_x^{-1}A'G_yA} \right| = 0. \tag{5.21}$$

First, we solve the system of normal equations which characterize $\mathbf{x}_{\ell m}$ $\mathbf{G_x}$, $\mathbf{G_y}$-MINOLESS of $\mathbf{x}$ for the case of a full rank matrix of the metric $\mathbf{G_x}$ of the parametric space $\mathbb{X}$, rk$\mathbf{G_x} = m$ in particular. The system of normal equations is solved for

$$\begin{bmatrix} \mathbf{x}_{\ell m} \\ \lambda_{\ell m} \end{bmatrix} = \begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix}^{-} \begin{bmatrix} \mathbf{0} \\ \mathbf{A'G_yy} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{A'G_yy} \end{bmatrix} \tag{5.22}$$

subject to

$$\begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{G_x} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & 0 \end{bmatrix} \tag{5.23}$$

as a postulate for the $\mathbf{g}$-inverse of the partitioned matrix. *Cayley multiplication* of the three partitioned matrices leads us to *four matrix identities*.

$$\mathbf{G_xC_1G_x} + \mathbf{G_xC_2A'G_yA} + \mathbf{A'G_yAC_3G_x} + \mathbf{A'G_yAC_4A'G_yA} = \mathbf{G_x} \tag{5.24}$$

$$\mathbf{G_xC_1A'G_yA} + \mathbf{A'G_yAC_3A'G_yA} = \mathbf{A'G_yA} \tag{5.25}$$

$$\mathbf{A'G_yAC_1G_x} + \mathbf{A'G_yAC_2A'G_yA} = \mathbf{A'G_yA} \tag{5.26}$$

$$\mathbf{A'G_yAC_1A'G_yA} = 0. \tag{5.27}$$

Multiply the *third identity* by $\mathbf{G_x^{-1}A'G_yA}$ from the right side and substitute the *fourth identity* in order to solve for $\mathbf{C}_2$.

$$\mathbf{A'G_yAC_2A'G_yAG_x^{-1}A'G_yA} = \mathbf{A'G_yAG_x^{-1}A'G_yA} \tag{5.28}$$

$$\mathbf{C}_2 = \mathbf{G_x^{-1}A'G_yA(A'G_yAG_x^{-1}A'G_yA)^{-}}$$

solves the *fifth equation*

$$\begin{aligned} &\mathbf{A'G_yAG_x^{-1}A'G_yA(A'G_yAG_x^{-1}A'G_yA)^{-}A'G_yAG_x^{-1}A'G_yA} \\ &= \mathbf{A'G_yAG_x^{-1}A'G_yA} \end{aligned} \tag{5.29}$$

by the axiom of a generalized inverse

$$\mathbf{x}_{\ell m} = \mathbf{C_2A'G_yy} \tag{5.30}$$

$$\mathbf{x}_{\ell m} = \mathbf{G_y^{-1}A'G_yA(A'G_yAG_x^{-1}A'G_yA)^{-}A'G_yy}. \tag{5.31}$$

We leave the proof for "$\mathbf{G_x^{-1}A'G_yA(A'G_yAG_x^{-1}A'G_yA)^{-}A'G_y}$ is the weighted pseudo-inverse or Moore Penrose inverse $\mathbf{A}^{+}_{\mathbf{G_yG_x}}$" as an exercise.

case (ii): $\mathbf{G_x}$ positive semidefinite

Second, we relax the condition $\operatorname{rk}\mathbf{G_x} = m$ by the *alternative* $\operatorname{rk}[\mathbf{G_x}, \mathbf{A'G_yA}] = m$ $\mathbf{G_x}$ positive semidefinite. Add the second normal equation to the first one in order to receive the *modified system of normal equations*

$$\begin{bmatrix} \mathbf{G_x + A'G_yA} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\ell m} \\ \lambda_{\ell m} \end{bmatrix} = \begin{bmatrix} \mathbf{A'G_yy} \\ \mathbf{A'G_yy} \end{bmatrix} \tag{5.32}$$

$$\operatorname{rk}(\mathbf{G_x + A'G_yA}) = \operatorname{rk}[\mathbf{G_x}, \mathbf{A'G_yA}] = m. \tag{5.33}$$

The condition $\operatorname{rk}[\mathbf{G_x}, \mathbf{A'G_yA}] = m$ follows from the identity

$$\mathbf{G_x + A'G_yA} = [\mathbf{G_x}, \mathbf{A'G_yA}] \begin{bmatrix} \mathbf{G_x^-} & \mathbf{0} \\ \mathbf{0} & (\mathbf{A'G_yA})^- \end{bmatrix} \begin{bmatrix} \mathbf{G_x} \\ \mathbf{A'G_yA} \end{bmatrix}, \tag{5.34}$$

namely $\left|\mathbf{G_x + A'G_yA}\right| \neq 0$. The *modified system of normal equations* is solved for

$$\begin{aligned} \begin{bmatrix} \mathbf{x}_{\ell m} \\ \lambda_{\ell m} \end{bmatrix} &= \begin{bmatrix} \mathbf{G_x + A'G_yA} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix}^- \begin{bmatrix} \mathbf{A'G_yy} \\ \mathbf{A'G_yy} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{A'G_yy} \\ \mathbf{A'G_yy} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1\mathbf{A'G_yy} + \mathbf{C}_2\mathbf{A'G_yy} \\ \mathbf{C}_3\mathbf{A'G_yy} + \mathbf{C}_4\mathbf{A'G_yy} \end{bmatrix} \end{aligned} \tag{5.35}$$

subject to

$$\begin{aligned} &\begin{bmatrix} \mathbf{G_x + A'G_yA} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} \mathbf{G_x + A'G_yA} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \\ &\qquad = \begin{bmatrix} \mathbf{G_x + A'G_yA} & \mathbf{A'G_yA} \\ \mathbf{A'G_yA} & \mathbf{0} \end{bmatrix} \end{aligned} \tag{5.36}$$

as a postulate for the g-inverse of the partitioned matrix. *Cayley multiplication* of the three partitioned matrices leads us to the *four matrix identities*

"*element* (1,1)"

$$\begin{aligned} &(\mathbf{G_x + A'G_yA})\mathbf{C}_1(\mathbf{G_x + A'G_yA}) + \mathbf{A'G_yAC}_3(\mathbf{G_x + A'G_yA}) \\ &+(\mathbf{G_x + A'G_yA})\mathbf{C}_2\mathbf{A'G_yA} + \mathbf{A'G_yAC}_4\mathbf{A'G_yA} = \mathbf{G_x + A'G_yA} \end{aligned} \tag{5.37}$$

"*element* (1,2)"

$$(\mathbf{G_x + A'G_yA})\mathbf{C}_1\mathbf{A'G_yA} + \mathbf{A'G_yAC}_3 = \mathbf{A'G_yA} \tag{5.38}$$

"*element* (2,1)"

$$\mathbf{A'G_yAC}_1(\mathbf{G_x + A'G_yA}) + \mathbf{A'G_yAC}_2\mathbf{A'G_yA} = \mathbf{A'G_yA} \tag{5.39}$$

"*element* (2,2)"

$$\mathbf{A'G_yAC_1A'G_yA} = 0. \tag{5.40}$$

First, we realize that the *right sides* of the matrix identities are symmetric matrices. Accordingly the *left sides* have to constitute symmetric matrices, too.

$$(1,1): \begin{array}{l} \mathbf{(G_x + A'G_yA)C'_1(G_x + A'G_yA) + (G_x + A'G_yA)C'_3A'G_yA} \\ \mathbf{+A'G_yAC'_2(G_x + A'G_yA) + A'G_yAC'_4A'G_yA = G_x + A'G_yA} \end{array}$$

$$(1,2): \mathbf{A'G_yAC'_1(G_x + A'G_yA) + C'_3A'G_yA = A'G_yA}$$

$$(2,1): \mathbf{(G_x + A'G_yA)C'_1A'G_yA + A'G_yAC'_2A'G_yA = A'G_yA}$$

$$(2,2): \mathbf{A'G_yAC'_1A'G_yA = A'G_yAC_1A'G_yA} = 0.$$

We conclude

$$\mathbf{C_1 = C'_1},\ \mathbf{C_2 = C'_3},\ \mathbf{C_3 = C'_2},\ \mathbf{C_4 = C'_4}. \tag{5.41}$$

Second, we are going to solve for $C_1, C_2, C_3 = C_2\, and\ C_4$.

$$\begin{array}{l} \mathbf{C_1 = (G_x + A'G_yA)^{-1}\{I_\mathit{m} - A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-} \\ \qquad\quad \mathbf{*A'G_yA(G_x + A'G_yA)^{-1}\}} \end{array} \tag{5.42}$$

$$\mathbf{C_2 = (G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-} \tag{5.43}$$

$$\mathbf{C_3 = [A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-A'G_yA(G_x + A'G_yA)^{-1}} \tag{5.44}$$

$$\mathbf{C_4 = -[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-}. \tag{5.45}$$

For the proof, we depart from (1,2) to be multiplied by $\mathbf{A'G_yA(G_x + A'G_yA)^{-1}}$ from the left and implement (2,2)

$$\mathbf{A'G_yAC_2A'G_yA(G_x + A'G_yA)^{-1}A'G_yA = A'G_yA(G_x + A'G_yA)^{-1}A'G_yA}.$$

Obviously, $\mathbf{C_2}$ solves the *fifth equation* on the basis of the $\mathbf{g}$-inverse

$$\mathbf{[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-}$$

or

$$\begin{array}{l} \mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]} \\ \mathbf{*A'G_yA(G_x + A'G_yA)^{-1}A'G_yA = A'G_yA(G_x + A'G_yA)^{-1}A'G_yA}. \end{array} \tag{5.46}$$

We leave the proof for

$$\text{“}\mathbf{(G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^-A'G_y}$$

is the weighted pseudo-inverse or Moore–Penrose inverse $\mathbf{A}^{+}_{\mathbf{G_y}(\mathbf{G_x}+\mathbf{A'G_yA})}$" as an *exercise*. Similarly,

$$\mathbf{C}_1 = (\mathbf{I}_m - \mathbf{C}_2\mathbf{A'G_yA})(\mathbf{G_x} + \mathbf{A'G_yA})^{-1} \tag{5.47}$$

solves (2,2) where we again take advantage of the axiom of the **g**-inverse, namely

$$\mathbf{A'G_yAC_1A'G_yA} = 0 \Leftrightarrow \tag{5.48}$$

$$\begin{aligned}&\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA(G_x + A'G_yA)^{-1}A'G_yA}\\ &-\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^{-}}\\ &\qquad *\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA(G_x + A'G_yA)^{-1}A'G_yA} = 0 \Leftrightarrow\end{aligned}$$

$$\begin{aligned}&\Leftrightarrow \mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA}\\ &-\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA(A'G_yA(G_x + A'G_yA)^{-1}A'G_yA)^{-}}\\ &\qquad *\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA} = 0.\end{aligned}$$

For solving the system of modified normal equations, we have to compute

$$\mathbf{C_1A'G_y} = 0 \Leftrightarrow \mathbf{C}_1 = \mathbf{A'G_yA} = 0 \Leftrightarrow \mathbf{A'G_yAC_1A'G_yA} = 0,$$

a *zone identity* due to (2,2). In consequence,

$$\mathbf{x}_{\ell m} = \mathbf{C_2A'G_yy} \tag{5.49}$$

has been proven. The element (1,1) holds the key to solve for $\mathbf{C}_4$. As soon as we substitute $\mathbf{C}_1, \mathbf{C}_2 = \mathbf{C'}_3, \mathbf{C}_3 = \mathbf{C'}_2$ into (1,1) and multiply

left by	and	right by
$\mathbf{A'G_yA(G_x + A'G_yA)^{-1}}$		$\mathbf{(G_x + A'G_yA)^{-1}A'G_yA}$,

we receive

$$\begin{aligned}&2\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^{-}A'G_yA}\\ &*\mathbf{(G_x+A'G_yA)^{-1}A'G_yA+A'G_yA(G_x+A'G_yA)^{-1}A'G_yAC_4A'G_yA(G_x+A'G_yA)^{-1}}\\ &\qquad \mathbf{A'G_yA} = \mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA}.\end{aligned}$$

Finally, substitute

$$\mathbf{C}_4 = -\mathbf{[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^{-}} \tag{5.50}$$

to conclude

$$\begin{aligned}&\mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA[A'G_yA(G_x + A'G_yA)^{-1}A'G_yA]^{-}A'G_yA}\\ &*\mathbf{(G_x + A'G_yA)^{-1}A'G_yA} = \mathbf{A'G_yA(G_x + A'G_yA)^{-1}A'G_yA},\end{aligned}$$

namely the axiom of the **g**-inverse. Obviously, $\mathbf{C}_4$ is a symmetric matrix such that $\mathbf{C}_4 = \mathbf{C}'_4$.

Here ends our elaborate proof.

The results of the constructive proof of *Lemma 5.2* are collected in *Lemma 5.3*.

Lemma 5.3. ($\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution: MINOLESS):

$\mathbf{x}_{\ell m} = \hat{\mathbf{L}}\mathbf{y}$ is $\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution of (5.1) subject to

$$r := \mathrm{rk}\mathbf{A} = \mathrm{rk}(\mathbf{A}'\mathbf{G_y}\mathbf{A}) < \min\{n,\ m\}$$
$$\mathrm{rk}(\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A}) = m$$

if and only if

$$\hat{\mathbf{L}} = A^{+}_{\mathbf{G_y}\mathbf{G_x}} = (\mathbf{A}^{-}_{\ell m})_{\mathbf{G_y}\mathbf{G_x}} \tag{5.51}$$

$$\hat{\mathbf{L}} = (\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}[\mathbf{A}'\mathbf{G_y}\mathbf{A}(\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}]^{-}\mathbf{A}'\mathbf{G_y} \tag{5.52}$$

$$\mathbf{x}_{\ell m} = (\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}[\mathbf{A}'\mathbf{G_y}\mathbf{A}(\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}]^{-}\mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{5.53}$$

where $A^{+}_{\mathbf{G_y}\mathbf{G_x}} = \mathbf{A}^{1,2,3,4}_{\mathbf{G_y}\mathbf{G_x}}$ is the $\mathbf{G_y}$, $\mathbf{G_x}$-weighted *Moore–Penrose inverse*. If $\mathrm{rk}\mathbf{G_x} = m$, then

$$\hat{\mathbf{L}} = \mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-}\mathbf{A}'\mathbf{G_y} \tag{5.54}$$

$$\mathbf{x}_{\ell m} = \mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A}(\mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{G}_{\mathbf{x}}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-}\mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{5.55}$$

is an alternative unique solution of type MINOLESS.

Perhaps the lengthy formulae which represent $\mathbf{G_y}$, $\mathbf{G_x}$-MINOLESS in terms of a g-inverse motivate to implement

explicit representations for the analyst

of the $\mathbf{G_x}$-minimum norm (seminorm), $\mathbf{G_y}$-least squares solution, if *multiplication rank partitioning*, also known as *rank factorization*, or *additive rank partitioning* of the first order design matrix $\mathbf{A}$ is available. Here, we highlight both representations of $\mathbf{A}^{+} = \mathbf{A}^{-}_{\ell m}$.

Lemma 5.4. ($\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution: MINOLESS, rank factorization):

$\mathbf{x}_{\ell m} = \hat{\mathbf{L}}\mathbf{y}$ is $\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution (MINOLLES) of (5.1)

$$\{\mathbf{A}\mathbf{x} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n\times m},\ r := \mathrm{rk}\mathbf{A} = \mathrm{rk}(\mathbf{A}'\mathbf{G_y}\mathbf{A}) < \min\{n,\ m\}\},$$

if it is represented by *multiplicative rank partitioning* or *rank factorization*

$$\boxed{\mathbf{A} = \mathbf{DE},\ \mathbf{D} \in^{n-r},\ \mathbf{E} \in \mathbb{R}^{r \times m}} \tag{5.56}$$

as

$$\boxed{case\ (i):\ \mathbf{G_y} = \mathbf{I}_n,\ \mathbf{G_x} = \mathbf{I}_m}$$

$$\hat{\mathbf{L}} = \mathbf{A}^-_{\ell m} = \mathbf{E}'(\mathbf{EE}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \tag{5.57}$$

$$\hat{\mathbf{L}} = \mathbf{E}^-_R \mathbf{D}^-_L \left[\begin{array}{ll} \mathbf{E}^-_R = \mathbf{E}_m = \mathbf{E}'(\mathbf{EE}')^{-1} & \textit{right inverse} \\ \mathbf{D}^-_L = \mathbf{D}^-_\ell = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' & \textit{left inverse} \end{array} \right. \tag{5.58}$$

$$\boxed{\mathbf{x}_{\ell m} = \mathbf{A}^-_{\ell m}\mathbf{y} = \mathbf{A}^+\mathbf{y} = \mathbf{E}'(\mathbf{EE}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y}.} \tag{5.59}$$

The unknown vector $\mathbf{x}_{\ell m}$ has the minimum Euclidean length

$$\boxed{||\mathbf{x}_{\ell m}||^2 = \mathbf{x}'_{\ell m}\mathbf{x}_{\ell m} = \mathbf{y}'(\mathbf{A}^+)'\mathbf{A}^+\mathbf{y} = \mathbf{y}'(\mathbf{D}^-_\ell)'(\mathbf{EE}')^{-1}\mathbf{D}^-_\ell\mathbf{y}.} \tag{5.60}$$

$$\mathbf{y} = \mathbf{y}_{\ell m} + \mathbf{i}_{\ell m} \tag{5.61}$$

is an orthogonal decomposition of the observation vector $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n$ into

$$\mathbf{Ax}_{\ell m} = \mathbf{y}_{\ell m} \in \mathcal{R}(\mathbf{A})\ \ and\ \ \mathbf{y} - \mathbf{Ax}_{\ell m} = \mathbf{i}_{\ell m} \in \mathcal{R}(\mathbf{A})^\perp, \tag{5.62}$$

the vector of inconsistency.

$$\begin{array}{lcl} \mathbf{y}_{\ell m} = \mathbf{Ax}_{\ell m} = \mathbf{AA}^+\mathbf{y} & & \mathbf{i}_{\ell m} = \mathbf{y} - \mathbf{y}_{\ell m} = (\mathbf{I}_n - \mathbf{AA}^+)\mathbf{y} \\ = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} = \mathbf{DD}^-_\ell\mathbf{y} & and & = [\mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{y} = (\mathbf{I}_n - \mathbf{DD}^-_\ell)\mathbf{y} \end{array}$$

$\mathbf{AA}^+\mathbf{y} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{y} = \mathbf{DD}^-_\ell\mathbf{y} = \mathbf{y}_{\ell m}$ is the projection $\mathbf{P}_{\mathcal{R}(\mathbf{A})}$ and $(\mathbf{I}_n - \mathbf{AA}^+)\mathbf{y} = [\mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{y} = (\mathbf{I}_n - \mathbf{DD}^-_\ell)\mathbf{y}$ is the projection $\mathbf{P}_{\mathcal{R}(\mathbf{A})^\perp}$. $\mathbf{i}_{\ell m}$ *and* $\mathbf{y}_{\ell m}$ are *orthogonal* in the sense of $\langle \mathbf{i}_{\ell m} | \mathbf{y}_{\ell m} \rangle = 0$ or $(\mathbf{I}_n - \mathbf{AA}^+)'\mathbf{A} = [\mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']'\mathbf{D} = 0$. The "*goodness of fit*" of MINOLESS is

$$\boxed{\begin{array}{l} ||\mathbf{y} - \mathbf{Ax}_{\ell m}||^2 = ||\mathbf{i}_{\ell m}||^2 = \mathbf{y}'(\mathbf{I}_n - \mathbf{AA}^+)\mathbf{y} \\ = \mathbf{y}'[\mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{y} = \mathbf{y}'(\mathbf{I}_n - \mathbf{DD}^{-1}_\ell)\mathbf{y}. \end{array}} \tag{5.63}$$

$$case\ (ii):\ \mathbf{G_x}\ and\ \mathbf{G_y}\ positive\ \ definite$$

$$\hat{\mathbf{L}} = (\mathbf{A}^-_{m\ell})\ (\textit{weighted}) = \mathbf{G}^-_\mathbf{x}\mathbf{E}'(\mathbf{EG}^{-1}_\mathbf{x}\mathbf{E}')^{-1}(\mathbf{D}'\mathbf{G_y}\mathbf{D})^{-1}\mathbf{D}'\mathbf{G_y} \tag{5.64}$$

$$\hat{\mathbf{L}} = \mathbf{E}^-_R\ (\textit{weighted})\, \mathbf{D}^-_L\ (\textit{weighted}) \left[\begin{array}{l} \mathbf{E}^-_R = \mathbf{E}_m \text{ weighted } \textit{right inverse} \\ \mathbf{E}^-_L = \mathbf{E}_\ell \text{ weighted } \textit{left inverse} \end{array} \right. \tag{5.65}$$

$$\boxed{\begin{aligned}\mathbf{x}_{\ell m} &= (\mathbf{A}^{-}_{\ell m})_{\mathbf{G_y G_x}}\mathbf{y} \leq \mathbf{A}^{+}_{\mathbf{G_y G_x}}\mathbf{y}\\ &= \mathbf{G_x^{-1}E'(EG_x^{-1}E')^{-1}(D'G_yD)^{-1}DG_y y}.\end{aligned}} \tag{5.66}$$

The unknown vector $\mathbf{x}_{\ell m}$ has the weighted minimum Euclidean length

$$\begin{aligned}||\mathbf{x}_{\ell m}||^2_{\mathbf{G_x}} &= \mathbf{x}'_{\ell m}\mathbf{G_x x}_{\ell m} = \mathbf{y}'(\mathbf{A}^+)'\mathbf{G_x A^+ y}\\ &= \mathbf{y'G_yD(D'G_yD)^{-1}(EG_x^{-1}E')^{-1}EE'(EG_x^{-1}E')^{-1}(D'G_yD)^{-1}D'G_y y'}.\end{aligned} \tag{5.67}$$

$$\mathbf{y} = \mathbf{y}_{\ell m} + \mathbf{i}_{\ell m} \tag{5.68}$$

is an orthogonal decomposition of the observation vector $\mathbf{y} \in \mathbb{Y} = \mathbb{R}^n$ into

$$\mathbf{Ax}_{\ell m} = \mathbf{y}_{\ell m} \in \mathcal{R}(\mathbf{A}) \ \ and \ \ \mathbf{y} - \mathbf{Ax}_{\ell m} =: \mathbf{i}_{\ell m} \in \mathcal{R}(\mathbf{A})^{\perp} \tag{5.69}$$

of inconsistency.

$$\boxed{\begin{array}{lcl}\mathbf{y}_{\ell m} = \mathbf{AA}^{+}_{\mathbf{G_y G_x}}\mathbf{y} & and & \mathbf{i}_{\ell} = (\mathbf{I}_n - \mathbf{AA}^{+}_{\mathbf{G_y G_x}})\mathbf{y}\\ \mathbf{AA}^{+}_{\mathbf{G_y G_x}} = \mathbf{P}_{\mathcal{R}(\mathbf{A})} & & \mathbf{I}_n - \mathbf{AA}^{+}_{\mathbf{G_y G_x}} = \mathbf{P}_{\mathcal{R}(\mathbf{A})^{\perp}}\end{array}} \tag{5.70}$$

are $\mathbf{G_y}$-orthogonal

$$\langle \mathbf{i}_{\ell m} | \mathbf{y}_{\ell m}\rangle_{\mathbf{G_y}} = 0 \; or \; (\mathbf{I}_n - \mathbf{AA}^{+}(weighted))'\mathbf{G_y A} = 0. \tag{5.71}$$

The "*goodness of fit*" of $\mathbf{G_x}$, $\mathbf{G_y}$-MINOLESS is

$$\boxed{\begin{aligned}&||\mathbf{y} - \mathbf{Ax}_{\ell m}||^2_{\mathbf{G_y}} = ||\mathbf{i}_{\ell m}||^2_{\mathbf{G_y}}\\ &= \mathbf{y}'[\mathbf{I}_n - \mathbf{AA}^{+}_{\mathbf{G_y G_x}}]'\mathbf{G_y}[\mathbf{I}_n - \mathbf{AA}^{+}_{\mathbf{G_y G_x}}]\mathbf{y}\\ &= \mathbf{y}'[\mathbf{I}_n - \mathbf{D(D'G_yD)^{-1}D'G_y}]'\mathbf{G_y}[\mathbf{I}_n - \mathbf{D(D'G_yD)^{-1}D'G_y}]\mathbf{y}\\ &= \mathbf{y'[G_y - G_yD(D'G_yD)^{-1}D'G_y]y}.\end{aligned}} \tag{5.72}$$

While *Lemma 5.4* took advantage of rank factorization, *Lemma 5.5* will alternatively focus on additive rank partitioning.

Lemma 5.5. ($\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution: MINOLESS, additive rank partitioning)

$\mathbf{x}_{\ell m} = \hat{\mathbf{L}}\mathbf{y}$ is $\mathbf{G_x}$-minimum norm, $\mathbf{G_y}$-least squares solution (MINOLESS) of (5.1)

$$\{\mathbf{Ax} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m}, \ r := \text{rk}\mathbf{A} = \text{rk}(\mathbf{A'G_yA}) < \min\{n, m\}\},$$

if it is represented by *additive rank* partitioning

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \begin{matrix} \mathbf{A}_{11} \in \mathbb{R}^{r \times r}, \ \mathbf{A}_{12} \in \mathbb{R}^{r \times (m-r)} \\ \mathbf{A}_{21} \in \mathbb{R}^{(n-r) \times r}, \ \mathbf{A}_{22} \in \mathbb{R}^{(n-r) \times (m-r)} \end{matrix} \tag{5.73}$$

subject to the rank identity

$$\mathrm{rk}\mathbf{A} = \mathrm{rk}\mathbf{A}_{11} = r \tag{5.74}$$

as

$$\boxed{\text{case (i): } \mathbf{G_y} = \mathbf{I}_n, \ \mathbf{G_x} = \mathbf{I}_m}$$

$$\hat{\mathbf{L}} = \mathbf{A}^{-}_{\ell m} = \begin{bmatrix} \mathbf{N'}_{11} \\ \mathbf{N'}_{12} \end{bmatrix} (\mathbf{N}_{12}\mathbf{N'}_{12} + \mathbf{N}_{11}\mathbf{N'}_{11})^{-1} [\mathbf{A'}_{11}, \ \mathbf{A'}_{21}] \tag{5.75}$$

subject to

$$\mathbf{N}_{11} := \mathbf{A'}_{11}\mathbf{A}_{11} + \mathbf{A'}_{21}\mathbf{A}_{21}, \ \mathbf{N}_{12} := \mathbf{A'}_{11}\mathbf{A}_{12} + \mathbf{A'}_{21}\mathbf{A}_{22} \tag{5.76}$$

$$\mathbf{N}_{21} := \mathbf{A'}_{12}\mathbf{A}_{11} + \mathbf{A'}_{22}\mathbf{A}_{21}, \ \mathbf{N}_{22} := \mathbf{A'}_{12}\mathbf{A}_{12} + \mathbf{A'}_{22}\mathbf{A}_{22} \tag{5.77}$$

or

$$\boxed{\mathbf{x}_{\ell m} = \begin{bmatrix} \mathbf{N'}_{11} \\ \mathbf{N'}_{12} \end{bmatrix} (\mathbf{N}_{12}\mathbf{N'}_{12} + \mathbf{N}_{11}\mathbf{N'}_{11})^{-1} [\mathbf{A'}_{11}, \ \mathbf{A'}_{21}] \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}. \tag{5.78}$$

The unknown vector $\mathbf{x}_{\ell m}$ has the minimum Euclidean length

$$\begin{aligned} ||\mathbf{x}_{\ell m}||^2 &= \mathbf{x'}_{\ell m}\mathbf{x}_{\ell m} \\ &= [\mathbf{y'}_1, \ \mathbf{y'}_2] \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{21} \end{bmatrix} (\mathbf{N}_{12}\mathbf{N'}_{12} + \mathbf{N}_{11}\mathbf{N'}_{11})^{-1} [\mathbf{A'}_{11}, \ \mathbf{A'}_{21}] \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}. \end{aligned} \tag{5.79}$$

$$\mathbf{y} = \mathbf{y}_{\ell m} + \mathbf{i}_{\ell m}$$

is an orthogonal decomposition of the observation vector $\mathbf{y} \in \mathbb{Y} = \mathbb{R}$ into

$$\mathbf{A}\mathbf{x}_{\ell m} = \mathbf{y}_{\ell m} \in \mathcal{R}(\mathbf{A}) \ \ and \ \ \mathbf{y} - \mathbf{A}\mathbf{x}_{\ell m} =: \mathbf{i}_{\ell m} \in \mathcal{R}(\mathbf{A})^{\perp}, \tag{5.80}$$

the vector of inconsistency.

$$\boxed{\begin{aligned} \mathbf{y}_{\ell m} = \mathbf{A}\mathbf{x}_{\ell m} = \mathbf{A}\mathbf{A}^{-}_{\ell m}\mathbf{y} \quad and \quad \mathbf{i}_{\ell m} &= \mathbf{y} - \mathbf{A}\mathbf{x}_{\ell m} \\ &= (\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-}_{\ell m})\mathbf{y} \end{aligned}}$$

are projections onto R(A) and $\mathcal{R}(\mathbf{A})^{\perp}$, respectively. $\mathbf{i}_{\ell m}$ and $\mathbf{y}_{\ell m}$ are orthogonal in the sense of $\langle \mathbf{i}_{\ell m} | \mathbf{y}_{\ell m} \rangle = 0$ or $(\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-}_{\ell m})'\mathbf{A} = 0$. The "*goodness of fit*" of MINOLESS is

$$||\mathbf{y} - \mathbf{A}\mathbf{x}_{\ell m}||^2 = ||\mathbf{i}_{\ell m}||^2 = \mathbf{y}'(\mathbf{I}_n - \mathbf{A}\mathbf{A}^-_{\ell m})\mathbf{y}.$$

$\mathbf{I}_n - \mathbf{A}\mathbf{A}^-_{\ell m}$, $\mathrm{rk}(\mathbf{I}_n - \mathbf{A}\mathbf{A}^-_{\ell m}) = n - \mathrm{rk}\mathbf{A} = n - r$, is the rank deficient a posteriori weight matrix $(\mathbf{G_y})_{\ell m}$.

$$\boxed{\text{case (ii): } \mathbf{G_x} \text{ and } \mathbf{G_y} \text{ positive definite}}$$

$$\hat{\mathbf{L}} = (\mathbf{A}^-_{\ell m})_{\mathbf{G_y}\mathbf{G_x}}.$$

5-22 $(\mathbf{G_x}, \mathbf{G_y})$-*MINOS and Its Generalized Inverse*

A more formal version of the generalized inverse which is characteristic for $\mathbf{G_x}$-MINOS, $\mathbf{G_y}$-LESS or $(\mathbf{G_x}, \mathbf{G_y})$-MINOS is presented by

Lemma 5.6. (characterization of $\mathbf{G_x}$, $\mathbf{G_y}$-MINOS):

$$\mathrm{rk}(\mathbf{A}'\mathbf{G_y}\mathbf{A}) = \mathrm{rk}\mathbf{A} \sim \mathcal{R}(\mathbf{A}'\mathbf{G_y}) = \mathcal{R}(\mathbf{A}') \tag{5.81}$$

is assumed. $\mathbf{x}_{\ell m} = \mathbf{L}_\mathbf{y}$ is $(\mathbf{G_x}, \mathbf{G_y})$-MINOLESS of (5.1) for all $\mathbf{y} \in \mathbb{R}^n$ if and only if the matrix $\mathbf{L} \in \mathbb{R}^{m \times n}$ fulfils the four conditions

$$\mathbf{G_y}\mathbf{A}\mathbf{L}\mathbf{A} = \mathbf{G_y}\mathbf{A} \tag{5.82}$$

$$\mathbf{G_x}\mathbf{L}\mathbf{A}\mathbf{L} = \mathbf{G_x}\mathbf{L} \tag{5.83}$$

$$\mathbf{G_y}\mathbf{A}\mathbf{L} = (\mathbf{G_y}\mathbf{A}\mathbf{L})' \tag{5.84}$$

$$\mathbf{G_x}\mathbf{L}\mathbf{A} = (\mathbf{G_x}\mathbf{L}\mathbf{A})'. \tag{5.85}$$

In this case $\mathbf{G_x}\mathbf{x}_{\ell m} = \mathbf{G_x}\mathbf{L}_\mathbf{y}$ is always unique. $\mathbf{L}$, fulfilling the four conditions, is called the weighted MINOS inverse or *weighted Moore–Penrose inverse*.

Proof. The equivalence of (5.81) follows from

$$\mathcal{R}(\mathbf{A}'\mathbf{G_y}) = \mathcal{R}(\mathbf{A}'\mathbf{G_y}\mathbf{A}).$$

$$\boxed{(i)\ \mathbf{G_y}\mathbf{A}\mathbf{L}\mathbf{A} = \mathbf{G_y}\mathbf{A}\ \mathit{and}\ \mathbf{G_y}\mathbf{A}\mathbf{L} = (\mathbf{G_y}\mathbf{A}\mathbf{L})'.}$$

Condition (i) $\mathbf{G_y}\mathbf{A}\mathbf{L}\mathbf{A} = \mathbf{G_y}\mathbf{A}$ and (iii) $\mathbf{G_y}\mathbf{A}\mathbf{L} = (\mathbf{G_y}\mathbf{A}\mathbf{L})'$ are a consequence of $\mathbf{G_y}$-LESS.

$$||\mathbf{i}||^2_{\mathbf{G_y}} = ||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathbf{G_y}} = \min_{\mathbf{x}} \Rightarrow \mathbf{A}'\mathbf{G_y}\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{G_y}\mathbf{y}.$$

If $\mathbf{G_x}$ is *positive definite*, we can represent the four conditions (i)–(iv) of $\mathbf{L}$ by $(\mathbf{G_x}, \mathbf{G_y})$-MINOS inverse of $\mathbf{A}$ by two alternative solutions $\mathbf{L}_1$ and $\mathbf{L}_2$, namely

$$\begin{aligned}
\mathbf{AL}_1 = \mathbf{A(A'G_yA)^-A'G_yAL}_1 &= \mathbf{A(A'G_yA)^-A'L'}_1\mathbf{G_y}\\
&= \mathbf{A(A'G_yA)^-A'G_y}\\
= \mathbf{A(A'G_yA)^-A'L'}_2\mathbf{A'G_y} &= \mathbf{A(A'G_yA)^-A'G_yAL}_2 = \mathbf{AL}_2
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{L}_2\mathbf{A} = \mathbf{G_x^{-1}(A'L'}_2\mathbf{G_x)} &= \mathbf{G_x^{-1}(A'L'}_2\mathbf{A'L'}_2\mathbf{G_x)} = \mathbf{G_x^{-1}(A'L'}_1\mathbf{A'L'}_2\mathbf{G_x)}\\
&= \mathbf{G_x^{-1}(A'L'}_1\mathbf{G_xL}_2\mathbf{A)} = \mathbf{G_x^{-1}(G_xL}_1\mathbf{AL}_2\mathbf{A)}\\
&= \mathbf{G_x^{-1}(G_xL}_1\mathbf{AL}_1\mathbf{A)} = \mathbf{L}_1\mathbf{A},\\
\mathbf{L}_1 = \mathbf{G_x^{-1}(G_xL}_1\mathbf{AL}_1) &= \mathbf{G_x^{-1}(G_xL}_2\mathbf{AL}_2) = \mathbf{L}_2
\end{aligned}$$

concludes our proof.

The inequality

$$\boxed{\begin{aligned}||\mathbf{x}_{\ell m}||^2_{\mathbf{G_x}} = ||\mathbf{Ly}||^2_{\mathbf{G_x}} \leq ||\mathbf{Ly}||^2_{\mathbf{G_y}} + 2\mathbf{y'L'G_x(I}_n - \mathbf{LA)z}\\ + ||(\mathbf{I}_m - \mathbf{LA)z}||^2_{\mathbf{G_x}} \quad \forall \mathbf{y} \in \mathbb{R}^n\end{aligned}} \tag{5.86}$$

is fulfilled if and only if the "*condition of* $\mathbf{G_x}$*-orthogonality*"

$$\boxed{\mathbf{L'G_x(I}_m - \mathbf{LA)} = 0} \tag{5.87}$$

applies. An equivalence is

$$\mathbf{L'G_x} = \mathbf{L'G_xLA} \qquad or \qquad \mathbf{L'G_xL} = \mathbf{L'G_xLAL},$$

which is produced by left multiplying with $\mathbf{L}$. The left side of this equation is a symmetric matrix. *Consequently*, the right side has to be a *symmetric matrix, too*.

$$\boxed{\mathbf{G_xLA} = (\mathbf{G_xLA})'.}$$

Such an identity agrees to *condition* (iv). As soon as we substitute in the "*condition of* $\mathbf{G_x}$*-orthogonality*" we are led to

$$\mathbf{L'G_x} = \mathbf{L'G_xLA} \Rightarrow \boxed{\mathbf{G_xL} = (\mathbf{G_xLA})'\mathbf{L} = \mathbf{G_xLAL}},$$

a result which agrees to *condition* (ii).

? How to prove uniqueness of $\mathbf{A}^{1,2,3,4} = \mathbf{A}_{\ell m} = \mathbf{A}^+$?

Uniqueness of $\mathbf{G_xx}_{\ell m}$ can be taken from *Lemma 1.4* (characterization of $\mathbf{G_x}$-MINOS).

Substitute $\mathbf{x}_\ell = \mathbf{Ly}$ and multiply the left side by $\mathbf{L}$.

$$\mathbf{A'G_yALy} = \mathbf{A'G_yy} \Leftrightarrow \mathbf{A'G_yAL} = \mathbf{A'G_y}$$

$$\mathbf{L'A'G_yAL} = \mathbf{L'A'G_y} \Leftrightarrow \boxed{\mathbf{G_yAL} = (\mathbf{G_yAL})' = \mathbf{L'A'G_y}.}$$

The left side of the equation $\mathbf{L'A'G_yAL} = \mathbf{L'A'G_y}$ is a *symmetric matrix*. Consequently the right side has to be symmetric, *too*. Indeed we have proven condition (iii) $(\mathbf{G_yAL})' = \mathbf{G_yAL}$. Let us transplant the symmetric condition (iii) into the original normal equations in order to benefit from

$$\mathbf{A'G_yAL} = \mathbf{A'G_y} \ or \boxed{\mathbf{G_yA} = \mathbf{L'A'G_yA} = (\mathbf{G_yAL})'\mathbf{A} = \mathbf{G_yALA}.}$$

Indeed, we have succeeded to have proven condition (i), in condition

$$\boxed{(ii)\, \mathbf{G_xLAL} = \mathbf{G_xL} \ \ and \ \ \mathbf{G_xLA} = (\mathbf{G_xLA})'.}$$

Condition (ii) $\mathbf{G_yLAL} = \mathbf{G_xL}$ *and* (iv) $\mathbf{G_xLA} = (\mathbf{G_xLA})'$ are a consequence of $\mathbf{G_x}$-MINOS.

The general solution of the normal equations $\mathbf{A'G_yAx_\ell} = \mathbf{A'G_yy}$ is

$$\boxed{\mathbf{x}_\ell = \mathbf{x}_{\ell m} + [\mathbf{I}_m - (\mathbf{A'G_yA})^-(\mathbf{A'G_yA})]\mathbf{z}} \tag{5.88}$$

for an arbitrary vector $\mathbf{z} \in \mathbb{R}^m$. $\mathbf{A'G_yALA} = \mathbf{A'G_yA}$ implies

$$\begin{aligned}\mathbf{x}_\ell &= \mathbf{x}_{\ell m} + [\mathbf{I}_m - (\mathbf{A'G_yA})^-\mathbf{A'G_yALA}]\mathbf{z}\\ &= \mathbf{x}_{\ell m} + [\mathbf{I}_m - \mathbf{LA}]\mathbf{z}.\end{aligned}$$

Note 1: The following conditions are equivalent:

$$\boxed{(1st)\left[\begin{array}{c} (1)\ \mathbf{AA^-A} = \mathbf{A}\\ (2)\ \mathbf{A^-AA^-} = \mathbf{A}\\ (3)\ (\mathbf{AA^-})'\mathbf{G_y} = \mathbf{G_yAA^-}\\ (4)\ (\mathbf{A^-A})'\mathbf{G_x} = \mathbf{G_xA^-A}\end{array}\right.}$$

$$\boxed{(2nd)\left[\begin{array}{c} \mathbf{A^\#G_yAA^-} = \mathbf{A'G_y}\\ (\mathbf{A^-})'\mathbf{G_xA^-A} = (\mathbf{A^-})'\mathbf{G_x}\end{array}\right.}$$

"if $\mathbf{G_x}$ and $\mathbf{G_y}$ are positive definite matrices, then

$$\begin{gathered}\mathbf{A^\#G_y} = \mathbf{G_xA^\#}\\ or\\ \mathbf{A^\#} = \mathbf{G_x^{-1}A'G_y}\end{gathered}$$

are representations for the *adjoint matrix*" "if $\mathbf{G_x}$ and $\mathbf{G_y}$ are positive definite matrices, then

$$\boxed{\begin{array}{c}(\mathbf{A}'\mathbf{G_y A})\mathbf{AA}^- = \mathbf{A}'\mathbf{G_y}\\ (\mathbf{A}^-)'\mathbf{G_x A}^-\mathbf{A} = (\mathbf{A}^-)'\mathbf{G}_\mathbf{x}''\end{array}}$$

$$\boxed{(3rd)\left[\begin{array}{l}\mathbf{AA}^- = \mathbf{P}_{\mathcal{R}(\mathbf{A})}\\ \mathbf{A}^-\mathbf{A} = \mathbf{P}_{\mathcal{R}(\mathbf{A}^-)}\,.\end{array}\right.}$$

The concept of a generalized inverse of an arbitrary matrix is originally due to *E.H. Moore* (1920) who used the 3rd definition. *R. Penrose* (1955), unaware of *E.H. Moores* work, defined a generalized inverse by the *1st definition* to $\mathbf{G_x} = \mathbf{I}_m$, $\mathbf{G_y} = \mathbf{I}_n$ of unit matrices which is the same as the *Moore inverse*. *Y. Tseng* (1949, a, b, 1956) defined a *generalized inverse* of a linear operator *between function spaces* by means of

$$\boxed{\mathbf{AA}^- = \mathbf{P}_{\overline{\mathcal{R}(\mathbf{A})}},\ \mathbf{A}^-\mathbf{A} = \mathbf{P}_{\overline{\mathcal{R}(\mathbf{A}^-)}},}$$

where $\overline{\mathcal{R}(\mathbf{A})}$, $\overline{\mathcal{R}(\mathbf{A}^-)}$, respectively are the *closure* of $\mathcal{R}(\mathbf{A})$, $\mathcal{R}(\mathbf{A}^-)$, respectively. The Tseng inverse has been reviewed by *B. Schaffrin, E. Heidenreich and E. Grafarend* (1977). *A. Bjerhammar* (1951, 1957, 1956) initiated the notion of the least-squares generalized inverse. *C.R. Rao* (1967) presented the first classification of **g**-*inverses*.

Note 2: Let $||\mathbf{y}||_{\mathbf{G_y}} = (\mathbf{y}'\mathbf{G_y y})^{1/2}$ and $||\mathbf{x}||_{\mathbf{G_x}} = (\mathbf{x}'\mathbf{G_x x})^{1/2}$, where $\mathbf{G_y}$ and $\mathbf{G_x}$ are *positive semidefinite*. If there exists a matrix $\mathbf{A}^-$ which satisfies the definitions of *Note 1*, then it is *necessary, but not sufficient* that

$$(1)\ \mathbf{G_y AA^-A} = \mathbf{G_y A}$$

$$(2)\ \mathbf{G_x A^-AA^-} = \mathbf{G_x A^-}$$

$$(3)\ (\mathbf{AA}^-)'\mathbf{G_y} = \mathbf{G_y AA}^-$$

$$(4)\ (\mathbf{A}^-\mathbf{A})'\mathbf{G_x} = \mathbf{G_x A^-A}.$$

Note 3: A **g**-inverse which satisfies the conditions of *Note 1* is denoted by $\mathbf{A}^+_{\mathbf{G_y G_x}}$ and referred to as $\mathbf{G_y}$, $\mathbf{G_x}$-MINOLESS **g**-inverse of **A**.

$\mathbf{A}^+_{\mathbf{G_y G_x}}$ is unique if $\mathbf{G_x}$ is positive definite. When both $\mathbf{G_x}$ and $\mathbf{G_y}$ are general positive semi definite matrices, $\mathbf{A}^+_{\mathbf{G_y G_x}}$ *may not be unique*. If $|\mathbf{G_x} + \mathbf{A}'\mathbf{G_y A}| \neq 0$ holds, $\mathbf{A}^+_{\mathbf{G_y G_x}}$ is unique.

Note 4: If the matrices of the metric are *positive definite*, $|\mathbf{G_x}| \neq 0$, $\left|\mathbf{G_y}\right| \neq 0$, then

(i) $(\mathbf{A}^+_{\mathbf{G_y G_x}})^+_{\mathbf{G_x G_y}} = \mathbf{A}$,

(ii) $(\mathbf{A}^+_{\mathbf{G_y G_x}})^\# = (\mathbf{A}')^+_{\mathbf{G_x^{-1} G_y^{-1}}}$.

5-23 *Eigenvalue Decomposition of* $(\mathbf{G_x}, \mathbf{G_y})$*-MINOLESS*

For the system analysis of an inverse problem the *eigenspace analysis* and *eigenspace synthesis* of $\mathbf{x}_{\ell m}(\mathbf{G_x}, \mathbf{G_y})$-MINOLESS of x is very useful and give some peculiar insight into a dynamical system. Accordingly we are confronted with the problem to develop "*canonical MINOLESS*", also called the *eigenvalue decomposition* of $(\mathbf{G_x}, \mathbf{G_y})$-MINOLESS.

First we refer again to the canonical representation of the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$ introduced to you in the first chapter, *Box 1.6* and *Box 1.9*. But we add here by means of *Box 5.8* the forward and backward transformation of the general bases versus the orthogonal bases spanning the parameter space X as well as the observation space $\mathbb{Y}$. In addition, we refer to *Definition 1.5* and *Lemma 1.6* where the adjoint operator $\mathbf{A}^*$ has been introduced and represented.

Box 5.8. General bases versus orthogonal bases spanning the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$.

"*left*"	"*right*"	
"*parameter space*"	"*observation space*"	
"*general left base*"	"*general right base*"	
$span\{a_1, \ldots, a_m\} = \mathbb{X}$	$\mathbb{Y} = span\{b_1, \ldots, b_n\}$	
: *matrix of the metric:*	: *matrix of the metric* :	
$\mathbf{aa}' = \mathbf{G_x}$	$\mathbf{bb}' = \mathbf{G_y}$	(5.89)
"*orthogonal left base*"	"*orthogonal right bsase*"	
$span\{\mathbf{e}_1^{\mathbf{x}}, \ldots, \mathbf{e}_m^{\mathbf{x}}\} = \mathbb{X}$	$\mathbb{Y} = span\{\mathbf{e}_1^{\mathbf{y}}, \ldots, \mathbf{e}_n^{\mathbf{y}}\}$	
: *matrix of the metric:*	:*matrix of the metric* :	
$\mathbf{e_x}\mathbf{e'}_{\mathbf{x}} = \mathbf{I}_m$	$\mathbf{e_y}\mathbf{e'}_{\mathbf{y}} = \mathbf{I}_n$	(5.90)
"*base transformation*"	"*base transformation*"	
$\mathbf{a} = \mathbf{\Lambda}_{\mathbf{x}}^{1/2}\mathbf{V}\mathbf{e_x}$	$\mathbf{b} = \mathbf{\Lambda}_{\mathbf{y}}^{1/2}\mathbf{U}\mathbf{e_y}$	(5.91)
versus	*versus*	
$\mathbf{e_x} = \mathbf{V}'\mathbf{\Lambda}_{\mathbf{x}}^{-1/2}\mathbf{a}$	$\mathbf{e_y} = \mathbf{U}'\mathbf{\Lambda}_{\mathbf{y}}^{-1/2}\mathbf{b}$	(5.92)
$span\{\mathbf{e}_1^{\mathbf{x}}, \ldots, \mathbf{e}_m^{\mathbf{x}}\} = \mathbb{X}$	$\mathbb{Y} = span\{\mathbf{e}_1^{\mathbf{y}}, \ldots, \mathbf{e}_n^{\mathbf{y}}\}.$	

Second, we are solving the general system of linear equations

$$\{\mathbf{y} = \mathbf{Ax} | \mathbf{A} \in \mathbb{R}^{n \times m},\ \text{rk}\mathbf{A} < \min\{n, m\}\}$$

by introducing

- The eigenspace of the rank deficient, rectangular matrix of rank $r := \text{rk}\mathbf{A} < \min\{n, m\} : \mathbf{A} \mapsto \mathbf{A}^*$
- The left and right canonical coordinates: $\mathbf{x} \mapsto \mathbf{x}^*$, $\mathbf{y} \mapsto \mathbf{y}^*$

as supported by *Box 5.9*. The transformations $\mathbf{x} \mapsto \mathbf{x}^*$ (5.92), $\mathbf{y} \mapsto \mathbf{y}^*$(5.93, left) *from* the original coordinates $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$ to the canonical coordinates $(\mathbf{x}_1^*, \ldots, \mathbf{x}_m^*)$, the *left star coordinates*, as well as from the original coordinates $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ to the canonical coordinates $(\mathbf{y}_1^*, \ldots, \mathbf{y}_n^*)$, the *right star coordinates*, are *polar decompositions*: a rotation $\mathbf{U}, \mathbf{V}$ is followed by a *general stretch* $\{\mathbf{G}_\mathbf{y}^{1/2}, \mathbf{G}_\mathbf{x}^{1/2}\}$. Those root matrices are generated by *product decompositions* of type $\mathbf{G}_\mathbf{y} = (\mathbf{G}_\mathbf{y}^{1/2})'\mathbf{G}_\mathbf{y}^{1/2}$ as well as $\mathbf{G}_\mathbf{x} = (\mathbf{G}_\mathbf{x}^{1/2})'\mathbf{G}_\mathbf{x}^{1/2}$. Let us substitute the *inverse transformations* (5.93, right) $\mathbf{x}^* \mapsto \mathbf{x} = \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}\mathbf{x}^*$ and (5.94, left) $\mathbf{y}^* \mapsto \mathbf{y} = \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}\mathbf{y}^*$ into the system of linear equations (5.1), (5.94, right) $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i}$, $\text{rk}\mathbf{A} < \min\{n, m\}$ or its *dual* (5.95, left) $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$. Such an operation leads us to (5.95, right) $\mathbf{y}^* = f(\mathbf{x}^*)$ as well as (5.96, left) $\mathbf{y} = f(\mathbf{x})$. Subject to the orthonormality condition (5.96, right) $\mathbf{U}'\mathbf{U} = \mathbf{I}_n$ and (5.97, left) $\mathbf{V}'\mathbf{V} = \mathbf{I}_m$ we have generated the left-right *eigenspace analysis* (5.97, right)

$$\boxed{\mathbf{\Lambda}^* = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3 \end{bmatrix}}$$

subject to the *rank partitioning* of the matrices $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$ and $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$. Alternatively, the left-right eigenspace synthesis (5.104, left)

$$\boxed{\mathbf{A} = \mathbf{G}_\mathbf{y}^{-1/2}[\mathbf{U}_1, \mathbf{U}_2] \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3 \end{bmatrix} \begin{bmatrix} \mathbf{V'}_1 \\ \mathbf{V'}_2 \end{bmatrix} \mathbf{G}_\mathbf{x}^{1/2}}$$

is based upon the left matrix (5.99) $\mathbf{L} := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}$ decomposed into (5.111) $\mathbf{L}_1 := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}_1$ and $\mathbf{L}_2 := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}_2$ and the *right matrix* (5.94, left) $\mathbf{R} := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}$ decomposed into $\mathbf{R}_1 := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}_1$ and $\mathbf{R}_2 := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}_2$. Indeed the left matrix $\mathbf{L}$ by means of (5.101, right) $\mathbf{L}\mathbf{L}' = \mathbf{G}_\mathbf{y}^{-1}$ reconstructs the inverse matrix of the metric of the *observation space* $\mathbb{Y}$. Similarly, the right matrix $\mathbf{R}$ by means of (5.102, left) $\mathbf{R}\mathbf{R}' = \mathbf{G}_\mathbf{x}^{-1}$ generates the inverse matrix of the metric of the *parameter space* $\mathbf{X}$. In terms of "$\mathbf{L}, \mathbf{R}$" we have summarized the eigenvalue decompositions (5.103, right)–(5.106, left).

Such an eigenvalue decomposition helps us to *canonically invert* $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$ by means of (5.106, right), namely the "*full rank partitioning*" of the system of canonical linear equations $\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^*$. The *observation vector* $\mathbf{y}^* \in \mathbb{R}^n$ is decomposed into $\mathbf{y}_1^* \in \mathbb{R}^{r\times 1}$ and $\mathbf{y}_2^* \in \mathbb{R}^{(n-r)\times 1}$ while the vector $\mathbf{x}^* \in \mathbb{R}^m$ of unknown parameters is decomposed into $\mathbf{x}_1^* \in \mathbb{R}^{r\times 1}$ and $\mathbf{x}_2^* \in \mathbb{R}^{(m-r)\times 1}$.

$$\boxed{(\mathbf{x}_1^*)_{\ell m} = \mathbf{\Lambda}^{-1}\mathbf{y}_1^*}$$

is *canonical MINOLESS* leaving $\mathbf{y}_2^*$ "*unrecognized*" and $\mathbf{x}_2^* = 0$ as a "*fixed datum*".

Box 5.9. (Canonical representation, the general case: overdetermined and underdetermined system without full rank):

$$\boxed{\text{"parameter space } \mathbb{X}\text{"}} \quad \textit{versus} \quad \boxed{\text{"observation space"}}$$

$$\mathbf{x}^* = \mathbf{V}'\mathbf{G}_\mathbf{x}^{1/2}\mathbf{x} \qquad \mathbf{y}^* = \mathbf{U}'\mathbf{G}_\mathbf{y}^{1/2}\mathbf{y} \tag{5.93}$$

$$\textit{and} \qquad \textit{and}$$

$$\mathbf{x} = \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}\mathbf{x}^*\mathbf{y} = \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}\mathbf{y}^* \tag{5.94}$$

$$\boxed{\text{"overdetermined and unterdetermined system without full rank''}}$$

$$\{\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} | \mathbf{A} \in \mathbb{R}^{n\times m},\ \text{rk}\mathbf{A} < \min\{n, m\}\}$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} \quad \textit{versus} \quad \mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^* \tag{5.95}$$

$$\begin{aligned}\mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}\mathbf{y}^* &= \mathbf{A}\mathbf{G}_\mathbf{x}^{-1/2}\mathbf{x}^* \\ &+\mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}\mathbf{i}^*\end{aligned} \qquad \begin{aligned}\mathbf{U}'\mathbf{G}_\mathbf{y}^{1/2}\mathbf{y} &= \mathbf{A}^*\mathbf{V}'\mathbf{G}_\mathbf{x}^{-1/2}\mathbf{x} \\ &+\mathbf{U}\mathbf{G}_\mathbf{y}^{1/2}\mathbf{i}\end{aligned}$$

$$\mathbf{y}^* = (\mathbf{U}'\mathbf{G}_\mathbf{y}^{1/2}\mathbf{A}\mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V})\mathbf{x}^* \quad \textit{versus} \quad \mathbf{y} = (\mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}\mathbf{A}^*\mathbf{V}'\mathbf{G}_\mathbf{x}^{-1/2})\mathbf{x} + \mathbf{i} \tag{5.96}$$

$$\boxed{\textit{subject to}} \qquad \boxed{\textit{subject to}}$$

$$\mathbf{U}'\mathbf{U} = \mathbf{U}\mathbf{U}' = \mathbf{I}_n \quad \textit{versus} \quad \mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}_m \tag{5.97}$$

$$\boxed{\text{"left and right eigenspace synthesis"}}$$

$$\begin{aligned}\mathbf{A}^* &= \begin{bmatrix}\mathbf{U}'_1 \\ \mathbf{U}'_2\end{bmatrix}\mathbf{G}_\mathbf{y}^{1/2}\mathbf{A}\mathbf{G}_\mathbf{x}^{-1/2}[\mathbf{V}_1, \mathbf{V}_2] \\ &= \begin{bmatrix}\mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3\end{bmatrix}\end{aligned} \tag{5.98}$$

$$\mathbf{A} = \mathbf{G}_\mathbf{y}^{-1/2}[\mathbf{U}_1, \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3\end{bmatrix}\begin{bmatrix}\mathbf{V}'_1 \\ \mathbf{V}'_2\end{bmatrix}\mathbf{G}_\mathbf{x}^{1/2} \tag{5.99}$$

"*dimension identities*"

$$\mathbf{\Lambda} \in \mathbb{R}^{r\times r},\ \mathbf{O}_1 \in \mathbb{R}^{r\times(m-r)},\ \mathbf{U}_1 \in \mathbb{R}^{n\times r},\ \mathbf{V}_1 \in \mathbb{R}^{m\times r}$$

$$\mathbf{O}_2 \in \mathbb{R}^{(n-r)\times r},\ \mathbf{O}_3 \in \mathbb{R}^{(n-r)\times(m-r)},\ \mathbf{U}_2 \in \mathbb{R}^{n\times(n-r)},\ \mathbf{V}_2 \in \mathbb{R}^{m\times(m-r)}$$

$$\boxed{\text{"left eigenspace"}} \qquad \boxed{\text{"right eigenspace"}}$$

$$\mathbf{L} := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U} \Rightarrow \mathbf{L}^{-1} = \mathbf{U}'\mathbf{G}_\mathbf{y}^{1/2} \qquad \mathbf{R} := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V} \Rightarrow \mathbf{R}^{-1} = \mathbf{V}'\mathbf{G}_\mathbf{x}^{1/2} \tag{5.100}$$

$$\mathbf{L}_1 := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}_1,\ \mathbf{L}_2 := \mathbf{G}_\mathbf{y}^{-1/2}\mathbf{U}_2 \qquad \mathbf{R}_1 := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}_1,\ \mathbf{R}_2 := \mathbf{G}_\mathbf{x}^{-1/2}\mathbf{V}_2 \tag{5.101}$$

$$\mathbf{L}\mathbf{L}' = \mathbf{G}_\mathbf{y}^{-1} \Rightarrow (\mathbf{L}^{-1})'\mathbf{L}^{-1} = \mathbf{G}_\mathbf{y} \qquad \mathbf{R}\mathbf{R}' = \mathbf{G}_\mathbf{x}^{-1} \Rightarrow (\mathbf{R}^{-1})'\mathbf{R}^{-1} = \mathbf{G}_\mathbf{x} \tag{5.102}$$

$$\mathbf{L}^{-1} = \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix} \mathbf{G}_\mathbf{y}^{1/2} =: \begin{bmatrix} \mathbf{L}_1^- \\ \mathbf{L}_2^- \end{bmatrix} \qquad \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \mathbf{G}_\mathbf{x}^{1/2} =: \begin{bmatrix} \mathbf{R}_1^- \\ \mathbf{R}_2^- \end{bmatrix} \tag{5.103}$$

$$\mathbf{A} = \mathbf{L}\mathbf{A}^*\mathbf{R}^{-1} \quad \textit{versus} \quad \mathbf{A}^* = \mathbf{L}^{-1}\mathbf{A}\mathbf{R} \tag{5.104}$$

$$\mathbf{A} = [\mathbf{L}_1, \mathbf{L}_2]\mathbf{A}^* \begin{bmatrix} \mathbf{R}_1^- \\ \mathbf{R}_2^- \end{bmatrix} \quad \textit{versus} \quad \begin{aligned} \mathbf{A}^* &= \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{L}_1^- \\ \mathbf{L}_2^- \end{bmatrix} \mathbf{A}[\mathbf{R}_1, \mathbf{R}_2] \end{aligned} \tag{5.105}$$

$$\left.\begin{matrix} \mathbf{A}\mathbf{A}^\#\mathbf{L}_1 = \mathbf{L}_1\mathbf{\Lambda}^2 \\ \mathbf{A}\mathbf{A}^\#\mathbf{L}_2 = 0 \end{matrix}\right] \quad \textit{versus} \quad \left[\begin{matrix} \mathbf{A}^\#\mathbf{A}\mathbf{R}_1 = \mathbf{R}_1\mathbf{\Lambda}^2 \\ \mathbf{A}^\#\mathbf{A}\mathbf{R}_2 = 0 \end{matrix}\right. \tag{5.106}$$

"*inconsistent system of linear equations without full rank*"

$$\mathbf{y}^* = \mathbf{A}^*\mathbf{x}^* + \mathbf{i}^* = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \end{bmatrix} + \begin{bmatrix} \mathbf{i}_1^* \\ \mathbf{i}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} \tag{5.107}$$

$$\mathbf{y}_1^* \in \mathbb{R}^{r\times 1},\ \mathbf{y}_2^* \in \mathbb{R}^{(n-r)\times 1},\ \mathbf{i}_1^* \in \mathbb{R}^{r\times 1},\ \mathbf{i}_2^* \in \mathbb{R}^{(n-r)\times 1}$$
$$\mathbf{x}_1^* \in \mathbb{R}^{r\times 1},\ \mathbf{x}_2^* \in \mathbb{R}^{(m-r)\times 1}$$

"if $(\mathbf{x}^*, \mathbf{i}^*)$ is MINOLESS, then $\mathbf{x}_2^* = 0$, $\mathbf{i}^* = 0$:
$(\mathbf{x}_1^*)_{\ell m} = \mathbf{\Lambda}^{-1}\mathbf{y}_1^*$."

Consult the commutative diagram of Fig. 5.6 for a shortened summary of the newly introduced transformation of coordinates, both of the parameter space $\mathbb{X}$ as well as the observation space $\mathbb{Y}$.

Third, we prepare ourselves for MINOLESS of the general system of linear equations

$$\{\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{i} | \mathbf{A} \in^{n\times m},\ \text{rk}\mathbf{A} < \min\{n, m\}$$
$$||\mathbf{i}||^2_{\mathbf{G}_\mathbf{y}} = \min,\ \textit{subject to}\ ||\mathbf{x}||^2_{\mathbf{G}_\mathbf{x}} = \min\}$$

by introducing *Lemmas 5.4 and 5.5*, namely the *eigenvalue-eigencolumn equations* of the matrices $\mathbf{A}^\#\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\#$, respectively, as well as *Lemma 5.6*, our basic result of "*canonical MINOLESS*", subsequently completed by proofs. Throughout we refer to the adjoint operator which has been introduced by *Definition 1.5* and *Lemma 1.6*.

Fig. 5.6 Commutative diagram of coordinate transformations

$$\begin{tikzcd}
\mathbb{X} \ni \mathbf{x} \arrow[r, "\mathbf{A}"] \arrow[d, "\mathbf{V}'\mathbf{G}_{\mathbf{x}}^{1/2}"'] & \mathcal{R}(\mathbf{A}) \subset \mathbb{Y} \arrow[d, "\mathbf{U}'\mathbf{G}_{\mathbf{y}}^{1/2}"] \\
\mathbb{X} \ni \mathbf{x}^* \arrow[r] & \mathbf{y}^* \in \mathcal{R}(\mathbf{A}) \subset \mathbb{Y}
\end{tikzcd}$$

Lemma 5.7. (eigenspace analysis versus eigenspace synthesis $\mathbf{A} \in \mathbb{R}^{n\times m}$, $r := \text{rk}\mathbf{A} < \min\{n, m\}$):

The pair of matrices $\mathbf{L}, \mathbf{R}$ for the *eigenspace analysis* and the *eigenspace synthesis* of the rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ of rank $r := \text{rk}\mathbf{A} < \min\{n, m\}$, namely

$$\mathbf{A}^* = \mathbf{L}^{-1}\mathbf{A}\mathbf{R} \qquad versus \qquad \mathbf{A} = \mathbf{L}\mathbf{A}^*\mathbf{R}^{-1}$$

$$or \qquad\qquad or$$

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O}_1 \\ \mathbf{O}_2 & \mathbf{O}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1^- \\ \mathbf{L}_2^- \end{bmatrix} \mathbf{A}[\mathbf{R}_1, \mathbf{R}_2] \qquad versus \qquad \mathbf{A} = [\mathbf{L}_1, \mathbf{L}_2]\mathbf{A}^* \begin{bmatrix} \mathbf{R}_1^- \\ \mathbf{R}_2^- \end{bmatrix}$$

are determined by the *eigenvalue-eigencolumn equations* (eigenspace equations)

$$\begin{aligned} \mathbf{A}^\#\mathbf{A}\mathbf{R}_1 &= \mathbf{R}_1\mathbf{\Lambda}^2 \\ \mathbf{A}^\#\mathbf{A}\mathbf{R}_2 &= 0 \end{aligned} \qquad versus \qquad \begin{aligned} \mathbf{A}\mathbf{A}^\#\mathbf{L}_1 &= \mathbf{L}_1\mathbf{\Lambda}^2 \\ \mathbf{A}\mathbf{A}^\#\mathbf{L}_2 &= 0 \end{aligned}$$

$$subject\ to$$

$$\begin{bmatrix} \lambda_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r^2 \end{bmatrix}, \ \mathbf{\Lambda} = \text{Diag}(+\sqrt{\lambda_1^2}, \ldots, +\sqrt{\lambda_r^2}).$$

5-24 Notes

The algebra of *eigensystems* is treated in varying degrees by most books on linear algebra, in particular tensor algebra. Special mention should be made of

R. Bellman's classic "*Introduction to matrix analysis*" (1970) and *Horn's and Johnson's* two books (1985, 1991) on introductory and advanced matrix analysis.

More or less systematic treatments of *eigensystem* are found in books on matrix computations. The classics of the field are *Householder's* "*Theory of matrices in numerical analysis*" (1964) and *Wilkinson's* "*The algebraic eigenvalue problem*" (1965). *G. Golub's and Van Loan's* "*Matrix computations*" (1996) is the currently definite survey of the field. *Trefethen's and Bau's* "*Numerical linear algebra*" (1997) is a high-level, insightful treatment with a welcomed stress on geometry. *G.W. Stewart's "Matrix algorithm: eigensystems"* (2001) is becoming a classic as well.

The term "*eigenvalue*" derives from the German *Eigenwert*, which was introduced by *D. Hilbert* (1904) to denote for integral equations the reciprocal of the matrix eigenvalue. At some point *Hilbert's Eigenwert* inverted themselves and became attached to matrices. *Eigenvalues* have been called many things in their day. The "*characteristic value*" is a reasonable translation of *Eigenwert*. However, "*characteristic*" has an inconveniently large number of syllables and survives only in the terms "*characteristic equation*" and "*characteristic polynomial*". For symmetric matrices the characteristic equation and its equivalent are also called the *secular equation* owing to its connection with the secular perturbations in the orbits of planets. Other terms are "*latent value*" and "*proper value*" from the French "*valeur propre*".

Indeed the day when purists and pedants could legitimately object to "*eigenvalue*" as a hybrid of German and English has long since passed. The German "eigen" has become thoroughly materialized English prefix meaning having to do with *eigenvalues* and *eigenvectors*. Thus we can use "*eigensystem*", "*eigenspace*" or "*eigenexpansion*" without fear of being misunderstood. The term "*eigenpair*" used to denote an eigenvalue and eigenvector is a recent innovation.

5-3 The Hybrid Approximation Solution: α-HAPS and *Tykhonov–Phillips* Regularization

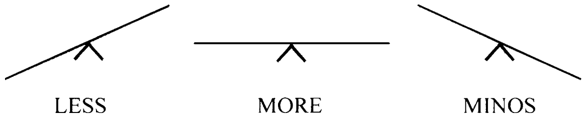
$\mathbf{G_x}, \mathbf{G_y}$-MINOLESS has been built on *sequential approximations*. First, the *surjectivity defect* was secured by $\mathbf{G_y}$-LESS. The corresponding normal equations suffered from the effect of the *injectivity defect*. Accordingly, *second* $\mathbf{G_x}$-LESS generated a unique solution the rank deficient normal equations. Alternatively, we may constitute a unique solution of the system of inconsistent, rank deficient equations

$$\{\mathbf{Ax} + \mathbf{i} = \mathbf{y} | \mathbf{A} \in \mathbb{R}^{n \times m}, r := \text{rk}\mathbf{A} < \min\{n, m\}\}$$

by the α-weighted hybrid norm of type "LESS" and "MINOS". Such a solution of a general algebraic regression problem is also called

- Tykhonov–Phillips regularization

Fig. 5.7 Balance of LESS and MINOS to general MORE



- Ridge estimator
- α – HAPS.

Indeed, $\alpha - HAPS$ is the most popular inversion operation, namely to *regularize improperly posed problems*. An example is the discretized version of an *integral equation of the first kind*.

Definition 5.2. *Definition 5.8* ($\alpha - HAPS$):

An $m \times 1$ vector $\mathbf{x}$ is called *weighted* $\alpha - HAPS$ (**H**ybrid **AP**proximative **S**olution) with respect to an α-weighted $\mathbf{G_x}$, $\mathbf{G_y}$-seminorm of (5.1), if

$$\begin{aligned} \mathbf{x}_h = \arg\{||\mathbf{y} - \mathbf{Ax}||^2_{\mathbf{G_y}} + \alpha||\mathbf{x}||^2_{\mathbf{G_x}} = \min|\mathbf{Ax} + \mathbf{i} = \mathbf{y}, \\ \mathbf{A} \in \mathbb{R}^{n \times m}; \mathrm{rk}\mathbf{A} \leq \min\{n, m\}\}. \end{aligned} \tag{5.108}$$

Note that we may apply weighted $\alpha - HAPS$ *even for* the case of rank identity $rk\mathbf{A} \leq \min\{n, m\}$. The factor $\alpha \in \mathbb{R}^+$ *balances* the least squares norm and the minimum norm of the unknown vector which is illustrated by *Fig. 5.7*.

Lemma 5.8. ($\alpha - HAPS$):

$\mathbf{x}_h$ is weighted $\alpha - HAPS$ of $\mathbf{x}$ of the general system of inconsistent, possibly weakly inconsistent, rank deficient system of linear equations (5.1) if and only if the system of normal equations

$$(\alpha\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})\mathbf{x}_h = \mathbf{A}'\mathbf{G_y}\mathbf{y} \; or \; \left(\mathbf{G_x} + \frac{1}{\alpha}\mathbf{A}'\mathbf{G_y}\mathbf{A}\right)\mathbf{x}_h = \frac{1}{\alpha}\mathbf{A}'\mathbf{G_y}\mathbf{y} \tag{5.109}$$

is fulfilled. $\mathbf{x_h}$ always exists and is uniquely determined if the matrix

$$\alpha\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A} \; is \;\; regular \;\; or \; \mathrm{rk}[\mathbf{G_x}, \mathbf{A}'\mathbf{G_y}\mathbf{A}] = m. \tag{5.110}$$

Proof. $\alpha - HAPS$ is constructed by means of the Lagrangean

$$\mathcal{L}(\mathbf{x}) := ||\mathbf{y} - \mathbf{Ax}||^2_{\mathbf{G_y}} + \alpha||\mathbf{x}||^2_{\mathbf{G_y}} = (\mathbf{y} - \mathbf{Ax})'\mathbf{G_y}(\mathbf{y} - \mathbf{Ax}) + \alpha(\mathbf{x}'\mathbf{G_y}\mathbf{x}) = \min_{\mathbf{x}},$$

such that the *first derivatives*

$$\frac{d\mathcal{L}}{d\mathbf{x}}(\mathbf{x_h}) = 2(\alpha\mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})\mathbf{x_h} - 2\mathbf{A}'\mathbf{G_y}\mathbf{y} = \mathbf{0}$$

constitute the *necessary conditions*. The *second derivatives*

$$\frac{\partial^2 \mathcal{L}}{\partial \mathbf{x} \partial \mathbf{x}'}(\mathbf{x_h}) = 2(\alpha \mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A}) \geq \mathbf{0}$$

generate the *sufficiency conditions* for obtaining the minimum of the unconstrained Lagrangean. If $\alpha \mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A}$ is regular of $\text{rk}[\mathbf{G_y}, \mathbf{A}'\mathbf{G_y}\mathbf{A}] = m$, there exists a unique solution.

Lemma 5.9. $(\alpha - HAPS)$:

If $\mathbf{x_h}$ is $\alpha - HAPS$ of $\mathbf{x}$ of the general system of inconsistent, possibly weakly inconsistent, possibly rank deficient system of linear equations (5.1) fulfilling the rank identity

$$\text{rk}[\mathbf{G_y}, \mathbf{A}'\mathbf{G_y}\mathbf{A}] = m \; or \; \det(\alpha \mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A}) \neq 0$$

then

$$\mathbf{x_h} = (\alpha \mathbf{G_x} + \mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}$$

or

$$\mathbf{x_h} = (\mathbf{G_x} + \frac{1}{\alpha}\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\frac{1}{\alpha}\mathbf{A}'\mathbf{G_y}\mathbf{y}$$

or

$$\mathbf{x_h} = (\alpha \mathbf{I} + \mathbf{G_x}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{G_x}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}$$

or

$$\mathbf{x_h} = (\mathbf{I} + \frac{1}{\alpha}\mathbf{G_x}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\frac{1}{\alpha}\mathbf{G_x}^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{y}$$

are four representations of the unique solution.

Chapter 6
The Third Problem of Probabilistic Regression

Special Gauss–Markov Model with Datum Defect

Setup of BLUMBE and BLE for the moments of first order and of BIQUUE and BIQE for the central moment of second order $\{\mathbf{y} = \mathbf{A}\xi + \mathbf{c_y}, \mathbf{A} \in \mathbb{R}^{n \times m}, \mathrm{rk}\mathbf{A} < \min\{n, m\}\}$

The Special Gauss-Markov model with datum defect – the stochastic analogue of *Minimum Norm Least-Squares*, is treated here first by the *Best Linear Minimum Bias Estimator* (BLUMBE), namely by *Theorem 6.3*, in the first section. *Theorem 6.5* offers the estimation of σ^2 by *Best Invariant Quadratic Uniformly Unbiased Estimation* to be compared with the result of *Theorem 6.6*, namely the estimation of σ^2 by *Best Invariant Quadratic Estimation*. Six numerical models are compared based on practical *leveling measurements*. In the *second section*, we setup the *Best Linear Estimation (hom BLE), (hom S-BLE)* as well as *(hom α-BLE)* for fixed effect based on the decomposition

$$\hat{\xi} - \xi = \mathbf{L}(\mathbf{y} - E\{\mathbf{y}\}) - [\mathbf{I}_m - \mathbf{LA}]\,\xi$$

This setup produces the weighting of two norms:

1. The *average variances* $\|\mathbf{L}'\|^2$
2. The *average bias* $\|(\mathbf{I}_m - \mathbf{LA})'\|^2$.

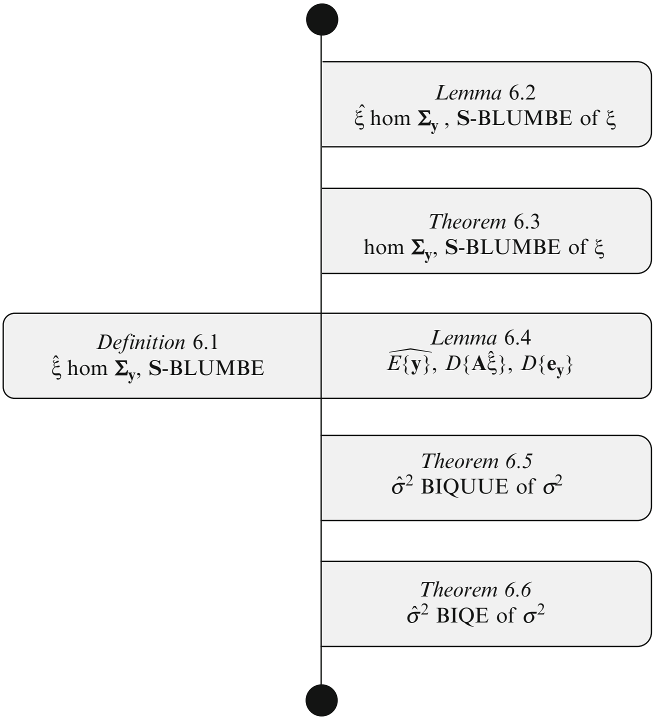
α-BLE balances the variance and the bias illustrated by *Fig. 6.1*. *Theorems 6.12 and 6.13* compare (*hom S-BLE*) and (*hom α-BLE*, also called *ridge estimator* or *Tykhonov-Phillips regularization published by many authors* (J. Cai, E. Grafarend and B. Schaffrin (2004, 92 references))).

A *key example* is our treatment of *continuous networks* in the *third section*, namely of *second derivative type* and finally approximated by a *discrete setup*.

Finally we intend to state our assumption of *C.R. Rao's design of the substitute matrix* **S** *or* α-**I** which solves the missing information of the variable ξ, perhaps by *Prior Information* ξ_0, subject of *Bayes-estimation*.

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_6,
© Springer-Verlag Berlin Heidelberg 2012

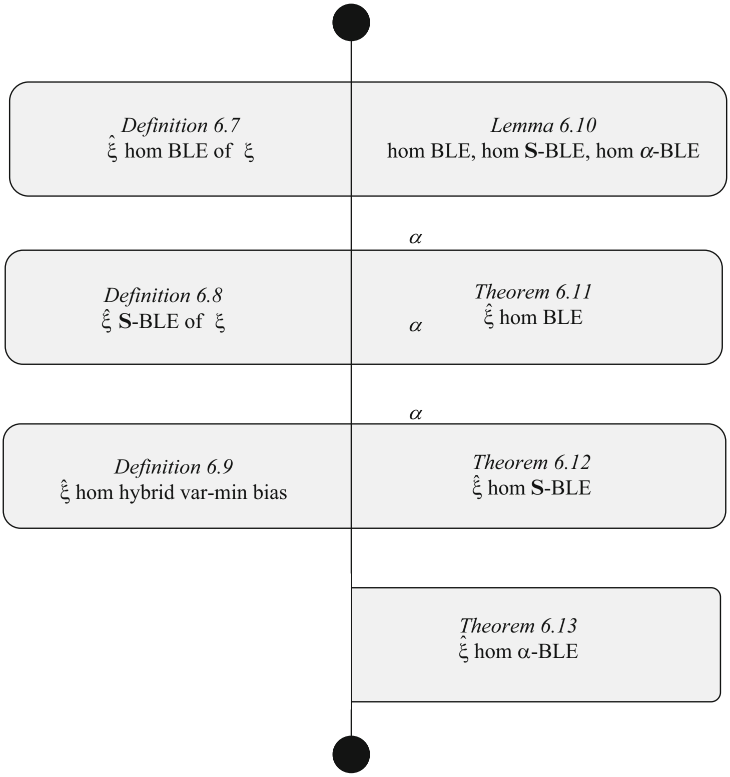
Read only *Definition 6.1, Theorem 6.3, Definition 6.4–6.6, Theorem 6.8–6.11*.



Definition 6.7 and Lemma 6.2, Theorem 6.3, Lemma 6.4, Theorems 6.5 and 6.6 review $\hat{\xi}$ of type hom $\mathbf{\Sigma_y}$, **S**-BLUMBE, BIQE, followed by the first example. Alternatively, estimators of type best linear, namely hom BLE, hom S-BLE and hom α-BLE are presented. *Definitions 6.7, 6.8* and *6.9* relate to various estimators followed by *Lemma 6.10*, *Theorems 6.11, 6.12* and *6.13*.

In Chap. 5, we have solved a *special algebraic regression problem*, namely the inversion of a system of inconsistent linear equations with a datum defect. By means of a hierarchic postulate of a minimum norm $||\mathbf{x}||^2 = \min$, least squares solution $||\mathbf{y} - \mathbf{Ax}||^2 = \min$ ("MINOLESS") we were able to determine m unknowns from n observations through the rank of the linear operator, $\text{rk}\mathbf{A} = r < \min\{n, m\}$, was *less* than the number of observations or less the number of unknowns. Though "MINOLESS" generates a rigorous solution, we were left with the problem to *interpret our results*.

The key for an evolution of "MINOLESS" is handed over to us by treating the special algebraic problem by means of a *special probabilistic regression problem*, namely as a *special Gauss–Markov model with datum defect*. The bias generated by any solution of a rank deficient system of linear equations will again be introduced



as a *decisive criterion* for evaluating "MINOLESS", now in the context of a *probabilistic regression problem*. In particular, a special form of "LUMBE" the *linear uniformly minimum bias estimator* $||\mathbf{LA} - \mathbf{I}|| =$ min, leads us to a solution which is equivalent to "MINOS". "*Best*" of LUMBE in the sense of the average variance $||D\{\hat{\xi}\}||^2 = \mathrm{tr}D\{\hat{\xi}\} =$ min also called BLUMBE, will give us a unique solution of $\hat{\xi}$ as a linear estimation of the observation vector $\mathbf{y} \in \{\mathbb{Y}, pdf\}$ with respect to the linear model $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi}$, $D\{\mathbf{y}\} = \boldsymbol{\Sigma}_\mathbf{y}$ of "*fixed effects*" $\xi \in \Xi$.

Alternatively, in Chap. 5, we had solved the ill-posed problem $\mathbf{y} = \mathbf{Ax} + \mathbf{i}$, $\mathbf{A} \in \mathbb{R}^{n \times m}$, $\mathrm{rk}\mathbf{A} < \min\{n, m\}$, by means of α-HAPS. Here with respect to a *special probabilistic regression problem* we succeed to compute α-BLE (α *weighted*, **S** *modified Best Linear Estimation*) as an equivalence to α-HAPS of a *special algebraic regression problem*. Most welcomed is the analytical optimization problem to determine the *regularization parameter* α by means of a special form of $||MSE\{\alpha\}||^2 =$ min, the *weighted Mean Square Estimation Error*. Such an optimal design of the regulator α is not possible in the *Tykhonov–Phillips regularization* in the context of α-HAPS, but definitely in the context of α-BLE.

6-1 Setup of the Best Linear Minimum Bias Estimator of Type BLUMBE

Box 6.1 is a definition of our *special linear Gauss–Markov model* with datum defect. We assume (6.1) $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi}$, $\text{rk}\mathbf{A} < \min\{n, m\}$ (1st moments) and (6.2) $D\{\mathbf{y}\} = \boldsymbol{\Sigma}_\mathbf{y}$, $\boldsymbol{\Sigma}_\mathbf{y}$ *positive definite*, $\text{rk}\boldsymbol{\Sigma}_\mathbf{y} = n$ (2nd moments). *Box 6.2* reviews the bias vector as well as the bias matrix including the related norms.

Box 6.1. (Special linear Gauss–Markov model with datum defect):

$$\{\mathbf{y} = \mathbf{A}\xi + \mathbf{c}_\mathbf{y},\ \mathbf{A} \in \mathbb{R}^{n\times m},\ \text{rk}\mathbf{A} < \min\{n, m\}\}$$

1st moments

$$E\{\mathbf{y}\} = \mathbf{A}\xi\,,\quad \text{rk}\mathbf{A} < \min\{n, m\} \tag{6.1}$$

2nd moments

$$D\{\mathbf{y}\} = :\boldsymbol{\Sigma}_\mathbf{y},\ \boldsymbol{\Sigma}_\mathbf{y}\ \textit{positive definite},\ \text{rk}\boldsymbol{\Sigma}_\mathbf{y} = \text{n}, \tag{6.2}$$

$$\xi \in \mathbb{R}^{\mathbf{m}}, \textit{vector of ``fixed effects''}, \textit{unknown},$$

$$\boldsymbol{\Sigma}_\mathbf{y}\ \textit{unknown or known from prior information.}$$

Box 6.2. (Bias vector, bias matrix Vector and matrix bias norms Special linear Gauss–Markov model of fixed effects subject to datum defect)

$$\begin{gathered}\mathbf{A} \in \mathbb{R}^{n\times m},\ \text{rk}\mathbf{A} < \min\{n, m\}\\ E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi},\ D\{\mathbf{y}\} = \boldsymbol{\Sigma}_\mathbf{y}\end{gathered} \tag{6.3}$$

"ansatz"

$$\hat{\xi} = \mathbf{L}\mathbf{y} \tag{6.4}$$

bias vector

$$\beta := E\{\hat{\xi} - \xi\} = E\{\hat{\xi}\} - \xi \neq 0 \tag{6.5}$$

$$\beta\ = \mathbf{L}E\{\mathbf{y}\} - \xi = -[\mathbf{I}_m - \mathbf{L}\mathbf{A}]\xi \neq 0 \tag{6.6}$$

bias matrix

$$\mathbf{B} := \mathbf{I}_n - \mathbf{LA} \tag{6.7}$$

"*bias norms*"

$$||\beta||^2 = \beta'\beta = \xi'[\mathbf{I}_m - \mathbf{LA}]'[\mathbf{I}_m - \mathbf{LA}]\xi \tag{6.8}$$

$$||\beta||^2 = \mathrm{tr}\beta\beta' = \mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\xi\xi'[\mathbf{I}_m - \mathbf{LA}]' = ||\mathbf{B}||^2_{\xi\xi'} \tag{6.9}$$

$$||\beta||^2_{\mathbf{S}} := \mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]' = ||\mathbf{B}||^2_{\mathbf{S}} \tag{6.10}$$

"*dispersion matrix*"

$$D\{\hat{\xi}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' = \mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' \tag{6.11}$$

"*dispersion norm, average variance*"

$$||D\{\hat{\xi}\}||^2 := \mathrm{tr}\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' = \mathrm{tr}\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' =: ||\mathbf{L}'||_{\boldsymbol{\Sigma}_{\mathbf{y}}} \tag{6.12}$$

"*decomposition*"

$$\hat{\xi} - \xi = (\hat{\xi} - E\{\hat{\xi}\}) + (E\{\hat{\xi}\} - \hat{\xi}) \tag{6.13}$$

$$\hat{\xi} - \xi = \mathbf{L}(\mathbf{y} - E\{\mathbf{y}\}) - [\mathbf{I}_m - \mathbf{LA}]\xi \tag{6.14}$$

"*Mean Square Estimation Error*"

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \tag{6.15}$$

$$MSE\{\hat{\xi}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + [\mathbf{I}_m - \mathbf{LA}]\xi\xi'[\mathbf{I}_m - \mathbf{LA}]' \tag{6.16}$$

"*modified Mean Square Estimation Error*"

$$MSE_{\mathbf{S}}\{\hat{\xi}\} := \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + [\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]' \tag{6.17}$$

"*MSE norms, average MSE*"

$$||MSE\{\hat{\xi}\}||^2 := \mathrm{tr}E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \tag{6.18}$$

$$\begin{gathered} ||MSE\{\hat{\xi}\}||^2 \\ = \mathrm{tr}\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\xi\xi'[\mathbf{I}_m - \mathbf{LA}]' \\ = ||\mathbf{L}'||^2_{\boldsymbol{\Sigma}_{\mathbf{y}}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'} \end{gathered} \tag{6.19}$$

$$\begin{aligned} &||MSE_{\mathbf{S}}\{\hat{\xi}\}||^2 : \\ : = &\mathrm{tr}\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}[\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]' \\ &= ||\mathbf{L}'||^2_{\mathbf{\Sigma_y}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'} \,. \end{aligned} \tag{6.20}$$

Definition 6.1 defines (1st) $\hat{\xi}$ as a *linear homogenous form*, (2nd) of type "minimum bias" and (3rd) of type "*smallest average variance*".

Section 6.11 is a collection of definitions and lemmas, theorems basic for the developments in the future.

6-11 Definitions, Lemmas and Theorems

Definition 6.1. ($\hat{\xi}$ hom $\mathbf{\Sigma_y}$, $\mathbf{S}$–BLUMBE):

An $m \times 1$ vector $\hat{\xi} = \mathbf{Ly}$ is called *homogeneous* $\mathbf{\Sigma_y}$, $\mathbf{S}$-BLUMBE (*homogeneous Best Linear Uniformly Minimum Bias Estimation*) of ξ in the special inconsistent *linear Gauss–Markov model of fixed effects of Box 6.1*, if (1st) $\hat{\xi}$ is a homogeneous linear form

$$\hat{\xi} = \mathbf{Ly} \tag{6.21}$$

(2nd) in comparison to all other linear estimations $\hat{\xi}$ has the *minimum bias* in the sense of

$$||\mathbf{B}||^2_{\mathbf{S}} : = ||(\mathbf{I_m} - \mathbf{LA})'||^2_{\mathbf{S}} = \min_{\mathbf{L}}, \tag{6.22}$$

(3rd) in comparison to all other *minimum bias estimations* $\hat{\xi}$ has the *smallest average variance* in the sense of

$$||D\{\hat{\xi}\}||^2 = \mathrm{tr}\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' = ||\mathbf{L}'||^2_{\mathbf{\Sigma_y}} = \min_{\mathbf{L}} \,. \tag{6.23}$$

The estimation $\hat{\xi}$ of type *hom* $\mathbf{\Sigma_y}$, $\mathbf{S}$-BLUMBE can be characterized by

Lemma 6.2. ($\hat{\xi}$ hom $\mathbf{\Sigma}_{\boldsymbol{y}}$, *S–BLUMBE of* ξ):

An $m \times 1$ vector $\hat{\xi} = \mathbf{Ly}$ is *hom* $\mathbf{\Sigma_y}$, $\mathbf{S}$-BLUMBE of ξ in the *special inconsistent linear Gauss–Markov model with fixed effects* of *Box 6.1*, if and only if the matrix $\mathbf{L}$ fulfils the *normal equations*

$$\begin{bmatrix} \mathbf{\Sigma_y} & \mathbf{ASA}' \\ \mathbf{ASA}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{L}' \\ \mathbf{\Lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{AS} \end{bmatrix} \tag{6.24}$$

with the $n \times n$ matrix $\mathbf{\Lambda}$ of "*Lagrange multipliers*".

: Proof :

First, we minimize the $\mathbf{S}$-modified bias matrix norm, second the MSE($\hat{\xi}$) matrix norm. All matrix norms have been chosen "*Frobenius*".

$$(i)\ ||(\mathbf{I_m} - \mathbf{LA})'||_{\mathbf{S}}^2 = \min_{\mathbf{L}}.$$

The $\mathbf{S}$-weighted *Frobenius matrix norm* $||(\mathbf{I_m} - \mathbf{LA})'||_{\mathbf{S}}^2$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) :\ = \mathrm{tr}(\mathbf{I_m} - \mathbf{LA})\mathbf{S}(\mathbf{I_m} - \mathbf{LA})' = \min_{\mathbf{L}}$$

for $\mathbf{S}$-BLUMBE.

$$\mathcal{L}(\mathbf{L}) = \min_{\mathbf{L}} \Leftrightarrow \left[\begin{array}{l} \mathbf{ASA'\hat{L}' - AS = 0} \\ (\mathbf{ASA'}) \otimes \mathbf{I_m} > \mathbf{0}, \end{array}\right.$$

according to Theorem 2.3.

$$\begin{bmatrix} \mathbf{\Sigma_y} & \mathbf{ASA'} \\ \mathbf{ASA'} & 0 \end{bmatrix} \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma_y} & \mathbf{ASA'} \\ \mathbf{ASA'} & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma_y} & \mathbf{ASA'} \\ \mathbf{ASA'} & 0 \end{bmatrix} \tag{6.25}$$

$$\mathbf{\Sigma_y} C_1 \mathbf{\Sigma_y} + \mathbf{\Sigma_y} C_2 \mathbf{ASA'} + \mathbf{ASA'} C_3 \mathbf{\Sigma_y} + \mathbf{ASA'} C_4 \mathbf{ASA'} = \mathbf{\Sigma_y} \tag{6.26}$$

$$\mathbf{\Sigma_y} C_1 \mathbf{ASA'} + \mathbf{ASA'} C_3 \mathbf{ASA'} = \mathbf{ASA'} \tag{6.27}$$

$$\mathbf{ASA'} C_1 \mathbf{\Sigma_y} + \mathbf{ASA'} C_2 \mathbf{ASA'} = \mathbf{ASA'} \tag{6.28}$$

$$\mathbf{ASA'} C_1 \mathbf{ASA'} = 0. \tag{6.29}$$

Let us multiply the *third identity* by $\mathbf{\Sigma_y^{-1} ASA'} = 0$ from the right side and substitute the *fourth identity* in order to solve for C_2.

$$\mathbf{ASA'} C_2 \mathbf{ASA' \Sigma_y^{-1} ASA'} = \mathbf{ASA' \Sigma_y^{-1} ASA'} \tag{6.30}$$

$$C_2 = \mathbf{\Sigma_y^{-1} ASA'(ASA' \Sigma_y^{-1} ASA')^-} \tag{6.31}$$

solves the *fifth equation*

$$\begin{gathered} \mathbf{A'SA\Sigma_y^{-1}ASA'(ASA'\Sigma_y^{-1}ASA')^- ASA'\Sigma_y^{-1}ASA'} \\ = \mathbf{ASA'\Sigma_y^{-1}ASA'} \end{gathered} \tag{6.32}$$

by the axiom of a generalized inverse.

$$(ii)\ ||\mathbf{L}'||_{\mathbf{\Sigma_y}}^2 = \min_{\mathbf{L}}.$$

The $\mathbf{\Sigma_y}$-weighted Frobenius matrix norm of $\mathbf{L}$ subject to the condition of LUMBE generates the *constrained Lagrangean*

$$\mathcal{L}(\mathbf{L},\mathbf{\Lambda}) = \mathrm{tr}\mathbf{L}\mathbf{\Sigma}_{\mathbf{y}}\mathbf{L}' + 2\mathrm{tr}\mathbf{\Lambda}'(\mathbf{ASA}'\mathbf{L}' - \mathbf{AS}) = \min_{\mathbf{L},\mathbf{\Lambda}}.$$

According to the theory of matrix derivatives outlined in *Appendix B*

$$\frac{\partial\mathcal{L}}{\partial\mathbf{L}}(\hat{\mathbf{L}},\hat{\mathbf{\Lambda}}) = 2(\mathbf{\Sigma}_{\mathbf{y}}\hat{\mathbf{L}}' + \mathbf{ASA}'\hat{\mathbf{\Lambda}})' = \mathbf{0},$$
$$\frac{\partial\mathcal{L}}{\partial\mathbf{\Lambda}}(\hat{\mathbf{L}},\hat{\mathbf{\Lambda}}) = 2(\mathbf{ASA}'\mathbf{L}' - \mathbf{AS}) = \mathbf{0},$$

at the "point" $(\hat{\mathbf{L}},\hat{\mathbf{\Lambda}})$ constitute the *necessary conditions*, while

$$\frac{\partial^2\mathcal{L}}{\partial(\mathrm{vec}\mathbf{L})\,\partial(\mathrm{vec}\mathbf{L}')}(\hat{\mathbf{L}},\hat{\mathbf{\Lambda}}) = 2(\mathbf{\Sigma}_{\mathbf{y}}\otimes\mathbf{I_m}) > \mathbf{0},$$

to be a *positive definite matrix*, the *sufficiency conditions*. Indeed, the first matrix derivations have been identified as the normal equations of the *sequential optimization problem*.

For an explicit representation of $\hat{\xi}$hom $\mathbf{\Sigma}_{\mathbf{y}}$, $\mathbf{S}$-BLUMBE of ξ we solve the *normal equations* for

$$\hat{\mathbf{L}} = \arg\{||D(\hat{\xi})|| = \min_{\mathbf{L}} |\mathbf{ASA}'\mathbf{L}' - \mathbf{AS} = \mathbf{0}\}.$$

In addition, we compute the dispersion matrix $D\{\hat{\xi}|$ hom BLUMBE$\}$ as well as the mean square estimation error $MSE\{\hat{\xi}|$ hom BLUMBE$\}$.

Theorem 6.3. (hom $\mathbf{\Sigma}_{\mathbf{y}}$, $\mathbf{S}$-BLUMBE of ξ):

Let $\hat{\xi} = \mathbf{Ly}$ be hom $\mathbf{\Sigma}_{\mathbf{y}}$, $\mathbf{S}$-BLUMBE in the *special Gauss–Markov model of Box 6.1*. Then *independent of the choice of the generalized inverse* $(\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}\mathbf{ASA}')^-$ the *unique solution* of the *normal equations* (6.24) is

$$\hat{\xi} = \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{ASA}')^-\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}, \tag{6.33}$$

completed by the dispersion matrix

$$\mathbf{D}(\hat{\xi}) = \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{ASA}')^-\mathbf{AS} \tag{6.34}$$

the bias vector

$$\beta = -[\mathbf{I_m} - \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{ASA}')^-\mathbf{ASA}'\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}]\,\xi \tag{6.35}$$

and the matrix MSE $\{\hat{\xi}\}$ of *mean estimation errors*

$$\mathbf{E}\{(\hat{\xi}-\xi)(\hat{\xi}-\xi)'\} = D\{\hat{\xi}\} + \beta\beta' \tag{6.36}$$

modified by

$$\begin{aligned}\mathbf{E}\{(\hat{\xi}-\xi)(\hat{\xi}-\xi)'\} &= D\{\hat{\xi}\} + [\mathbf{I_m} - \mathbf{LA}]\mathbf{S}[\mathbf{I_m} - \mathbf{LA}]' \\ &= D\{\hat{\xi}\} + [\mathbf{S} - \mathbf{SA}'(\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{AS}],\end{aligned} \tag{6.37}$$

based upon the solution of $\xi\xi'$ by $\mathbf{S}$.

$$\mathrm{rk}MSE\{\hat{\xi}\} = \mathrm{rk}\mathbf{S} \tag{6.38}$$

is the corresponding rank identity.

:Proof:

$$(i)\ \hat{\xi} hom\ \boldsymbol{\Sigma}_\mathbf{y},\ \mathbf{S} - \text{BLUMBE}\ of\ \xi.$$

First, we prove that the matrix of the normal equations

$$\begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{y} & \mathbf{ASA}' \\ \mathbf{ASA}' & 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{y} & \mathbf{ASA}' \\ \mathbf{ASA}' & 0 \end{bmatrix} = 0$$

is singular.

$$\begin{vmatrix} \boldsymbol{\Sigma}_\mathbf{y} & \mathbf{ASA}' \\ \mathbf{ASA}' & 0 \end{vmatrix} = |\boldsymbol{\Sigma}_\mathbf{y}|\ \ |-\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}'| = 0,$$

due to $\mathrm{rk}[\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}'] = \mathrm{rk}\mathbf{A} < \min\{n, m\}$ assuming $\mathrm{rk}\mathbf{S} = m$, $\mathrm{rk}\boldsymbol{\Sigma}_\mathbf{y} = n$. Note

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}|\ \ |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|\ if \left[\begin{array}{l} \mathbf{A}_{11} \in^{m_1\times m_1} \\ \mathrm{rk}\mathbf{A}_{11} = m_1 \end{array}\right.$$

if with reference to Appendix A.7. Thanks to the rank deficiency of the partitioned normal equation matrix, we are forced to compute *secondly* its generalized inverse.

The system of normal equations is solved for

$$\begin{bmatrix} \hat{\mathbf{L}}' \\ \hat{\boldsymbol{\Lambda}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{y} & \mathbf{ASA}' \\ \mathbf{ASA}' & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{AS} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \\ \mathbf{C}_3 & \mathbf{C}_4 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{AS} \end{bmatrix} \tag{6.39}$$

$$\hat{\mathbf{L}}' = \mathbf{C}_2\mathbf{AS} \tag{6.40}$$

$$\Leftrightarrow \hat{\mathbf{L}} = \mathbf{SA}'\mathbf{C}'_2 \tag{6.41}$$

$$\hat{\mathbf{L}} = \mathbf{SA}(\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1} \tag{6.42}$$

such that

$$\hat{\xi} = \hat{\mathbf{L}}\mathbf{y} = \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{y}. \tag{6.43}$$

We leave the proof for

"$\mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}$

is a weighted pseudo-inverse or Moore-Penrose inverse" as an exercise.

(ii) *Dispersion matrix* $D\{\hat{\xi}\}$.

The residual vector

$$\hat{\xi} - E\{\mathbf{y}\} = \hat{\mathbf{L}}(\mathbf{y} - E\{\mathbf{y}\}) \tag{6.44}$$

leads to the variance-covariance matrix

$$\begin{aligned} D\{\hat{\xi}\} &= \hat{\mathbf{L}}\mathbf{\Sigma}_\mathbf{y}\hat{\mathbf{L}}' \\ &= \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{AS} \\ &= \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{AS}\,. \end{aligned} \tag{6.45}$$

(iii) *Bias vector* β

$$\begin{aligned} \beta := E\{\hat{\xi} - \xi\} &= -(\mathbf{I}_m - \hat{\mathbf{L}}\mathbf{A})\xi \\ &= -[\mathbf{I}_m - \mathbf{SA}'(\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A}]\xi. \end{aligned} \tag{6.46}$$

Such a bias vector is *not* accessible to observations since ξ is unknown. Instead it is common practice to replace ξ by $\hat{\xi}$ (BLUMBE), the estimation $\hat{\xi}$ of ξ of type BLUMBE.

(iv) Mean Square Estimation Error $MSE\{\hat{\xi}\}$

$$\begin{aligned} MSE\{\hat{\xi}\} := E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} &= D\{\hat{\xi}\} + \beta\beta' \\ &= \hat{\mathbf{L}}\mathbf{\Sigma}_\mathbf{y}\hat{\mathbf{L}}' + (\mathbf{I}_m - \hat{\mathbf{L}}\mathbf{A})\xi\xi'(\mathbf{I}_m - \hat{\mathbf{L}}\mathbf{A})'\,. \end{aligned} \tag{6.47}$$

Neither $D\{\hat{\xi}|\mathbf{\Sigma}_\mathbf{y}\}$, nor $\beta\beta'$ are accessible to measurements. $\xi\xi'$ is replaced by *C.R. Rao's* substitute matrix S, $\mathbf{\Sigma}_\mathbf{y} = \mathbf{V}\sigma^2$ by a *one variance component* model σ^2 by $\hat{\sigma}^2$ (BIQUUE) or $\hat{\sigma}^2$(BIQE), for instance.

Lemma 6.4. ($\widehat{E\{\mathbf{y}\}}$, $D\{\mathbf{A}\hat{\boldsymbol{\xi}}\}$, $\tilde{\mathbf{e}}_\mathbf{y}$, $D\{\mathbf{e}_\mathbf{y}\}$ for $\hat{\xi}$ hom $\mathbf{\Sigma}_\mathbf{y}$, S of ξ):

(i) With respect to the model (1st) $\mathbf{A}\boldsymbol{\xi} = E\{\mathbf{y}\}$, $E\{\mathbf{y}\} \in \mathbf{R}(\mathbf{A})$, rk$\mathbf{A} =: r \leq m$ and $\mathbf{V}\sigma^2 = D\{\mathbf{y}\}$, $\mathbf{V}$ *positive definite*, $rk\,V = n$ under the condition $\dim \mathcal{R}(\mathbf{SA}') = \mathrm{rk}(\mathbf{SA}') = \mathrm{rk}\mathbf{A} = r$, namely $\mathbf{V}$, $\mathbf{S}$-BLUMBE, is given by

$$\widehat{E\{\mathbf{y}\}} = \mathbf{A}\hat{\boldsymbol{\xi}} = \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}\mathbf{y} \tag{6.48}$$

with the related singular dispersion matrix

$$D\{\mathbf{A}\hat{\boldsymbol{\xi}}\} = \sigma^2\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}' \tag{6.49}$$

for any choice of the *generalized inverse* $(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}$.

(ii) The empirical error vector $\mathbf{e_y} = \mathbf{y} - E\{\mathbf{y}\}$ results in the residual error vector $\tilde{\mathbf{e}}_\mathbf{y} = \mathbf{y} - \mathbf{A}\hat{\boldsymbol{\xi}}$ of type

$$\tilde{\mathbf{e}}_\mathbf{y} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})\mathbf{A}'\mathbf{V}^{-1}\mathbf{B}\mathbf{y}] \tag{6.50}$$

with the related singular dispersion matrices

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \sigma^2[\mathbf{V} - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'] \tag{6.51}$$

for any choice of the generalized inverse $(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}$.

(iii) The various dispersion matrices are related by

$$\begin{aligned} D\{\mathbf{y}\} &= D\{\mathbf{A}\hat{\boldsymbol{\xi}} + \tilde{\mathbf{e}}_\mathbf{y}\} = D\{\mathbf{A}\hat{\boldsymbol{\xi}}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\} \\ &= D\{\tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}, \end{aligned} \tag{6.52}$$

where the dispersion matrices

$$\tilde{\mathbf{e}}_\mathbf{y}\ and\ \mathbf{A}\hat{\boldsymbol{\xi}} \tag{6.53}$$

are uncorrected, in particularly,

$$C\{\tilde{\mathbf{e}}_\mathbf{y}, \mathbf{A}\hat{\boldsymbol{\xi}}\} = C\{\tilde{\mathbf{e}}_\mathbf{y}, \tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} = 0. \tag{6.54}$$

When we compute the solution by $\hat{\sigma}$ of type BIQUUE and of type BIQE we arrive at *Theorems* 6.5 and 6.6.

Theorem 6.5. ($\hat{\sigma}^2$ BIQUUE of σ^2, *special Gauss–Markov model*: $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi}$, $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$, $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathrm{rk}\mathbf{A} = r \leq m$, $\mathbf{V} \in \mathbb{R}^{n\times m}$, $\mathrm{rk}\mathbf{V} = n$):

Let $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\mathrm{vec}\mathbf{M})'\mathbf{y}\otimes\mathbf{y} = \mathbf{y}'\otimes\mathbf{y}'(\mathrm{vec}\mathbf{M})$ be BIQUUE with respect to the *special Gauss–Markov model* 6.1. Then

$$\hat{\sigma}^2 = (n-r)^{-1}\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y} \tag{6.55}$$

$$\hat{\sigma}^2 = (n-r)^{-1}\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{ASA}'(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{V}^{-1}]\mathbf{y} \tag{6.56}$$

$$\hat{\sigma}^2 = (n-r)^{-1}\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_\mathbf{y} = (n-r)^{-1}\tilde{\mathbf{e}}_\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_\mathbf{y} \tag{6.57}$$

are *equivalent representations* of the BIQUUE variance component $\hat{\sigma}^2$ which are independent of the generalized inverses

$$(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-} \ or \ (\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{A}'\mathbf{SA})^{-}.$$

The residual vector $\tilde{\mathbf{e}}_y$, namely

$$\tilde{\mathbf{e}}_{\mathbf{y}}(\text{BLUMBE}) = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y} \tag{6.58}$$

is of type BLUMBE. The variance of $\hat{\sigma}^2$ BIQUUE of σ^2

$$D\{\hat{\sigma}^2\} = 2(n-r)^{-1}\sigma^4 = 2(n-r)^{-1}(\sigma^2)^2 \tag{6.59}$$

can be substituted by the estimation

$$D\{\hat{\sigma}^2\} = 2(n-r)^{-1}(\hat{\sigma}^2)^2 = 2(n-r)^{-1}(\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}})^2 \tag{6.60}$$

Theorem 6.6. ($\hat{\sigma}^2$ BIQE of σ^2, *special Gauss–Markov model*: $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi}$, $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$, $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\text{rk}\mathbf{A} = r \leq m$, $\mathbf{V} \in \mathbb{R}^{n\times m}$, $\text{rk}\mathbf{V} = n$):

Let $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\text{vec}\mathbf{M})'\mathbf{y} \otimes \mathbf{y} = \mathbf{y}' \otimes \mathbf{y}'(\text{vec}\mathbf{M})$ be BIQE with respect to the *special Gauss–Markov model* 6.1. Independent of the matrix $\mathbf{S}$ and of the generalized inverses

$$(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-} \ or \ (\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{A}'\mathbf{SA})^{-},$$

$$\hat{\sigma}^2 = (n-r+2)^{-1}\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y} \tag{6.61}$$

$$\hat{\sigma}^2 = (n-r+2)^{-1}\mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{ASA}'(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-1}\mathbf{ASA}'\mathbf{V}^{-1}]\mathbf{y} \tag{6.62}$$

$$\hat{\sigma}^2 = (n-r+2)^{-1}\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} = (n-r+2)^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} \tag{6.63}$$

are *equivalent representations* of the BIQE variance component $\hat{\sigma}^2$. The *residual vector* $\tilde{\mathbf{e}}_{\mathbf{y}}$, namely

$$\tilde{\mathbf{e}}_{\mathbf{y}}(\text{BLUMBE}) = [\mathbf{I}_m - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}, \tag{6.64}$$

is of type BLUMBE. The variance of $\hat{\sigma}^2$ BIQE of σ^2

$$D\{\hat{\sigma}^2\} = 2(n-r)(n-r+2)^{-2}\sigma^4 = 2(n-r)[(n-r+2)^{-1}\sigma^2]^2 \tag{6.65}$$

can be substituted by the estimation

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-r)(n-r+2)^{-4}(\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}})^2. \tag{6.66}$$

The special bias

$$\beta_{\sigma^2} := E\{\hat{\sigma}^2 - \sigma^2\} = -2(n-r+2)^{-1}\sigma^2 \tag{6.67}$$

can be substituted by the estimation

$$\hat{\boldsymbol{\beta}}_{\sigma^2} = \hat{E}\{\hat{\sigma}^2 - \sigma^2\} = -2(n-r+2)^{-2}\tilde{\mathbf{e}}'_{\mathbf{y}}\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}. \tag{6.68}$$

Its $MSE(\hat{\sigma}^2)$ (Mean Square Estimation Error)

$$\begin{aligned} MSE\{\hat{\sigma}^2\} &:= \hat{E}\{(\hat{\sigma}^2 - \sigma^2)^2\} = D\{\hat{\sigma}^2\} + (\sigma^2 - E\{\hat{\sigma}^2\})^2 \\ &= 2(n-r+2)^{-1}(\sigma^2)^2 \end{aligned} \tag{6.69}$$

can be substituted by the estimation

$$\begin{aligned} \widehat{MSE}\{\hat{\sigma}^2\} &= \hat{E}\{(\hat{\sigma}^2 - \sigma^2)^2\} = \hat{D}\{\hat{\sigma}^2\} + (\hat{E}\{\sigma^2\})^2 \\ &= 2(n-r+2)^{-3}(\tilde{\mathbf{e}}'_{\mathbf{y}}\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}). \end{aligned} \tag{6.70}$$

6-12 The First Example: BLUMBE Versus BLE, BIQUUE Versus BIQE, Triangular Leveling Network

The first example for the *special Gauss–Markov model with datum defect*

$$\begin{aligned} \{E\{\mathbf{y}\} &= \mathbf{A}\xi,\ \mathbf{A} \in \mathbb{R}^{n\times m},\ \mathrm{rk}\mathbf{A} < \min\{n,m\}, \\ D\{\mathbf{y}\} &= \mathbf{V}\sigma^2,\ \mathbf{V} \in \mathbb{R}^{n\times m},\ \sigma^2 \in \mathbb{R}^+, \mathrm{rk}\mathbf{V} = n\} \end{aligned}$$

is taken from a *triangular leveling network*. 3 modal points are connected, by leveling measurements $[\underline{h}_{\alpha\beta},\ \underline{h}_{\beta\gamma},\ \underline{h}_{\gamma\alpha}]'$, also called potential differences of absolute potential heights $[h_\alpha,\ h_\beta,\ h_\gamma]'$ of "*fixed effects*". Alternative estimations of type

$$(i)\ \mathbf{I},\mathbf{I}\text{–}\mathtt{BLUMBE\ of}\ \boldsymbol{\xi} \in \mathbb{R}^m$$

$$(ii)\ \mathbf{V},\mathbf{S}\text{–}\mathtt{BLUMBE\ of}\ \boldsymbol{\xi} \in \mathbb{R}^m$$

$$(iii)\ \mathbf{I},\mathbf{I}\text{–}\mathtt{BLE\ of}\ \boldsymbol{\xi} \in \mathbb{R}^m$$

$$(iv)\ \mathbf{V},\mathbf{S}\text{–}\mathtt{BLE\ of}\ \boldsymbol{\xi} \in \mathbb{R}^m$$

$$(v)\ \mathtt{BIQUUE\ of}\ \sigma^2 \in \mathbb{R}^+$$

$$(vi)\ \mathtt{BIQE\ of}\ \sigma^2 \in \mathbb{R}^+$$

will be considered. In particular, we use *consecutive results* of Appendix A.7, namely from solving linear system of equations based upon *generalized inverse*, in short **g**-inverses. For the *analyst*, the *special Gauss–Markov model* with datum defect constituted by the problem of estimating absolute heights $[h_\alpha, h_\beta, h_\gamma]$ of points $\{P_\alpha, P_\beta, P_\gamma\}$ from height differences is formulated in *Box 6.3*.

Box 6.3. (The first example)

$$E\left\{\begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix}\right\} = \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \\ +1 & 0 & -1 \end{bmatrix} \begin{bmatrix} h_\alpha \\ h_\beta \\ h_\gamma \end{bmatrix}$$

$$\mathbf{y} := \begin{bmatrix} \underline{h}_\beta \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix}, \ A := \begin{bmatrix} -1 & +1 & 0 \\ 0 & -1 & +1 \\ +1 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{3\times 3}, \ \xi := \begin{bmatrix} h_\alpha \\ h_\beta \\ h_\gamma \end{bmatrix}$$

$$D\left\{\begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix}\right\} = D\{\mathbf{y}\} = \mathbf{V}\sigma^2, \ \sigma^2 \in^+$$

:dimensions:

$$\boldsymbol{\xi} \in \mathbb{R}^3, \ \dim \boldsymbol{\xi} = 3, \ \mathbf{y} \in \mathbb{R}^3, \ \dim\{\mathbb{Y}, \ pdf\} = 3$$
$$m = 3, \ n = 3, \ \mathrm{rk}\mathbf{A} = 2, \ rk\mathbf{V} = 3.$$

6-121 The First Example: $\mathbf{I}_3$, $\mathbf{I}_3$-BLUMBE

In the first case, we assume

a dispersion matrix $D\{\mathbf{y}\} = \mathbf{I}_3\sigma^2$ of i.i.d. *observations* $[\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3]'$	and	a *unity* substitute matrix $\mathbf{S} = \mathbf{I}_3$, in short u.s..

Under such a specification $\hat{\xi}$ is $\mathbf{I}_3, \mathbf{I}_3$-BLUMBE of ξ in the *special Gauss–Markov model* with datum defect.

$$\hat{\xi} = \mathbf{A}'(\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-}\mathbf{A}\mathbf{A}'\mathbf{y}$$

$$\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}' = 3\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \ (\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-} = \frac{1}{9}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

? How did we compute the g-inverse $(\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-}$?

The computation of the g-inverse $(\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-}$ has been based upon bordering.

$$(\mathbf{AA}'\mathbf{AA}')^- = \begin{bmatrix} 6 & -3 & -3 \\ -3 & 6 & -3 \\ -3 & -3 & 6 \end{bmatrix}^- = \begin{bmatrix} \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Please, check by yourself the axiom of a **g**-inverse:

$$\begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix}\begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix}^-\begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix} = \begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix}$$

or

$$\begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix}\frac{1}{9}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}^-\begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix} = \begin{bmatrix} +6 & -3 & -3 \\ -3 & +6 & -3 \\ -3 & -3 & +6 \end{bmatrix}$$

$$\hat{\xi} = \begin{bmatrix} h_\alpha \\ h_\beta \\ h_\gamma \end{bmatrix}(\mathbf{I}_3,\ \mathbf{I}_3\text{–BLUMBE}) = \frac{1}{3}\begin{bmatrix} -y_1 + y_3 \\ y_1 - y_2 \\ y_2 - y_3 \end{bmatrix}$$

$$\hat{\xi}_1 + \hat{\xi}_2 + \hat{\xi}_3 = 0.$$

Dispersion matrix $D\{\hat{\xi}\}$ of the unknown vector of "*fixed effects*"

$$D\{\hat{\xi}\} = \sigma^2\mathbf{A}'(\mathbf{AA}'\mathbf{AA}')^-\mathbf{A}$$

$$D\{\hat{\xi}\} = \frac{\sigma^2}{9}\begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix}$$

replace σ^2 by $\hat{\sigma}^2$ (BIQUUE):"

$$\hat{\sigma}^2 = (n - \mathrm{rk}\mathbf{A})^{-1}\tilde{\mathbf{e}}_\mathbf{y}'\tilde{\mathbf{e}}_\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{AA}')^-\mathbf{A}']\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{1}{3}(y_1 + y_2 + y_3)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_\mathbf{y}'\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{9}\mathbf{y}'\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y}$$

$$
\begin{aligned}
\tilde{\mathbf{e}}'_{\mathbf{y}}\tilde{\mathbf{e}}_{\mathbf{y}} &= \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)\\
\hat{\sigma}^2(\text{BIQUUE}) &= \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)\\
D\{\hat{\xi}\} &= \frac{1}{9}\begin{bmatrix} +2 & -1 & -1\\ -1 & +2 & -1\\ -1 & -1 & +2 \end{bmatrix}\hat{\sigma}^2(\text{BIQUUE})
\end{aligned}
$$

replace σ^2 by $\hat{\sigma}^2$ (BIQE):

$$
\begin{aligned}
\hat{\sigma}^2 &= (n + 2 - \text{rk}\mathbf{A})^{-1}\tilde{\mathbf{e}}'_{\mathbf{y}}\tilde{\mathbf{e}}_{\mathbf{y}}\\
\tilde{\mathbf{e}}_{\mathbf{y}}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE}) &= [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}\mathbf{A}')^{-}\mathbf{A}']\mathbf{y}\\
\tilde{\mathbf{e}}_{\mathbf{y}} &= \frac{1}{3}\begin{bmatrix} 1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{1}{3}(y_1 + y_2 + y_3)\begin{bmatrix} 1\\ 1\\ 1 \end{bmatrix}\\
\hat{\sigma}^2(\text{BIQE}) &= \frac{1}{9}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)\\
D\{\hat{\xi}\} &= \frac{1}{9}\begin{bmatrix} +2 & -1 & -1\\ -1 & +2 & -1\\ -1 & -1 & +2 \end{bmatrix}\hat{\sigma}^2(\text{BIQE}).
\end{aligned}
$$

For practice, we recommend $D\{\hat{\xi}(\text{BLUMBE}),\ \hat{\sigma}^2(\text{BIQE})\}$, since the dispersion matrix $D\{\hat{\xi}\}$ is remarkably smaller when compared to $D\{\hat{\xi}(\text{BLUMBE}), \hat{\sigma}^2(\text{BIQUUE})\}$, a result which seems to be unknown!

Bias vector β(BLUMBE) of the unknown vector of "*fixed effects*"

$$
\begin{aligned}
\beta &= -[\mathbf{I}_3 - \mathbf{A}'(\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-}\mathbf{A}\mathbf{A}'\mathbf{A}]\xi,\\
\beta &= -\frac{1}{3}\begin{bmatrix} 1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1 \end{bmatrix}\xi
\end{aligned}
$$

"*replace* ξ which is inaccessible by $\hat{\xi}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE})$"

$$
\beta = -\frac{1}{3}\begin{bmatrix} 1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1 \end{bmatrix}\hat{\xi}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE}),
$$

$$
\beta = 0
$$

$$
(\text{due to } \hat{\xi}_1 + \hat{\xi}_2 + \hat{\xi}_3 = 0).
$$

Mean Square Estimation Error $MSE\{\hat{\xi}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE})\}$

$$MSE\{\hat{\xi}\} = D\{\hat{\xi}\} + [\mathbf{I}_3 - \mathbf{A}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{AA}'\mathbf{A}]\sigma^2,$$

$$MSE\{\hat{\xi}\} = -\frac{\sigma^2}{9}\begin{bmatrix} 5 & 2 & 2 \\ 2 & 5 & 2 \\ 2 & 2 & 5 \end{bmatrix}.$$

replace σ^2 by $\hat{\sigma}^2$ (BIQUUE): $\hat{\sigma}^2 = (n - \text{rk}\mathbf{A})^{-1}\mathbf{e}'_\mathbf{y}\mathbf{e}_\mathbf{y}$

$$\hat{\sigma}^2(\text{BIQUUE}) = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1),$$

$$MSE\{\hat{\xi}\} = -\frac{1}{9}\begin{bmatrix} 5 & 2 & 2 \\ 2 & 5 & 2 \\ 2 & 2 & 5 \end{bmatrix}\hat{\sigma}^2(\text{BIQUUE}).$$

"*replace* σ^2 by $\hat{\sigma}^2$ (BIQE): $\hat{\sigma}^2 = (n + 2 - \text{rk}\mathbf{A})^{-1}\mathbf{e}'_\mathbf{y}\mathbf{e}_\mathbf{y}$"

$$\hat{\sigma}^2(\text{BIQE}) = \frac{1}{9}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)$$

$$MSE\{\hat{\xi}\} = -\frac{1}{9}\begin{bmatrix} 5 & 2 & 2 \\ 2 & 5 & 2 \\ 2 & 2 & 5 \end{bmatrix}\hat{\sigma}^2(\text{BIQE}).$$

Residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$ of the "random effect" $\mathbf{e}_\mathbf{y}$

$$\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{I}_3, \mathbf{I}_3\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}']\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{1}{3}(y_1 + y_2 + y_3)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \sigma^2[\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}']$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{\sigma^2}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

"*replace* σ^2 by $\hat{\sigma}^2$ (BIQUUE) or $\hat{\sigma}^2$ (BIQE)":

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\hat{\sigma}^2(\text{BIQUUE})$$

or

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \hat{\sigma}^2(\text{BIQE}).$$

Finally note that $\hat{\xi}$ $(\mathbf{I}_3, \mathbf{I}_3-\text{BLUMBE})$ corresponds $\mathbf{x}_{\mathbf{lm}}(\mathbf{I}_3, \mathbf{I}_3-\mathit{MINOLESS})$ discussed in Chapter 5. In addition, $D\{\tilde{\mathbf{e}}_{\mathbf{y}}|\mathbf{I}_3, \mathbf{I}_3-\mathit{BLUUE}\} = D\{\tilde{\mathbf{e}}_{\mathbf{y}}|\mathbf{I}_3, \mathbf{I}_3-\text{BLUMBE}\}$.

6-122 The First Example: V, S-BLUMBE

In the second case, we assume

a dispersion matrix $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$ of weighted observations$[y_1, y_2, y_3]$	*and*	*a weighted substitute* matrix **S**, in short w.s.

Under such a specification $\hat{\xi}$ is V, S-BLUMBE of ξ in the *special Gauss–Markov model* with datum defect.

$$\hat{\xi} = \mathbf{SA}'(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-1}\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{y}.$$

As dispersion matrix $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$ we choose

$$\mathbf{V} = \frac{1}{2}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \ \text{rk}\mathbf{V} = 3 = n$$

$$\mathbf{V}^{-1} = \frac{1}{2}\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}, \ \mathit{but}$$

$$\mathbf{S} = \mathit{Diag}(0, 1, 1), \ \text{rk}\mathbf{S} = 2$$

as the *bias semi-norm*. The matrix **S** fulfils the *condition*

$$\text{rk}(\mathbf{SA}') = \text{rk}\mathbf{A} = r = 2.$$

?How did we compute the g-inverse $(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-}$?

The computation of the g-inverse $(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-}$ has been based upon bordering.

$$\mathbf{V}^{-1} = \frac{1}{2}\begin{bmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{bmatrix}, \ \mathbf{S} = Diag(0,1,1), \ \mathrm{rk}\mathbf{S} = 2$$

$$\hat{\xi} = \mathbf{SA}'(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-}\mathbf{ASA}'\mathbf{V}^{-1}$$

$$\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}' = 2\begin{bmatrix} 2 & -3 & 1 \\ -3 & 6 & -3 \\ 1 & -3 & 2 \end{bmatrix}$$

$$(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-} = \frac{1}{6}\begin{bmatrix} 2 & 0 & -1 \\ 0 & 0 & -3 \\ -1 & 0 & 2 \end{bmatrix}$$

$$\hat{\xi} = \begin{bmatrix} h_\alpha \\ h_\beta \\ h_\gamma \end{bmatrix}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}} = \frac{1}{3}\begin{bmatrix} 0 \\ 2y_1 - y_2 - y_3 \\ y_1 + y_2 - 2y_3 \end{bmatrix}.$$

Dispersion matrix $D\{\hat{\xi}\}$ of the unknown vector of "*fixed effects*"

$$D\{\hat{\xi}\} = \sigma^2\mathbf{SA}'(\mathbf{ASA}'\mathbf{V}^{-1}\mathbf{ASA}')^{-}\mathbf{AS}$$

$$D\{\hat{\xi}\} = \frac{\sigma^2}{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

"replace σ^2 by $\hat{\sigma}^2$ (BIQUUE): $\hat{\sigma}^2 = (n - \mathrm{rk}\mathbf{A})^{-1}\tilde{\mathbf{e}}_\mathbf{y}'\tilde{\mathbf{e}}_\mathbf{y}$"

$$\tilde{\mathbf{e}}_\mathbf{y} = (\mathbf{V},\mathbf{S}\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{y_1 + y_2 + y_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_\mathbf{y}'\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{9}\mathbf{y}'\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y}'\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)$$

$$\hat{\sigma}^2(\mathrm{BIQUUE}) = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)$$

$$D\{\hat{\xi}\} = [\mathbf{V} + \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}']\hat{\sigma}^2(\text{BIQUUE})$$

$$D\{\hat{\xi}\} = \frac{2}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\hat{\sigma}^2(\text{BIQUUE}).$$

"replace σ^2 by $\hat{\sigma}^2$ (BIQE): $\hat{\sigma}^2 = (n + 2 - \text{rk}\mathbf{A})^{-1}\tilde{\mathbf{e}}'_{\mathbf{y}}\tilde{\mathbf{e}}_{\mathbf{y}}$"

$$\tilde{\mathbf{e}}_{\mathbf{y}}(\mathbf{V},\mathbf{S}\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{y_1 + y_2 + y_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\hat{\sigma}^2(\text{BIQE}) = \frac{1}{9}(y_1^2 + y_2^2 + y_3^2 + 2y_1y_2 + 2y_2y_3 + 2y_3y_1)$$

$$D\{\hat{\xi}\} = \frac{1}{9}\begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix}\hat{\sigma}^2(\text{BIQE}).$$

We repeat the statement that we recommend the use of $D\{\hat{\xi}(\text{BLUMBE}),\ \hat{\sigma}(\text{BIQE})\}$ since the dispersion matrix $D\{\hat{\xi}\}$ is remarkably smaller when compared to $D\{\hat{\xi}\ (\text{BLUMBE}),\ \hat{\sigma}^2(\text{BIQUUE})\}$!

Bias vector β(BLUMBE) of the unknown vector of "*fixed effects*"

$$\beta = -[\mathbf{I}_3 - \mathbf{S}\mathbf{A}'(\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}')^{-}\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}]\xi$$

$$\beta = -\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\xi = -\begin{bmatrix} \xi_1 \\ \xi_1 \\ \xi_1 \end{bmatrix},$$

"replace ξ which is inaccessible by $\hat{\xi}$ (V,S-BLUMBE)"

$$\beta = -\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\hat{\xi},\ (\mathbf{V},\mathbf{S}\text{–BLUMBE}) \neq 0.$$

Mean Square Estimation Error $MSE\{\hat{\xi}(\mathbf{V},\mathbf{S}\text{–BLUMBE})\}$

$$MSE\{\hat{\xi}\} = D\{\hat{\xi}\} + [\mathbf{S} - \mathbf{S}\mathbf{A}'(\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}')^{-}\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{V}^{-1}\mathbf{A}\mathbf{S}]\sigma^2$$

$$MSE\{\hat{\xi}\} = \frac{\sigma^2}{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = D\{\hat{\xi}\}.$$

"replace σ^2 by $\hat{\sigma}^2$ (BIQUUE): $\hat{\sigma}^2 = (n - \text{rk}\mathbf{A})^{-1}\mathbf{e}'_\mathbf{y}\mathbf{e}_\mathbf{y}$"

$$\hat{\sigma}^2(\text{BIQUUE}) = 3\sigma^2$$

$$MSE\{\hat{\xi}\} = D\{\hat{\xi}\}.$$

Residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$ of the "random effect" $\mathbf{e}_\mathbf{y}$

$$\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{V}, \mathbf{S}\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{y_1 + y_2 + y_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \sigma^2[\mathbf{V} - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}']$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{2}{3}\sigma^2\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

"*replace* σ^2 by $\hat{\sigma}^2$ (BIQE): $\hat{\sigma}^2 = (n + 2 - \text{rk}\mathbf{A})^{-1}\mathbf{e}'_\mathbf{y}\mathbf{e}_\mathbf{y}$"

$$\hat{\sigma}^2(\text{BIQE}) versus$$

$$MSE\{\hat{\xi}\} = \frac{1}{6}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}\sigma^2(\text{BIQE}).$$

Residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$ of the "random effects" $\mathbf{e}_\mathbf{y}$

$$\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{V}, \mathbf{S}\text{–BLUMBE}) = [\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{y} = \frac{y_1 + y_2 + y_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \sigma^2[\mathbf{V} - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}']$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{2}{3}\sigma^2\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

"*replace* σ^2 by $\hat{\sigma}^2$ (BIQUUE) or $\hat{\sigma}^2$ (BIQE)":

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{2}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\hat{\sigma}^2(\text{BIQUUE})$$

or

$$D\{\tilde{\mathbf{e}}_\mathbf{y}\} = \frac{2}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\hat{\sigma}^2(\text{BIQE}).$$

6-123 The First Example: $\mathbf{I}_3, \mathbf{I}_3$-BLE

In the third case, we assume

a dispersion matrix $D\{\mathbf{y}\} = \mathbf{I}_3\sigma^2$ *observations* $[y_1, y_2, y_3]$.	*and*	a *unity substitute* matrix $\mathbf{S} = \mathbf{I}_3$, in short u.s.

Under such a specification $\hat{\xi}$ is $\mathbf{I}_3, \mathbf{I}_3$-BLE of ξ in the *special Gauss–Markov model* with datum defect.

$$\hat{\xi}(\text{BLE}) = (\mathbf{I}_3 + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$$

$$\mathbf{I}_3 + \mathbf{A}'\mathbf{A} = \begin{bmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{bmatrix}, \qquad (\mathbf{I}_3 + \mathbf{A}'\mathbf{A})^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$\hat{\xi}(\text{BLE}) = \frac{1}{4}\begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}\mathbf{y} = \frac{1}{4}\begin{bmatrix} -y_1 + y_3 \\ +y_1 - y_2 \\ +y_2 - y_3 \end{bmatrix}.$$

$$\hat{\xi}_1 + \hat{\xi}_2 + \hat{\xi}_3 = 0.$$

Dispersion matrix $D\{\hat{\xi}|\text{BLE}\}$ of the unknown vector of "*fixed effects*"

$$D\{\hat{\xi}|\text{BLE}\} = \sigma^2\mathbf{A}'\mathbf{A}(\mathbf{I}_3 + \mathbf{A}'\mathbf{A})^{-2}$$

$$D\{\hat{\xi}|\text{BLE}\} = \frac{\sigma^2}{16}\begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix}.$$

Bias vector $\beta(\text{BLE})$ of the unknown vector of "*fixed effects*"

$$\beta = -[\mathbf{I}_3 + \mathbf{A}'\mathbf{A}]^{-1}\xi$$

$$\beta = -\frac{1}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xi = -\frac{1}{4}\begin{bmatrix} 2\xi_1 + \xi_2 + \xi_3 \\ \xi_1 + 2\xi_2 + \xi_3 \\ \xi_1 + \xi_2 + 2\xi_3 \end{bmatrix}.$$

Mean Square Estimation Error $MSE\{\xi(\text{BLE})\}$

$$MSE\{\hat{\xi}(\text{BLE})\} = \sigma^2[\mathbf{I}_3 + \mathbf{A}'\mathbf{A}]^{-1}$$

$$MSE\{\hat{\xi}(\text{BLE})\} = \frac{\sigma^2}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$ of the "*random effect*" $\mathbf{e}_\mathbf{y}$

$$\tilde{\mathbf{e}}_\mathbf{y}(\text{BLE}) = [\mathbf{A}\mathbf{A}' + \mathbf{I}_3]^{-1}\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y}(\text{BLE}) = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}\mathbf{y} = \frac{1}{4}\begin{bmatrix} 2y_1 + y_2 + y_3 \\ y_1 + 2y_2 + y_3 \\ y_1 + y_2 + 2y_3 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}(\text{BLE})\} = \sigma^2[\mathbf{I}_3 + \mathbf{A}\mathbf{A}']^{-2}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}(\text{BLE})\} = \frac{\sigma^2}{16}\begin{bmatrix} 6 & 5 & 5 \\ 5 & 6 & 5 \\ 5 & 5 & 6 \end{bmatrix}.$$

Correlations

$$C\{\tilde{\mathbf{e}}_\mathbf{y}, \mathbf{A}\hat{\boldsymbol{\xi}}\} = \sigma^2[\mathbf{I}_3 + \mathbf{A}\mathbf{A}']^{-2}\mathbf{A}\mathbf{A}'$$

$$C\{\tilde{\mathbf{e}}_\mathbf{y}, \mathbf{A}\hat{\boldsymbol{\xi}}\} = \frac{\sigma^2}{16}\begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix}.$$

Comparisons BLUMBE-BLE

$$(i)\ \hat{\xi}_{\text{BLUMBE}} - \hat{\xi}_{\text{BLE}}$$

$$\hat{\xi}_{\text{BLUMBE}} - \hat{\xi}_{\text{BLE}} = \mathbf{A}'(\mathbf{A}\mathbf{A}'\mathbf{A}\mathbf{A}')^{-}\mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{I}_3)^{-1}\mathbf{y}$$

$$\hat{\xi}_{\text{BLUMBE}} - \hat{\xi}_{\text{BLE}} = \frac{1}{12}\begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}\mathbf{y} = \frac{1}{12}\begin{bmatrix} -y_1 + y_3 \\ +y_1 - y_2 \\ +y_2 - y_3 \end{bmatrix}.$$

$$(ii)\ D\{\hat{\xi}_{\text{BLUMBE}}\} - D\{\hat{\xi}_{\text{BLE}}\}$$

$$\begin{aligned} &D\{\hat{\xi}_{\text{BLUMBE}}\} - D\{\hat{\xi}_{\text{BLE}}\} \\ &= \sigma^2 \mathbf{A}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{AA}'(\mathbf{AA}' + \mathbf{I}_3)^{-1}\mathbf{AA}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{A} \\ &+ \sigma^2 \mathbf{A}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{AA}'(\mathbf{AA}' + \mathbf{I}_3)^{-1}\mathbf{AA}'(\mathbf{AA}' + \mathbf{I}_3)^{-1}\mathbf{AA}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{A} \end{aligned}$$

$$D\{\hat{\xi}_{\text{BLUMBE}}\} - D\{\hat{\xi}_{\text{BLE}}\} = \frac{7}{144}\sigma^2 \begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix} \textit{ positive semidefinite}.$$

$$(iii)\ MSE\{\hat{\xi}_{\text{BLUMBE}}\} - MSE\{\hat{\xi}_{\text{BLE}}\}$$

$$\begin{aligned} &MSE\{\hat{\xi}_{\text{BLUMBE}}\} - MSE\{\hat{\xi}_{\text{BLE}}\} \\ &= \sigma^2 \mathbf{A}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{AA}'(\mathbf{AA}' + \mathbf{I}_3)^{-1}\mathbf{AA}'(\mathbf{AA}'\mathbf{AA}')^{-}\mathbf{A} \end{aligned}$$

$$MSE\{\hat{\xi}_{\text{BLUMBE}}\} - MSE\{\hat{\xi}_{\text{BLE}}\} = \frac{\sigma^2}{36} \begin{bmatrix} +2 & -1 & -1 \\ -1 & +2 & -1 \\ -1 & -1 & +2 \end{bmatrix} \textit{ positive semidefinite}.$$

6-124 The First Example: V, S-BLE

In the fourth case, we assume

a dispersion matrix $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$ of	and	a *weighted substitute* matrix $\mathbf{S}$,
weighted observations $[y_1, y_2, y_3]$		in short w.s. .

We choose

$$\mathbf{V} = \frac{1}{2}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \textit{ positive definite}, \ \text{rk}\mathbf{V} = 3 = n,$$

$$\mathbf{V}^{-1} = \frac{1}{2}\begin{bmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{bmatrix},$$

and

$$\mathbf{S} = Diag(0, 1, 1), \ \ \text{rk}\mathbf{S} = 2,$$

$$\hat{\xi} = (\mathbf{I}_3 + \mathbf{SA}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{SA}'\mathbf{V}^{-1}\mathbf{y},$$

$$\mathbf{I}_3 + \mathbf{SA}'\mathbf{V}^{-1}\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 5 & -2 \\ -2 & -2 & 5 \end{bmatrix}, (\mathbf{I}_3 + \mathbf{SA}'\mathbf{V}^{-1}\mathbf{A})^{-1} = \frac{1}{21}\begin{bmatrix} 21 & 0 & 0 \\ 14 & 5 & 2 \\ 14 & 2 & 5 \end{bmatrix},$$

$$\hat{\xi}(\mathbf{V}, \mathbf{S}\text{–BLE}) = \begin{bmatrix} h_\alpha \\ h_\beta \\ h_\gamma \end{bmatrix}_{\mathbf{V},\mathbf{S}-\text{BLE}}$$

$$= \tfrac{1}{21}\begin{bmatrix} 0 & 0 & 0 \\ 14 & -6 & -4 \\ 4 & 6 & -10 \end{bmatrix}\mathbf{y} = \tfrac{1}{21}\begin{bmatrix} 0 \\ 10y_1 - 6y_2 - 4y_3 \\ 4y_1 + 6y_2 - 10y_3 \end{bmatrix}.$$

Dispersion matrix $D\{\hat{\xi}|\mathbf{V}, \mathbf{S}\text{–BLE}\}$ of the unknown vector of "*fixed effects*"

$$D\{\hat{\xi}|\mathbf{V}, \mathbf{S}\text{–BLE}\} = \sigma^2\mathbf{SA}'\mathbf{V}^{-1}\mathbf{A}[\mathbf{I}_3 + \mathbf{SA}'\mathbf{V}^{-1}\mathbf{A}]^{-1}\mathbf{S},$$

$$D\{\hat{\xi}|\mathbf{V}, \mathbf{S}\text{–BLE}\} = \frac{\sigma^2}{441}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 76 & 22 \\ 0 & 22 & 76 \end{bmatrix}.$$

Bias vector $\beta(\mathbf{V}, \mathbf{S}\text{–BLE})$ of the unknown vector of "*fixed effects*"

$$\beta = -[\mathbf{I}_3 + \mathbf{SA}'\mathbf{V}^{-1}\mathbf{A}]^{-1}\xi$$

$$\beta = -\frac{1}{21}\begin{bmatrix} 21 & 0 & 0 \\ 14 & 5 & 2 \\ 14 & 2 & 5 \end{bmatrix}\xi = -\frac{1}{21}\begin{bmatrix} 21\xi_1 \\ 14\xi_1 + 5\xi_2 + 2\xi_3 \\ 14\xi_1 + 2\xi_2 + 5\xi_3 \end{bmatrix}.$$

Mean Square Estimation Error $MSE\{\xi|\mathbf{V}, \mathbf{S}\text{–BLE}\}$

$$MSE\{\xi|\mathbf{V}, \mathbf{S}\text{–BLE}\} = \sigma^2[\mathbf{I}_3 + \mathbf{SA}'\mathbf{VA}]^{-1}\mathbf{S}$$

$$MSE\{\xi|\mathbf{V}, \mathbf{S}\text{–BLE}\} = \frac{\sigma^2}{21}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix}.$$

Residual vector $\tilde{\mathbf{e}}_\mathbf{y}$ and dispersion matrix $D\{\tilde{\mathbf{e}}_\mathbf{y}\}$ of the "*random effect*" $\mathbf{e}_\mathbf{y}$

$$\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{V}, \mathbf{S}\text{–BLE}) = [\mathbf{I}_3 + \mathbf{ASA}'\mathbf{V}^{-1}]^{-1}\mathbf{y}$$

$$\tilde{\mathbf{e}}_\mathbf{y}\{\mathbf{V}, \mathbf{S}\text{–BLE}\} = \frac{1}{21}\begin{bmatrix} 11 & 6 & 4 \\ 6 & 9 & 6 \\ 4 & 6 & 11 \end{bmatrix}\mathbf{y} = \frac{1}{21}\begin{bmatrix} 11y_1 + 6y_2 + 4y_3 \\ 6y_1 + 9y_2 + 6y_3 \\ 4y_1 + 6y_2 + 11y_3 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_\mathbf{y}(\mathbf{V}, \mathbf{S}\text{–BLE})\} = \sigma^2[\mathbf{I}_3 + \mathbf{ASA}'\mathbf{V}^{-1}]^{-2}\mathbf{V}$$

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}(\mathbf{V},\mathbf{S}{-}\mathrm{BLE})\} = \frac{\sigma^2}{882}\begin{bmatrix} 614 & 585 & 565 \\ 585 & 594 & 585 \\ 565 & 585 & 614 \end{bmatrix}.$$

Correlations

$$C\{\tilde{\mathbf{e}}_{\mathbf{y}}, \mathbf{A}\hat{\boldsymbol{\xi}}\} = \sigma^2(\mathbf{I}_3 + \mathbf{ASA'V}^{-1})^{-2}\mathbf{ASA'}$$

$$C\{\tilde{\mathbf{e}}_{\mathbf{y}}, \mathbf{A}\hat{\boldsymbol{\xi}}\} = \frac{\sigma^2}{441}\begin{bmatrix} 29 & -9 & -20 \\ -9 & 18 & -9 \\ -20 & -9 & 29 \end{bmatrix}.$$

Comparisons BLUMBE-BLE

$$(i)\ \hat{\xi}_{\mathrm{BLUMBE}} - \hat{\xi}_{\mathrm{BLE}}$$

$$\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}} - \hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}} = \mathbf{SA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{ASA'}(\mathbf{ASA'}+\mathbf{V})^{-1}\mathbf{y}$$

$$\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}} - \hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}} = \frac{1}{21}\begin{bmatrix} 0 & 0 & 0 \\ 4 & -1 & 3 \\ 3 & 1 & -4 \end{bmatrix}\mathbf{y} = \frac{1}{21}\begin{bmatrix} 0 \\ 4y_1 - y_2 - 3y_3 \\ 3y_1 + y_2 - 4y_3 \end{bmatrix}.$$

$$(ii)\ D\{\hat{\xi}_{\mathrm{BLUMBE}}\} - D\{\hat{\xi}_{\mathrm{BLE}}\}$$

$$\begin{gathered} D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\} \\ = \sigma^2\mathbf{SA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{ASA'}(\mathbf{ASA'}+\mathbf{V})^{-1} \\ \mathbf{ASA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{AV} + \\ \sigma^2\mathbf{SA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{ASA'}(\mathbf{ASA'}+\mathbf{V})^{-1}\mathbf{ASA'}(\mathbf{ASA'}+\mathbf{V})^{-1} \\ \mathbf{ASA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{AS} \end{gathered}$$

$$D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\} = \frac{\sigma^2}{882}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 142 & 103 \\ 0 & 103 & 142 \end{bmatrix}\ \textit{positive semidefinite.}$$

$$(iii)\ MSE\{\hat{\xi}_{\mathrm{BLUMBE}}\} - MSE\{\hat{\xi}_{\mathrm{BLE}}\}$$

$$\begin{gathered} MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\} \\ = \sigma^2\mathbf{SA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{ASA'}(\mathbf{ASA'}+\mathbf{V})^{-1}\mathbf{ASA'}(\mathbf{ASA'V}^{-1}\mathbf{ASA'})^{-}\mathbf{AS} \end{gathered}$$

$$MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\} = \frac{\sigma^2}{42}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 3 \\ 0 & 3 & 4 \end{bmatrix}\ \textit{positive semidefinite}.$$

Summarizing, let us compare

$$\mathbf{I},\mathbf{I}-\texttt{BLUMBE versus}\,\mathbf{I},\mathbf{I}-\mathrm{BLE}$$
$$and$$
$$\mathbf{V},\mathbf{S}-\texttt{BLUMBE versus}\,\mathbf{V},\mathbf{S}-\mathrm{BLE!}$$

$\hat{\xi}_{\mathrm{BLUMBE}} - \hat{\xi}_{\mathrm{BLE}}$, $D\{\hat{\xi}_{\mathrm{BLUMBE}}\} - D\{\hat{\xi}_{\mathrm{BLE}}\}$ and $MSE\{\hat{\xi}_{\mathrm{BLUMBE}}\} - MSE\{\hat{\xi}_{\mathrm{BLE}}\}$ as well as $\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}} - \hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}$, $D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - D\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\}$ and $MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLUMBE}}\} - MSE\{\hat{\xi}_{\mathbf{V},\mathbf{S}-\mathrm{BLE}}\}$ result positive semidefinite: In consequence, for three different measures of distortions

BLE *is in favor of* BLIMBE:

BLE produces smaller errors in comparing with BLIMBE! Finally let us compare weighted BIQUUE and weighted BIQE:
(i) Weighted BIQUUE $\hat{\sigma}^2$ and weighted BIQE $\hat{\sigma}^2$

$$\begin{array}{lcl} \hat{\sigma}^2 = (n-r)^{-1}\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} & & \hat{\sigma}^2 = (n-r+2)\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} = \\ = (n-r)^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} & versus & = (n-r+2)\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} \end{array}$$

$$(\tilde{\mathbf{e}}_{\mathbf{y}})_{\mathbf{V}},\mathbf{S}\text{–BLUMBE} = \frac{1}{6}\begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix}\mathbf{y}$$

$$r = \mathrm{rk}\mathbf{A} = 2,\ n = 3,\ \mathbf{V}^{-1} = \frac{1}{2}\begin{bmatrix} +3 & -1 & -1 \\ -1 & +3 & -1 \\ -1 & -1 & +3 \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{2}(y_1^2 + y_2^2 + y_3^2)\ \ versus\ \ \hat{\sigma}^2 = \frac{1}{6}(y_1^2 + y_2^2 + y_3^2)$$

$$(ii)\ D\{\hat{\sigma}^2|\mathrm{BIQUUE}\}\ \ versus\ \ D\{\hat{\sigma}^2|\mathrm{BIQE}\}$$

$$D\{\hat{\sigma}^2\} = 2(n-r)^{-1}\sigma^4\ \ versus\ \ D\{\hat{\sigma}^2\} = 2(n-r)(n-r+2)^{-1}\sigma^4$$

$$D\{\hat{\sigma}^2\} = 2\sigma^4\ \ versus\ \ D\{\hat{\sigma}^2\} = \frac{2}{9}\sigma^4$$

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-r)^{-1}(\hat{\sigma}^2)^2\ \ versus\ \ \begin{array}{l} \hat{E}\{\hat{\sigma}^2 - \sigma^2\} = \\ = -2(n-r+2)^{-1}\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}} \end{array}$$

$$\hat{D}\{\hat{\sigma}^2\} = \frac{1}{2}(y_1^2 + y_2^2 + y_3^2)\ \ versus\ \ \begin{array}{l} \hat{E}\{\hat{\sigma}^2 - \sigma^2\} = -\frac{1}{9}(y_1^2 + y_2^2 + y_3^2) \\ \hat{E}\{\hat{\sigma}^2 - \sigma^2\} = 2(n-r+2)^{-1}(\tilde{\mathbf{e}}_{\mathbf{y}}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}})^2 \\ \hat{E}\{(\hat{\sigma}^2 - \sigma^2)\} = \frac{1}{54}(y_1^2 + y_2^2 + y_3^2). \end{array}$$

$$(iii)\ \ (\tilde{\mathbf{e}}_{\mathbf{y}})_{\mathrm{BLUMBE}} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}\mathbf{V}^{-1}](\tilde{\mathbf{e}}_{\mathbf{y}})_{\mathrm{BLE}}$$

$$(\hat{\sigma}^2)_{\mathrm{BIQUUE}} = (n-r)(\tilde{\mathbf{e}}_{\mathbf{y}}')_{\mathrm{BLE}}[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-}\mathbf{A}\mathbf{V}^{-1}](\tilde{\mathbf{e}}_{\mathbf{y}})_{\mathrm{BLE}}$$

$$\hat{\sigma}^2_{\text{BIQUUE}} - \hat{\sigma}^2_{\text{BIQE}} = \frac{1}{3}(y_1^2 + y_2^2 + y_3^2)\ positive.$$

We repeat that the difference $\hat{\sigma}^2_{\text{BIQUUE}} - \hat{\sigma}^2_{\text{BIQE}}$ is in favor of $\hat{\sigma}^2_{\text{BIQE}} < \hat{\sigma}^2_{\text{BIQUUE}}$.

6-2 Setup of the Best Linear Estimators of Type hom BLE, hom S-BLE and hom a-BLE for Fixed Effects

The topic of this section has been the subject of *Jianqing Cai's* Ph.D. Thesis, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaft, Report C 377, München 2004.

Numerical tests have documented that $\hat{\xi}$ of type $\boldsymbol{\Sigma}$-BLUUE of ξ is *not* robust against *outliers* in the stochastic vector $\mathbf{y}$ of observations. It is for this reason that *we give up* the postulate of unbiasedness, but keeping the set up of a *linear estimation* $\hat{\xi} = \mathbf{L}\mathbf{y}$ of homogeneous type. The matrix $\mathbf{L}$ is uniquely determined by the α-*weighted hybrid norm* of type minimum variance $||D\{\hat{\xi}\}||^2$ and minimum bias $||\mathbf{I} - \mathbf{LA}||^2$. For such a homogeneous linear estimation (2.21) by means of *Box 6.4* let us specify the real-valued, nonstochastic *bias vector* $\beta : = E\{\hat{\xi} - \xi\} = E\{\hat{\xi}\} - \xi$ of type (6.11), (6.12), (6.13) and the real-valued, non-stochastic *bias matrix* $\mathbf{I}_m - \mathbf{LA}$ (6.74) in more detail.

First, let us discuss why a setup of an inhomogeneous linear estimation is not suited to solve our problem. In the case of an unbiased estimator, the setup of an inhomogeneous linear estimation $\hat{\xi} = \mathbf{L}\mathbf{y} + \boldsymbol{\kappa}$ led us to $E\{\hat{\xi}\} = \xi$ the postulate of unbiasedness if and only if $E\{\hat{\xi}\} - \xi \ : \ = \mathbf{L}E\{\mathbf{y}\} - \xi + \boldsymbol{\kappa} = -(\mathbf{I}_m - \mathbf{LA})\xi + \boldsymbol{\kappa} = 0$ for all $\xi \in \mathbb{R}^m$ or $\mathbf{LA} = \mathbf{I}$ and $\boldsymbol{\kappa} = 0$. Indeed the postulate of unbiasedness restricted the linear operator L to be the (non-unique) *left inverse* $\mathbf{L} = \mathbf{A}^-_{\mathbf{L}}$ as well as the vector $\boldsymbol{\kappa}$ of inhomogeneity to zero. In contrast the bias vector $\beta : = E\{\hat{\xi} - \xi\} = E\{\hat{\xi}\} - \xi = \mathbf{L}E\{\mathbf{y}\} - \xi = -(\mathbf{I}_m - \mathbf{LA})\xi + \boldsymbol{\kappa}$ for a setup of an inhomogeneous linear estimation should approach zero if $\xi = 0$ is chosen as a special case. In order to include this case in the linear biased estimation procedure we set $\boldsymbol{\kappa} = 0$.

Second, we focus on the norm (2.79) namely $||\beta||^2 \ : \ = E\{(\hat{\xi} - \xi)'(\hat{\xi} - \xi)\}$ of the *bias vector* β, also called *Mean Square Error* $MSE\{\hat{\xi}\}$ of $\hat{\xi}$. In terms of a setup of a homogeneous linear estimation, $\hat{\xi} = \mathbf{L}\mathbf{y}$, the norm of the *bias vector* is represented by $(\mathbf{I}_m - \mathbf{LA})'\xi\xi'(\mathbf{I}_m - \mathbf{LA})$ or by the *weighted Frobenius matrix norm* $||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'}$ where the weight matrix $\xi\xi'$, $\text{rk}\xi\xi' = 1$, has *rank one*. Obviously $||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'}$ is only a semi-norm. In addition, $\xi\xi'$ is not accessible since ξ is unknown. In this problematic case we replace the matrix $\xi\xi'$ by a fixed, *positive definite* $m \times m$ matrix $\mathbf{S}$ and define the $\mathbf{S}$-*weighted Frobenius matrix norm* $||(\mathbf{I}_m - \mathbf{LA})'||^2_{\mathbf{S}}$ of type (6.6) of the *bias matrix* $\mathbf{I}_m - \mathbf{LA}$. Indeed by means of the rank identity, $rk\mathbf{S} = m$ we have chosen a weight matrix of maximal rank. Now we are prepared to understand intuitively the following.

Here we focus on best linear estimators of type hom BLE, hom S-BLE and hom a-BLE of *fixed effects* ξ, which turn out to be better than the best linear uniformly unbiased estimator of type hom BLUUE, but suffer from the effect to be biased. At first let us begin with a discussion about the bias vector and the bias matrix as well as the *Mean Square Estimation Error* $MSE\{\hat{\xi}\}$ with respect to a homogeneous linear estimation $\hat{\xi} = \mathbf{Ly}$ of fixed effects ξ based upon *Box 6.4*.

Box 6.4. (Bias vector, bias matrix Mean Square Estimation Error in the special Gauss–Markov model with fixed effects):

$$E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi} \tag{6.71}$$

$$D\{\mathbf{y}\} = \boldsymbol{\Sigma}_{\mathbf{y}} \tag{6.72}$$

"*ansatz*"

$$\hat{\xi} = \mathbf{Ly} \tag{6.73}$$

bias vector

$$\beta := E\{\hat{\xi} - \xi\} = E\{\hat{\xi}\} - \xi \tag{6.74}$$

$$\beta = \mathbf{L}E\{\mathbf{y}\} - \xi = -[\mathbf{I}_m - \mathbf{LA}]\,\xi \tag{6.75}$$

bias matrix

$$\mathbf{B} := \mathbf{I}_m - \mathbf{LA} \tag{6.76}$$

decomposition

$$\hat{\xi} - \xi = (\hat{\xi} - E\{\hat{\xi}\}) + (E\{\hat{\xi}\} - \xi) \tag{6.77}$$

$$\hat{\xi} - \xi = \mathbf{L}(\mathbf{y} - E\{\mathbf{y}\}) - [\mathbf{I}_m - \mathbf{LA}]\,\xi \tag{6.78}$$

Mean Square Estimation Error

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \tag{6.79}$$

$$MSE\{\hat{\xi}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + [\mathbf{I}_m - \mathbf{LA}]\,\xi\xi'\,[\mathbf{I}_m - \mathbf{LA}]' \tag{6.80}$$

$$(E\{\hat{\xi} - E\{\hat{\xi}\}\} = 0)$$

modified Mean Square Estimation Error

$$MSE_{\mathbf{S}}\{\hat{\xi}\} := \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + [\mathbf{I}_m - \mathbf{LA}]\,\mathbf{S}\,[\mathbf{I}_m - \mathbf{LA}]' \tag{6.81}$$

Frobenius matrix norms

$$||MSE\{\hat{\xi}\}||^2 : = \text{tr}\, E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \tag{6.82}$$

$$\begin{aligned} &||MSE\{\hat{\xi}\}||^2 \\ &= \text{tr}\, \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \text{tr}\, [\mathbf{I}_m - \mathbf{LA}]\, \xi\xi'\, [\mathbf{I}_m - \mathbf{LA}]' \\ &= ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'} \end{aligned} \tag{6.83}$$

$$\begin{aligned} &||MSE_\mathbf{S}\{\hat{\xi}\}||^2 \\ : &= \text{tr}\, \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \text{tr}\, [\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]' \\ &= ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_\mathbf{S} \end{aligned} \tag{6.84}$$

hybrid minimum variance – minimum bias norm α-weighted

$$\mathcal{L}(\mathbf{L}) : = ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + \tfrac{1}{\alpha}||(\mathbf{I}_m - \mathbf{LA})'||^2_\mathbf{S} \tag{6.85}$$

special model

$$\dim\, \mathcal{R}(\mathbf{SA}') = \text{rk}\mathbf{SA}' = \text{rk}\mathbf{A} = m. \tag{6.86}$$

The bias vector β is conventionally defined by $E\{\hat{\xi}\} - \xi$ subject to the homogeneous estimation form $\hat{\xi} = \mathbf{Ly}$. Accordingly the bias vector can be represented by (6.75) $\beta = -[\mathbf{I}_m - \mathbf{LA}]\, \xi$. Since the vector ξ of *fixed effects* is unknown, there has been made the proposal to use instead the matrix $\mathbf{I}_m - \mathbf{LA}$ as a *matrix-valued measure of bias*. A measure of the estimation error is the *Mean Square estimation error* $\text{MSE}\{\hat{\xi}\}$ of type (6.79). $\text{MSE}\{\hat{\xi}\}$ can be decomposed into two basic parts:

- the dispersion matrix $D\{\hat{\xi}\} = \mathbf{L} D\{\mathbf{y}\}\mathbf{L}'$
- the bias product $\beta\beta'$.

Indeed the vector $\hat{\xi} - \xi$ can be decomposed as well into two parts of type (6.77), (6.78), namely (i) $\hat{\xi} - E\{\hat{\xi}\}$ and (ii) $E\{\hat{\xi}\} - \xi$ which may be called estimation error and bias, respectively. The double decomposition of the vector $\hat{\xi} - \xi$ leads straightforward to the double representation of the matrix $MSE\{\hat{\xi}\}$ of type (6.80). Such a representation suffers from two effects: Firstly the vector ξ of fixed effects is unknown, secondly the matrix $\xi\xi'$ has only rank 1. In consequence, the matrix $[\mathbf{I}_m - \mathbf{LA}]\, \xi\xi'\, [\mathbf{I}_m - \mathbf{LA}]'$ has only rank 1, too. In this situation there has been made a proposal to modify $MSE\{\hat{\xi}\}$ with respect to $\xi\xi'$ by the regular matrix $\mathbf{S}$. $MSE_\mathbf{S}\{\hat{\xi}\}$ has been defined by (6.81). A scalar measure of $MSE\{\hat{\xi}\}$ as well as $MSE_\mathbf{S}\{\hat{\xi}\}$ are the *Frobenius norms* (6.82), (6.83), (6.84). Those scalars constitute the optimal risk in *Definition 6.7* (hom BLE) and *Definition 6.8* (hom **S**-BLE). Alternatively a homogeneous α-weighted hybrid minimum variance–minimum bias estimation (hom α-BLE) is presented in *Definition 6.9* (hom α-BLE) which is based upon the weighted sum of two norms of type (6.85), namely

- Average variance $||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} = tr\mathbf{L}\Sigma_\mathbf{y}\mathbf{L}'$
- Average bias $||(\mathbf{I}_m - \mathbf{LA})'||^2_\mathbf{S} = tr\, [\mathbf{I}_m - \mathbf{LA}]\, \mathbf{S}\, [\mathbf{I}_m - \mathbf{LA}]'$.

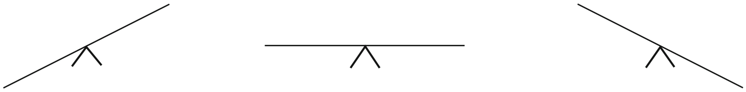


Fig. 6.1 Balance of variance and bias by the weight factor α

The very important estimator α-BLE is *balancing variance* and bias by the weight factor a which is illustrated by *Fig. 6.1*.

Definition 6.7. ($\hat{\xi}$ hom BLE of ξ):

An $m1$ vector $\hat{\xi}$ is called homogeneous BLE of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*, if
(1st) $\hat{\xi}$ is a homogeneous linear form

$$\hat{\xi} = \mathbf{Ly} \tag{6.87}$$

(2nd) in comparison to all other linear estimations $\hat{\xi}$ has the minimum **M**ean **S**quare **E**stimation **E**rror in the sense of

$$\begin{aligned} &||MSE\{\hat{\xi}\}||^2 \\ &= \operatorname{tr} \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \operatorname{tr} [\mathbf{I}_m - \mathbf{LA}]\, \xi\xi'\, [\mathbf{I}_m - \mathbf{LA}]' \\ &= ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_{\xi\xi'}. \end{aligned} \tag{6.88}$$

Definition 6.8. ($\hat{\xi}$ **S**-BLE of ξ):

An $m1$ vector $\hat{\xi}$ is called homogeneous **S**-BLE of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*, if
(1st) $\hat{\xi}$ is a homogeneous linear form

$$\hat{\xi} = \mathbf{Ly} \tag{6.89}$$

(2nd) in comparison to all other linear estimations $\hat{\xi}$ has the minimum **S**-modified **M**ean **S**quare **E**stimation **E**rror in the sense of

$$\begin{aligned} &||MSE_\mathbf{S}\{\hat{\xi}\}||^2 \\ &:= \operatorname{tr} \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \operatorname{tr} [\mathbf{I}_m - \mathbf{LA}]\mathbf{S}[\mathbf{I}_m - \mathbf{LA}]' \\ &= ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_m - \mathbf{LA})'||^2_\mathbf{S} = \min_\mathbf{L}. \end{aligned} \tag{6.90}$$

Definition 6.9. ($\hat{\xi}$ hom hybrid min var-min bias solution, a-weighted or hom a-BLE):

An $m \times 1$ vector $\hat{\xi}$ is called homogeneous α-weighted hybrid minimum variance–minimum bias estimation (hom a-BLE) of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*, if

(1st) $\hat{\xi}$ is a homogeneous linear form

$$\hat{\xi} = \mathbf{L}\mathbf{y} \tag{6.91}$$

(2nd) in comparison to all other linear estimations $\hat{\xi}$ has the minimum variance-minimum bias in the sense of the α-weighted hybrid norm

$$\begin{aligned} &\operatorname{tr} \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \tfrac{1}{\alpha}\operatorname{tr}(\mathbf{I}_m - \mathbf{L}\mathbf{A})\,\mathbf{S}\,(\mathbf{I}_m - \mathbf{L}\mathbf{A})' \\ &= \ ||\mathbf{L}'||^2_{\Sigma_{\mathbf{y}}} + \tfrac{1}{\alpha}||(\mathbf{I}_m - \mathbf{L}\mathbf{A})'||^2_{\mathbf{S}} = \min_{\mathbf{L}}, \end{aligned} \tag{6.92}$$

in particular with respect to the special model

$$\alpha \in \mathbb{R}^+, \ \dim \mathcal{R}(\mathbf{S}\mathbf{A}') = \operatorname{rk}\mathbf{S}\mathbf{A}' = \operatorname{rk}\mathbf{A} = m.$$

The estimations $\hat{\xi}$ hom BLE, hom **S**-BLE and hom α-BLE can be characterized as following:

Lemma 6.10. (hom BLE, hom **S**-BLE and hom α-BLE):

An m1 vector $\hat{\xi}$ is hom BLE, hom **S**-BLE or hom α-BLE of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*, if and only if the matrix $\hat{\mathbf{L}}$ fulfils the normal equations

(1st) hom BLE:

$$(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}')\hat{\mathbf{L}}' = \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}' \tag{6.93}$$

(2nd) hom **S**-BLE:

$$(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\mathbf{S}\mathbf{A}')\hat{\mathbf{L}}' = \mathbf{A}\mathbf{S} \tag{6.94}$$

(3rd) hom α-BLE:

$$(\boldsymbol{\Sigma}_{\mathbf{y}} + \tfrac{1}{\alpha}\mathbf{A}\mathbf{S}\mathbf{A}')\hat{\mathbf{L}}' = \tfrac{1}{\alpha}\mathbf{A}\mathbf{S} \tag{6.95}$$

:Proof:

(i) hom BLE:

The hybrid norm $||MSE\{\hat{\xi}\}||^2$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) : = \operatorname{tr} \mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' + \operatorname{tr}(\mathbf{I}_m - \mathbf{L}\mathbf{A})\,\xi\xi'\,(\mathbf{I}_m - \mathbf{L}\mathbf{A})' = \min_{\mathbf{L}}$$

for $\hat{\xi}$ hom BLE of ξ. The *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$ are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) : = 2[\Sigma_{\mathbf{y}}\hat{\mathbf{L}}' + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}'\hat{\boldsymbol{L}}' - \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'] = 0$$

which agree to the normal equations (6.93). (The theory of matrix derivatives is reviewed in Appendix A.7 (Facts: derivative of a scalar-valued function of a matrix: *trace*).

The second derivatives

$$\frac{\partial^2 \mathcal{L}}{\partial(\text{vec}\mathbf{L})\partial(\text{vec}\mathbf{L})'}(\hat{\mathbf{L}}) > 0$$

at the "point" $\hat{\mathbf{L}}$ constitute the *sufficiency conditions*. In order to compute such an $mn \times mn$ matrix of second derivatives we have to vectorize the matrix normal equation

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) := 2\hat{\mathbf{L}}(\Sigma_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}') - 2\xi\xi' A',$$

$$\frac{\partial \mathcal{L}}{\partial(\text{vec}\mathbf{L})}(\hat{\mathbf{L}}) := \text{vec}[2\hat{\mathbf{L}}(\Sigma_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}') - 2\xi\xi' A'].$$

(ii) hom S-BLE:

The hybrid norm $||MSE_{\mathbf{s}}\{\hat{\xi}\}||^2$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) : = \text{tr}\,\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' + \text{tr}\,(\mathbf{I}_m - \mathbf{LA})\,\mathbf{S}\,(\mathbf{I}_m - \mathbf{LA})' = \min_{\mathbf{L}}$$

for $\hat{\xi}$ *hom* **S**-BLE of ξ. Following the first part of the proof we are led to the *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) : = 2[\Sigma_{\mathbf{y}}\hat{\mathbf{L}}' + \mathbf{ASA}'\hat{\mathbf{L}}' - \mathbf{AS}]' = 0$$

as well as to the *sufficiency conditions*

$$\frac{\partial^2 \mathcal{L}}{\partial(\text{vec}\mathbf{L})\partial(\text{vec}\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}') \otimes \mathbf{I}_m] > 0.$$

The *normal equations* of hom **S**-BLE $\partial\mathcal{L}/\partial\mathbf{L}(\hat{\mathbf{L}}) = 0$ agree to (6.92).

(iii) hom α-BLE:

The *hybrid norm* $||\mathbf{L}'||^2_{\Sigma_{\mathbf{y}}} + \frac{1}{\alpha}||(\mathbf{I}_m - \mathbf{LA})'||^2_{\mathbf{S}}$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) : = \text{tr}\,\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' + \frac{1}{\alpha}\text{tr}\,(\mathbf{I}_m - \mathbf{LA})\mathbf{S}(\mathbf{I}_m - \mathbf{LA})' = \min_{\mathbf{L}}$$

for $\hat{\xi}$ *hom* α-BLE of ξ. Following the first part of the proof we are led to the *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) = 2[(\boldsymbol{\Sigma}_\mathbf{y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}') \otimes \mathbf{I}_m]\text{vec}\hat{\mathbf{L}} - 2\text{vec}(\xi\xi' A').$$

The *Kronecker–Zehfuss product* $\mathbf{A} \otimes \mathbf{B}$ of two arbitrary matrices as well as $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{B} + \mathbf{B} \otimes \mathbf{C}$ of three arbitrary matrices subject to dim $\mathbf{A} =$ dim $\mathbf{B}$ is introduced in Appendix A.7, "Definition of Matrix Algebra: multiplication of matrices of the same dimension (internal relation) and Laws". The vec operation (vectorization of an array) is reviewed in Appendix A.7, too, "Definition, Facts: $\text{vec}\mathbf{AB} = (\mathbf{B}' \otimes \mathbf{I}'_\ell)\text{vec}\mathbf{A}$ for suitable matrices $\mathbf{A}$ and $\mathbf{B}$". Now we are prepared to compute

$$\frac{\partial^2 \mathcal{L}}{\partial(vec\mathbf{L})\partial(vec\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\Sigma_\mathbf{y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}') \otimes \mathbf{I}_m] > 0$$

as a *positive definite matrix*. The theory of matrix derivatives is reviewed in Appendix A.7 "Facts: Derive of a matrix-valued function of a matrix, namely $\partial(\text{vec}\mathbf{X})/\partial(\text{vec}\mathbf{X})'$".

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) = 2[\tfrac{1}{\alpha}\mathbf{ASA}'\hat{\mathbf{L}}' + \boldsymbol{\Sigma}_\mathbf{y}\hat{\mathbf{L}}' - \tfrac{1}{\alpha}\mathbf{AS}]'\alpha$$

as well as to the *sufficiency conditions*

$$\frac{\partial^2 \mathcal{L}}{\partial(\text{vec}\mathbf{L})\partial(\text{vec}\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\tfrac{1}{\alpha}\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y}) \otimes \mathbf{I}_m] > 0\,.$$

The *normal equations* of hom α-BLE $\partial\mathcal{L}/\partial\mathbf{L}(\hat{\mathbf{L}}) = 0$ agree to (6.93).

For an *explicit representation* of $\hat{\xi}$ as hom BLE, hom **S**-BLE and hom α-BLE of ξ in the *special Gauss–Markov model with fixed effects* of *Box 6.3*, we solve the normal equations (6.94), (6.95) and (6.96) for

$$\hat{\mathbf{L}} = \arg\{\mathcal{L}(\mathbf{L}) = \min_{\mathbf{L}}\}.$$

Beside the *explicit representation* of $\hat{\xi}$ of type hom BLE, hom **S**-BLE and hom α-BLE we compute the related dispersion matrix $D\{\hat{\xi}\}$, the **M**ean **S**quare **E**stimation **E**rror $MSE\{\hat{\xi}\}$, the modified Mean Square Estimation Error $MSE_\mathbf{S}\{\hat{\xi}\}$ and $MSE_{\alpha,\mathbf{S}}\{\hat{\xi}\}$ in

Theorem 6.11. ($\hat{\xi}$ hom BLE):

Let $\hat{\xi} = \mathbf{Ly}$ be hom BLE of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*. Then equivalent representations of the solutions of the normal equations (6.93) are

$$\hat{\xi} = \xi\xi' A'[\boldsymbol{\Sigma}_\mathbf{y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\mathbf{y} \tag{6.96}$$

(if $[\Sigma_\mathbf{y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}$ exists)
and completed by the dispersion matrix

$$D\{\hat{\xi}\} = \xi\xi' A'[\mathbf{\Sigma_y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\mathbf{\Sigma_y} \times[\mathbf{\Sigma_y} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi'}, \tag{6.97}$$

by the *bias vector*

$$\begin{aligned}\beta &:= E\{\hat{\xi}\} - \xi \\ &= -[\mathbf{I}_m - \xi\xi' A'(\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}' + \mathbf{\Sigma_y})^{-1}\mathbf{A}]\,\xi\end{aligned} \tag{6.98}$$

and by the matrix of the Mean Square Estimation Error $MSE\{\hat{\xi}\}$:

$$\begin{aligned}MSE\{\hat{\xi}\} &:= E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \\ &= D\{\hat{\xi}\} + \beta\beta'\end{aligned} \tag{6.99}$$

$$\begin{aligned}&MSE\{\hat{\xi}\} \\ &= D\{\hat{\xi}\} + [\mathbf{I}_m - \xi\xi' A'(\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}' + \mathbf{\Sigma_y})^{-1}\mathbf{A}] \\ &\times \xi\xi'\,[\mathbf{I}_m - \mathbf{A}'(\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}' + \mathbf{\Sigma_y})^{-1}\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi'}].\end{aligned} \tag{6.100}$$

At this point we have to comment what *Theorem 6.11* tells us. *hom* BLE has generated the estimation $\hat{\xi}$ of type (6.96), the dispersion matrix $D\{\hat{\xi}\}$ of type (6.97), the bias vector of type (6.98) and the *Mean Square Estimation Error* of type (6.100) which all depend on the vector ξ and the matrix $\xi\xi'$, respectively. We already mentioned that ξ and the matrix $\xi\xi'$ are not accessible from measurements. The situation is similar to the one in *hypothesis testing*. As shown later in this section we can produce only an estimator $\widehat{\xi}$ and consequently can setup a hypothesis ξ_0 of the "*fixed effect*" ξ. Indeed, a similar argument applies to the *second central moment* $D\{\mathbf{y}\} \sim \mathbf{\Sigma_y}$ of the "*random effect*" $\mathbf{y}$, the observation vector. Such a dispersion matrix has to be known in order to be able to compute $\hat{\xi}$ $D\{\hat{\xi}\}$, and $MSE\{\hat{\xi}\}$. Again we have to apply the argument that we are only able to construct an estimate $\hat{\mathbf{\Sigma}}_\mathbf{y}$ and to setup a hypothesis about $D\{\mathbf{y}\} \sim \Sigma_\mathbf{y}$.

Theorem 6.12. ($\hat{\xi}$ hom S-BLE):

Let $\hat{\xi} = \mathbf{Ly}$ be hom **S**-BLE of ξ in the *special linear Gauss–Markov model with fixed effects* of *Box 6.3*. Then equivalent representations of the solutions of the normal equations (6.94) are

$$\hat{\xi} = \mathbf{SA}'(\mathbf{\Sigma_y} + \mathbf{ASA}')^{-1}\mathbf{y} \tag{6.101}$$

$$\hat{\xi} = (\mathbf{A}'\mathbf{\Sigma_y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\mathbf{\Sigma_y}^{-1}\mathbf{y} \tag{6.102}$$

$$\hat{\xi} = (\mathbf{I}_m + \mathbf{SA}'\mathbf{\Sigma_y}^{-1}\mathbf{A})^{-1}\mathbf{SA}'\Sigma_\mathbf{y}^{-1}\mathbf{y} \tag{6.103}$$

(if $\mathbf{S}^{-1}$, $\mathbf{\Sigma}_\mathbf{y}^{-1}$ exist) are completed by the dispersion matrices

$$D\{\hat{\xi}\} = \mathbf{SA}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{\Sigma}_\mathbf{y}(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{AS} \tag{6.104}$$

$$D\{\hat{\xi}\} = (\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1} \tag{6.105}$$

(if $\mathbf{S}^{-1}$, $\mathbf{\Sigma}_\mathbf{y}^{-1}$ exist) by the *bias vector*

$$\begin{aligned} \beta &:= E\{\hat{\xi}\} - \xi \\ &= -[\mathbf{I}_m - \mathbf{SA}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{A}]\,\xi \\ \beta &= -[\mathbf{I}_m - (\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\Sigma_\mathbf{y}^{-1}\mathbf{A}]\,\xi \end{aligned} \tag{6.106}$$

(if $\mathbf{S}^{-1}$, $\mathbf{\Sigma}_\mathbf{y}^{-1}$ exist)

and by the matrix of the modified *Mean Square Estimation Error* $MSE\{\hat{\xi}\}$:

$$\begin{aligned} MSE_\mathbf{S}\{\hat{\xi}\} &:= E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \\ &= D\{\hat{\xi}\} + \beta\beta' \end{aligned} \tag{6.107}$$

$$\begin{aligned} MSE_\mathbf{S}\{\hat{\xi}\} &= \mathbf{SA}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{\Sigma}_\mathbf{y}(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{AS} \\ &\quad + [\mathbf{I}_m - \mathbf{SA}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{A}]\,\xi\xi'\,[\mathbf{I}_m - \mathbf{A}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{AS}] \\ &= \mathbf{S} - \mathbf{SA}'(\mathbf{ASA}' + \mathbf{\Sigma}_\mathbf{y})^{-1}\mathbf{AS} \end{aligned} \tag{6.108}$$

$$\begin{aligned} MSE_\mathbf{S}\{\hat{\xi}\} &= (\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1} \\ &\quad + [\mathbf{I}_m - (\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A}]\,\xi\xi' \\ &\quad \times [\mathbf{I}_m - \mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}] \\ &= (\mathbf{A}'\mathbf{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1} \end{aligned} \tag{6.109}$$

(if $\mathbf{S}^{-1}$, $\mathbf{\Sigma}_\mathbf{y}^{-1}$ exist).

The interpretation of *hom* **S**-BLE is even more complex. In extension of the comments to *hom* BLE we have to live with another matrix-valued degree of freedom, $\hat{\xi}$ of type (6.101), (6.102), (6.103) and $D\{\hat{\xi}\}$ of type (6.104), (6.105) do no longer depend on the inaccessible matrix $\xi\xi'$, $\mathrm{rk}(\xi\xi')$, but on the "*bias weight matrix*" **S**, $rk\mathbf{S} = m$. Indeed we can associate any element of the bias matrix with a particular weight which can be "*designed*" by the analyst. Again the bias vector β of type (6.106) as well as the *Mean Square Estimation Error* of type (6.107), (6.108), (6.109) depend on the vector ξ which is inaccessible. Beside the "*bias weight matrix* **S**" $\hat{\xi}$, $D\{\hat{\xi}\}$, β and $MSE_\mathbf{s}\{\hat{\xi}\}$ are vector-valued or matrix-valued functions of the dispersion matrix $D\{\mathbf{y}\} \sim \Sigma_\mathbf{y}$ of the stochastic observation vector which is inaccessible. By hypothesis testing we may decide upon the construction of $D\{\mathbf{y}\} \sim \Sigma_\mathbf{y}$ from an estimate $\widehat{\Sigma}_\mathbf{y}$.

Theorem 6.13. ($\hat{\xi}$ hom α-BLE):

Let $\hat{\xi} = \mathbf{Ly}$ be *hom* α-BLE of ξ in the *special linear Gauss–Markov model with fixed effects Box 6.3*. Then equivalent representations of the solutions of the normal equations (6.95) are

$$\hat{\xi} = \tfrac{1}{\alpha}\mathbf{SA}'(\boldsymbol{\Sigma}_\mathbf{y} + \tfrac{1}{\alpha}\mathbf{ASA}')^{-1}\mathbf{y} \tag{6.110}$$

$$\hat{\xi} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y} \tag{6.111}$$

$$\hat{\xi} = (\mathbf{I}_m + \tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A})^{-1}\tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{y} \tag{6.112}$$

(if $\mathbf{S}^{-1}$, $\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ exist)
are completed by the dispersion matrix

$$D\{\hat{\xi}\} = \tfrac{1}{\alpha}\mathbf{SA}'(\boldsymbol{\Sigma}_\mathbf{y} + \tfrac{1}{\alpha}\mathbf{ASA}')^{-1}\boldsymbol{\Sigma}_\mathbf{y}(\boldsymbol{\Sigma}_\mathbf{y} + \tfrac{1}{\alpha}\mathbf{ASA}')^{-1}\mathbf{AS}\tfrac{1}{\alpha} \tag{6.113}$$

$$D\{\hat{\xi}\} = (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1} \tag{6.114}$$

(if $\mathbf{S}^{-1}$, $\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ exist),
by the bias vector

$$\begin{aligned}\beta &:= E\{\hat{\xi}\} - \xi \\ &= -[\mathbf{I}_m - \tfrac{1}{\alpha}\mathbf{SA}'(\tfrac{1}{\alpha}\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\mathbf{A}]\,\xi \\ \beta &= -[\mathbf{I}_m - (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}]\,\xi \end{aligned} \tag{6.115}$$

(if $\mathbf{S}^{-1}$, $\boldsymbol{\Sigma}_\mathbf{y}^{-1}$ exist) and by the matrix of the *Mean Square Estimation Error* $MSE\{\hat{\xi}\}$:

$$\begin{aligned} MSE\{\hat{\xi}\} &:= E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} \\ &= D\{\hat{\xi}\} + \beta\beta' \end{aligned} \tag{6.116}$$

$$\begin{aligned} MSE_{\alpha,\mathbf{S}}\{\hat{\xi}\} &= \mathbf{SC}'(\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\boldsymbol{\Sigma}_\mathbf{y}(\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\mathbf{AS} \\ &\quad + [\mathbf{I}_m - \tfrac{1}{\alpha}\mathbf{SA}'(\tfrac{1}{\alpha}\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\mathbf{A}]\,\xi\xi' \\ &\quad \times [\mathbf{I}_m - \mathbf{A}'(\tfrac{1}{\alpha}\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\mathbf{AS}\tfrac{1}{\alpha}] \\ &= \tfrac{1}{\alpha}\mathbf{S} - \tfrac{1}{\alpha}\mathbf{SA}'(\tfrac{1}{\alpha}\mathbf{ASA}' + \boldsymbol{\Sigma}_\mathbf{y})^{-1}\tfrac{1}{\alpha}\mathbf{AS} \end{aligned} \tag{6.117}$$

$$\begin{aligned} MSE_{\alpha,\mathbf{S}}\{\hat{\xi}\} &= (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1} \\ &\quad + [\mathbf{I}_m - (\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}]\,\xi\xi' \\ &\quad \times [\mathbf{I}_m - \mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_\mathbf{y}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}] \end{aligned} \tag{6.118}$$

$$= (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}$$

$$(\text{if } \mathbf{S}^{-1},\ \ \boldsymbol{\Sigma}_{\mathbf{y}}^{-1} \text{ exist}).$$

The interpretation of the very important estimator hom α-BLE $\hat{\xi}$ of ξ is as follows: $\hat{\xi}$ of type (6.111), also called *ridge estimator* or *Tykhonov–Phillips regulator*, contains the *Cayley inverse* of the normal equation matrix which is *additively decomposed* into $\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}$ and $\alpha\,\mathbf{S}^{-1}$. The weight factor α balances the *first inverse dispersion part* and the *second inverse bias part*. While the experiment informs us of the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$, say $\widehat{\boldsymbol{\Sigma}}_{\mathbf{y}}$, the *bias weight matrix* $\mathbf{S}$ and the *weight factor* α are at the disposal of the analyst. For instance, by the choice $\mathbf{S} = Diag(s_1, \ldots, s_\ell)$ we may emphasize increase or decrease of certain bias matrix elements. The choice of an equally weighted bias matrix is $\mathbf{S} = \mathbf{I}_m$. In contrast the weight factor α can be determined by the *A-optimal design* of type

- $trD\{\hat{\xi}\} = \min\limits_{\alpha}$
- $\beta\beta' = \min\limits_{\alpha}$
- $trMSE_{\alpha,\mathbf{S}}\{\hat{\xi}\} = \min\limits_{\alpha}.$

In the *first case* we optimize the *trace of the variance-covariance matrix* $D\{\hat{\xi}\}$ of type (6.113), (6.114). Alternatively by means of $\beta\beta' = \min\limits_{\alpha}$ we optimize the *quadratic bias* where the bias vector β of type (6.115) is chosen, regardless of the dependence on ξ. Finally for the *third case* – the most popular one – we minimize the trace of the *Mean Square Estimation Error* $MSE_{\alpha,\mathbf{S}}\{\hat{\xi}\}$ of type (6.118), regardless of the dependence on $\xi\xi'$. But beforehand let us present the *proof of Theorem 6.10*, *Theorem 6.11* and *Theorem 6.8*.

Proof:

$$(i)\ \hat{\xi} = \xi\xi' A'[\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\mathbf{y}$$

If the matrix $\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}'$ of the normal equations of type *hom* BLE is of full rank, namely $\text{rk}(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}') = n$, then a straightforward solution of (6.93) is

$$\hat{\mathbf{L}} = \xi\xi' A'[\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}.$$

$$(ii)\ \hat{\xi} = \mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1}\mathbf{y}$$

If the matrix $\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}'$ of the normal equations of type *hom* $\mathbf{S}$-BLE is of full rank, namely $\text{rk}(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}') = n$, then a straightforward solution of (6.94) is

$$\hat{\mathbf{L}} = \mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1}.$$

$$(iii)\ \tilde{\mathbf{z}} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A.7 (Fact: *Cayley inverse*: sum of two matrices, s(10), *Duncan–Guttman matrix identity*) the fundamental matrix identity

$$\mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1},$$

if $\mathbf{S}^{-1}$ and $\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(iv)\ \hat{\xi} = (\mathbf{I}_m + \mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A.7 (Fact: Cayley inverse: sum of two matrices, s(9)) the fundamental matrix identity

$$\mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1} = (\mathbf{I}_m + \mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1},$$

if $\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exists. Such a result concludes this part of the proof.

$$(v)\ \hat{\xi} = \tfrac{1}{\alpha}\mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \tfrac{1}{\alpha}\mathbf{ASA}')^{-1}\mathbf{y}$$

If the matrix $\boldsymbol{\Sigma}_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{ASA}'$ of the normal equations of type *hom* α-BLE is of full rank, namely $\mathrm{rk}(\boldsymbol{\Sigma}_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{ASA}') = n$, then a straightforward solution of (6.95) is

$$\hat{\mathbf{L}} = \tfrac{1}{\alpha}\mathbf{SA}'[\boldsymbol{\Sigma}_{\mathbf{y}} + \tfrac{1}{\alpha}\mathbf{ASA}']^{-1}.$$

$$(vi)\ \hat{\xi} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(10), Duncan–Guttman matrix identity) the fundamental matrix identity

$$\tfrac{1}{\alpha}\mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1} = (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$$

if $\mathbf{S}^{-1}$ and $\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(vii)\hat{\xi} = (\mathbf{I}_m + \tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(9), Duncan–Guttman matrix identity) the fundamental matrix identity

$$\tfrac{1}{\alpha}\mathbf{SA}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{ASA}')^{-1} = (\mathbf{I}_m + \tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\tfrac{1}{\alpha}\mathbf{SA}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$$

if $\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(viii)\ hom\ \ BLE{:}\, D\{\hat{\xi}\}$$

$$D\{\hat{\xi}\} := E\{[\hat{\xi} - E\{\hat{\xi}\}][\hat{\xi} - E\{\hat{\xi}\}]'\}$$

$$= \xi\xi' A'[\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\boldsymbol{\Sigma}_{\mathbf{y}}[\boldsymbol{\Sigma}_{\mathbf{y}} + \mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{A}']^{-1}\mathbf{A}\boldsymbol{\xi}\boldsymbol{\xi}'.$$

By means of the definition of the dispersion matrix $D\{\hat{\xi}\}$ and the substitution of $\hat{\xi}$ of type *hom* BLE the proof has been straightforward.

$$(ix)\ hom\ S\text{-BLE} : D\{\hat{\xi}\}\ (1st\ representation)$$

$$D\{\hat{\xi}\} := E\{[\hat{\xi} - E\{\hat{\xi}\}][\hat{\xi} - E\{\hat{\xi}\}]'\}$$
$$= \mathbf{S}\mathbf{A}'(\mathbf{A}\mathbf{S}\mathbf{A}' + \boldsymbol{\Sigma}_{\mathbf{y}})^{-1}\boldsymbol{\Sigma}_{\mathbf{y}}(\mathbf{A}\mathbf{S}\mathbf{A}' + \boldsymbol{\Sigma}_{\mathbf{y}})^{-1}\mathbf{A}\mathbf{S}.$$

By means of the definition of the dispersion matrix $D\{\hat{\xi}\}$ and the substitution of $\hat{\xi}$ of type *hom* **S**-BLE the proof of the first representation has been straightforward.

$$(x)\ hom\ \text{S-BLE} : D\{\hat{\xi}\}\ (2nd\ representation)$$

$$D\{\hat{\xi}\} := E\{[\hat{\xi} - E\{\hat{\xi}\}][\hat{\xi} - E\{\hat{\xi}\}]'\}$$
$$= (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \mathbf{S}^{-1})^{-1},$$

if $\mathbf{S}^{-1}$and $\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exist. By means of the definition of the dispersion matrix $D\{\hat{\xi}\}$ and the substitution of $\hat{\xi}$ of type *hom* **S**-BLE the proof of the second representation has been straightforward.

$$(xi)\ hom\ \alpha\text{-BLE} : D\{\hat{\xi}\}\ (1st\ representation)$$

$$D\{\hat{\xi}\} := E\{[\hat{\xi} - E\{\hat{\xi}\}][\hat{\xi} - E\{\hat{\xi}\}]'\}$$
$$= \tfrac{1}{\alpha}\mathbf{S}\mathbf{A}'(\boldsymbol{\Sigma}_{\mathbf{y}} + \tfrac{1}{\alpha}\mathbf{A}\mathbf{S}\mathbf{A}')^{-1}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\Sigma}_{\mathbf{y}} + \tfrac{1}{\alpha}\mathbf{A}\mathbf{S}\mathbf{A}')^{-1}\mathbf{A}\mathbf{S}\tfrac{1}{\alpha}.$$

By means of the definition of the dispersion matrix $D\{\hat{\xi}\}$ and the substitution of $\hat{\xi}$ of type *hom* α-BLE the proof of the first representation has been straightforward.

$$(xii)\ hom\alpha\text{-BLE} :\ D\{\hat{\xi}\}(2nd\ representation)$$

$$D\{\hat{\xi}\} := E\{[\hat{\xi} - E\{\hat{\xi}\}][\hat{\xi} - E\{\hat{\xi}\}]'\}$$
$$= (\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A} + \alpha\mathbf{S}^{-1})^{-1},$$

if $\mathbf{S}^{-1}$and$\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}$ exist. By means of the definition of the dispersion matrix and the $D\{\hat{\xi}\}$ substitution of $\hat{\xi}$ of type *hom* α-BLE the proof of the second representation has been straightforward.

$$(xiii)\ \text{bias } \boldsymbol{\beta} \text{ for } hom \text{ BLE, } hom\ \mathbf{S}\text{-BLE and } hom\ \alpha\text{-BLE}$$

As soon as we substitute into the bias $\beta\ : \ = E\{\hat{\xi}\} - \xi = \ - \xi\ +\ E\{\hat{\xi}\}$ the various estimators $\hat{\xi}$ of the type *hom* BLE, *hom* **S**-BLE and *hom* α-BLE we are directly led to various bias representations $\boldsymbol{\beta}$ of type *hom* BLE, *hom* **S**-BLE and *hom* α-BLE.
(*xiv*) $MSE\{\hat{\xi}\}$ of type *hom* BLE, *hom* **S**-BLE and *hom* α-BLE

$$MSE\{\hat{\xi}\} := E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\}$$
$$\hat{\xi} - \xi = \hat{\xi} - E\{\hat{\xi}\} + (E\{\hat{\xi}\} - \xi)$$

$$\begin{aligned} &E\{(\hat{\xi} - \xi)(\hat{\xi} - \xi)'\} = \\ &E\{(\hat{\xi} - E\{\hat{\xi}\})((\hat{\xi} - E\{\hat{\xi}\})'\} \\ &+(E\{\hat{\xi}\} - \xi)(E\{\hat{\xi}\} - \xi)' \end{aligned}$$

$$MSE\{\hat{\xi}\} = D\{\hat{\xi}\} + \beta\beta'.$$

At first we have defined the *Mean Square Estimation Error* $MSE\{\hat{\xi}\}$ of $\hat{\xi}$. Secondly we have decomposed the difference $\hat{\xi} - \xi$ into the two terms

- $\hat{\xi} - E\{\hat{\xi}\}$
- $E\{\hat{\xi}\} - \xi$

in order to derive thirdly the decomposition of $MSE\{\hat{\xi}\}$, namely

- The dispersion matrix of $\hat{\xi}$, namely $D\{\hat{\xi}\}$
- The quadratic bias $\beta\beta'$.

As soon as we substitute $MSE\{\hat{\xi}\}$ the dispersion matrix $D\{\hat{\xi}\}$ and the bias vector β of various estimators $\hat{\xi}$ of the type *hom* BLE, *hom* S-BLE and *hom* α-BLE we are directly led to various representations β of the *Mean Square Estimation Error* $MSE\{\hat{\xi}\}$.

Here is our proof's end.

6-3 Continuous Networks

Criterion matrices/ideal variance-covariance matrices are constructed from continuous networks being observed by signal derivatives of first and second order. The method of constrained least-squares leads to differential equations up to fourth order, to characteristic boundary values and constraints according to the datum choice. Here only one-dimensional networks on a line and on a circle are discussed. Their characteristic (modified) *Green function* is constructed and used for the computation of the variance-covariance function of adjusted mean signal functions. Finally numerical aspects originating from the discrete nature of real observational series are discussed. In detail, the transformation of a criterion matrix into a network datum and its comparison with the variance-covariance matrix of an ideally configurated network is presented.

The aim of a study of continuous geodetic network is to find idealized variance-covariance matrices, also called criterion matrices, for various types of geodetic networks. From standard geodesy textbooks, e.g. *W. Jordan, O. Eggert and M. Kneissl* (1956 p. 123) it is well known for instance, that the error propagation on a leveling line is proportional to its length or in a two-dimensional leveling or distance network proportional to the logarithm of the distance between points. Here we concentrated on one-dimensional geodetic networks, especially on variance-covariance matrices of first and second order derivative observations on a line and on a circle. Higher dimensional continuous networks are introduced by *B. Benciolini* in the following contribution. For the related Fourier analysis of geodetic networks we refer to the contribution of *H. Sünkel*. In addition, more extensive studies on continuous geodetic networks are referenced to *K. Borre and T. Kraup* (1974), *K Borre* (1977,1978, 1979i-iii, 1980), *E. Grafarend* (1977) and *E. Grafarend et al* (1979).

At first we shall review continuous network of *first* derivative type, namely on a line and on a circle, by means of the least-squares variational principle and (modified) *Green functions*. *Secondly*, we give a similar analysis of one-dimensional geodetic networks of *second* derivative type. *Thirdly* we summarize the network analysis based on boundary value problems and (modified) Green functions and introduce discrete approximation techniques. In addition we discuss *Taylor Karman* structured criterion matrices in the context of ideally configurated discrete networks.

6-31 Continuous Networks of Second Derivatives Type

This treatment is based on the contribution of *E. Grafarend and F. Krumm (1985)*. *Green's* function approach can be studied more deeply from reading R.P. Kanwal (1971) under the title "*Linear Integral Equations, Theory and Technique*", in particular "*integration by parts*."

At first we discuss one-dimensional networks assuming *second derivatives* of signal functions to be observed. For network on a *line* we compute the characteristic differential equations, boundary values and constraints by the method of least-squares constrained by the datum choice of fix and free type. The solutions are constructed by direct integration leading to the characteristic (modified) *Green function*. The important result that up to a variance factor the variance-covariance function of the adjusted signal function coincides with the (modified) Green function is proven. Two exercises extend the problems.

Assume an observational series $\mathbf{s}_i''(\mathbf{x}) = \frac{d^2}{d\mathbf{x}^2}\mathbf{s}_i(\mathbf{x})$, $i = 1(1)n$, $0 \leq \mathbf{x} \leq 1$, of second derivatives of signal functions $\mathbf{s}_i(\mathbf{x})$ on a line n is the total number of measurement series. As an example think of acceleration measurements on a line. The random function $\mathbf{s}_i''(\mathbf{x})$ will be characterized by the first two moments, the mean value and the variance-covariance matrix, namely

$$E\{\mathbf{s}_i''(\mathbf{x})\} = e_i \mu''(\mathbf{x}) \qquad \forall_i = 1(1)n \tag{6.119}$$

$$E\{[\mathbf{s}_i''(\mathbf{x}) - E\{s_i''(\mathbf{x})\}][s_j''(\mathbf{y}) - E\{s_j''(\mathbf{y})\}]\} = \mathbf{\Sigma}_{ij}(\mathbf{x},\mathbf{y}) \tag{6.120}$$

$$= \delta_{ij}\delta(\mathbf{x},\mathbf{y})\sigma^2 \qquad \forall i,j = 1(1)n$$

where $\mathbf{e}_i$ is the vector of ones, δ_{ij} the Krockener or unit matrix, $\delta(x, y)$ the *Dirac* or unit function, $\mu(\mathbf{x})$ the unknown mean value function of continuity class $C^4[0, 1]$ and σ^2 the unknown variance of the observational series. Thus we have assumed no correlation between observations within a series and no correlation of observations between series.

As an example consider a one-dimensional network of a second derivatives of signal functions $\mathbf{s}_i(\mathbf{x})$ where we choose $\int_0^1 \mu(\mathbf{x})d\mathbf{x} = 0$ and $\int_0^1 \mu'(\mathbf{x})d\mathbf{x} = 0$ as the network datum of *free type*.

Then the constrained least-squares variational principle is used to determine the mean value function $\mu(\mathbf{x})$ of continuity class $C^4[0, 1]$:

$$\begin{aligned} L(\mu, \lambda_1, \lambda_2) &= \frac{1}{2}\int_0^1 \varepsilon_i''(\mathbf{x})\varepsilon_i''(\mathbf{x})d\mathbf{x} + \lambda_1 \int_0^1 \mu(\mathbf{x})d\mathbf{x} + \lambda_2 \int_0^1 \mu'(\mathbf{x})d\mathbf{x} \\ &= \frac{1}{2}\int_0^1 [\mathbf{s}_i''(\mathbf{x}) - \mathbf{e}_i\mu''(\mathbf{x})][\mathbf{s}_i''(\mathbf{x}) - \mathbf{e}_i\mu''(\mathbf{x})]d\mathbf{x} \\ &\quad + \lambda_1 \int_0^1 \mu(\mathbf{x})d\mathbf{x} + \lambda_2 \int_0^1 \mu'(\mathbf{x})d\mathbf{x} = \min_{\mu,\lambda_1,\lambda_2} \end{aligned} \tag{6.121}$$

$$\begin{aligned} \delta\mathbf{L}(\mu, \lambda_1, \lambda_2) &= \int_0^1 [-\mathbf{e}_i\mathbf{s}_i''(\mathbf{x}) + n\hat{\mu}''(\mathbf{x})]\delta\mu''(\mathbf{x})d\mathbf{x} + \lambda_1 \int_0^1 \delta\mu(\mathbf{x})d\mathbf{x} \\ &\quad + \lambda_2 \int_0^1 \delta\mu'(\mathbf{x})d\mathbf{x} = 0 \end{aligned} \tag{6.122}$$

$$\begin{aligned} \delta^2\mathbf{L}(\mu, \lambda_1, \lambda_2) &= \int_0^1 n\delta\mu''^2(\mathbf{x})d\mathbf{x} + \lambda_1 \int_0^1 \delta\mu^2(\mathbf{x})d\mathbf{x} \\ &\quad + \lambda_2 \int_0^1 \delta\mu'^2(\mathbf{x})d\mathbf{x} + \int_0^1 [-\mathbf{e}_i\mathbf{s}_i''(\mathbf{x}) \\ &\quad + n\hat{\mu}''(\mathbf{x})]\delta^2\mu''(\mathbf{x})d\mathbf{x} \end{aligned} \tag{6.123}$$

$\boxed{\text{“integration by parts”:}}$

(IBPB):

$$\delta\mathbf{L}(\mu, \lambda_1, \lambda_2) = [-\mathbf{e}_i\mathbf{s}_i''(1) + n\hat{\mu}''(1)]\delta\mu'(1)$$

$$
\begin{aligned}
&- [-\mathbf{e}_i \mathbf{s}_i''(0) + n\hat{\mu}''(0)]\delta\mu'(0) \\
&- [-\mathbf{e}_i \mathbf{s}_i'''(1) + n\hat{\mu}'''(1) - \lambda_2]\delta\mu(1) \qquad (6.124) \\
&+ [-\mathbf{e}_i \mathbf{s}_i'''(0) + n\hat{\mu}'''(0) - \lambda_2]\delta\mu(0) \\
&+ \int_0^1 [-\mathbf{e}_i \mathbf{s}_i(x) + n\mu(x) + \lambda_1]\delta\mu(\mathbf{x})d\mathbf{x} = 0 \\
\forall\ &\delta\mu(\mathbf{x}), \delta\mu'(0), \delta\mu'(1)
\end{aligned}
$$

$\boxed{\text{Boundary Value Problem}}$

(i) differential equation

$$\hat{\mu}^{IV}(\mathbf{x}) = \frac{d^2}{d\mathbf{x}^2}\hat{\mathbf{s}}''(\mathbf{x}) - \hat{\lambda}_1, \hat{\lambda}_1 := \frac{1}{n}\lambda_1 \tag{6.125}$$

(ii) boundary values

$$\hat{\mu}'''(0) = \frac{d}{d\mathbf{x}}\hat{\mathbf{s}}''(0) + \hat{\lambda}_2, \hat{\lambda}_2 := \frac{1}{n}\lambda_2 \tag{6.126}$$

$$\hat{\mu}'''(1) = \frac{d}{d\mathbf{x}}\hat{\mathbf{s}}''(0) + \hat{\lambda}_2 \tag{6.127}$$

$$\hat{\mu}''(0) = \hat{\mathbf{s}}''(0) \tag{6.128}$$

$$\hat{\mu}''(1) = \hat{\mathbf{s}}''(1) \tag{6.129}$$

(iii) constraints

$$\int_0^1 \mu(\mathbf{x})d\mathbf{x} = 0 \qquad \int_0^1 \mu'(\mathbf{x})d\mathbf{x} = 0 \tag{6.130}$$

$\boxed{\text{Integration}}$

- *1st step*

$$\int_0^{\mathbf{x}} \hat{\mu}^{IV}(\mathbf{x}_1)d\mathbf{x}_1 = \left[\hat{\mu}'''(\mathbf{x}_1)\right]_0^{\mathbf{x}} = -\hat{\lambda}_1\mathbf{x} + \int_0^{\mathbf{x}} d\mathbf{x}_1\hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.131}$$

$$\Rightarrow \hat{\mu}'''(\mathbf{x}) = \hat{\mu}'''(0) - \hat{\lambda}_1\mathbf{x} + \int_0^{\mathbf{x}} d\mathbf{x}_1\hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.132}$$

• *2nd step*

$$\int_0^{\mathbf{x}} \hat{\mu}'''(\mathbf{x}_2)d\mathbf{x}_2 = \left[\hat{\mu}''(\mathbf{x}_2)\right]_0^{\mathbf{x}} = \hat{\mathbf{s}}'''(0)\mathbf{x} + \hat{\lambda}_2\mathbf{x} - \tfrac{1}{2}\hat{\lambda}_1\mathbf{x}^2 \tag{6.133}$$

$$+\int_0^{\mathbf{x}} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{s}^{IV}(\mathbf{x}_1)$$

$$\Rightarrow \hat{\mu}''(\mathbf{x}) = \hat{\mathbf{s}}''(0) + \hat{\mathbf{s}}'''(0)\mathbf{x} + \hat{\lambda}_2\mathbf{x} + \tfrac{1}{2}\hat{\lambda}_1\mathbf{x}^2$$

$$+\int_0^{\mathbf{x}} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.134}$$

• *3rd step*

$$\int_0^{\mathbf{x}} \hat{\mu}''(\mathbf{x}_3)d\mathbf{x}_3 = \left[\hat{\mu}'(\mathbf{x}_3)\right]_0^{\mathbf{x}} = \hat{\mathbf{s}}''(0)\mathbf{x} + \frac{1}{2}\hat{\mathbf{s}}'''(0)\mathbf{x}^2 + \frac{1}{2}\hat{\lambda}_2\mathbf{x}^2$$

$$-\frac{1}{6}\hat{\lambda}_1\mathbf{x}^3 + \int_0^{\mathbf{x}} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.135}$$

$$\Rightarrow$$

$$\hat{\mu}'(\mathbf{x}) = \hat{\mu}'(\mathbf{x}) + \hat{\mathbf{s}}''(0)\mathbf{x} + \frac{1}{2}\hat{\mathbf{s}}'''(0)\mathbf{x}^2 + \frac{1}{2}\hat{\lambda}_2\mathbf{x}^2 - \frac{1}{6}\hat{\lambda}_1\mathbf{x}^3$$

$$+\int_0^{\mathbf{x}} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.136}$$

• *4th step*

$$\int_0^{\mathbf{x}} \hat{\mu}'(\mathbf{x}_4)dx_4 = \left[\hat{\mu}(\mathbf{x}_4)\right]_0^{\mathbf{x}} = \hat{\mu}'(0)\mathbf{x} + \frac{1}{2}\hat{\mathbf{s}}''(0)\mathbf{x}^2 + \frac{1}{6}\hat{\mathbf{s}}'''(0)\mathbf{x}^3$$

$$+\frac{1}{6}\hat{\lambda}_2\mathbf{x}^3 - \frac{1}{24}\hat{\lambda}_1\mathbf{x}^4 + \int_0^{\mathbf{x}} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.137}$$

$$\Rightarrow$$

$$\hat{\mu}(\mathbf{x}) = \hat{\mu}(0) + \hat{\mu}'(0)\mathbf{x} + \frac{1}{2}\hat{\mathbf{s}}''(0)\mathbf{x}^2 + \frac{1}{6}\hat{\mathbf{s}}'''(0)\mathbf{x}^3 + \frac{1}{6}\hat{\lambda}_2\mathbf{x}^3$$

$$-\frac{1}{24}\hat{\lambda}_1\mathbf{x}^4 + \int_0^{\mathbf{x}} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} dx_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.138}$$

$$\hat{\mu}'''(1) = \hat{\mathbf{s}}'''(1) + \hat{\lambda}_2 = \hat{\mathbf{s}}'''(0) + \hat{\lambda}_2 - \hat{\lambda}_1 + \int_0^1 d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.139}$$

$$\hat{\lambda}_1 = \hat{\mathbf{s}}'''(0) - \hat{\mathbf{s}}'''(1) + \int_0^1 d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.140}$$

$$\hat{\mu}''(1) = \hat{\mathbf{s}}''(1) = \hat{\mathbf{s}}''(0) + \hat{\mathbf{s}}'''(0) + \hat{\lambda}_2 - \frac{1}{2}\hat{\lambda}_1$$

$$+ \int_0^1 d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{s}^{IV}(\mathbf{x}_1)$$

$$\Rightarrow$$

$$\hat{\lambda}_2 = \hat{\mathbf{s}}''(1) - \hat{\mathbf{s}}''(0) - \hat{\mathbf{s}}'''(0) + \frac{1}{2}\hat{\lambda}_1 - \int_0^1 d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.141}$$

$$\hat{\lambda}_2 = \hat{\mathbf{s}}''(1) - \hat{\mathbf{s}}''(0) - \frac{1}{2}\hat{\mathbf{s}}'''(0) - \frac{1}{2}\hat{\mathbf{s}}'''(1) + \frac{1}{2}\int_0^1 d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(x_1)$$

$$- \int_0^1 d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.142}$$

$$\int_0^1 \hat{\mu}(\mathbf{x}) d\mathbf{x} = \hat{\mu}(0) + \frac{1}{2}\hat{\mu}'(0) + \frac{1}{6}\hat{\mathbf{s}}''(0) + \frac{1}{24}\hat{\mathbf{s}}'''(0) + \frac{1}{24}\hat{\lambda}_2$$

$$-\frac{1}{120}\hat{\lambda}_1 + \int_0^1 d\mathbf{x}_5 \int_0^{\mathbf{x}_5} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) = 0 \tag{6.143}$$

$$\Rightarrow$$

$$\hat{\mu}(0) = \frac{1}{12}\hat{\mathbf{s}}''(0) + \frac{1}{24}\hat{\mathbf{s}}'''(0) + \frac{1}{24}\hat{\lambda}_2 - \frac{1}{80}\hat{\lambda}_1$$

$$+\frac{1}{2}\int_0^{\mathbf{x}} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1)$$

$$- \int_0^1 d\mathbf{x}_5 \int_0^{\mathbf{x}_5} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.144}$$

$$\hat{\mu}'(0) = \frac{1}{2}\hat{\mathbf{s}}''(0) - \frac{1}{6}\hat{\mathbf{s}}'''(0) - \frac{1}{6}\hat{\lambda}_2 + \frac{1}{24}\hat{\lambda}_1$$
$$- \int_0^1 d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 \hat{\mathbf{s}}^{IV}(\mathbf{x}_1) \tag{6.145}$$

Following *R.P. Kanwal* (1971 p. 285) by means of the *Heaviside function* H(x-y) we can rewrite

$$\int_0^1 d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 f(\mathbf{x}_1) = \int_0^1 (1 - \mathbf{x}_1) f(\mathbf{x}_1) d\mathbf{x}_1 \tag{6.146}$$

$$\int_0^1 d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 f(\mathbf{x}_1) = \frac{1}{2}\int_0^1 (1 - \mathbf{x}_1)^2 f(\mathbf{x}_1) d\mathbf{x}_1 \tag{6.147}$$

$$\int_0^1 d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 f(\mathbf{x}_1) = \frac{1}{6}\int_0^{\mathbf{x}} (\mathbf{x} - \mathbf{x}_1)^3 f(\mathbf{x} - 1) d\mathbf{x}_1$$

$$= \frac{1}{6}\int_0^1 (\mathbf{x} - \mathbf{x}_1)^3 \mathbf{H}(\mathbf{x} - \mathbf{x}_1) f(\mathbf{x}_1) d\mathbf{x}_1 \tag{6.148}$$

$$\int_0^1 d\mathbf{x}_5 \int_0^{\mathbf{x}_5} d\mathbf{x}_4 \int_0^{\mathbf{x}_4} d\mathbf{x}_3 \int_0^{\mathbf{x}_3} d\mathbf{x}_2 \int_0^{\mathbf{x}_2} d\mathbf{x}_1 f(\mathbf{x}_1) = \frac{1}{24}\int_0^1 (\mathbf{x} - \mathbf{x}_1)^4 f(\mathbf{x}_1) d\mathbf{x}_1 \tag{6.149}$$

such that

$$\boxed{\begin{aligned} \hat{\mu}(\mathbf{x}) &= -\mathbf{G}'(\mathbf{x}, 0)\hat{\mathbf{s}}''(0) + \mathbf{G}'(\mathbf{x}, 1)\hat{\mathbf{s}}''(1) - \mathbf{G}(\mathbf{x}, 1)\hat{\mathbf{s}}'''(1) \\ &\quad + \mathbf{G}(\mathbf{x}, 0)\hat{\mathbf{s}}'''(0) + \int_0^1 \mathbf{G}(\mathbf{x}, \mathbf{y})\hat{\mathbf{s}}^{IV}(\mathbf{y}) d\mathbf{y} \end{aligned}}$$

where the *modified Green function* is defined by

$$\boxed{\begin{aligned}\mathbf{G}(\mathbf{x},\mathbf{y}) = {} & \frac{1}{120} - \frac{1}{24}(\mathbf{x}+\mathbf{y}) - \frac{1}{12}(\mathbf{x}^3-\mathbf{y}^3) - \frac{1}{24}(\mathbf{x}^4+\mathbf{y}^4) \\ & + \mathbf{x}\left(\frac{1}{3}\mathbf{y} - \frac{1}{2}\mathbf{y}^2 + \frac{1}{6}\mathbf{y}^3\right) + \frac{1}{6}\mathbf{x}^3\mathbf{y} + \frac{1}{6}(\mathbf{x}-\mathbf{y})^3\mathbf{H}(\mathbf{x}-\mathbf{y})\end{aligned}}$$

$$\left.\begin{aligned}
\mathbf{G}(\mathbf{x},\mathbf{x}) &= \frac{1}{120} - \frac{1}{12}\mathbf{x} + \frac{1}{3}\mathbf{x}^2 - \frac{1}{2}\mathbf{x}^3 + \frac{1}{4}\mathbf{x}^4 \\
\mathbf{x} > \mathbf{y} : \mathbf{G}(\mathbf{x},\mathbf{y}) &= \frac{1}{120}(1 - 5\mathbf{y} - 10\mathbf{y}^3 - 5\mathbf{y}^4 - 5\mathbf{x} + 40\mathbf{x}\mathbf{y} \\
&\quad + 20\mathbf{x}\mathbf{y}^3 + 10\mathbf{x}^3 - 60\mathbf{x}^2\mathbf{y} + 20\mathbf{x}^3\mathbf{y} - 5\mathbf{x}^4) \\
\mathbf{x} < \mathbf{y} : \mathbf{G}(\mathbf{x},\mathbf{y}) &= \frac{1}{120}(1 - 5\mathbf{y} + 10\mathbf{y}^3 - 5\mathbf{y}^4 - 5\mathbf{x} + 40\mathbf{x}\mathbf{y} \\
&\quad - 60\mathbf{x}\mathbf{y}^2 + 20\mathbf{x}\mathbf{y}^3 - 10\mathbf{x}^3 + 20\mathbf{x}^3\mathbf{y} - 5\mathbf{x}^4) \\
\mathbf{G}(\mathbf{x},0) &= \frac{1}{120} - \frac{1}{24}\mathbf{x} + \frac{1}{12}\mathbf{x}^3 - \frac{1}{24}\mathbf{x}^4 \\
\mathbf{G}(\mathbf{x},1) &= \frac{1}{120} - \frac{1}{24}\mathbf{x} + \frac{1}{12}\mathbf{x}^3 - \frac{1}{24}\mathbf{x}^4 \\
\mathbf{G}'(\mathbf{x},0) &= -\frac{1}{24} + \frac{1}{3}\mathbf{x} - \frac{1}{2}\mathbf{x}^2 + \frac{1}{6}\mathbf{x}^3 \\
\mathbf{G}'(\mathbf{x},1) &= \frac{1}{24} - \frac{1}{6}\mathbf{x} + \frac{1}{6}\mathbf{x}^3 \\
\mathbf{G}'(\mathbf{x},\mathbf{y}) &= -\frac{1}{24} + \frac{1}{4}\mathbf{y}^2 - \frac{1}{6}\mathbf{y}^3 + \mathbf{x}(\frac{1}{3} - \mathbf{y} + \frac{1}{2}\mathbf{y}^2) \\
&\quad + \frac{1}{6}\mathbf{x}^3 - \frac{1}{2}(\mathbf{x}-\mathbf{y})^2\mathbf{H}(\mathbf{x}-\mathbf{y}) \\
\mathbf{G}''(\mathbf{x},\mathbf{y}) &= \frac{1}{2}\mathbf{y} - \frac{1}{2}\mathbf{y}^2 - \mathbf{x} + \mathbf{x}\mathbf{y} + (\mathbf{x}-\mathbf{y})\mathbf{H}(\mathbf{x}-\mathbf{y}) \\
\mathbf{G}'''(\mathbf{x},\mathbf{y}) &= \frac{1}{2} - \mathbf{y} + \mathbf{x} - \mathbf{H}(\mathbf{x}-\mathbf{y}) \\
\mathbf{G}^{IV}(\mathbf{x},\mathbf{y}) &= -1 + \delta(\mathbf{x}-\mathbf{y})
\end{aligned}\right] \tag{6.150}$$

and plotted in *Figs. 6.2* and *6.3* The result can be rewritten in a simpler form by applying integration by part.

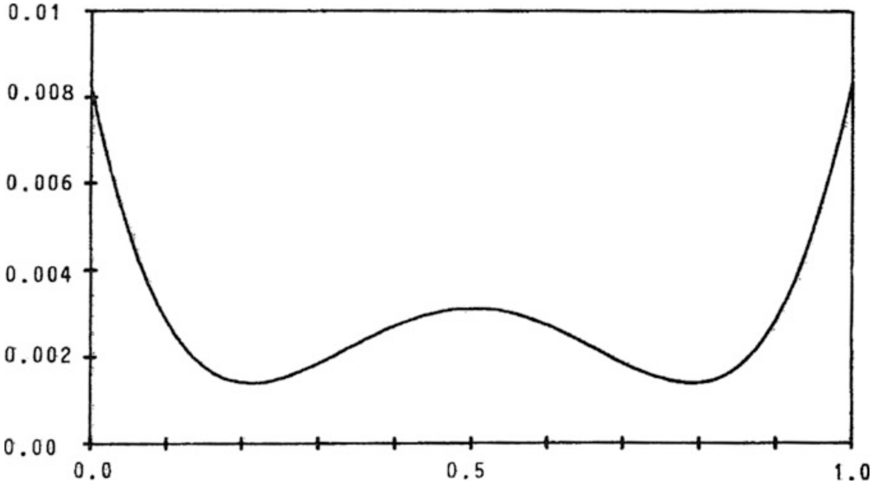


Fig. 6.2 $\mathbf{G(x) = \frac{1}{120} - \frac{1}{12}x + \frac{1}{3}x^2 - \frac{1}{2}x^3 + \frac{1}{4}x^4}$

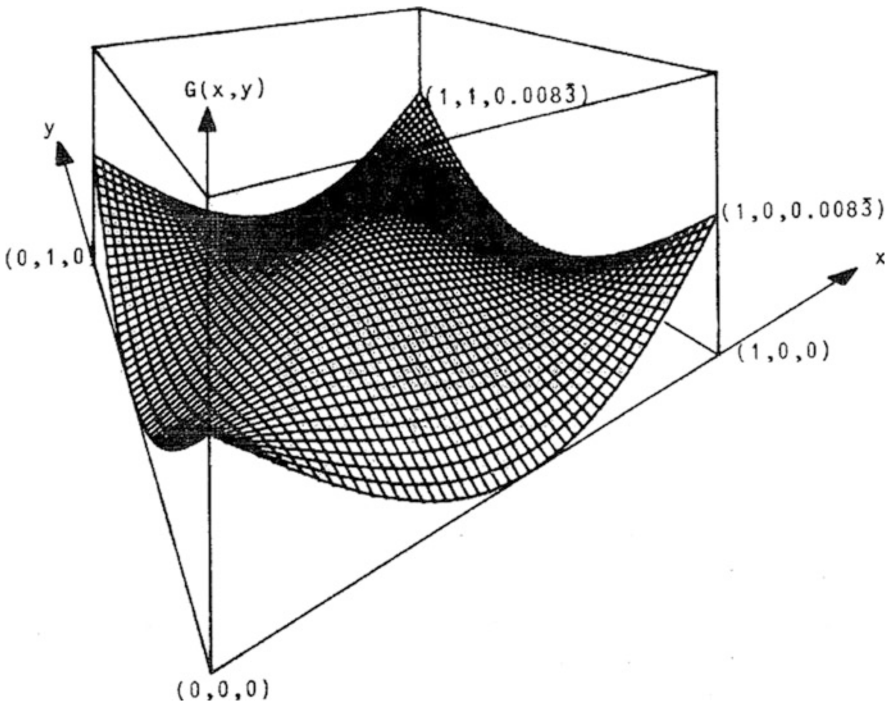


Fig. 6.3 $\mathbf{G(x,y) = \frac{1}{120} - \frac{1}{24}(x+y) - \frac{1}{12}(x^3 - y^3) - \frac{1}{24}(x^4 + y^4) + \frac{1}{6}(x^3y + xy^3) + x(\frac{1}{3}y - \frac{1}{2}y^2) + \frac{1}{6}(x-y)^3H(x-y)}$

$$\begin{aligned}\int_0^1 d\mathbf{y}\mathbf{G}(\mathbf{x},\mathbf{y})\tfrac{d^2}{d\mathbf{y}^2}\hat{\mathbf{s}}''(\mathbf{y}) &= \left[\mathbf{G}(\mathbf{x},\mathbf{y})\tfrac{d}{d\mathbf{y}}\hat{\mathbf{s}}''(\mathbf{y})\right]_0^1 - \int_0^1 \tfrac{\partial\mathbf{G}}{\partial\mathbf{y}}(\mathbf{x},\mathbf{y})\tfrac{d}{d\mathbf{y}}\hat{\mathbf{s}}''(\mathbf{y})\\
&= \left[\mathbf{G}(\mathbf{x},\mathbf{y})\tfrac{d}{d\mathbf{y}}\hat{\mathbf{s}}''(\mathbf{y})\right]_0^1 - \left[\tfrac{\partial\mathbf{G}}{\partial\mathbf{y}}(\mathbf{x},\mathbf{y})\hat{\mathbf{s}}''(\mathbf{y})\right]_0^1\\
&\quad + \int_0^1 d\mathbf{y}\tfrac{\partial^2\mathbf{G}}{\partial\mathbf{y}^2}(\mathbf{x},\mathbf{y})\hat{\mathbf{s}}''(\mathbf{y})\\
&= \mathbf{G}(\mathbf{x},1)\hat{\mathbf{s}}'''(1) - \mathbf{G}(\mathbf{x},0)\hat{\mathbf{s}}'''(0)\\
&\quad - \mathbf{G}'(\mathbf{x},1)\hat{\mathbf{s}}''(1) + \mathbf{G}'(\mathbf{x},0)\hat{\mathbf{s}}''(0) + \int_0^1 d\mathbf{y}\tfrac{\partial^2\mathbf{G}}{\partial\mathbf{y}^2}(\mathbf{x},\mathbf{y})\hat{\mathbf{s}}''(\mathbf{y})
\end{aligned} \tag{6.151}$$

such that

$$\boxed{\hat{\mu}(\mathbf{x}) = \int_0^1 \tfrac{\partial^2\mathbf{G}}{\partial\mathbf{y}^2}(\mathbf{x},\mathbf{y})\hat{\mathbf{s}}''(\mathbf{y})d\mathbf{y}}$$

Now once again we base the *error propagation study* on ... : In general

$$\mathbf{\Sigma}_{\hat{\mu}}(\mathbf{x}_1,\mathbf{x}_2) = \frac{1}{n^2}\sum_{i,j=1}^{n}\int_0^1 d\mathbf{y}_1\int_0^1 d\mathbf{y}_2\frac{\partial^2\mathbf{G}}{\partial\mathbf{y}_1^2}(\mathbf{x}_1,\mathbf{y}_1)\frac{\partial^2\mathbf{G}}{\partial\mathbf{y}_2^2}(\mathbf{x}_2,\mathbf{y}_2)\sum_{ij}^{\mathbf{s}''}(\mathbf{y_1},\mathbf{y_2}) \tag{6.152}$$

is the transformation of the variance-covariance function of $\mathbf{s}_i''(\mathbf{x})$ into the variance-covariance function of $\hat{\mu}(\mathbf{x})$. Due to the dispersion matrix model.

$$\sum_{ij}^{\mathbf{s}''}(\mathbf{y_1},\mathbf{y_2}) = \delta_{ij}\delta(\mathbf{y_1},\mathbf{y_2})\sigma^2_{\mathbf{s}''} \tag{6.153}$$

can be transformed into

$$\boxed{\mathbf{\Sigma}_{\hat{\mu}}(\mathbf{x_1},\mathbf{x_2}) = \tfrac{1}{n}\sigma^2_{\mathbf{s}''}\int_0^1 \tfrac{\partial^2\mathbf{G}}{\partial\mathbf{y}^2}(\mathbf{x_1},\mathbf{y})\tfrac{\partial^2\mathbf{G}}{\partial\mathbf{y}^2}(\mathbf{x_2},\mathbf{y})}$$

Again we integrated by parts:

$$\begin{aligned}\int_0^1 d\mathbf{y}\mathbf{G}''(\mathbf{x_1},\mathbf{y})\mathbf{G}''(\mathbf{x_2},\mathbf{y}) &= [\mathbf{G}''(\mathbf{x_1},\mathbf{y}),\mathbf{G}'(\mathbf{x_2},\mathbf{y})]\Big|_0^1\\
-\int_0^1 d\mathbf{y}\mathbf{G}'''(\mathbf{x_1},\mathbf{y})\mathbf{G}'(\mathbf{x_2},\mathbf{y}) &= [\mathbf{G}''(\mathbf{x_1},\mathbf{y}),\mathbf{G}'(\mathbf{x_2},\mathbf{y})]\Big|_0^1\\
-[\mathbf{G}'''(\mathbf{x_1},\mathbf{y})\mathbf{G}(\mathbf{x_2},\mathbf{y})]\Big|_0^1 &+ \int_0^1 d\mathbf{y}\frac{\partial^4\mathbf{G}}{\partial\mathbf{y}^4}(\mathbf{x_1},\mathbf{y})\mathbf{G}(\mathbf{x_2},\mathbf{y})
\end{aligned} \tag{6.154}$$

Since $\mathbf{G}^{IV}(\mathbf{x},\mathbf{y}) = -1 + \delta(\mathbf{x}-\mathbf{y})$, $\mathbf{G}''(\mathbf{x},1) = 0$, $\mathbf{G}''(\mathbf{x},0) = 0$, $\mathbf{G}'''(x,0) = \mathbf{G}'''(\mathbf{x},\mathbf{1})$, $\mathbf{G}(\mathbf{x},\mathbf{0}) = \mathbf{G}(\mathbf{x},\mathbf{1})$ hold, takes the simple form

$$\int_0^1 d\mathbf{y}\mathbf{G}^{IV}(\mathbf{x_1},\mathbf{y})\mathbf{G}(\mathbf{x_2},\mathbf{y}) = -\int_0^1 \mathbf{G}(\mathbf{x_2},\mathbf{y})d\mathbf{y} + \mathbf{G}(\mathbf{x_1},\mathbf{x_2}) \tag{6.155}$$

A lengthy computation proves

$$\int_0^1 d\mathbf{y}\mathbf{G}(\mathbf{x_2},\mathbf{y}) = \int_0^{\mathbf{x}_2} d\mathbf{y}\mathbf{G}(\mathbf{x_2},\mathbf{y}) + \int_{\mathbf{x_2}}^1 d\mathbf{y}\mathbf{G}(\mathbf{x_2},\mathbf{y}) = 0 \tag{6.156}$$

$$\boxed{\mathbf{\Sigma}_{\hat{\mu}}(\mathbf{x_1},\mathbf{x_2}) = \tfrac{1}{n}\sigma^2_{\mathbf{s}''}\mathbf{G}(\mathbf{x_1},\mathbf{x_2})}$$

$$\sigma^2_{\hat{\mu}}(\mathbf{x}) = \frac{1}{n}\sigma^2_{\mathbf{s}''}\left(\frac{1}{120} - \frac{1}{12}\mathbf{x} + \frac{1}{3}\mathbf{x}^2 - \frac{1}{2}\mathbf{x}^3 + \frac{1}{4}\mathbf{x}^4\right) \tag{6.157}$$

$$\begin{aligned}\Sigma_{\hat{\mu}}(\mathbf{x_1},\mathbf{x_2}) = \frac{1}{n}\sigma^2_{\mathbf{s}''}\bigg[&\frac{1}{120} - \frac{1}{24}(\mathbf{x_1}+\mathbf{x_2}) - \frac{1}{12}(\mathbf{x_1^3}+\mathbf{x_2^3}) \\ &-\frac{1}{24}(\mathbf{x_1^4}+\mathbf{x_2^4}) + x_1\left(\frac{1}{3}\mathbf{x}_2 - \frac{1}{2}\mathbf{x}_2^2 + \frac{1}{6}\mathbf{x}_2^3\right) \\ &+ \frac{1}{6}\mathbf{x}_1^3\mathbf{x}_2 + \frac{1}{6}(\mathbf{x_1}-\mathbf{x_2})^3\mathbf{H}(\mathbf{x_1}-\mathbf{x_2})\bigg]\end{aligned} \tag{6.158}$$

Again we have to discuss the contradictory boundary value problem. What are the condition of existence, especially with respect to the observation that the unobserved derivatives $\hat{\mathbf{s}}'''(0)$, $\hat{\mathbf{s}}'''(1)$ appear in ... Thus let us compute the *general solution* of the boundary value problem of the *homogenous* differential equation, namely

$$\begin{aligned}&(i) \quad \hat{\mu}_1^{IV}(\mathbf{x}) = 0 \\ &(ii) \quad \hat{\mu}_1'''(0) = \hat{\mathbf{s}}'''(0) + \hat{\lambda}_2 \qquad ; \hat{\mu}_1'''(1) = \hat{\mathbf{s}}'''(1) + \hat{\lambda}_2 \\ &\qquad\ \ \hat{\mu}_1''(0) = \hat{\mathbf{s}}''(0) \qquad\qquad ; \hat{\mu}_1''(1) = \hat{\mathbf{s}}''(1) \\ &(iii) \quad \int_0^1 \hat{\mu}(\mathbf{x})d\mathbf{x} = 0 \qquad\qquad ; \int_0^1 \hat{\mu}'(\mathbf{x})d\mathbf{x} = 0\end{aligned}$$

The *fundamental solution*

$$\hat{\mu}_1(\mathbf{x}) = \mathbf{A} + \mathbf{B}\mathbf{x} + \mathbf{C}\mathbf{x^2} + \mathbf{D}\mathbf{x^3}$$

leads to

$$\hat{\mu}_1^{'''}(\mathbf{x}) = 6\mathbf{D} \Rightarrow \hat{\mu}_1^{'''}(1) = \hat{\mu}_1^{'''}(0) = 6\mathbf{D}) \Rightarrow$$

$$\boxed{\hat{\mathbf{s}}^{'''}(1) = \hat{\mathbf{s}}^{'''}(0)} \tag{6.159}$$

$$\boxed{\int_0^1 \hat{\mathbf{s}}^{IV}(\mathbf{x})d\mathbf{x} = 0} \tag{6.160}$$

and, in addition,

$$\hat{\mu}_1^{''}(\mathbf{x}) = \mathbf{2C} + \mathbf{6D} \Rightarrow \begin{bmatrix} \hat{\mu}_1^{''}(1) = \mathbf{2C} + \mathbf{6D} \\ \hat{\mu}_1^{''}(0) = 2\mathbf{C} \end{bmatrix} \Rightarrow$$

$$6\mathbf{D} = \hat{\mu}_1^{''}(1) - \hat{\mu}_1^{''}(0) = \hat{\mathbf{s}}^{''}(1) - \hat{\mathbf{s}}^{''}(0)$$

$$\boxed{\hat{\mathbf{s}}^{''}(1) - \hat{\mathbf{s}}^{''}(0) = \hat{\mathbf{s}}^{'''}(1) = \hat{\mathbf{s}}^{'''}(0)} \tag{6.161}$$

$$\boxed{\int_0^1 \hat{\mathbf{s}}^{'''}(\mathbf{x})d\mathbf{x} = \hat{\mathbf{s}}^{'''}(1) = \hat{\mathbf{s}}^{'''}(0)} \tag{6.162}$$

has therefore to be transformed into

$$\boxed{\begin{aligned} \hat{\mu}(x) \;=\; & [-\mathbf{G}^{'}(\mathbf{x},\mathbf{0}) + \mathbf{G}(\mathbf{x},\mathbf{1}) - \mathbf{G}(\mathbf{x},\mathbf{0})]\hat{\mathbf{s}}^{''}(0) \\ & + [\mathbf{G}^{'}(\mathbf{x},\mathbf{1}) - \mathbf{G}(\mathbf{x},\mathbf{1}) + \mathbf{G}(\mathbf{x},\mathbf{0})]\hat{\mathbf{s}}^{''}(1) \\ & + \int_0^1 \mathbf{G}(\mathbf{x},\mathbf{y})\mathbf{s}^{IV}(\mathbf{y})d\mathbf{y} \end{aligned}} \tag{6.163}$$

Exercise

Where are the minima and where is the maximum of the variance function $\sigma^2_{\hat{\mu}}(\mathbf{x})$ located?

$$\frac{d}{d\mathbf{x}}\sigma^2_{\hat{\mu}}(\mathbf{x}) = \frac{1}{n}\sigma^2_{\mathbf{s}''}(-\frac{1}{12} + \frac{2}{3}\mathbf{x} - \frac{3}{2}x^2 + x^3) = 0 \tag{6.164}$$

$$\Leftrightarrow\ \hat{\mathbf{x}}_1 = 0.211,324,865\ ; \hat{\mathbf{x}}_2 = 0.5\ ; \hat{\mathbf{x}}_3 = 0.788,675,135 \tag{6.165}$$

Table 6.1 Summary of (modified) Green functions

Type of continuous network	(modified) Green function
1st order derivative, "fixed"	$\mathbf{x} - (\mathbf{x} - \mathbf{y})\mathbf{H}(\mathbf{x} - \mathbf{y})$
2nd order derivative, "free"	$\frac{1}{3} - y + \frac{\mathbf{x}^2 + \mathbf{y}^2}{2} - (\mathbf{x} - \mathbf{y})\mathbf{H}(\mathbf{x} - \mathbf{y})$
1st order derivative, "circle", "free"	$\frac{2}{3}\pi - \phi_y + \frac{1}{4\pi}(\phi_{\mathbf{x}}^2 + \phi_{\mathbf{y}}^2) - (\phi_{\mathbf{x}} - \phi_{\mathbf{y}})\mathbf{H}(\phi_{\mathbf{x}} - \phi_{\mathbf{y}})$
2nd order derivative, "free"	$\frac{1}{120} + \frac{1}{24}(\mathbf{x} + \mathbf{y}) - \frac{1}{12}(\mathbf{x}^3 - \mathbf{y}^3) - \frac{1}{24}(\mathbf{x}^4 - \mathbf{y}^4) + x\left(\frac{1}{3}\mathbf{y} - \frac{1}{2}\mathbf{y}^2 + \frac{1}{6}\mathbf{y}^3\right) + \frac{1}{6}\mathbf{x}^3\mathbf{y} + \frac{1}{6}(\mathbf{x} - \mathbf{y})^3\mathbf{H}(\mathbf{x} - \mathbf{y})$

$$\frac{d^2}{d\mathbf{x}^2}\sigma_{\hat{\mu}}^2(\mathbf{x}) = \frac{1}{n}\sigma_{\mathbf{s}''}^2\left(\frac{2}{3} - \mathbf{3x} + \mathbf{3x^2}\right) \tag{6.166}$$

$$\frac{d^2}{d\mathbf{x}^2}\sigma_{\hat{\mu}}^2(\hat{\mathbf{x}}_1) \sim \frac{1}{6} > 0 \quad \text{(minimum!)} \tag{6.167}$$

$$\frac{d^2}{d\mathbf{x}^2}\sigma_{\hat{\mu}}^2(\hat{\mathbf{x}}_2) \sim -\frac{1}{12} < 0 \quad \text{(maximum!)} \tag{6.168}$$

$$\frac{d^2}{d\mathbf{x}^2}\sigma_{\hat{\mu}}^2(\hat{\mathbf{x}}_3) \sim \frac{1}{6} > 0 \quad \text{(minimum!)} \tag{6.169}$$

Finally note *Table* 6.1 where all (modified) *Green functions* we have derived so far are collected.

6-32 Discrete Versus Continuous Geodetic Networks

K. Borre and *T. Krarup* (1974) developed the general theory using the method of *finite elements* to perform the transition from the actual discrete network to a two-dimensional continuum. Of course, we can consider the real network situation from another point of view: Once the characteristic continuous boundary value problem is formulated for the observed signal derivative function and solved by the method of *(modified) Green functions* we can interpret discrete measurement series as approximations of the continuous approach. (Compare the geodetic boundary value problem in physical geodesy, namely the Stokes integral, which can be discretized by means of finite elements, splines.) In using the concept of approximation to the solution of the continuous boundary value problem we avoid the inversion of the discrete normal equations as the following example will demonstrate.

Example

Consider the solution of the adjustment of the free, first derivative network on a finite line. The modified *Green function* is given by Table 6.1. Now assume the Dirac series of discrete first derivative data of type

$$\mathbf{s}_i'(\mathbf{y}) = \sum_{\alpha=1}^{p-1} \delta(\mathbf{y}, \mathbf{y}_\alpha)\mathbf{s}_i'(\mathbf{y}_\alpha) \tag{6.170}$$

where p is the number of points where the data are given. $\delta(\mathbf{y}, \mathbf{y}_\alpha)$ is the chosen base function of minimum local support. We are led to the solution

$$\boxed{\hat{\mu}(\mathbf{x}) = \int_0^1 \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{y})\hat{\mathbf{s}}'(\mathbf{y})d\mathbf{y} = \sum_{\alpha=1}^{p-1} G'(\mathbf{x}, \mathbf{y}_\alpha)\mathbf{s}_i'(\mathbf{y}_\alpha)} \tag{6.171}$$

The *error propagation study* based on the modified Green function (...) and the dispersion matrix model

$$\mathbf{\Sigma}_{ij}^{\mathbf{s}'}(\mathbf{y}_\alpha, \mathbf{y}_\beta) = \delta_{ij}\delta_{\alpha,\beta}\sigma_{\mathbf{s}'}^2 \tag{6.172}$$

leads to the variance-covariance function of $\hat{\mu}(\mathbf{x})$:

$$\mathbf{\Sigma}_{\hat{\mu}}(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{n}\sigma_{\mathbf{s}'}^2 \sum_{\alpha=1}^{p-1} \mathbf{G}'(\mathbf{x_1}, \mathbf{y}_\alpha)\mathbf{G}'(\mathbf{x_2}, \mathbf{y}_\alpha) \tag{6.173}$$

$$= \frac{1}{n}\sigma_{\mathbf{s}'}^2 \sum_{\alpha=1}^{p-1} [-1 + \mathbf{y}_\alpha + \mathbf{H}(\mathbf{x_1} - \mathbf{y}_\alpha)][-1 + \mathbf{y}_\alpha + \mathbf{H}(\mathbf{x_2} - \mathbf{y}_\alpha)] \tag{6.174}$$

For more elaborate treatment of discrete networks and their finite element continuation we refer to the literature. We finally discuss *Taylor-Karman structured networks*, namely for didactic reasons on a line.

Example

Consider an absolute signal network whose variance -covariance function is assumed to be *homogeneous*:

$$\mathbf{\Sigma}_{\hat{\mu}}(\alpha, \beta) = \begin{bmatrix} f(0) = \text{const}\ \forall \alpha = \beta, \\ f(|\alpha - \beta|) \quad \forall \alpha \neq \beta, \end{bmatrix} \alpha = 1(1)p, \beta = 1(1)p \tag{6.175}$$

p again is the number of discrete points equally distant on a line. We are interested to transform the *ideal* variance-covariance function into the one with respect to a network datum is fixed or free type. Assume an inconsistent linear observational equation $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A_1}\mathbf{x_1} + \mathbf{A_2}\mathbf{x_2}$ whose design matrix $\mathbf{A}$ and unknown vector $\mathbf{x}$ is partitioned according to the rank of the matrix $\mathbf{A}$. The method of least-squares leads to the normal equations

$$\left.\begin{aligned} \mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1\hat{\mathbf{x}}_1 + \mathbf{A}_1^{\mathrm{T}}\mathbf{A}_2\hat{\mathbf{x}}_2 = \mathbf{A}_1^{\mathrm{T}}\mathbf{y} \\ \hat{\mathbf{x}}_2 = 0 \end{aligned}\right] \tag{6.176}$$

where $\hat{\mathbf{x}}_2 = 0$ is the fixed datum choice.

$$\hat{\mathbf{x}}_1 = (\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)^{-1}\mathbf{A}_1^{\mathrm{T}}\mathbf{y} \quad \hat{\mathbf{x}}_2 = 0 \tag{6.177}$$

$$\mathbf{x_1} := \begin{bmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)^{-1}\mathbf{A}_1^{\mathrm{T}} \\ 0 \end{bmatrix}\mathbf{y} = \begin{bmatrix} (\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)^{-1}\mathbf{A}_1^{\mathrm{T}} \\ 0 \end{bmatrix}\mathbf{A}\mathbf{x}_{II} \tag{6.178}$$

$$\mathbf{\Sigma}_I = \begin{bmatrix} (\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)^{-1}\mathbf{A}_1^{\mathrm{T}} \\ 0 \end{bmatrix}\mathbf{A}\mathbf{\Sigma}_{II}\mathbf{A}^{\mathrm{T}}[\mathbf{A}_1(\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)^{-1}, 0] \tag{6.179}$$

$\mathbf{\Sigma}_I$ will be identify with the variance-covariance matrix of the network in the fixed datum. $\mathbf{\Sigma}_{II}$ with the one of the homogeneous type $\mathbf{\Sigma}_{\hat{\mu}}(\alpha, \beta)$.
For a discrete network of observed signal differences the $(p-1) \times p$ design matrix $\mathbf{A}$ is given by

$$\mathbf{A} = (a_{ij}) = \left[\begin{array}{ll} -1\ \forall i = j & i = 1(1)p-1 \\ \ \ 1\ \forall j = i+1 & \\ \ \ 0 \text{ otherwise} & \end{array}\right. \tag{6.180}$$

Let us choose the point p as the fix-point. The rank partitioning of the design matrix leads to

$$\mathbf{A}_1 = (a_{ij})_1 = \left[\begin{array}{ll} -1\ \forall i = j & i = 1(1)p-1 \\ \ \ 1\ \forall j = i+1 & i = 1(1)p-2 \\ \ \ 0 \text{ otherwise} & \end{array}\right. \tag{6.181}$$

$$\mathbf{A}_2 = (a_{ij})_2 = \begin{bmatrix} 0\ \forall i = 1(1)p-2 \\ 1\ \forall i = p-1 \end{bmatrix} \quad j = p \tag{6.182}$$

The $(p-1) \times (p-1)$ normal equation matrix $\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1$ is structured according

$$(\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1)_{ij} = \left[\begin{array}{ll} 1\ \forall i = j = 1 & \\ 2\ \forall i = j \neq 1 & i = 2(1)p-1 \\ -1\ \forall j = i+1 & i = 1(1)p-2 \\ \qquad i = j+1 & j = 1(1)p-2 \\ 0 \text{ otherwise} & \end{array}\right.$$

Thus the inverse $(\mathbf{A_1^T A}_1)^{-1}$ enjoys the simple structure

$$(\mathbf{A_1^T A}_1)^{-1}_{ij} = \begin{bmatrix} p-i \ \forall i=j,\, i=1(1)p-1 \\ \qquad i>j,\ j=1(1)p-1 \\ p-j \quad j>i. \end{bmatrix}$$

$$[(\mathbf{A_1^T A}_1)^{-1}\mathbf{A_1}]_{ij} = \begin{bmatrix} -1 \ j \geq i, \quad i,j=1(1)p-1 \\ 0 \text{ otherwise} \end{bmatrix}$$

$$\mathbf{A\Sigma}_{\hat{\mu}}\mathbf{A^T} = 2f(|j-i|) - f(|i+1-j|) - f(|i-j-1|)$$

$$\begin{aligned}\{(\mathbf{A_1^T A}_1)^{-1}\mathbf{A_1^T A}\Sigma_{\hat{\mu}}\mathbf{A^T}\}_{ij} &= -f(|i-j|) - f(|j-p+1|) + f(|i-j-1|) \\ &\quad + f(|p-j|) + f(|i-j-1|) \forall i,j=1(1)p-1\end{aligned}$$

$$\boxed{\Sigma_I(\alpha,\beta) = f(0) - f(|\alpha-p|) - f(|\beta-p|) + f(|\alpha-\beta|)} \tag{6.183}$$

Note that the point p was the datum point. In order to determine the unknown function $f(\alpha,\beta) = f(|\alpha-\beta|)$ we compare $(\mathbf{A}_1^{\mathbf{TA}_1})^{-1}$ and $\mathbf{\Sigma}_I(\alpha,\beta)\ \forall\alpha,\ \beta = 1(1)$ $p-1$. It is easily seen that

$$\boxed{\begin{array}{l} f(|\alpha-\beta|) = \mathbf{C} - \frac{1}{2}|\alpha-\beta| \ \ \forall\alpha,\beta = 1(1)p-1, \ \ C\epsilon\mathbf{R}^+ \\ f(0) = C \end{array}} \tag{6.184}$$

or equivalently

$$\boxed{\Sigma_{\hat{\mu}}(\alpha,\beta) = f(0) - \frac{1}{2}|\alpha-\beta|} \tag{6.185}$$

Obviously to the ideally configured discrete network on a line there corresponds an ideal variance-covariance function of homogeneous type. The result can be generalized into homogeneous and isotropic networks of more than one dimension.

Chapter 7
Overdetermined System of Nonlinear Equations on Curved Manifolds

The spherical problem of algebraic regression – inconsistent system of directional observational equations

Here we review a special issue of distributions on manifolds, in particular the spherical problem of algebraic regression or analyse the inconsistent of directional observational equations. The first section introduces loss functions on longitudinal data ($\Phi = 1$) and ($p = 2$) on the *circle* or on the *sphere* as a differential manifold of dimension $\Phi = 1$ and $p = 2$. Section 6.2 introduces the minimal distance mapping on S^1 and S^2 and constructs the related normal equations. Section 6.3 reviews the transformation from the *circular normal distribution* to an *oblique normal distribution* including a historical note to *von Mises* analyzing data on a circle, *namely atomic weights*. We conclude with note on the "angular metric." As a *case study* in section four we analysze 3D angular observations with two different theodolites, namely Theodolite I and Theodolite II. The main practical result is the set of data from ($\tan \wedge^{\mathrm{n}}$, $\tan \Phi^{\mathrm{n}}$), its solution ($\wedge^{\mathrm{n}}$, Φ^{n}) is very *different* from the *Least Squares Solution*.

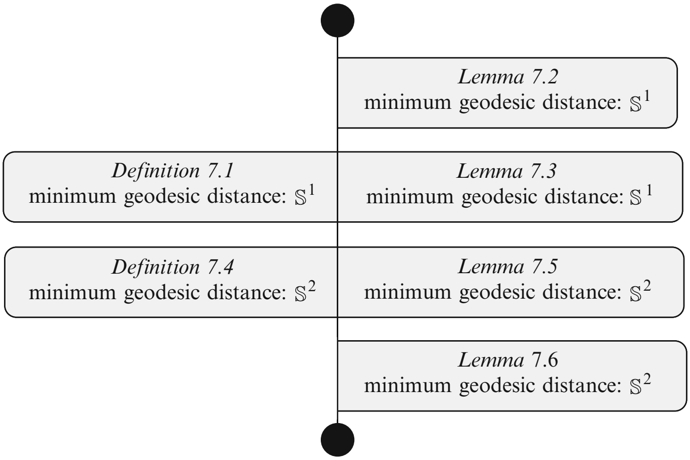
Finally we discuss the von Mises-Fisher distribution for dimension 1, 2, 3 and 4, namely *Relativity* and *Earth Sciences* (for instance *phase observations* for the *Global Positioning System* (GPS)), *Biology, Meteorology, Psychology, Image Analysis* and *Astronomy*. A detailed reference list stays at the end.

Read only *Sect.* 7-2 for a careful treatment(HAPS).

"*Least squares regression is not appropriate when the response variable is circular, and can lead to erroneous results. The reason for this is that the squared difference is not an appropriate measure of distance on the circle.*" – U. Lund (1999)

A typical example of a nonlinear model is the inconsistent system of nonlinear observational equations generated by directional measurements (angular observations, longitudinal data). Here the observation space Y as well as the parameter space $\mathbb{X}$ is the *hypersphere* $\mathbb{S}^p \subset \mathbb{R}^{p+1}$: the *von Mises circle* $\mathbb{S}^1$, $p = 2$ the *Fisher sphere* $\mathbb{S}^2$ in general the *Langevin sphere* $\mathbb{S}^p$. For instance, assume *repeated measurements* of horizontal directions to one target which are distributed as *polar coordinates* on a unit circle clustered around a central direction. Alternatively, assume *repeated measurements* of horizontal and vertical directions to one target which are similarly distributed as *spherical coordinates* (longitude, latitude) on a unit sphere clustered around a central direction. By means of a properly chosen *loss function* we aim at a determination of the central direction. Let us connect all points

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_7,
© Springer-Verlag Berlin Heidelberg 2012



on $\mathbb{S}^1$, $\mathbb{S}^2$ or in general $\mathbb{S}^p$ the measurement points, by a *geodesic*, here the great circle, to the point of the central direction. Indeed the loss function will be optimal at a point on $\mathbb{S}^1$, $\mathbb{S}^2$ or in general $\mathbb{S}^p$ called the *central point*. The result for such a *minimum geodesic distance mapping* will be presented.

7-1 Introduction

Directional data, also called "longitudinal data" or "angular data", arise in several situations, notable geodesy, geophysics, geology, oceanography, atmospheric science, meteorology and others. The *von Mises* or circular normal distribution $\mathcal{CN}(\mu, \kappa)$ with *mean direction parameter* μ $(0 \leq \mu \leq 2\pi)$ and *concentration parameter* κ $(\kappa > 0)$, the reciprocal of a dispersion measure, plays the role in circular data parallel to that of the *Gauss normal distribution* in linear data. A natural extension of the $\mathcal{CN}$ distribution to the distribution on a p-dimensional sphere $\mathbb{S}^P \subset \mathbb{R}^{P+1}$ leads to the *Fisher-von Mises* or *Langevin distribution* $\mathcal{L}(\mu, \kappa)$. For $p = 2$, namely for spherical data (spherical longitude, spherical latitude), this distribution has been studied by *R. A. Fisher* (1953), generalizing the result of R. von Mises (1918) for $p = 1$, and is often quoted as the Fisher distribution. Further details can be taken from *K. V. Mardia* (1972), *K. V. Mardia and P.E. Jupp* (2000), *G. S. Watson* (1986, 1998) and *A. Sen Gupta and R. Maitra* (1998).

Box 7.1. (Fisher-von Mises or Langevin distribution):

$$\boxed{p = 1\,(R.\,von\ \ Mises\,1918)}$$

$$f(\Lambda|\mu, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\Lambda - \mu_\Lambda)] \tag{7.1}$$

$$f(\Lambda|\mu,\,\kappa) = [2\pi I_0(\kappa)]^{-1}\exp\kappa < \mu|\mathbf{X} > \tag{7.2}$$

$$\cos\Psi := \tag{7.3}$$

$$\begin{gathered} :=< \mu|\mathbf{X} >= \mu_x X + \mu_y Y = \cos\mu_\Lambda \cos\Lambda + \sin\mu_\Lambda \sin\Lambda \\ \cos\Psi = \cos(\Lambda - \mu_\Lambda) \end{gathered} \tag{7.4}$$

$$\mu = \mathbf{e}_1 \cos\mu_\Lambda + \mathbf{e}_2 \sin\mu_\Lambda \in \mathbb{S}^1 \tag{7.5}$$

$$\mathbf{X} = \mathbf{e}_1 \cos\Lambda \quad + \mathbf{e}_2 \sin\Lambda \quad \in \mathbb{S}^1 \tag{7.6}$$

$$\boxed{p = 2\,(R.\,A.\,Fisher\,1953)}$$

$$\begin{aligned} &f(\Lambda, \Phi|\mu_\Lambda,\, \mu_\Phi, \kappa) \\ &= \frac{\kappa}{4\pi\,\sinh\kappa}\exp[\cos\Phi\cos\mu_\Phi\cos(\Lambda-\mu_\Lambda) + \sin\Phi\sin\mu_\Phi] \\ &= \frac{\kappa}{4\pi\,\sinh\kappa}\exp\kappa < \mu|\mathbf{X} > \end{aligned} \tag{7.7}$$

$$\begin{gathered} \cos\Psi :=< \mu|\mathbf{X} >= \mu_x X + \mu_y Y + \mu_z Z \\ = \cos\mu_\Phi \cos\mu_\Lambda \cos\Phi \cos\Lambda + \cos\mu_\Phi \sin\mu_\Lambda \cos\Phi \sin\Lambda + \sin\mu_\Phi \sin\Phi \\ \cos\Psi = \cos\Phi\cos\mu_\Phi\cos(\Lambda-\mu_\Lambda) + \sin\Phi\sin\mu_\Phi \end{gathered} \tag{7.8}$$

$$\begin{gathered} \mu = \mathbf{e}_1\mu_x + \mathbf{e}_2\mu_y + \mathbf{e}_3\mu_z \\ = \mathbf{e}_1 \cos\mu_\Phi \cos\mu_\Lambda + \mathbf{e}_2 \cos\mu_\Phi \sin\mu_\Lambda + \mathbf{e}_3 \sin\mu_\Phi \in \mathbb{S}^2 \end{gathered} \tag{7.9}$$

$$\begin{gathered} \mathbf{X} = \mathbf{e}_1 X + \mathbf{e}_2 Y + \mathbf{e}_3 Z \\ = \mathbf{e}_1 \cos\Phi\cos\Lambda + \mathbf{e}_2\cos\Phi\sin\Lambda + \mathbf{e}_3\sin\Phi \in S^2\,. \end{gathered} \tag{7.10}$$

Box 7.1 is a review of the *Fisher-von Mises or Langevin distribution*. *First*, we setup the *circular normal distribution* on $\mathbb{S}^1$ with longitude Λ as the stochastic variable and (μ_Λ, κ) the distributional parameters called "*mean direction* μ" and "*concentration measure*", the reciprocal of a dispersion measure. Due to the normalization of the circular *probability density function* ("pdf") $I_0(\kappa)$ as the zero order *modified Bessel function* of the first kind of κ appears. The circular distance between the circular mean vector $\mu \in \mathbb{S}^1$ and the placement vector $\mathbf{X} \in \mathbb{S}^1$ is measured by "$\cos\Psi$", namely the *inner product* $< \mu|\mathbf{X} >$, both μ and $\mathbf{X}$ represented in *polar coordinates* $(\mu_\Lambda,\ \Lambda)$, respectively. In summary, (7.1) is the circular normal pdf, namely an element of the exponential class.

Second, we refer to the spherical normal pdf on $\mathbb{S}^2$ with spherical longitude Λ, spherical latitude Φ as the stochastic variables and $(\mu_\Lambda, \mu_\Phi, \kappa)$ the distributional parameters called "*longitudinal mean direction, lateral mean direction* (μ_Λ, μ_Φ)" and "*concentration measure* κ", the reciprocal of a dispersion measure. Here the normalization factor of the spherical pdf is $\kappa/(4\pi\sinh\kappa)$. The spherical distance between the spherical mean vector $\mu \in \mathbb{S}^2$ and the placement vector $\mathbf{X} \in \mathbb{S}^2$ is measured by "$\cos\Psi$", namely the *inner product* $< \mu|\mathbf{X} >$, both μ and $\mathbf{X}$ represented in polar coordinates – spherical coordinates $(\mu_\Lambda, \mu_\Phi,\ \Lambda, \Phi)$, respectively.

In summary, (7.7) is the spherical normal pdf, namely an element of the *exponential class*.

Box 7.2. (Loss function):

p = 1: longitudinal data

$$\textit{type}\ 1: \quad \sum_{i=1}^{n} \cos \Psi_i = \max \quad \sim \quad \mathbf{1}' \cos \boldsymbol{\Psi} = \max \tag{7.11}$$

$$\textit{type}\ 2: \quad \sum_{i=1}^{n} (1 - \cos \Psi_i) = \min \quad \sim \quad \mathbf{1}'(\mathbf{1} - \cos \boldsymbol{\Psi}) = \min \tag{7.12}$$

$$\textit{type}\ 3: \quad \sum_{i=1}^{n} \sin^2 \Psi_i/2 = \min \quad \sim \quad \left(\sin \frac{\boldsymbol{\Psi}}{2}\right)' \left(\sin \frac{\boldsymbol{\Psi}}{2}\right) = \min \tag{7.13}$$

transformation

$$1 - \cos \Psi = 2 \sin^2 \Psi/2 \tag{7.14}$$

$$cos \Psi_i = \cos(\Lambda_i - x) = \cos \Lambda_i \cos x + \sin \Lambda_i \sin x \tag{7.15}$$

$$2 \sin^2 \Psi_i/2 = 1 - \cos \Psi_i = 1 - \cos \Lambda_i \cos x + \sin \Lambda_i \sin x \tag{7.16}$$

$$\cos \boldsymbol{\Psi} = \begin{bmatrix} \cos \Psi_1 \\ \vdots \\ \cos \Psi_n \end{bmatrix}, \quad \cos \mathbf{y} := \cos \boldsymbol{\Lambda} := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \cos \Lambda_1 \\ \vdots \\ \cos \Lambda_n \end{bmatrix} \tag{7.17}$$

$$\cos \boldsymbol{\Psi} = \cos \mathbf{y} \cos x + \sin \mathbf{y} \sin x. \tag{7.18}$$

How to generate a loss function substituting least squares ?

Obviously the von Mises pdf ($p = 1$) has maximum likelihood if $\Sigma_{i=1}^{n} \cos \Psi_i = \Sigma_{i=1}^{n} \cos(\Lambda_i - x)$ is *maximal*. Equivalently $\Sigma_{i=1}^{n}(1 - \cos \Psi_i)$ is minimal. By transforming $1 - \cos \Psi_i$ by (7.14) into $2 \sin^2 \Psi/2$, an equivalent loss function is $\Sigma_{i=1}^{n} \sin^2 \Psi/2$ to be postulated minimal. According to *Box 7.2* the geodetic distance is represented as a nonlinear function of the unknown mean direction $\mu \sim x$. (7.17) constitutes the observation vector $\mathbf{y} \in \mathbb{S}^1$.

Similarly the Fisher pdf ($p \equiv 2$) has maximum likelihood if $\Sigma_{i=1}^{n} \cos \Psi_i$ is *maximal*. Equivalent postulates are (7.20) $\Sigma_{i=1}^{n}(1 - \cos \Psi_i) = \min$ and (7.21) $\Sigma_{i=1}^{n} \sin^2 \Psi_i/2 = \min$. According to *Box 7.3* the geodetic distance (7.23) is represented as a nonlinear function of the unknown mean direction $(\mu_\Lambda, \mu_\Phi) \sim (x_1, x_2)$. (7.24), (7.25) constitute the nonlinear observational equations for direct observations of type "longitude latitude" $(\Lambda_i, \Phi_i) \in \mathbb{S}^2$, the *observation space* $\mathbb{Y}$, and unknown parameters of type mean longitudinal, lateral direction $(\mu_\Lambda, \mu_\Phi) \in \mathbb{S}^2$, the *parameter space* $\mathbb{X}$.

Box 7.3. (Loss function):

p = 2: longitudinal data

$$type\ 1: \quad \sum_{i=1}^{n} \cos \Psi_i = \max \quad \sim \quad \mathbf{1}' \cos \boldsymbol{\Psi} = \max \tag{7.19}$$

$$type\ 2: \quad \sum_{i=1}^{n} (1 - \cos \Psi_i) = \min \quad \sim \quad \mathbf{1}'(\mathbf{1} - \cos \boldsymbol{\Psi}) = \min \tag{7.20}$$

$$type\ 3: \quad \sum_{i=1}^{n} \sin^2 \Psi_i/2 = \min \quad \sim \quad (\sin \frac{\boldsymbol{\Psi}}{2})' (\sin \frac{\boldsymbol{\Psi}}{2}) = \min \tag{7.21}$$

transformation

$$1 - \cos \Psi = 2 \sin^2 \Psi/2 \tag{7.22}$$

$$\begin{aligned} cos\Psi_i &= \cos \Phi_i \cos x_2 \cos(\Lambda_i - x_1) + \sin \Phi_i \sin x_2 \\ &= \cos \Phi_i \cos \Lambda_i \cos x_1 \cos x_2 + \cos \Phi_i \sin \Lambda_i \sin x_1 \cos x_2 + \sin \Phi_i \sin x_2 \end{aligned} \tag{7.23}$$

$$\cos \boldsymbol{\Psi} := \begin{bmatrix} \cos \Psi_1 \\ \vdots \\ \cos \Psi_n \end{bmatrix}, \quad \cos \mathbf{y}_1 := \cos \boldsymbol{\Lambda} := \begin{bmatrix} \cos \Lambda_1 \\ \vdots \\ \cos \Lambda_n \end{bmatrix} \tag{7.24}$$

$$\cos \mathbf{y}_2 := \cos \boldsymbol{\Phi} := \begin{bmatrix} \cos \Phi_1 \\ \vdots \\ \cos \Phi_n \end{bmatrix}, \quad \sin \mathbf{y}_1, \ \sin \mathbf{y}_2 \text{ correspondingly}$$

$$\begin{aligned} \cos \boldsymbol{\Psi} = &\begin{bmatrix} \cos \Phi_1 \cos \Lambda_1 \\ \vdots \\ \cos \Phi_n \cos \Lambda_n \end{bmatrix} \cos x_1 \cos x_2 + \begin{bmatrix} \cos \Phi_1 \sin \Lambda_1 \\ \vdots \\ \cos \Phi_n \sin \Lambda_n \end{bmatrix} \sin x_1 \cos x_2 \\ &+ \begin{bmatrix} \sin \Phi_1 \\ \vdots \\ \sin \Phi_n \end{bmatrix} \sin x_2 \,. \end{aligned} \tag{7.25}$$

7-2 Minimal Geodesic Distance: MINGEODISC

By means of *Definition 7.1* we define the *minimal geodetic distance solution* (MINGEODISC) on $\mathbb{S}^2$. *Lemma 7.2* presents you the corresponding nonlinear normal equation whose close form solution is explicitly given by *Lemma 7.5* in terms of

Gauss brackets (special summation symbols). In contrast *Definition 7.4* confronts us with the definition of the *minimal geodetic distance solution* (MINGEODISC) on $\mathbb{S}^2$. *Lemma 7.7* relates to the corresponding nonlinear normal equations which are solved in a closed form via *Lemma 7.8*, again taking advantage of the *Gauss brackets*.

Definition 7.1. (minimum geodesic distance: $\mathbb{S}^1$):

A point $\Lambda_g \in \mathbb{S}^1$ is called at *minimum geodesic distance* to other points $\Lambda_i \in \mathbb{S}^1$, $i \in \{1, \cdots, n\}$ if the circular distance function

$$\mathcal{L}(\Lambda_g) := \sum_{i=1}^{n} 2(1 - \cos \Psi_i) = \sum_{i=1}^{n} 2[1 - \cos(\Lambda_i - \Lambda_g)] = \min_{\Lambda_g} \tag{7.26}$$

is *minimal*.

$$\Lambda_g = \arg \left\{ \sum_{i=1}^{n} 2(1 - \cos \Psi_i) = \min \,|\, \cos \Psi_i = \cos(\Lambda_i - \Lambda_g) \right\} \tag{7.27}$$

Lemma 7.2. (minimum geodesic distance, normal equation: $\mathbb{S}^1$):

A point $\Lambda_g \in \mathbb{S}^1$ is at *minimum geodesic distance* to other points $\Lambda_i \in \mathbb{S}^1$, $i \in \{1, \cdots, n\}$ if $\Lambda_g = x$ fulfils the normal equation

$$-\sin x \left(\sum_{i=1}^{n} \cos \Lambda_i \right) + \cos x \left(\sum_{i=1}^{n} \sin \Lambda_i \right) = 0. \tag{7.28}$$

Proof. Λ_g is generated by means of the *Lagrangean* (*loss function*)

$$\begin{aligned} \mathcal{L}(x) &:= \sum_{i=1}^{n} 2[1 - \cos(\Lambda_i - x)] \\ &= 2n - 2\cos x \sum_{i=1}^{n} \cos \Lambda_i - 2\sin x \sum_{i=1}^{n} \sin \Lambda_i = \min_x . \end{aligned}$$

The *first derivatives*

$$\frac{d\mathcal{L}(x)}{dx}(\Lambda_g) = 2\sin \Lambda_g \sum_{i=1}^{n} \cos \Lambda_i - 2\cos \Lambda_g \sum_{i=1}^{n} \sin \Lambda_i = 0$$

constitute the *necessary conditions*. The *second derivative*

$$\frac{d^2\mathcal{L}(x)}{dx^2}(\Lambda_g) = 2\cos\Lambda_g \sum_{i=1}^{n} \cos\Lambda_i + 2\sin\Lambda_g \sum_{i=1}^{n} \sin\Lambda_i > 0$$

builds up the sufficiency condition for the minimum at Λ_g.

Lemma 7.3. (minimum geodesic distance, solution of the normal equation: $\mathbb{S}^1$):

Let the point $\Lambda_g \in \mathbb{S}^1$ be at *minimum geodesic distance* to other points $\Lambda_i \in \mathbb{S}^1$, $i \in \{1, \cdots, n\}$. Then the *corresponding normal equation* (7.28) is uniquely solved by

$$\tan\Lambda_g = [\sin\Lambda]/[\cos\Lambda], \tag{7.29}$$

such that the circular solution point is

$$\mathbf{X}_g = \mathbf{e}_1\cos\Lambda_g \quad + \mathbf{e}_2\sin\Lambda_g = \frac{1}{\sqrt{[\sin\Lambda]^2 + [\cos\Lambda]^2}}\{\mathbf{e}_1[\cos\Lambda] + \mathbf{e}_2[\sin\Lambda]\} \tag{7.30}$$

with respect to the Gauss brackets

$$[\sin\Lambda]^2 := \left(\sum_{i=1}^{n} \sin\Lambda_i\right)^2 \tag{7.31}$$

$$[\cos\Lambda]^2 := \left(\sum_{i=1}^{n} \cos\Lambda_i\right)^2 . \tag{7.32}$$

Next we generalize MINGEODISC ($p = 1$) on $\mathbb{S}^1$ to MINGEODISC ($p = 2$) on $\mathbb{S}^1$.

Definition 7.4. (minimum geodesic distance: $\mathbb{S}^2$):

A point $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$ is called at *minimum geodesic distance* to other points $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$, $i \in \{1, \cdots, n\}$ if the spherical distance function

$$\mathcal{L}(\Lambda_g,\ \Phi_g) := \sum_{i=1}^{n} 2(1 - \cos\Psi_i) \tag{7.33}$$

$$= \sum_{i=1}^{n} 2[1 - \cos\Phi_i \cos\Phi_g \cos(\Lambda_i - \Lambda_g) - \sin\Phi_i \sin\Phi_g] = \min_{\Lambda_g, \Phi_g}$$

is *minimal*.

$$(\Lambda_g,\ \Phi_g) = \arg\{\sum_{i=1}^{n} 2(1-\cos\Psi_i) = \min| \tag{7.34}$$
$$|\cos\Psi_i = \cos\Phi_i\cos\Phi_g\cos(\Lambda_i-\Lambda_g)+\sin\Phi_i\sin\Phi_g\}.$$

Lemma 7.5. (minimum geodesic distance, normal equation: $\mathbb{S}^2$):

A point $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$ is called at *minimum geodesic distance* to other points $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$, $i \in \{1,\cdots,n\}$ *if* $\Lambda_g = x_1, \Phi_g = x_2$ fulfils the normal equations

$$\begin{aligned}&-\sin x_2\cos x_1 \sum_{i=1}^{n}\cos\Phi_i\cos\Lambda_i - \sin x_2\sin x_1\sum_{i=1}^{n}\cos\Phi_i\sin\Lambda_i\\ &+\cos x_2\sum_{i=1}^{n}\sin\Phi_i = 0\end{aligned} \tag{7.35}$$

$$\cos x_2\cos x_1\sum_{i=1}^{n}\cos\Phi_i\sin\Lambda_i - \cos x_2\sin x_1\sum_{i=1}^{n}\cos\Phi_i\cos\Lambda_i = 0. \tag{7.36}$$

Proof. (Λ_g,Φ_g) is generated by means of the *Lagrangean* (*loss function*)

$$\begin{aligned}\mathcal{L}(x_1,\ x_2) :&= \sum_{i=1}^{n} 2[1-\cos\Phi_i\cos\Lambda_i\cos x_1\cos x_2\\ &\quad -\cos\Phi_i\sin\Lambda_i\sin x_1\cos x_2 - \sin\Phi_i\sin x_2]\\ &= 2n - 2\cos x_1\cos x_2\sum_{i=1}^{n}\cos\Phi_i\cos\Lambda_i\\ &\quad -2\sin x_1\cos x_2\sum_{i=1}^{n}\cos\Phi_i\sin\Lambda_i - 2\sin x_2\sum_{i=1}^{n}\sin\Phi_i\,.\end{aligned}$$

The *first derivatives*

$$\begin{aligned}\frac{\partial\mathcal{L}(x)}{\partial x_1}(\Lambda_g,\Phi_g) &= 2\sin\Lambda_g\cos\Phi_g\sum_{i=1}^{n}\cos\Phi_i\cos\Lambda_i\\ &\quad -2\cos\Lambda_g\cos\Phi_g\sum_{i=1}^{n}\cos\Phi_i\sin\Lambda_i = 0\\ \frac{\partial\mathcal{L}(x)}{\partial x_2}(\Lambda_g,\Phi_g) &= 2\cos\Lambda_g\sin\Phi_g\sum_{i=1}^{n}\cos\Phi_i\cos\Lambda_i\end{aligned}$$

$$+2 \sin \Lambda_g \sin \Phi_g \sum_{i=1}^{n} \cos \Phi_i \sin \Lambda_i$$

$$-2 \cos \Phi_g \sum_{i=1}^{n} \sin \Phi_i = 0$$

constitute the *necessary conditions*. The matrix of *second derivative*

$$\frac{\partial^2 \mathcal{L}(x)}{\partial \mathbf{x} \partial \mathbf{x}'}(\Lambda_g, \Phi_g) \geq 0$$

builds up the sufficiency condition for the minimum at $(\Lambda_g,\ \Phi_g)$.

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(x)}{\partial x_1^2}(\Lambda_g, \Phi_g) &= 2 \cos \Lambda_g \cos \Phi_g [\cos \Phi \cos \Lambda] \\
&\quad +2 \sin \Lambda_g \cos \Phi_g [\cos \Phi \sin \Lambda] \\
\frac{\partial^2 \mathcal{L}(x)}{\partial x_1 x_2}(\Lambda_g, \Phi_g) &= -2 \sin \Lambda_g \sin \Phi_g [\cos \Phi \cos \Lambda] \\
&\quad +2 \cos \Lambda_g \sin \Phi_g [\cos \Phi \sin \Lambda] \\
\frac{\partial^2 \mathcal{L}(x)}{\partial x_2^2}(\Lambda_g, \Phi_g) &= 2 \cos \Lambda_g \cos \Phi_g [\cos \Phi \cos \Lambda] \\
&\quad +2 \sin \Lambda_g \cos \Phi_g [\cos \Phi \sin \Lambda] \\
&\quad + \sin \Phi_g [\sin \Phi].
\end{aligned}$$

Lemma 7.6. (minimum geodesic distance, solution of the normal equation: $\mathbb{S}^2$):

Let the point $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$ be at *minimum geodesic distance* to other points $(\Lambda_g,\ \Phi_g) \in \mathbb{S}^2$, $i \in \{1, \cdots, n\}$. Then the *corresponding normal equations* (7.35), (7.36) are uniquely solved by

$$\tan \Lambda_g = [\cos \Phi \sin \Lambda] / [\cos \Phi \cos \Lambda] \tag{7.37}$$

$$\tan \Phi_g = \frac{[\sin \Phi]}{\sqrt{[\cos \Phi \cos \Lambda]^2 + [\cos \Phi \sin \Lambda]^2}}$$

such that the circular solution point is

$$\begin{aligned}
\mathbf{X}_g &= \mathbf{e}_1 \cos \Phi_g \cos \Lambda_g + \mathbf{e}_2 \cos \Phi_g \sin \Lambda_g + \mathbf{e}_3 \sin \Phi_g \\
&= \frac{1}{\sqrt{[\cos \Phi \cos \Lambda]^2 + [\cos \Phi \sin \Lambda]^2 + [\sin \Phi]^2}} * \\
&\quad * \{\mathbf{e}_1 [\cos \Phi \cos \Lambda]^2 + \mathbf{e}_2 [\cos \Phi \sin \Lambda]^2 + \mathbf{e}_3 [\sin \Phi]\}
\end{aligned} \tag{7.38}$$

subject to

$$[\cos \Phi \cos \Lambda] := \sum_{i=1}^{n} \cos \Phi_i \cos \Lambda_i \tag{7.39}$$

$$[\cos \Phi \sin \Lambda] := \sum_{i=1}^{n} \cos \Phi_i \sin \Lambda_i \tag{7.40}$$

$$[\sin \Phi] := \sum_{i=1}^{n} \sin \Phi_i \ .$$

7-3 Special Models: From the Circular Normal Distribution to the Oblique Normal Distribution

First, we present a historical note about the von *Mises distribution on the circle*. Second, we aim at constructing a two-dimensional generalization of the *Fisher circular normal distribution* to its *elliptic counterpart*. We present 5 lemmas of different type. *Third*, we intend to prove that an angular metric fulfils the *four axioms of a metric*.

7-31 A Historical Note of the von Mises Distribution

Let us begin with a historical note:

The von Mises Distribution on the Circle

In the early part of the last century, *Richard von Mises* (1918) considered the table of the atomic weights of elements, seven entries of which are as follows (Table 7.1):

He asked the question "*Does a typical element in some sense have integer atomic weight* ?" A natural interpretation of the question is "*Do the fractional parts of the weight cluster near 0 and 1*?" The atomic weight W can be identified in a natural way with points on the unit circle, in such a way that equal fractional parts correspond to identical points. This can be done under the mapping

$$W \to \underset{\sim}{\mathbf{x}} = \begin{bmatrix} \cos \theta_1 \\ \sin \theta_1 \end{bmatrix}, \theta_1 = 2\pi(W - [W]),$$

where $[u]$ is the largest integer not greater than u. *Von Mises*' question can now be seen to be equivalent to asking "Do this points on the circle cluster near $\underset{\sim}{\mathbf{e}} = [1\ 0]'$?". Incidentally, the mapping $W \to \underset{\sim}{\mathbf{x}}$ can be made in another way:

$$W \to \underset{\sim}{\mathbf{x}} = \begin{bmatrix} \cos \theta_2 \\ \sin \theta_2 \end{bmatrix}, \theta_2 = 2\pi(W - [W + \frac{1}{2}]).$$

Table 7.1

Element		Al	Sb	Ar	As	Ba	Be	Bi
Atomic Weight	W	26.98	121.76	39.93	74.91	137.36	9.01	209.00

Table 7.2

Element	Al	Sb	Ar	As	Ba	Be	Bi	Average
$\theta_1/2\pi$	0.98	−0.76	−0.93	−0.91	0.36	0.01	0.00	$\bar{\theta}_1/2\pi = 0.566$
$\theta_2/2\pi$	−0.02	−0.24	−0.06	−0.09	0.36	0.01	0.00	$\bar{\theta}_2/2\pi = 0.006$

The two sets of angles for the two mappings are then as follows (Table 7.2):

We note from the discrepancy between the averages in the final column that our usual ways of describing data, e.g., means and standard deviations, are likely to fail us when it comes to measurements of direction.

If the points do cluster near $\underset{\sim}{\mathbf{e}}_1$ then the resultant vector $\Sigma_{j=1}^{N}\underset{\sim}{\mathbf{x}}_j$ (here $N = 7$) should point in that direction, i.e., we should have approximately $\bar{\underset{\sim}{\mathbf{x}}}/||\underset{\sim}{\mathbf{x}}|| = \underset{\sim}{\mathbf{e}}_1$, where $\bar{\underset{\sim}{\mathbf{x}}} = \Sigma\underset{\sim}{\mathbf{x}}_j/N$ and $||\underset{\sim}{\mathbf{x}}|| = (\underset{\sim}{\mathbf{x}}'\underset{\sim}{\mathbf{x}})^{1/2}$ is the length of $\underset{\sim}{\mathbf{x}}$. For the seven elements whose weights are considered here, we find

$$\bar{\underset{\sim}{\mathbf{x}}}/||\underset{\sim}{\mathbf{x}}|| = \begin{bmatrix} 0.9617 \\ -0.2741 \end{bmatrix} = \begin{bmatrix} \cos 344.09^\circ \\ \sin 344.09^\circ \end{bmatrix} = \begin{bmatrix} \cos -15.91^\circ \\ \sin -15.91^\circ \end{bmatrix},$$

a direction not far removed from $\underset{\sim}{\mathbf{e}}_1$. Von Mises then asked "*For what distribution of the unit circle is the unit vector* $\hat{\mu} = [\cos\hat{\theta}_0 \ \ \sin\hat{\theta}_0]' = \bar{\underset{\sim}{\mathbf{x}}}/||\underset{\sim}{\mathbf{x}}||$ *a maximum likelihood estimator (MLE) of a direction* θ_0 *of clustering or concentration* ?" The answer is the distribution now known as the *von Mises* or *circular normal distribution*. It has density, expressed in terms of random angle θ,

$$\frac{\exp\{k\cos(\theta-\theta_0)\}}{I_0(k)}\frac{d\theta}{2\pi},$$

where θ_0 is the direction of concentration and the normalizing constant $I_0(k)$ is a *Bessel function*. An alternative expression is

$$\frac{\exp(k\underset{\sim}{\mathbf{x}}'\underset{\sim}{\mu})}{I_0(k)}\frac{dS}{2\pi}, \quad \underset{\sim}{\mu} = \begin{bmatrix} \cos\hat{\theta}_0 \\ \sin\hat{\theta}_0 \end{bmatrix}, ||\mathbf{x}|| = 1.$$

Von Mises' question clearly has to do with the truth of the hypothesis

$$H_0 : \theta_0 = 0 \, or \, \mu = \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

It is worth mentioning that *Fisher* found the same distribution in another context (*Fisher*, 1956, SMSI, pp. 133–138) as the conditional distribution of $\underset{\sim}{x}$, given $\|\underset{\sim}{x}\| = 1$ when $\underset{\sim}{x}$ is $\mathcal{N}_2(\underset{\sim}{\zeta}, k^{-1}I_2)$.

7-32 Oblique Map Projection

A special way to derive the general representation of the *two-dimensional generalized Fisher sphere* is by forming the general map of $\mathbb{S}^2$ onto $\mathbb{R}^2$. In order to follow a systematic approach, let us denote

by $\{A, \Psi\}$ the "real" meta-longitude/meta-colatitude representing a point on the sphere given by $\{\mu_\Lambda, \mu_\Phi\}$	versus	by $\{\alpha, r\}$ resp. $\{x, y\}$, the meta-longitude/meta-latitude representing the plane mapped values on the tangential given by $\{\mu_\Lambda, \mu_\Phi\}$, alternatively by its polar coordinates $\{x, y\}$

At *first*, we want to derive the equations generating an oblique map projection of $\mathbb{S}^2_R$ onto $\mathbb{T}_0\mathbb{S}^2_R$ of *equiareal type*.

$$\alpha = A, r = 2R \sin \Psi/2$$

versus

$$\begin{aligned} x &= 2R \sin \Psi/2 \cos A \\ y &= 2R \sin \Psi/2 \sin A. \end{aligned}$$

Second, we intend to derive the transformation from the local surface element $dA d\Psi \sin \Psi$ to the alternate local surface element $|\mathbf{J}| d\alpha dr \sin \Psi(\alpha, r)$ by means of the *inverse Jacobian determinant*

$$\begin{bmatrix} \frac{d\alpha}{dA} & 0 \\ 0 & \frac{dr}{d\Psi} \end{bmatrix} = |\mathbf{J}^{-1}| \sim |\mathbf{J}| = \begin{bmatrix} \frac{dA}{d\alpha} & 0 \\ 0 & \frac{d\Psi}{dr} \end{bmatrix}$$

$$\mathbf{J}^{-1} = \begin{bmatrix} \frac{\partial\alpha}{\partial A} & \frac{\partial\alpha}{\partial\Psi} \\ \frac{\partial r}{\partial A} & \frac{\partial r}{\partial\Psi} \end{bmatrix} = \begin{bmatrix} D_A\alpha & 0 \\ 0 & D_\Psi r \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} D_\alpha A & D_r A \\ D_\alpha \Psi & D_r \Psi \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{R\cos\Psi/2} \end{bmatrix}, \; |\mathbf{J}| = \frac{1}{R\cos\Psi/2}.$$

Third, we read the inverse equations of an oblique map projection of $\mathbb{S}^2_{\mathbb{R}}$ of equiareal type.

$$A = \alpha,\ \sin\frac{\Psi}{2} = \frac{r}{2R},\ \cos\frac{\Psi}{2} = \sqrt{1 - \frac{r^2}{4R^2}}$$

$$\sin\Psi = 2\sin\frac{\Psi}{2}\cos\frac{\Psi}{2} = 2\sin\frac{\Psi}{2}\sqrt{1-\sin^2\frac{\Psi}{2}} = \frac{r}{R} - \sqrt{1-\frac{r^2}{4R^2}}.$$

We collect our basic results in a few lemmas.

Lemma 7.7.

$$dAd\Psi\sin\Psi = \frac{rd\alpha dr}{R^2}$$

Lemma 7.8. (oblique azimuthal map projection of $\mathbb{S}^2_{\mathbb{R}}$) of equiareal type):

direct equations

$$\alpha = A, r = 2R\sin\Psi/2$$

inverse equations

$$A = \alpha,\ \sin\Psi = \frac{r}{R}\sqrt{1-\left(\frac{r}{2R}\right)^2}.$$

Lemma 7.9.

$$dAd\Psi\sin\Psi = \frac{1}{R^2}dxdy = \frac{d\alpha\, rdr}{R^2}.$$

Lemma 7.10. (oblique azimuthal map projection of $\mathbb{S}^2_{\mathbb{R}}$) of equiareal type):

direct equations

$$x = 2R\sin\frac{\Psi}{2}\cos A,\ y = 2\mathrm{R}\sin\frac{\Psi}{2}\sin A$$

inverse equations

$$\tan A = \frac{y}{x},\ \sin\frac{\Psi}{2} = \frac{1}{2R}\sqrt{x^2+y^2}$$

$$\sin\Psi = \frac{1}{R}\sqrt{x^2+y^2}\sqrt{1-\frac{x^2+y^2}{4R^2}} = \frac{1}{2R^2}\sqrt{x^2+y^2}\sqrt{4R^2-(x^2+y^2)}.$$

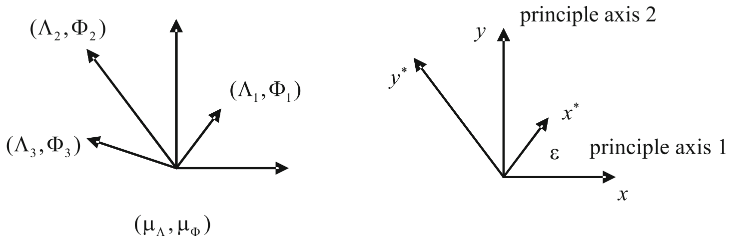


Fig. 7.1 *Left*: Tangential plane. *Right*: Tangential plane

Lemma 7.11. (change from one chart in another chart ("*cha-cha-cha*"), Kartenwechsel):

The *direct equations* of the transformation of spherical longitude and spherical latitude $\{\Lambda^*, \Phi^*\}$ into spherical meta-longitude and spherical meta-colatitude $\{A, \ \Psi\}$ are established by

$$\left[\begin{array}{l}\cot A = [-\cos\Phi\tan\Phi^* + \sin\Phi\cos(\Lambda^* - \Lambda)]/\sin(\Lambda^* - \Lambda)\\ \cos\Psi = \cos\Phi\cos\Phi^*\cos(\Lambda^* - \Lambda) + \sin\Phi\sin\Phi^*)\end{array}\right.$$

with respect to a meta-North pole $\{\Lambda, \Phi\}$. In contrast, the *inverse equations* of the transformation of spherical meta-longitude and spherical meta-colatitude $\{A, \Psi\}$ into spherical longitude and spherical latitude $\{\Lambda^*, \Phi^*\}$ read

$$\left[\begin{array}{l}\cot(\Lambda^* - \Lambda) = [-\sin\Phi\cos A + \cos\Phi\cot\Psi]/\sin A\\ \sin\Phi^* = \sin\Phi\cos\Psi + \cos\Phi\sin\Psi\cos A\,.\end{array}\right.$$

We report two key problems (Fig. 7.1).

First, in the plane located at (μ_Λ, μ_Φ) we place the *circular normal distribution*

$$x = 2(1 - \cos\Psi)\cos A = r(\Psi)\cos\alpha$$
$$y = 2(1 - \cos\Psi)\sin A = r(\Psi)\sin\alpha$$

or

$$A = \alpha, r = 2(1 - \cos\Psi)$$

as an alternative. A natural generalization towards an *oblique normal distribution* will be given by

$$x = x^*\cos\varepsilon - y^*\sin\varepsilon \qquad x^* = x\cos\varepsilon + y\sin\varepsilon$$

or

$$y^* = -x\sin\varepsilon + y\cos\varepsilon \qquad y = x^*\sin\varepsilon + y^*\cos\varepsilon$$

and

$$\begin{aligned}
(x^*)^2\chi_x + (y^*)^2\chi_y &= (x\cos\varepsilon + y\sin\varepsilon)^2\chi_x + (-x\sin\varepsilon + y\cos\varepsilon)^2\chi_y \\
&= x^2\cos^2\varepsilon\,\chi_x + x^2\sin^2\varepsilon\,\chi_y + y^2\sin^2\varepsilon\,\chi_x + y^2\cos^2\varepsilon\,\chi_y \\
&\quad + xy(\sin\varepsilon\cos\varepsilon\,\chi_x - \sin\varepsilon\cos\varepsilon\,\chi_y) \\
&= x^2(\cos^2\varepsilon\,\chi_x + \sin^2\varepsilon\,\chi_y) + y^2(\sin^2\varepsilon\,\chi_x + \cos^2\varepsilon\,\chi_y) \\
&\quad + xy\sin\varepsilon\cos\varepsilon(\chi_x - \chi_y).
\end{aligned}$$

The parameters (χ_x, χ_y) determine the initial values of the *elliptic curve* representing the canonical data set, namely $(1/\sqrt{\chi_x},\ 1/\sqrt{\chi_y})$. The *circular normal distribution* is achieved for the data set $\chi_x = \chi_y = 1$.

Second, we intend to transform the representation of coordinates of the oblique normal distribution *from* Cartesian coordinates in the oblique equatorial plane to curvilinear coordinates in the spherical reference frame:

$$\begin{aligned}
& x^2(\cos^2\varepsilon\chi_x + \sin^2\varepsilon\chi_y) + y^2(\sin^2\varepsilon\chi_x + \cos^2\varepsilon\,\chi_y) + xy\sin\varepsilon\cos\varepsilon(\,\chi_x - \,\chi_y) \\
&= r^2\cos^2\alpha(\cos^2\varepsilon\,\chi_x + \sin^2\varepsilon\,\chi_y) + r^2\sin^2\alpha(\sin^2\varepsilon\,\chi_x + \cos^2\varepsilon\,\chi_y) \\
&\quad + r^2\sin\alpha\cos\alpha\sin\varepsilon\cos\varepsilon(\,\chi_x - \,\chi_y) \\
&= r^2(\Psi, A)\cos^2 A(\cos^2\varepsilon\chi_x + \sin^2\varepsilon\chi_y) + r^2(\Psi, A)\sin^2 A(\sin^2\varepsilon\chi_x + \cos^2\varepsilon\chi_y) \\
&\quad + r^2(\Psi, A)\sin A\cos A\sin\varepsilon\cos\varepsilon(\chi_x - \chi_y).
\end{aligned}$$

Characteristically, the radical component $r(\Psi, A)$ be a function of the colatitude Ψ and of the azimuth A. The angular coordinate is preserved, namely $\alpha = A$. Here, our comments on the topic of the *oblique normal distribution* are finished.

7-33 A Note on the Angular Metric

We intend to prove that an angular metric fulfils all the *four axioms of a metric*. Let us begin with these axioms of a metric, namely

$$\cos\alpha = \frac{<\mathbf{x}|\mathbf{y}>}{||\mathbf{x}||\,||\mathbf{y}||}, \forall 0 \le \alpha \le \pi \quad \Leftrightarrow \quad \alpha = \arccos^{-1}\frac{<\mathbf{x}|\mathbf{y}>}{||\mathbf{x}||\,||\mathbf{y}||}$$

based on *Euclidean metric forms*. Let us introduce the distance function

$$d\left(\frac{\mathbf{x}}{||\mathbf{x}||}, \frac{\mathbf{y}}{||\mathbf{y}||}\right) = \arccos\frac{<\mathbf{x}|\,\mathbf{y}>}{||\mathbf{x}||\,||\mathbf{y}||}$$

to fulfill

$$
\begin{aligned}
&M1 : d(\mathbf{x}, \mathbf{y}) \geq 0\\
&M2 : d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}\\
&M3 : d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})\\
&M4 : d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{z}) : \text{Triangular Inequality.}
\end{aligned}
$$

Assume $||\mathbf{x}||_2 = ||\mathbf{y}||_2 = ||\mathbf{z}||_2 = 1$: *Axioms M1 and M3* are easy to prove, *Axiom M 2* is not complicated, but the *Triangular Inequality* requires work. Let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathbb{X}, \alpha = d(\mathbf{x}, \mathbf{y}), \beta = d(\mathbf{y}, \mathbf{z})and\gamma = d(\mathbf{x}, \mathbf{z})$, i.e.

$$
\begin{gathered}
\alpha,\ \beta,\ \gamma \in [0, \pi],\\
\cos\alpha =< \mathbf{x},\ \mathbf{y} >, \cos\beta =< \mathbf{y},\ \mathbf{z} >, \cos\gamma =< \mathbf{x},\ \mathbf{z} > .
\end{gathered}
$$

We wish to prove $\gamma \leq \alpha + \beta$. This result is trivial in the case $\cos \geq \pi$, so we may assume $\alpha + \beta \in [0, \pi]$. The third desired inequality is equivalent to $\cos\gamma \geq \cos(\alpha + \beta)$. The proof of the basic formulas relies heavily on the properties of the *inverse product*:

$$
\left.
\begin{array}{l}
< \mathbf{u} + \mathbf{u}', \mathbf{v} >=< \mathbf{u}, \mathbf{v} > + < \mathbf{u}', \mathbf{v} >\\
< \mathbf{u}, \mathbf{v} + \mathbf{v}' >=< \mathbf{u}, \mathbf{v} > + < \mathbf{u}, \mathbf{v}' >\\
< \lambda\mathbf{u}, \mathbf{v} > \ = \lambda < \mathbf{u}, \mathbf{v} > \ = \ < \mathbf{u}, \lambda\mathbf{v} >
\end{array}
\right] \textit{for all}\ \mathbf{u}, \mathbf{u}', \mathbf{v}, \mathbf{v}' \in \mathbb{R}^3,\ \text{and}\ \lambda \in \mathbb{R}.
$$

Define $\mathbf{x}', \mathbf{z}' \in \mathbb{R}^3$ by $\mathbf{x} = (\cos\alpha)\mathbf{y} + \mathbf{x}', \mathbf{z} = (\cos\beta)\mathbf{y} - \mathbf{z}'$, then

$$
\begin{gathered}
< \mathbf{x}',\ \mathbf{z}' >=< \mathbf{x} - (\cos\alpha)\mathbf{y}, -\mathbf{z} + (\cos\beta)\mathbf{y} >\\
= - < \mathbf{x},\ \mathbf{z} > +(\cos\alpha) < \mathbf{y}, \mathbf{z} > +(\cos\beta) < \mathbf{x}, \mathbf{y} > -(\cos\alpha)(\cos\beta) < \mathbf{y}, \mathbf{y} >\\
= -\cos\gamma + \cos\alpha\cos\beta + \cos\alpha\cos\beta - \cos\alpha\cos\beta = \cos\gamma + \cos\alpha\cos\beta.
\end{gathered}
$$

In the same way

$$
||\mathbf{x}'||^2 =< \mathbf{x},\ \mathbf{x}' >= 1 - \cos^2\alpha = \sin^2\alpha
$$

so that, since $0 \leq \alpha \leq \pi, ||\mathbf{x}'|| = \sin\alpha$. Similarly, $||\mathbf{z}'|| = \sin\beta$. But by *Schwarz' Inequality*,$< \mathbf{x}',\ \mathbf{z}' > \leq ||\mathbf{x}'||\,||\mathbf{z}'||$. It follows that $\cos\gamma \geq \cos\alpha\cos\beta - \sin\alpha\sin\beta = = \cos(\alpha + \beta)$ and *we are done*!

7-4 Case Study

Table 7.3 collects 30 angular observations with two different theodolites. The *first column* contains the number of the *directional observations* Λ_i, $i \in \{1, \ldots, 30\}$, $n = 30$. The *second column* lists in fractions of seconds the directional data, while the *third column/fourth column* is a printout of $\cos\Lambda_i / \sin\Lambda_i$. *Table 7.4* is a comparison of Λ_g and $\hat{\Lambda}$ as the arithmetic mean. Obviously, on the level of concentration of data, Λ_g and $\hat{\Lambda}$ are nearly the same.

Table 7.3 (The directional observation data using two theodolites and its calculation):

No.	Theodolite 1			Theodolite 2		
	Valueof Observation (Λ_i)	$\cos \Lambda_i$	$\sin \Lambda_i$	**Valueof observation (Λ_i)**	$\cos \Lambda_i$	$\sin \Lambda_i$
1	76°42′ 17.2″	0.229969	0.973198	76°42′19.5″	0.229958	0.973201
2	19.5	0.229958	0.973201	19.0	0.229960	0.973200
3	19.2	0.229959	0.973200	18.8	0.229961	0.973200
4	16.5	0.229972	0.973197	16.9	0.229970	0.973198
5	19.6	0.229957	0.973201	18.6	0.229962	0.973200
6	16.4	0.229972	0.973197	19.1	0.229960	0.973200
7	15.5	0.229977	0.973196	18.2	0.229964	0.973199
8	19.9	0.229956	0.973201	17.7	0.229966	0.973199
9	19.2	0.229959	0.973200	17.5	0.229967	0.973198
10	16.8	0.229970	0.973198	18.6	0.229962	0.973200
11	15.0	0.229979	0.973196	16.0	0.229974	0.973197
12	16.9	0.229970	0.973198	17.3	0.229968	0.973198
13	16.6	0.229971	0.973197	17.2	0.229969	0.973198
14	20.4	0.229953	0.973202	16.8	0.229970	0.973198
15	16.3	0.229973	0.973197	18.8	0.229961	0.973200
16	16.7	0.229971	0.973197	17.7	0.229966	0.973199
17	16.0	0.229974	0.973197	18.6	0.229962	0.973200
18	15.5	0.229977	0.973196	18.8	0.229961	0.973200
19	19.1	0.229960	0.973200	17.7	0.229966	0.973199
20	18.8	0.229961	0.973200	17.1	0.229969	0.973198
22	18.7	0.229962	0.973200	16.9	0.229970	0.973198
22	19.2	0.229959	0.973200	17.6	0.229967	0.973198
23	17.5	0.229967	0.973198	17.0	0.229970	0.973198
24	16.7	0.229971	0.973197	17.5	0.229967	0.973198
25	19.0	0.229960	0.973200	18.2	0.229964	0.973199
26	16.8	0.229970	0.973198	18.3	0.229963	0.973199
27	19.3	0.229959	0.973200	19.8	0.229956	0.973201
28	20.0	0.229955	0.973201	18.6	0.229962	0.973200
29	17.4	0.229968	0.973198	16.9	0.229970	0.973198
30	16.2	0.229973	0.973197	16.7	0.229971	0.973197
Σ	$\hat{\Lambda}_L =$ 76°42′17.73″ $\hat{s} = \pm 1.55''$	6.898982	29.195958	$\hat{\Lambda}_L =$ 76°42′17.91″ $\hat{s} = \pm 0.94''$	6.898956	29.195968

"The precision of the theodolite two is higher compared to the theodolite one". Alternatively, let us present a second example. Let there be given observed azimuths Λ_i and vertical directions Φ_i, by *Table 7.3* in detail. *First*, we compute the solution of the optimization problem (Table 7.5)

Table 7.4 (Computation of theodolite data Comparison of $\hat{\Lambda}\, and\ \Lambda_g$):

Left data set versus Right data set

$$Theodolite\,1: \ \hat{\Lambda} = 76°42'17.73'', \hat{s} = 1.55''$$

$$Theodolite\,2: \ \hat{\Lambda} = 76°42'17.91'', \hat{s} = 0.94''$$

Table 7.5 (Data of type azimuth Λ_i and vertical direction Φ_i):

Λ_i	Φ_i	Λ_i	Φ_i
124°9	88°1	125°0	88°0
125°2	88°3	124°9	88°2
126°1	88°2	124°8	88°1
125°7	88°1	125°1	88°0

$$\sum_{i=1}^{n} 2(1-\cos\Psi_i) = \sum_{i=1}^{n} 2[1-\cos\Phi_i\cos\Phi_\mu\cos(\Lambda_i-\Lambda_\mu)-\sin\Phi_i\sin\Phi_\mu]$$
$$= \min_{\Lambda_\mu,\ \Phi_\mu}$$

subject to values of the central direction

$$\tan\hat{\Lambda} = \frac{\sum\limits_{i=1}^{n}\cos\Phi_i\sin\Lambda_i}{\sum\limits_{i=1}^{n}\cos\Phi_i\cos\Lambda_i},\ \tan\hat{\Phi} = \frac{\sum\limits_{i=1}^{n}\sin\Phi_i}{\sqrt{\left(\sum\limits_{i=1}^{n}\cos\Phi_i\cos\Lambda_i\right)^2+\left(\sum\limits_{i=1}^{n}\cos\Phi_i\sin\Lambda_i\right)^2}}.$$

This accounts for measurements of data on the horizontal circle and the vertical circle being *Fisher normal distributed*. We want to tackle two problems:

Problem 1: Compare $(\hat{\Lambda},\hat{\Phi})$ with the arithmetic mean $(\bar{\Lambda},\bar{\Phi})$ of the data set. Why do the results *not* coincide?
Problem 2: In which case $(\hat{\Lambda},\hat{\Phi})$ and $(\bar{\Lambda},\bar{\Phi})$ do *coincide*?

Solving Problem 1

Let us compute

$$(\bar{\Lambda},\bar{\Phi}) = (125°.212,5,\ 88°.125)\, and\, (\hat{\Lambda},\hat{\Phi})$$
$$= (125°.206,664,5,\ 88°.125,050,77)$$
$$|\Delta\Lambda| = 0°.005,835,5 = 21''.007,8,$$
$$|\Delta\Phi| = 0°.000,050,7 = 0''.18.$$

The results do *not* coincide due to the fact that the *arithmetic means* are obtained by adjusting direct observations with least-squares technology.

Solving Problem 2

The results do coincide if the following conditions are met.

- All vertical directions are zero
- $\hat{\Lambda} = \bar{\Lambda}$ if the observations Λ_i, Φ_i fluctuate only "a little" around the constant value Λ_0, Φ_0
- $\hat{\Lambda} = \bar{\Lambda}$ if $\Phi_i = \text{const.}$
- $\hat{\Phi} = \bar{\Phi}$ if the fluctuation of Λ_i around Λ_0 is considerably smaller than the fluctuation of Φ_i around Φ_0.

Note the values

$$\sum_{i=1}^{8} \cos\Phi_i \sin\Lambda_i = 0.213,866,2; \left(\sum_{i=1}^{8} \cos\Phi_i \sin\Lambda_i\right)^2 = 0.045,378,8$$

$$\sum_{i=1}^{8} \cos\Phi_i \cos\Lambda_i = -0.750,903,27; \left(\sum_{i=1}^{8} \cos\Phi_i \cos\Lambda_i\right)^2 = 0.022,771,8$$

$$\sum_{i=1}^{8} \cos\Phi_i = 7.995,705,3$$

and

$$\begin{gathered}
\Lambda_i = \Lambda_0 + \delta\Lambda_i \text{ versus } \Phi_i = \Phi_0 + \delta\Phi_i \\
\bar{\delta}\bar{\Lambda} = \frac{1}{n}\sum_{i=1}^{n} \delta\Lambda_i \text{ versus } \bar{\delta}\bar{\Phi} = \frac{1}{n}\sum_{i=1}^{n} \delta\Phi_i
\end{gathered}$$

$$\begin{aligned}
\sum_{i=1}^{n} \cos\Phi_i \sin\Lambda_i &= n\cos\Phi_0 \sin\Lambda_0 + \cos\Phi_0 \cos\Lambda_0 \sum_{i=1}^{n} \delta\Lambda_i \\
&\quad - \sin\Phi_0 \sin\Lambda_0 \sum_{i=1}^{n} \delta\Phi_i \\
&= n\cos\Phi_0(\sin\Lambda_0 + \overline{\delta\Lambda}\cos\Lambda_0 - \overline{\delta\Phi}\tan\Lambda_0 \sin\Lambda_0) \\
\sum_{i=1}^{n} \cos\Phi_i \cos\Lambda_i &= n\cos\Phi_0 \cos\Lambda_0 - \cos\Phi_0 \sin\Lambda_0 \sum_{i=1}^{n} \delta\Lambda_i \\
&\quad - \sin\Phi_0 \sin\Lambda_0 \sum_{i=1}^{n} \delta\Phi_i \\
&= n\cos\Phi_0(\cos\Lambda_0 - \overline{\delta\Lambda}\sin\Lambda_0 - \overline{\delta\Phi}\tan\Lambda_0 \cos\Lambda_0)
\end{aligned}$$

$$\tan \hat{\Lambda} = \frac{\sin \Lambda_0 + \bar{\delta}\bar{\Lambda} \cos \Lambda_0 - \bar{\delta}\bar{\Phi} \tan \Phi_0 \sin \Lambda_0}{\cos \Lambda_0 - \bar{\delta}\bar{\Lambda} \sin \Lambda_0 - \bar{\delta}\bar{\Phi} \tan \Phi_0 \cos \Lambda_0}$$

$$\tan \bar{\Lambda} = \frac{\sin \Lambda_0 + \bar{\delta}\bar{\Lambda} \cos \Lambda_0}{\cos \Lambda_0 - \bar{\delta}\bar{\Lambda} \sin \Lambda_0}$$

$$\begin{aligned}(\sum_{i=1}^{n} \cos \Phi_i \sin \Lambda_i)^2 + (\sum_{i=1}^{n} \cos \Phi_i \cos \Lambda_i)^2 \\ = n^2(\cos^2 \Phi_0 + \cos^2 \overline{\delta\Lambda}^2 + \sin^2 \overline{\delta\Phi} - 2 \sin \Phi_0 \cos \overline{\delta\Phi})\end{aligned}$$

$$\sum_{i=1}^{n} \sin \Phi_i = n \sin \Phi_0 + \bar{\delta}\bar{\Phi} \cos \Phi_0$$

$$\tan \hat{\Phi} = \frac{n \sin \Phi_0 + \bar{\delta}\bar{\Phi} \cos \Phi_0}{n\sqrt{\cos^2 \Phi_0 + \bar{\delta}\bar{\Lambda}^2 \cos^2 \Phi_0 + \bar{\delta}\bar{\Phi}^2 \sin^2 \Phi_0 - 2\bar{\delta}\bar{\Phi} \sin \Phi_0 \cos \Phi_0}}$$

$$\tan \bar{\Phi} = \frac{\sin \Phi_0 + \bar{\delta}\bar{\Phi} \cos \Phi_0}{\cos \Phi_0 - \bar{\delta}\bar{\Phi} \sin \Phi_0}.$$

In consequence, $\hat{\Lambda} \neq \bar{\Lambda}, \hat{\Phi} \neq \bar{\Phi}$ holds in general.

At the end we will summarize to additional references like *E. Batschelet* (1965), *T.D. Downs and A.L. Gould* (1967), *E.W. Grafarend* (1970), *E.J. Gumbel et al* (1953), *P. Hartmann et al* (1974), and *M.A. Stephens* (1969).

Comments

Data on the *circle* or on the *sphere* are examples for measurements of a *Riemann manifold* with curvature called *RIEMANN* of dimension p. A *circle* is a *Riemann manifold* if dimension $p = 1$, but a *sphere* a *Riemann manifold* of dimension $p = 2$. A minimum atlas of the *circle* can be constructed by *two charts*, a minimum atlas of the *sphere* again by two charts. In simple words, we need *two charts* to cover the circle or the sphere. For instance, by the *stereographic projection* of the following type.

(*i*) *First Chart*

Stereographic projection with projection centre at the *South pole* and projection plane at the *North pole*.

(*ii*) *Second Chart*

Stereographic projection with projection centre at the *North pole* and projection plane at the *South pole*.

An illustration is referred to *E. Grafarend and F. Krumm* (pages 168f, 211f and 217f (2006)). These projections are classified as *conformal*, they establish a *conformism*. A p-dimension of type von *Mises–Fisher* (or *Langevin*, in *G.S. Watson's* terminology) distribution $\mathbb{M}_p(\mu\kappa)$ is a uniform distribution if

$$f(\mathbf{x};\mu,\kappa)=\left(\frac{\kappa}{2}\right)^{(p-1)/2}\frac{1}{T(\frac{p-1}{2})I(\frac{p-1}{2})(\kappa)}exp(\kappa<\mu|\mathbf{x}>)$$

holds. The embedding of a p-dimension Riemann manifold into a $p(p+1)$-dimensional Euclidean manifold is an important issue. For instance

$$\boxed{\begin{array}{c} p=1\Rightarrow \texttt{Riemann } \mathbb{M}_1 \Leftrightarrow \texttt{Euclid } \mathbb{E}_1 \\ p=2\Rightarrow \texttt{Riemann } \mathbb{M}_2 \Leftrightarrow \texttt{Euclid } \mathbb{E}_3 \\ p=3\Rightarrow \texttt{Riemann } \mathbb{M}_3 \Leftrightarrow \texttt{Euclid } \mathbb{E}_4 \\ p=4\Rightarrow \texttt{Riemann } \mathbb{M}_4 \Leftrightarrow \texttt{Euclid } \mathbb{E}_{10}. \\ (Ex: relativity) \end{array}}$$

K.V. Mardia and P.E. Jupp (2000) mention applications of the von *Mises–Fisher distribution* in

Earth Sciences, meteorology, biology, physics, psychology, image analysis and astronomy.

For instance, GPS (*Global Positioning System*) for the analysis of phase observations (*J. Cai and E. Grafarend* 2007) is a very important area of research as well as the diagnosis of directional data in gravitostatics, magnetostatics, and electrostatics and in problems dealing with *spherical harmonics*, both in synthetic as well as in analysis, namely *Fourier* and *Fourier–Legendre series*.

An open question is still the type of distribution on *other Riemann manifolds* like *ellipse, hyperbola* or *ellipsoid hyperboloids*, *one sided or two sided*, for instance.

References

Anderson, T.W. and M.A. Stephens (1972), Arnold, K.J. (1941), Barndorff-Nielsen, O. (1978), Batschelet, E. (1965), Batschelet, E. (1971), Batschelet, E. (1981), Beran, R.J. (1968), Beran, R.J. (1979), Blingham, C. (1964), Blingham, C. (1974), Chang, T. (1986) Downs, T. D. and Gould, A. L. (1967), Durand, D. and J.A. Greenwood (1957), Enkin, R. and Watson, G. S. (1996), Fisher, R. A. (1953), Fisher, N.J. (1985), Fisher, N.I. (1993), Fisher, N.I. and Lee A. J. (1983), Fisher, N.I. and Lee A. J. (1986), Fujikoshi, Y. (1980), Girko, V.L. (1985), Goldmann, J. (1976), Gordon, L. and M. Hudson (1977), Grafarend, E. W. (1970), Greenwood, J.A. and D. Durand (1955), Gumbel, E. J., Greenwood, J. A. and Durand, D. (1953), Hammersley, J.M. (1950), Hartman, P. and G. S. Watson (1974), Hetherington, T.J. (1981), Jensen, J.L. (1981), Jupp, P.E. and (1980), Jupp, P. E. and Mardia, K. V. , Kanwal, R.P. (1971), Kariya, T. (1989), (1989), Kendall, D.G. (1974), Kent, J. (1976), Kent, J.T. (1982), Kent, J.T. (1983), Krumbein, W.C. (1939), Langevin, P. (1905), Laycock, P.J.

(1975), Lenmitz, C. (1995), Lenth, R.V. (1981), Lord, R.D. (1948), Lund, U. (1999), Mardia, K.V. (1972), Mardia, K.V. (1975), Mardia, K. V. (1988), Mardia, K. V. and Jupp, P. E. (1999), Mardia, K.V. et al. (2000), Mhaskar, H.N., Narcowich, F.J. and J.D. Ward (2001), Muller, C. (1966), Neudecker, H. (1968), Okamoto, M. (1973), Parker, R.L. et al (1979), Pearson, K. (1905), Pearson, K. (1906), Pitman, J. and M. Yor (1981), Presnell, B., Morrison, S.P. and R.C. Littell (1998), Rayleigh, L. (1880), Rayleigh, L. (1905), Rayleigh, R. (1919), Rivest, L.P. (1982), Rivest, L.P. (1988), Roberts, P.H. and H.D. Ursell (1960), Sander, B. (1930), Saw, J.G. (1978), Saw, J.G. (1981), Scheidegger, A.E. (1965), Schmidt-Koenig, K. (1972), Selby, B. (1964), Sen Gupta, A. and R. Maitra (1998), Sibuya, M. (1962), Stam, A.J. (1982), Stephens, M.A. (1963), Stephens, M.A. (1964), Stephens, M. A. (1969), Stephens, M.A. (1979), Tashiro, Y. (1977), Teicher, H. (1961), Von Mises, R. (1918), Watson, G.S. (1956a, 1956b), Watson, G.S. (1960), Watson, G.S.(1961), Watson, G.S. (1962), Watson, G.S. (1965), Watson, G.S. (1966), Watson, G.S. (1967a, 1967b), Watson, G.S. (1968), Watson, G.S. (1969), Wat-son, G.S. (1970), Watson, G.S. (1974), Watson, G.S. (1981a, 1981b), Watson, G.S. (1982a, 1982b, 1982c, 1982d), Watson, G.S. (1983), Watson, G.S. (1986), Watson, G.S. (1988), Watson, G.S. (1998), Watson, G.S. and E.J. Williams (1956), Watson, G.S. and E..Irving (1957), Watson, G.S. and M.R. Leadbetter (1963), Watson, G.S. and S. Wheeler (1964), Watson, G.S. and R.J. Beran (1967), Watson, G.S., R. Epp and J.W. Tukey (1971), Wellner, J. (1979), Wood, A. (1982), Xu, P.L. (1999), Xu, P.L. (2001), Xu, P. (2002), Xu, P.L. et al. (1996a, 1996b), Xu, P.L., and Shimada, S.(1997).

Chapter 8
The Fourth Problem of Probabilistic Regression

Special Gauss–Markov model with random effects

Setup of BLIP and VIP for the central moments of first order

The *random effect model* as a special Gauss-Markov model with random effects is an extension of the classical *Gauss-Markov model*: both effect, namely the vector $\mathbf{y}$ of *observations* as well as the vector of the *regressor* $\mathbf{z}$ (derived from the German "*Zufall*") are random. Box 1 is a review of the model. Solutions are offered as (a) homogeneous best linear, minimum mean square *predictor* (*hom BLIP*), (b) homogeneous S-weighted best linear, minimum mean square *predictor* (*hom S-BLIP*) and (c) homogeneous α-weighted linear *predictor* (*hom α-VIP*). Over we take advantage of *C.R. Rao's* definition of "bias in the mean" by his substitute matrix $\mathbf{S}$. Essential is the *decomposition*.

$$\tilde{\mathbf{Z}} - \mathbf{Z} = \mathbf{L}(\mathbf{y} - \mathbf{E}\{\mathbf{y}\}) - (\mathbf{z} - \mathbf{E}\{\mathbf{z}\}) - [\mathbf{I}_l - \mathbf{LC}]E\{\mathbf{z}\}$$

As well as the *Mean Square Prediction Error*, the *modified Mean Square Prediction Error* and the *related Frobenius matrix norms*. Another decomposition is essential: *MSPE* $\{\mathbf{z}^{\mathrm{N}}\}$ can be divided in three components:

1. The *dispersion matrix* $\mathrm{D}\{\mathbf{z}^{\mathrm{N}}\}$
2. The *dispersion matrix* $\mathrm{D}\{\mathbf{z}\}$
3. The *bias product* $\boldsymbol{\beta}\boldsymbol{\beta}'$

$\mathbf{z}^{\mathrm{N}}$ *hom BLIP* of z, $\mathbf{z}^{\mathrm{N}}$ ***S***-*hom BLIP* and $\tilde{\mathbf{z}}$ *normhybrid min var-min bias*, α-*weighted* or *hom α-VIP* are defined in *Definitions* 8.1–8.3. *Theorems* 8.5–8.7 collect the related representations of the solutions for the *moments of first order*. The open problem which is left, namely the *choice of weight of the bias matrix*, manifest of the *weight factor*. We discuss the *A-optimal design* of

- $\mathrm{tr}\,\mathrm{D}(\mathbf{z}^{\mathrm{N}}) = \min$
- $\boldsymbol{\beta}\boldsymbol{\beta}' = \min$
- $\mathrm{tr}\,\mathrm{MSPE}\{\mathbf{z}^{\mathrm{N}}\} = \min_\alpha$

While the theory of the random effect model is well established, this is not true for various applications. Therefore, we pay attention to three *examples*.

The *first example* concentrates on three cases of *nonlinear scalar error propagation with the random effect model*. In *case 1*, we assume the mean value $\mu_{\mathbf{z}}$ is known, while for *case 2* this value is *unknown*. *Case 3* is built on $\mu_{\mathbf{z}}$ unknown, but $\mathbf{z}_0$ known as a random effect approximation.

The *second example* is oriented to *nonlinear vector-valued error propagation with random effect model* taken from a geoinformatics, the quality design of a nearly rectangular planar surface element. We pose two questions:

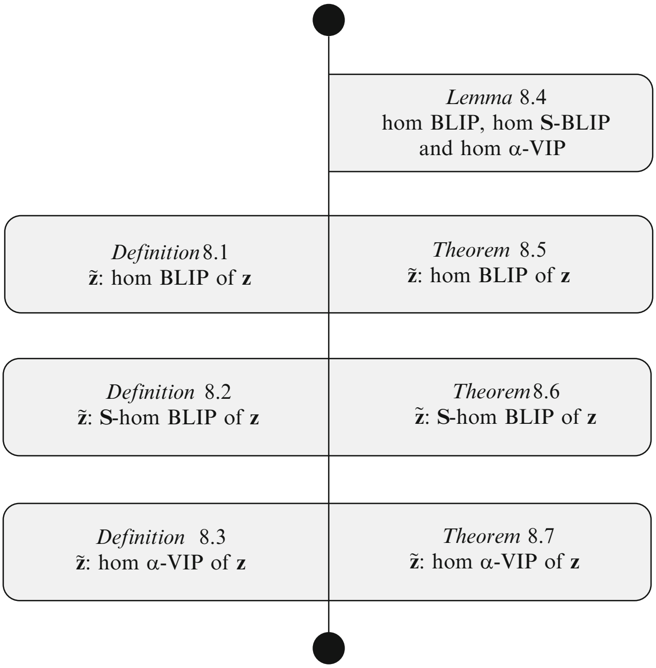
E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_8,
© Springer-Verlag Berlin Heidelberg 2012

1. What is the structure of the variance-covariance matrix?
2. How to approximate the chosen criterion matrix in terms of absolute coordinates?

   We have chosen a statical homogeneous and isotropic network of Taylor-Karman type within two dimensions.

The *third example* treats nonlinear vector valued error propagation with random effects applied to *two-dimensional distance* network. We have computed the *Jacobi* as well as the *Hesse* matrix for the *Taylor-Karman structured criterion matrix*.

Fast track reading : Read only *Theorems* 8.5, 8.6 and 8.7.

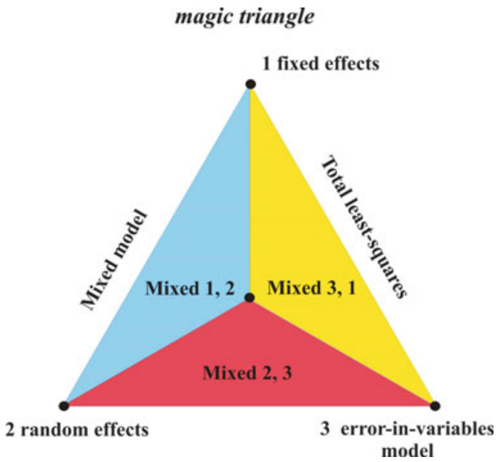


The general model of type "*fixed effects*", "*random effects*" and "*error-in-variables*" will be presented in our final chapter: Here we focus on "*random effects*" (Fig. 8.1).

8-1 The Random Effect Model

Let us introduce the *special Gauss–Markov model with random effects* $\mathbf{y} = \mathbf{Cz} + \mathbf{e}_y - \mathbf{Ce}_z$ specified in *Box 8.1*. Such a model is governed by two identities, namely the first identity $\mathbf{C}E\{\mathbf{z}\} = E\{\mathbf{y}\}$ of moments of first order and the second identity $D\{\mathbf{y} - \mathbf{Cz}\} + \mathbf{C}D\{\mathbf{z}\}\mathbf{C}' = D\{\mathbf{y}\}$ of central moments of second order.

Fig. 8.1 Magic triangle



The *first order moment identity* $\mathbf{C}E\{\mathbf{z}\} = E\{\mathbf{y}\}$ relates the expectation $E\{\mathbf{z}\}$ of the *stochastic, real-valued vector* $\mathbf{z}$ of *unknown random effects* ("Zufallseffekte") to the expectation $E\{\mathbf{y}\}$ of the stochastic, real-valued vector $\mathbf{y}$ of observations by means of the *non-stochastic* ("*fixed*") *real-valued matrix* $\mathbf{C} \in \mathbb{R}^{n\times l}$ of rank $\mathbf{C} = 1, n == dim\mathbb{Y}$ is the dimension of the observation space $\mathbb{Y}$, $l = dim\,\mathbb{Z}$ the dimension of the *parameter space* $\mathbb{Z}$ of random effects $\mathbf{z}$. The *second order central moment identity* $\boldsymbol{\Sigma}_{\mathbf{y}-\mathbf{Cz}} + \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{z}}\mathbf{C}' = \boldsymbol{\Sigma}_{\mathbf{y}}$ relates the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}-\mathbf{Cz}}$ of the random vector $\mathbf{y}-\mathbf{Cz}$, also called dispersion matrix $D\{\mathbf{y}-\mathbf{Cz}\}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{z}}$ of the random vector $\mathbf{z}$, also called dispersion matrix $D\{\mathbf{z}\}$, to the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$ of the random vector y of the observations, also called dispersion matrix $D\{\mathbf{y}\}$. In the simple random effect model we shall assume (i) rk $\boldsymbol{\Sigma}_{\mathbf{y}} = n$ and (ii) $C\{\mathbf{y}, \mathbf{z}\} = 0$, namely *zero correlation* between the random vector $\mathbf{y}$ of observations and the vector $\mathbf{z}$ of random effects. (In the random effect model of type *Kolmogorov–Wiener* we shall give up such a zero correlation.) There are three types of unknowns within the simple *special Gauss–Markov model with random effects*: (i) the vector z of random effects is unknown, (ii) the fixed vectors $E\{\mathbf{y}\},\ E\{\mathbf{z}\}$ of expectations of the vector $\mathbf{y}$ of observations and of the vector z of random effects (first moments) are unknown and (iii) the fixed matrices $\boldsymbol{\Sigma}_{\mathbf{y}}$, $\boldsymbol{\Sigma}_{\mathbf{z}}$ of dispersion matrices $D\{\mathbf{y}\},\ D\{\mathbf{z}\}$ (*second central moments*) are unknown.

Box 8.1. (Special Gauss–Markov model with random effects):

$$\mathbf{y} = \mathbf{Cz} + \mathbf{e}_y - \mathbf{Ce}_z$$
$$E\{\mathbf{y}\} = \mathbf{C}E\{\mathbf{z}\} \in \mathbb{R}^n$$
$$D\{\mathbf{y}\} = D\{\mathbf{y}-\mathbf{Cz}\} + \mathbf{C}D\{\mathbf{z}\}\mathbf{C}' \in \mathbb{R}^{n\times n}$$
$$C\{\mathbf{y}, \mathbf{z}\} = 0$$
$$\mathbf{z},\ E\{\mathbf{z}\},\ E\{\mathbf{y}\},\ \boldsymbol{\Sigma}_{\mathbf{y}-\mathbf{Cz}},\ \boldsymbol{\Sigma}_{\mathbf{z}}\ unknown$$
$$\dim\,\mathcal{R}(\mathbf{C}') = \mathrm{rk}\mathbf{C} = l.$$

Here we focus on best linear predictors of type *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP of *random effects* **z**, which turn out to be better than the best linear uniformly unbiased predictor of type hom BLUUP. At first let us begin with a discussion of the *bias vector* and the *bias matrix* as well as of the *Mean Square Prediction Error* MSPE$\{\tilde{\mathbf{z}}\}$ with respect to a homogeneous linear prediction $\tilde{\mathbf{z}} = \mathbf{Ly}$ of random effects z based upon *Box 8.2*.

Box 8.2. (Bias vector, bias matrix Mean Square Prediction Error in the special Gauss–Markov model with random effects)

$$E\{\mathbf{y}\} = \mathbf{C}E\{\mathbf{z}\} \tag{8.1}$$

$$D\{\mathbf{y}\} = D\{\mathbf{y} - \mathbf{Cz}\} + \mathbf{C}D\{\mathbf{z}\}\mathbf{C}' \tag{8.2}$$

"*ansatz*"

$$\tilde{\mathbf{z}} = \mathbf{Ly} \tag{8.3}$$

bias vector

$$\beta := E\{\tilde{\mathbf{z}} - \mathbf{z}\} = E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\} \tag{8.4}$$

$$\beta = \mathbf{L}E\{\mathbf{y}\} - E\{\mathbf{z}\} = -[\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\} \tag{8.5}$$

bias matrix

$$\mathbf{B} := \mathbf{I}_\ell - \mathbf{LC} \tag{8.6}$$

decomposition

$$\begin{gathered}\tilde{\mathbf{z}} - \mathbf{z} = \\ \tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\} - (\mathbf{z} - E\{\mathbf{z}\}) + (E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\})\end{gathered} \tag{8.7}$$

$$\begin{gathered}\tilde{\mathbf{z}} - \mathbf{z} = \\ \mathbf{L}(\mathbf{y} - E\{\mathbf{y}\}) - (\mathbf{z} - E\{\mathbf{z}\}) - [\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}\end{gathered} \tag{8.8}$$

Mean Square Prediction Error

$$\text{MSPE}\{\tilde{\mathbf{z}}\} := E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} \tag{8.9}$$

$$\text{MSPE}\{\tilde{\mathbf{z}}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + D\{\mathbf{z}\} + [\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_\ell - \mathbf{LC}]' \tag{8.10}$$

$$(C\{\mathbf{y}, \mathbf{z}\} = 0,\ E\{\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}\} = 0,\ E\{\mathbf{z} - E\{\mathbf{z}\}\} = 0)$$

modified Mean Square Prediction Error

$$\text{MSPE}_{\mathbf{S}}\{\tilde{\mathbf{z}}\} := \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + D\{\mathbf{z}\} + [\mathbf{I}_\ell - \mathbf{LC}]\,\mathbf{S}\,[\mathbf{I}_\ell - \mathbf{LC}]' \tag{8.11}$$

Frobenius matrix norms

$$||\mathrm{MSPE}\{\tilde{\mathbf{z}}\}||^2 := \mathrm{tr}\, E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} \tag{8.12}$$

$$\begin{aligned} &||\mathrm{MSPE}\{\tilde{\mathbf{z}}\}||^2 \\ &= \mathrm{tr}\, \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}\, D\{\mathbf{z}\} + \mathrm{tr}\, [\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_\ell - \mathbf{LC}]' \\ &= \ ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_\ell - \mathbf{LC})'||^2_{E\{(\mathbf{z})E(\mathbf{z})'} + \mathrm{tr}\, E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\} \end{aligned} \tag{8.13}$$

$$\begin{aligned} &||\mathrm{MSPE}_\mathbf{S}\{\tilde{\mathbf{z}}\}||^2 : \\ &:= \mathrm{tr}\, \mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}\, [\mathbf{I}_\ell - \mathbf{LC}]\mathbf{S}[\mathbf{I}_\ell - \mathbf{LC}]' + \mathrm{tr}\, D\{\mathbf{z}\} \\ &= \ ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_\ell - \mathbf{LC})'||^2_\mathbf{S} + \mathrm{tr}\, E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\} \end{aligned} \tag{8.14}$$

hybrid minimum variance – minimum bias norm
α-weighted

$$\mathcal{L}(\mathbf{L}) := \ ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + \frac{1}{\alpha}||(\mathbf{I}_\ell - \mathbf{LC})'||^2_\mathbf{S} \tag{8.15}$$

special model

$$\dim \mathcal{R}(\mathbf{SC}') = \mathrm{rk}\mathbf{SC}' = \mathrm{rk}\mathbf{C} = l. \tag{8.16}$$

The *bias vector* $\boldsymbol{\beta}$ is conventionally defined by $E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\}$ subject to the homogeneous prediction form $\tilde{\mathbf{z}}=\mathbf{Ly}$. Accordingly the bias vector can be represented by (8.5) $\boldsymbol{\beta} = -[\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}$. Since the expectation $E\{\mathbf{z}\}$ of the vector $\mathbf{z}$ of random effects is unknown, there has been made the proposal to use instead the matrix $\mathbf{I}_\ell - \mathbf{LC}$ as a matrix-valued measure of bias. A measure of the prediction error is the Mean Square prediction error $\mathrm{MSPE}\{\tilde{\mathbf{z}}\}$ of type (8.9). $MSPE\{\tilde{\mathbf{z}}\}$ can be decomposed into three basic parts:

- The dispersion matrix $D\{\tilde{\mathbf{z}}\} = \mathbf{L}D\{\mathbf{y}\}\mathbf{L}'$
- The dispersion matrix $D\{\mathbf{z}\}$
- The bias product $\beta\beta'$.

Indeed the vector $\tilde{\mathbf{z}} - \mathbf{z}$ can also be decomposed into three parts of type (8.7), (8.8) namely (i) $\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}$, (ii) $\mathbf{z} - E\{\mathbf{z}\}$ and (iii) $E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\}$ which may be called prediction error, random effect error and bias, respectively. The triple decomposition of the vector $\tilde{\mathbf{z}} - \mathbf{z}$ leads straightforward to the triple representation of the matrix $MSPE\{\tilde{\mathbf{z}}\}$ of type (8.10). Such a representation suffers from two effects: *Firstly* the expectation $E\{\mathbf{z}\}$ of the vector $\mathbf{z}$ of random effects is unknown, *secondly* the matrix $E\{\mathbf{z}\}E\{\mathbf{z}'\}$ has only rank 1. In consequence, the matrix $[\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_\ell - \mathbf{LC}]'$ has only rank 1, too. In this situation there has made the proposal to modify $MSPE\{\tilde{\mathbf{z}}\}$ by the matrix $E\{\mathbf{z}\}E\{\mathbf{z}'\}$ and by the regular matrix $\mathbf{S}$. $MSPE\{\tilde{\mathbf{z}}\}$ has been defined by (8.11). A scalar measure of $MSPE\{\tilde{\mathbf{z}}\}$ as well as $MSPE\{\tilde{\mathbf{z}}\}$ are the *Frobenius norms* (8.12), (8.13), (8.14). Those scalars constitute the optimal risk in *Definition 8.1* (*hom* BLIP) and *Definition 8.2* (*hom* **S**-BLIP). Alter-natively a homogeneous α-weighted hybrid minimum variance–minimum

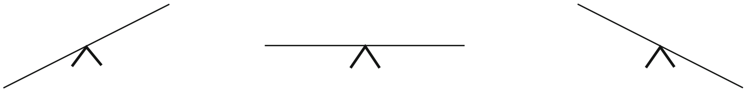


Fig. 8.2 Balance of variance and bias by the weight factor α

bias prediction (*hom* VIP) is presented in *Definition 8.3* (*hom* α-VIP) which is based upon the weighted sum of two norms of type (8.15), namely

- Average variance $||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} = tr\mathbf{L}\Sigma_\mathbf{y}\mathbf{L}'$
- Average bias $||(\mathbf{I}_\ell - \mathbf{LC})'||^2_\mathbf{S} = tr\,[\mathbf{I}_\ell - \mathbf{LC}]\,\mathbf{S}\,[\mathbf{I}_\ell - \mathbf{LC}]'$.

The very important predictor α-VIP is balancing variance and bias by the weight factor α which is illustrated by *Fig. 8.2*.

Definition 8.1. ($\tilde{\mathbf{z}}$ *hom* BLIP of $\mathbf{z}$):

An $l1$ vector $\tilde{\mathbf{z}}$ is called homogeneous BLIP of $\mathbf{z}$ in the *special linear Gauss–Markov model with random effects of Box 8.1*, if
(1st) $\tilde{\mathbf{z}}$ is a homogeneous linear form

$$\tilde{\mathbf{z}} = \mathbf{Ly}, \tag{8.17}$$

(2nd) in comparison to all other linear predictions $\tilde{\mathbf{z}}$ has the minimum Mean Square Prediction Error in the sense of

$$\begin{aligned}||MSPE\{\tilde{\mathbf{z}}\}||^2 &= \mathrm{tr}\,\mathbf{L}D\{\mathbf{y}\}\mathbf{L}' + \mathrm{tr}\,D\{\mathbf{z}\} + \mathrm{tr}\,[\mathbf{I}_\ell - \mathbf{LC}]E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_\ell - \mathbf{LC}]'\\ &= ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_\ell - \mathbf{LC})'||^2_{E\{(\mathbf{z})E(\mathbf{z})'} + \mathrm{tr}\,E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\}.\end{aligned} \tag{8.18}$$

Definition 8.2. ($\tilde{z}$ S-hom BLIP of z):

An $l1$ vector $\tilde{\mathbf{z}}$ is called homogeneous **S**-BLIP of $\mathbf{z}$ in the *special linear Gauss–Markov model with random effects of Box 8.1*, if
(1st) $\tilde{\mathbf{z}}$ is a homogeneous linear form

$$\tilde{\mathbf{z}} = \mathbf{Ly}, \tag{8.19}$$

(2nd) in comparison to all other linear predictions $\tilde{\mathbf{z}}$ has the minimum S-modified Mean Square Prediction Error in the sense of

$$
\begin{aligned}
&||MSPE_{\mathbf{S}}\{\tilde{\mathbf{z}}\}||^2 \\
:=& \operatorname{tr} \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \operatorname{tr}[\mathbf{I}_\ell - \mathbf{LC}]\mathbf{S}[\mathbf{I}_\ell - \mathbf{LC}]' + \operatorname{tr} E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\} \\
=& \ ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + ||(\mathbf{I}_\ell - \mathbf{LC})'||^2_{\mathbf{S}} + \operatorname{tr} E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\} = \min_{\mathbf{L}}. \qquad (8.20)
\end{aligned}
$$

Definition 8.3. ($\tilde{\mathbf{z}}$ *hom* hybrid min var-min bias solution, α-weighted or *hom* α-VIP):

An $l1$ vector $\tilde{\mathbf{z}}$ is called homogeneous α-weighted hybrid minimum variance–minimum bias prediction (hom α-VIP) of $\mathbf{z}$ in the *special linear Gauss–Markov model with random effects of Box 8.1*, if
(1st) $\tilde{\mathbf{z}}$ is a homogeneous linear form

$$\tilde{\mathbf{z}} = \mathbf{Ly}, \qquad (8.21)$$

(2nd) in comparison to all other linear predictions $\tilde{\mathbf{z}}$ has the minimum variance-minimum bias in the sense of the α-weighted hybrid norm

$$
\begin{aligned}
&\operatorname{tr} \mathbf{L} D\{\mathbf{y}\}\mathbf{L}' + \frac{1}{\alpha}\operatorname{tr}(\mathbf{I}_\ell - \mathbf{LC})\,\mathbf{S}\,(\mathbf{I}_\ell - \mathbf{LC})' \\
=& \ ||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + \frac{1}{\alpha}||(\mathbf{I}_\ell - \mathbf{LC})'||^2_{\mathbf{S}} = \min_{\mathbf{L}}
\end{aligned}
\qquad (8.22)
$$

in particular with respect to the special model

$$\alpha \in \mathbb{R}^+, \ \dim \mathcal{R}(\mathbf{SC}') = \operatorname{rk}\mathbf{SC}' = \operatorname{rk}\mathbf{C} = l.$$

The predictions $\tilde{\mathbf{z}}$ *hom* BLIP, *hom* $\mathbf{S}$-BLIP and *hom* α-VIP can be characterized as follows:

Lemma 8.4. (*hom* BLIP, *hom* $\mathbf{S}$-BLIP and hom α-VIP):

An $l1$ vector $\tilde{\mathbf{z}}$ is *hom* BLIP, *hom* $\mathbf{S}$-BLIP or *hom* α-VIP of $\mathbf{z}$ in the *special linear Gauss–Markov model with random effects of Box 8.1*, if and only if the matrix $\hat{\mathbf{L}}$ fulfils the normal equations
(1st) *hom* BLIP:

$$(\Sigma_\mathbf{y} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}')\hat{\mathbf{L}}' = \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}' \qquad (8.23)$$

(2nd) *hom* $\mathbf{S}$-BLIP:

$$(\Sigma_\mathbf{y} + \mathbf{CSC}')\hat{\mathbf{L}}' = \mathbf{CS} \qquad (8.24)$$

(3rd) *hom* α-VIP:

$$\left(\Sigma_\mathbf{y} + \frac{1}{\alpha}\mathbf{CSC}'\right)\hat{\mathbf{L}}' = \frac{1}{\alpha}\mathbf{CS} \qquad (8.25)$$

Proof. (i) hom BLIP:

The hybrid norm $||MSPE\{\tilde{\mathbf{z}}\}||^2$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) := \operatorname{tr} \mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' + \operatorname{tr} (\mathbf{I}_l - \mathbf{L}\mathbf{C})E\{\mathbf{z}\}E\{\mathbf{z}\}'(\mathbf{I}_l - \mathbf{L}\mathbf{C})' + \operatorname{tr} \Sigma_{\mathbf{z}} = \min_{\mathbf{L}}$$

for $\tilde{\mathbf{z}}$ *hom* BLIP of $\mathbf{z}$. The *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$ are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) := 2[\Sigma_{\mathbf{y}}\hat{\mathbf{L}}' + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'\hat{\mathbf{L}}' - \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'] = 0,$$

which agree to the normal equations (8.23). (The theory of matrix derivatives is reviewed in Appendix A.7 (Facts: derivative of a scalar-valued function of a matrix: *trace*)). The second derivatives

$$\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\mathbf{L})\partial(\operatorname{vec}\mathbf{L})'}(\hat{\mathbf{L}}) > 0$$

at the "point" $\hat{\mathbf{L}}$ constitute the sufficiency conditions. In order to compute such an $lnln$ matrix of second derivatives we have to vectorize the matrix normal equation

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) := 2\hat{\mathbf{L}}(\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}') - 2E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'$$

$$\frac{\partial \mathcal{L}}{\partial(\operatorname{vec}\mathbf{L})}(\hat{\mathbf{L}}) := \operatorname{vec}[2\hat{\mathbf{L}}(\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}') - 2E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'] .$$

(ii) hom S-BLIP:

The hybrid norm $||MSPE_{\mathbf{s}}\{\tilde{\mathbf{z}}\}||^2$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) := \operatorname{tr} \mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' + \operatorname{tr} (\mathbf{I}_\ell - \mathbf{L}\mathbf{C})\mathbf{S}(\mathbf{I}_\ell - \mathbf{L}\mathbf{C})' + \operatorname{tr} \Sigma_{\mathbf{z}} = \min_{\mathbf{L}}$$

for $\tilde{\mathbf{z}}$ *hom* **S**-BLIP of $\mathbf{z}$. Following the first part of the proof we are led to the *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) := 2[\Sigma_{\mathbf{y}}\hat{\mathbf{L}}' + \mathbf{C}\mathbf{S}\mathbf{C}'\hat{\mathbf{L}}' - \mathbf{C}\mathbf{S}]' = 0$$

as well as to the *sufficiency conditions*

$$\frac{\partial^2 \mathcal{L}}{\partial(\operatorname{vec}\mathbf{L})\partial(\operatorname{vec}\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\Sigma_{\mathbf{y}} + \mathbf{C}\mathbf{S}\mathbf{C}') \otimes \mathbf{I}_\ell] > 0 .$$

The *normal equations* of *hom* **S**-BLIP $\partial\mathcal{L}/\partial\mathbf{L}(\hat{\mathbf{L}}) = 0$ agree to (8.24).

(iii) hom α-VIP:

The hybrid norm $||\mathbf{L}'||^2_{\Sigma_\mathbf{y}} + \dfrac{1}{\alpha}||(\mathbf{I}_\ell - \mathbf{LC})'||^2_\mathbf{S}$ establishes the *Lagrangean*

$$\mathcal{L}(\mathbf{L}) := \operatorname{tr} \mathbf{L}\Sigma_\mathbf{y}\mathbf{L}' + \frac{1}{\alpha}\operatorname{tr}(\mathbf{I}_\ell - \mathbf{LC})\mathbf{S}(\mathbf{I}_\ell - \mathbf{LC})' = \min_\mathbf{L}$$

for $\tilde{\mathbf{z}}$ *hom* α-VIP of $\mathbf{z}$. Following the first part of the proof we are led to the *necessary conditions* for the minimum of the *quadratic Lagrangean* $\mathcal{L}(\mathbf{L})$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) = 2[(\Sigma_\mathbf{y} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}') \otimes \mathbf{I}_\ell]\text{vec}\hat{\mathbf{L}} - 2\text{vec}(E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}').$$

The *Kronecker–Zehfuss Product* $\mathbf{A} \otimes \mathbf{B}$ of two arbitrary matrices as well as $(\mathbf{A}+\mathbf{B}) \otimes \mathbf{C} = \mathbf{A}\otimes\mathbf{B}+\mathbf{B}\otimes\mathbf{C}$ of three arbitrary matrices subject to $dim\mathbf{A} = dim\mathbf{B}$ is introduced in Appendix A. (Definition of Matrix Algebra: multiplication matrices of the same dimension (internal relation) and multiplication of matrices (internal relation) and Laws). The vec operation (vectorization of an array) is reviewed in Appendix A, too. (Definition, Facts: $\text{vec}\mathbf{AB} = (\mathbf{B}' \otimes \mathbf{I}'_\ell)\text{vec}\mathbf{A}$ for suitable matrices A and B.) Now we are prepared to compute

$$\frac{\partial^2 \mathcal{L}}{\partial(vec\mathbf{L})\partial(vec\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\Sigma_\mathbf{y} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}\mathbf{C}') \otimes \mathbf{I}_\ell] > 0$$

as a *positive definite matrix*. (The theory of matrix derivatives is reviewed in Appendix B (Facts: derivative of a matrix-valued function of a matrix, namely $\partial(\text{vec}\mathbf{X})/\partial(\text{vec}\mathbf{X})'$).

$$\frac{\partial \mathcal{L}}{\partial \mathbf{L}}(\hat{\mathbf{L}}) = 2[\tfrac{1}{\alpha}\mathbf{CSC}'\hat{\mathbf{L}}' + \Sigma_\mathbf{y}\hat{\mathbf{L}}' - \tfrac{1}{\alpha}\mathbf{CS}]'\alpha = 0$$

as well as to the *sufficiency conditions*

$$\frac{\partial^2 \mathcal{L}}{\partial(\text{vec}\mathbf{L})\partial(\text{vec}\mathbf{L})'}(\hat{\mathbf{L}}) = 2[(\tfrac{1}{\alpha}\mathbf{CSC}' + \Sigma_\mathbf{y}) \otimes \mathbf{I}_\ell] > 0.$$

The *normal equations* of *hom* α-VIP $\partial\mathcal{L}/\partial\mathbf{L}(\hat{\mathbf{L}}) = 0$ agree to (8.25). ♣

For an *explicit representation* of $\tilde{\mathbf{z}}$ as *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP of z in the *special Gauss–Markov model with random effects of Box 8.1*, we solve the normal equations (8.23), (8.24) and (8.25) for

$$\hat{\mathbf{L}} = \arg\{\mathcal{L}(\mathbf{L}) = \min_\mathbf{L}\}.$$

Beside the *explicit representation* of $\tilde{\mathbf{z}}$ of type *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP we compute the related dispersion matrix $D\{\tilde{\mathbf{z}}\}$, the *Mean Square Prediction*

Error $MSPE\{\tilde{\mathbf{z}}\}$, the modified the *Mean Square Prediction Error* $MSPE_{\mathbf{S}}\{\tilde{\mathbf{z}}\}$ and $MSPE_{\alpha,\mathbf{S}}\{\tilde{\mathbf{z}}\}$ and the covariance matrices $C\{\mathbf{z},\, \tilde{\mathbf{z}} - \mathbf{z}\}$ in

Theorem 8.5. ($\tilde{\mathbf{z}}$ hom BLIP):

Let $\tilde{\mathbf{z}} = \mathbf{Ly}$ be *hom* BLIP of z in the *special linear Gauss–Markov model with random effects of Box 8.1*. Then equivalent representations of the solutions of the normal equations (8.23) are

$$\begin{aligned} \tilde{\mathbf{z}} &= E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\mathbf{y} \\ &= E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}-\mathbf{Cz}} + \mathbf{C}\Sigma_{\mathbf{z}}\mathbf{C}' + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\mathbf{y} \end{aligned} \tag{8.26}$$

$$(\text{if } [\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1} \text{ exists})$$

are completed by the dispersion matrix

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} = E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\,\Sigma_{\mathbf{y}} \\ \times[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}' \end{aligned} \tag{8.27}$$

by the *bias vector* (8.5)

$$\begin{aligned} \boldsymbol{\beta} &:= E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\} \\ &= -[\mathbf{I}_\ell - E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'(\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}]E\{\mathbf{z}\} \end{aligned} \tag{8.28}$$

and by the matrix of the *Mean Square Prediction Error* $MSPE\{\tilde{\mathbf{z}}\}$:

$$\begin{aligned} MSPE\{\tilde{\mathbf{z}}\} &:= E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} \\ &= D\{\tilde{\mathbf{z}}\} + D\{\mathbf{z}\} + \beta\beta' \end{aligned} \tag{8.29}$$

$$\begin{aligned} MSPE\{\tilde{\mathbf{z}}\} := D\{\tilde{\mathbf{z}}\} + D\{\mathbf{z}\} + [\mathbf{I}_\ell - E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'(\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}] \\ \times\; E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_\ell - \mathbf{C}'(\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}']. \end{aligned} \tag{8.30}$$

At this point we have to comment what *Theorem 8.5* tells us. *hom* BLIP has generated the prediction$\tilde{\mathbf{z}}$ of type (8.26), the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ of type (8.27), the bias vector of type (8.28) and the *Mean Square Prediction Error* of type (8.30) *which all depend on the vector* $E\{\mathbf{z}\}$ and the matrix $E\{\mathbf{z}\}E\{\mathbf{z}\}'$, respectively. We already mentioned that $E\{\mathbf{z}\}$ and $E\{\mathbf{z}\}E\{\mathbf{z}\}'$ are not accessible from measurements. The situation is similar to the one in hypothesis theory. As shown later in this section we can produce only an estimator $\widehat{E\{\mathbf{z}\}}$ and consequently can setup a hypothesis first moment $E\{\mathbf{z}\}$ of the "*random effect*" $\mathbf{z}$. Indeed, a similar argument applies to the *second central moment* $D\{\mathbf{y}\} \sim \Sigma_{\mathbf{y}}$ of the "*random effect*" $\mathbf{y}$, the observation vector. Such a dispersion matrix has to be known in order to be able to compute $\tilde{\mathbf{z}}$, $D\{\tilde{\mathbf{z}}\}$, and $MSPE\{\tilde{\mathbf{z}}\}$. Again we have to apply the argument that we are only able to construct an estimate $\hat{\Sigma}'_{\mathbf{y}}$ and to setup a hypothesis about $D\{\mathbf{y}\} \sim \Sigma_{\mathbf{y}}$.

Theorem 8.6. ($\tilde{\mathbf{z}}$ hom S-BLIP):

Let $\tilde{\mathbf{z}} = \mathbf{L}\mathbf{y}$ be *hom* $\mathbf{S}$-BLIP of $\mathbf{z}$ in the *special linear Gauss–Markov model with random effects of Box 8.1*. Then equivalent representations of the solutions of the normal equations (8.24) are

$$\begin{aligned}\tilde{\mathbf{z}} &= \mathbf{S}\mathbf{C}'(\Sigma_{\mathbf{y}} + \mathbf{C}\mathbf{S}\mathbf{C}')^{-1}\mathbf{y} \\ &= \mathbf{S}\mathbf{C}'(\Sigma_{\mathbf{y}-\mathbf{C}\mathbf{z}} + \mathbf{C}\Sigma_{\mathbf{z}}\mathbf{C}' + \mathbf{C}\mathbf{S}\mathbf{C}')^{-1}\mathbf{y}\end{aligned} \tag{8.31}$$

$$\tilde{\mathbf{z}} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y} \tag{8.32}$$

$$\tilde{\mathbf{z}} = (\mathbf{I}_\ell + \mathbf{S}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{S}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y} \tag{8.33}$$

$$(\text{if } \mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1} \text{ exist})$$

are completed by the dispersion matrices

$$D\{\tilde{\mathbf{z}}\} = \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\Sigma_{\mathbf{y}}(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}\mathbf{S} \tag{8.34}$$

$$D\{\tilde{\mathbf{z}}\} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1} \tag{8.35}$$

$$(\text{if } \mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1} \text{ exist})$$

by the bias vector (8.5)

$$\begin{aligned}\boldsymbol{\beta} &:= E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\} \\ &= -[\mathbf{I}_\ell - \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}]E\{\mathbf{z}\}\end{aligned}$$

$$\boldsymbol{\beta} = -[\mathbf{I}_l - (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}]E\{\mathbf{z}\} \tag{8.36}$$

$$(\text{if } \mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1} \text{ exist})$$

and by the matrix of the modified *Mean Square Prediction Error MSPE*$\{\tilde{\mathbf{z}}\}$:

$$\begin{aligned}MSPE_{\mathbf{S}}\{\tilde{\mathbf{z}}\} &:= E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} \\ &= D\{\tilde{\mathbf{z}}\} + D\{\mathbf{z}\} + \beta\beta'\end{aligned} \tag{8.37}$$

$$\begin{aligned}&MSPE_{\mathbf{S}}\{\tilde{\mathbf{z}}\} = \Sigma_{\mathbf{z}} + \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\Sigma_{\mathbf{y}}(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}\mathbf{S} \\ &+[\mathbf{I}_\ell - \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}]E\{\mathbf{z}\}E\{\mathbf{z}\}'[\mathbf{I}_l - \mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{C}\mathbf{S}]\end{aligned} \tag{8.38}$$

$$
\begin{aligned}
MSPE_{\mathbf{S}}\{\tilde{\mathbf{z}}\} = {} & \Sigma_{\mathbf{z}} + (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{CS} \\
& + [\mathbf{I}_\ell - (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}]E\{\mathbf{z}\}E\{\mathbf{z}\}' \\
& \times [\mathbf{I}_\ell - \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}] \qquad (8.39)
\end{aligned}
$$

$$(\text{if } \mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1} \text{ exist}).$$

The interpretation of *hom* **S**-BLIP is even more complex. In extension of the comments to *hom* BLIP we have to live with another matrix-valued degree of freedom, $\tilde{\mathbf{z}}$ of type (8.31), (8.32), (8.33) and $D\{\tilde{\mathbf{z}}\}$ of type (8.34), (8.35) do no longer depend on the inaccessible matrix $E\{\mathbf{z}\}E\{\mathbf{z}\}'$, $\text{rk}(E\{\mathbf{z}\}E\{\mathbf{z}\}')$, but on the "*bias weight matrix*" **S**, rk **S** = l. Indeed we can associate any element of the bias matrix with a particular weight which can be "*designed*" by the analyst. Again the *bias vector* $\boldsymbol{\beta}$ of type (8.36) as well as the *Mean Square Prediction Error* of type (8.37), (8.38), (8.39) depend on the vector $E\{\mathbf{z}\}$ which is inaccessible. Beside the "*bias weight matrix* **S**" $\tilde{\mathbf{z}}$, $D\{\tilde{\mathbf{z}}\}$, $\boldsymbol{\beta}$ and $MSPE\{\tilde{\mathbf{z}}\}$ are vector-valued or matrix-valued functions of the *dispersion matrix* $D\{\mathbf{y}\} \sim \Sigma_{\mathbf{y}}$ of the stochastic observation vector which is inaccessible. By *hypothesis testing* we may decide upon the construction of $D\{\mathbf{y}\} \sim \Sigma_{\mathbf{y}}$ from an estimate $\widehat{\Sigma}_{\mathbf{y}}$.

Theorem 8.7. ($\tilde{\mathbf{z}}$ hom a-VIP):

Let $\tilde{\mathbf{z}} = \mathbf{Ly}$ be *hom* α-VIP of z in the *special linear Gauss–Markov model with random effects Box 8.1*. Then equivalent representations of the solutions of the normal equations (8.25) are

$$
\begin{aligned}
\tilde{\mathbf{z}} &= \frac{1}{\alpha}\mathbf{SC}'\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\mathbf{y} \\
&= \frac{1}{\alpha}\mathbf{SC}'\left(\Sigma_{\mathbf{y}-\mathbf{Cz}} + \mathbf{C}\Sigma_{\mathbf{z}}\mathbf{C}' + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\mathbf{y}
\end{aligned} \qquad (8.40)
$$

$$\tilde{\mathbf{z}} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y} \qquad (8.41)$$

$$\tilde{\mathbf{z}} = (\mathbf{I}_\ell + \frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y} \qquad (8.42)$$

$$(\text{if } \mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1} \text{ exist})$$

are completed by the dispersion matrix

$$D\{\tilde{\mathbf{z}}\} = \frac{1}{\alpha}\mathbf{SC}'\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\Sigma_{\mathbf{y}}\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\mathbf{CS}\frac{1}{\alpha} \qquad (8.43)$$

$$D\{\tilde{\mathbf{z}}\} = \left(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1}\right)^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}\left(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1}\right)^{-1} \qquad (8.44)$$

(if $\mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1}$ exist)
by the *bias vector* (8.5)

$$\boldsymbol{\beta} := E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\} = -\left[\mathbf{I}_\ell - \frac{1}{\alpha}\mathbf{SC}'\left(\frac{1}{\alpha}\mathbf{CSC}' + \Sigma_{\mathbf{y}}\right)^{-1}\mathbf{C}\right]E\{\mathbf{z}\}$$

$$\boldsymbol{\beta} = -[\mathbf{I}_\ell - (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}]E\{\mathbf{z}\} \tag{8.45}$$

(if $\mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1}$ exist)

and by the matrix of the *Mean Square Prediction Error* $MSPE\{\tilde{\mathbf{z}}\}$:

$$\begin{aligned} MSPE\{\tilde{\mathbf{z}}\} &:= E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} \\ &= D\{\tilde{\mathbf{z}}\} + D\{\mathbf{z}\} + \beta\beta' \end{aligned} \tag{8.46}$$

$$\begin{aligned} MSPE\{\tilde{\mathbf{z}}\} = {} & \Sigma_{\mathbf{z}} + \mathbf{SC}'(\mathbf{CSC}' + \Sigma_{\mathbf{y}})^{-1}\Sigma_{\mathbf{y}}(\mathbf{CSC}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{CS} \\ & + \left[\mathbf{I}_\ell - \frac{1}{\alpha}\mathbf{SC}'\left(\frac{1}{\alpha}\mathbf{CSC}' + \Sigma_{\mathbf{y}}\right)^{-1}\mathbf{C}\right]E\{\mathbf{z}\}E\{\mathbf{z}\}' \\ & \times\left[\mathbf{I}_\ell - \mathbf{C}'\left(\frac{1}{\alpha}\mathbf{CSC}' + \Sigma_{\mathbf{y}}\right)^{-1}\mathbf{CS}\frac{1}{\alpha}\right] \end{aligned} \tag{8.47}$$

$$\begin{aligned} MSPE\{\tilde{\mathbf{z}}\} = {} & \Sigma_{\mathbf{z}} + (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{CS} \\ & + [\mathbf{I}_\ell - (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}]E\{\mathbf{z}\}E\{\mathbf{z}\}' \\ & \times[\mathbf{I}_\ell - \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}] \end{aligned} \tag{8.48}$$

(if $\mathbf{S}^{-1},\ \Sigma_{\mathbf{y}}^{-1}$ exist).

The interpretation of the very important predictors *hom* α-VIP $\tilde{\mathbf{z}}$ of $z$is as follows: $\tilde{\mathbf{z}}$ of type (8.41), also called *ridge estimator* or *Tykhonov–Phillips regulator*, contains the *Cayley inverse* of the normal equation matrix which is *additively decomposed* into $\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}$ and $\alpha\,\mathbf{S}^{-1}$. The weight factor α balances the *first inverse dispersion* part and the *second inverse bias part*. While the experiment informs us of the variance-covariance matrix $\Sigma_{\mathbf{y}}$, say $\widehat{\Sigma}_{\mathbf{y}}$, the *bias weight matrix* $\mathbf{S}$ and the *weight factor* α are at the disposal of the analyst. For instance, by the choice $\mathbf{S} = Diag(s_1, ..., s_\ell)$ we may emphasize increase or decrease of certain bias matrix elements. The choice of an equally weighted bias matrix is $\mathbf{S} = \mathbf{I}_\ell$. In contrast the weight factor α can be determined by the *A-optimal design* of type

- $trD\{\tilde{\mathbf{z}}\} = \min_{\alpha}$
- $\beta\beta' = \min_{\alpha}$
- $trMSPE\{\tilde{\mathbf{z}}\} = \min_{\alpha}$.

In the *first case* we optimize the *trace of the variance-covariance matrix* $D\{\tilde{\mathbf{z}}\}$ of type (8.43), (8.44). Alternatively by means of $\beta\beta' = \min_{\alpha}$ we optimize the *quadratic bias* where the bias vector β of type (8.45) is chosen, regardless of the dependence on $E\{\mathbf{z}\}$. Finally for the *third case* – the most popular one – we minimize the trace of the Mean Square Prediction Error $MSPE\{\tilde{\mathbf{z}}\}$ of type (8.48), regardless of the dependence on $E\{\mathbf{z}\}E\{\mathbf{z}\}'$. But beforehand let us present the *proof* of *Theorem 8.5*, *Theorem 8.6* and *Theorem 8.7*.

Proof.

$$(i)\, \tilde{\mathbf{z}} = E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\mathbf{y}$$

If the matrix $\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'$ of the normal equations of type *hom* BLIP is of full rank, namely $\text{rk}(\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}') = n$, then a straightforward solution of (8.23) is

$$\hat{\mathbf{L}} = E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}.$$

$$(ii)\, \tilde{\mathbf{z}} = \mathbf{SC}'(\Sigma_{\mathbf{y}} + \mathbf{CSC}')^{-1}\mathbf{y}$$

If the matrix $\Sigma_{\mathbf{y}} + \mathbf{CSC}'$ of the normal equations of type *hom* **S**-BLIP is of full rank, namely $\text{rk}(\Sigma_{\mathbf{y}} + \mathbf{CSC}') = n$, then a straightforward solution of (8.24) is

$$\hat{\mathbf{L}} = \mathbf{SC}'[\Sigma_{\mathbf{y}} + \mathbf{CSC}']^{-1}.$$

$$(iii)\, \tilde{\mathbf{z}} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(10), Duncan–Guttman matrix identity) the fundamental matrix identity

$$\mathbf{SC}'(\Sigma_{\mathbf{y}} + \mathbf{CSC}')^{-1} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1},$$

if $\mathbf{S}^{-1}$ and $\Sigma_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(iv)\, \tilde{\mathbf{z}} = (\mathbf{I}_{\ell} + \mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(9)) the fundamental matrix identity

$$\mathbf{SC}'(\Sigma_{\mathbf{y}} + \mathbf{CSC}')^{-1} = (\mathbf{I}_{\ell} + \mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1},$$

if $\Sigma_{\mathbf{y}}^{-1}$ exists. Such a result concludes this part of the proof.

$$(v)\,\tilde{\mathbf{z}} = \frac{1}{\alpha}\mathbf{SC}'\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\mathbf{y}$$

If the matrix $\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'$ of the normal equations of type *hom* α-VIP is of full rank, namely $\mathrm{rk}\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right) = n$, then a straightforward solution of (8.25) is

$$\hat{\mathbf{L}} = \frac{1}{\alpha}\mathbf{SC}'\left[\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right]^{-1}.$$

$$(vi)\,\tilde{\mathbf{z}} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(10), Duncan–Guttman matrix identity) the fundamental matrix identity

$$\frac{1}{\alpha}\mathbf{SC}'(\Sigma_{\mathbf{y}} + \mathbf{CSC}')^{-1} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1},$$

if $\mathbf{S}^{-1}$ and $\Sigma_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(vii)\,\tilde{\mathbf{z}} = \left(\mathbf{I}_\ell + \frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}\right)^{-1}\frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{y}$$

Let us apply by means of Appendix A (Fact: Cayley inverse: sum of two matrices, s(9), Duncan–Guttman matrix identity) the fundamental matrix identity

$$\frac{1}{\alpha}\mathbf{SC}'(\Sigma_{\mathbf{y}} + \mathbf{CSC}')^{-1} = \left(\mathbf{I}_l + \frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}\right)^{-1}\frac{1}{\alpha}\mathbf{SC}'\Sigma_{\mathbf{y}}^{-1},$$

if $\Sigma_{\mathbf{y}}^{-1}$ exist. Such a result concludes this part of the proof.

$$(viii)\,hom\ BLIP : D\{\tilde{\mathbf{z}}\}$$

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} :&= E\{[\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}][\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}]'\} \\ &= E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}'[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\Sigma_{\mathbf{y}} \\ &\quad \times[\Sigma_{\mathbf{y}} + \mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'\mathbf{C}']^{-1}\mathbf{C}E\{\mathbf{z}\}E\{\mathbf{z}\}'. \end{aligned}$$

By means of the definition of the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ and the substitution of $\tilde{\mathbf{z}}$ of type *hom* BLIP the proof has been straightforward.

$$(ix)\,hom\,\mathbf{S}\text{-BLIP} : D\{\tilde{\mathbf{z}}\}\ (\mathit{1st\ representation})$$

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} :&= E\{[\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}][\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}]'\} \\ &= \mathbf{SC}'(\mathbf{CSC}' + \Sigma_{\mathbf{y}})^{-1}\Sigma_{\mathbf{y}}(\mathbf{CSC}' + \Sigma_{\mathbf{y}})^{-1}\mathbf{CS}. \end{aligned}$$

By means of the definition of the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ and the substitution of $\tilde{\mathbf{z}}$ of type *hom* **S**-BLIP the proof of the first representation has been straightforward.

(x) *hom* **S**-BLIP : $D\{\tilde{\mathbf{z}}\}$ (*2nd representation*)

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} :&= E\{[\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}][\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}]'\} \\ &= (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \mathbf{S}^{-1})^{-1}, \end{aligned}$$

if $\mathbf{S}^{-1}$ and $\Sigma_{\mathbf{y}}^{-1}$ exist. By means of the definition of the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ and the substitution of $\tilde{\mathbf{z}}$ of type *hom* **S**-BLIP the proof the second representation has been straightforward.

(xi) *hom* α-VIP : $D\{\tilde{\mathbf{z}}\}$ (*1st representation*)

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} :&= E\{[\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}][\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}]'\} = \\ &= \frac{1}{\alpha}\mathbf{SC}'\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\Sigma_{\mathbf{y}}\left(\Sigma_{\mathbf{y}} + \frac{1}{\alpha}\mathbf{CSC}'\right)^{-1}\mathbf{CS}\frac{1}{\alpha}. \end{aligned}$$

By means of the definition of the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ and the substitution of $\tilde{\mathbf{z}}$ of type *hom* α-VIP the proof the first representation has been straightforward.

(xii) *hom* α-VIP : $D\{\tilde{\mathbf{z}}\}$ (*2nd representation)*

$$\begin{aligned} D\{\tilde{\mathbf{z}}\} :&= E\{[\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}][\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}]'\} \\ &= (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} + \alpha\mathbf{S}^{-1})^{-1}, \end{aligned}$$

if $\mathbf{S}^{-1}$and$\Sigma_{\mathbf{y}}^{-1}$ exist. By means of the definition of the dispersion matrix and the substitution of $\tilde{\mathbf{z}}$ of type *hom* α-VIP the proof of the second representation has been straightforward.

(xiii) bias β for *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP

As soon as we substitute into the bias $\beta := E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\} = -E\{\mathbf{z}\} + E\{\tilde{\mathbf{z}}\}$ the various predictors $\tilde{\mathbf{z}}$ of the type *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP we are directly led to various bias representations β of type *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP.

(xiv) *MSPE*$\{\tilde{\mathbf{z}}\}$ of type *hom* BLIP, *hom* **S**-BLIP and hom α-VIP

$$\text{MSPE}\{\tilde{\mathbf{z}}\} := E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\}$$
$$\tilde{\mathbf{z}} - \mathbf{z} = \tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\} - (\mathbf{z} - E\{\mathbf{z}\}) = \tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\} - (\mathbf{z} - E\{\mathbf{z}\}) - (E\{\mathbf{z}\} - E\{\tilde{\mathbf{z}}\})$$

$$\begin{gathered} E\{(\tilde{\mathbf{z}} - \mathbf{z})(\tilde{\mathbf{z}} - \mathbf{z})'\} = \\ E\{(\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\})((\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\})'\} + E\{(\mathbf{z} - E\{\mathbf{z}\})(\mathbf{z} - E\{\mathbf{z}\})'\} + \\ +(E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\})(E\{\tilde{\mathbf{z}}\} - E\{\mathbf{z}\})' \end{gathered}$$

$$MSPE\{\tilde{\mathbf{z}}\} = D\{\tilde{\mathbf{z}}\} + D\{\mathbf{z}\} + \beta\beta'.$$

At first we have defined the *Mean Square Prediction Error* $MSPE\{\tilde{\mathbf{z}}\}$ of $\tilde{\mathbf{z}}$. Secondly we have decomposed the difference $\tilde{\mathbf{z}} - \mathbf{z}$ into the *three terms*

- $\tilde{\mathbf{z}} - E\{\tilde{\mathbf{z}}\}$
- $\mathbf{z} - E\{\mathbf{z}\}$
- $E\{\mathbf{z}\} - E\{\tilde{\mathbf{z}}\}$

in order to derive *thirdly* the decomposition of $MSPE\{\tilde{\mathbf{z}}\}$, namely

- The dispersion matrix of $\tilde{\mathbf{z}}$, namely $D\{\tilde{\mathbf{z}}\}$
- The dispersion matrix of $\mathbf{z}$, namely $D\{\mathbf{z}\}$
- The quadratic bias $\beta\beta'$.

As soon as we substitute $MSPE\{\tilde{\mathbf{z}}\}$ the dispersion matrix $D\{\tilde{\mathbf{z}}\}$ and the bias vector β of various predictors $\tilde{\mathbf{z}}$ of the type *hom* BLIP, *hom* **S**-BLIP and *hom* α-VIP we are directly led to various representations β of the *Mean Square Prediction Error* $MSPE\{\tilde{\mathbf{z}}\}$. *Here is our proof's end.* ♣

8-2 Examples

Example 8.1 Nonlinear error propagation with random effect models

Consider a function $y = f(z)$ where y is a scalar valued observation and z a *random effect*. Three cases can be specified as follows:
Case 1 (μ_z assumed to be known):
By *Taylor series expansion* we have

$$f(z) = f(\mu_z) + \frac{1}{1!} f'(\mu_z)(z - \mu_z) + \frac{1}{2!} f''(\mu_z)(z - \mu_z)^2 + \mathcal{O}(3)$$

$$E\{y\} = E\{f(z)\} = f(\mu_z) + \frac{1}{2!} f''(\mu_z) E\{(z - \mu_z)^2\} + \mathcal{O}(3)$$

leading to (cf. *E. Grafarend* and *B. Schaffrin* 1983, p. 470)

$$E\{y\} = f(\mu_z) + \frac{1}{2!} f''(\mu_z)\sigma_z^2 + \mathcal{O}(3)$$

$$E\{(y - E\{y\})^2\} = E\{[f'(\mu_z)(z - \mu_z) + \frac{1}{2!}f''(\mu_z)(z - \mu_z)^2 + \mathcal{O}(3)$$
$$-\frac{1}{2!}f''(\mu_z)\sigma_z^2 - \mathcal{O}(3)]^2\},$$

hence $E\{[y - E\{y\}][[y - E\{y\}]\}$ is given by

$$\sigma_y^2 = f'^2(\mu_z)\sigma_z^2 - \frac{1}{4}f''^2(\mu_z)\sigma_z^4 + f'f''(\mu_z)E\{(z - \mu_z)^3\}$$
$$+\frac{1}{4}f''^2 E\{(z - \mu_z)^4\} + \mathcal{O}(3).$$

Finally if z is quasi-normally distributed, we have

$$\sigma_y^2 = f'^2(\mu_z)\sigma_z^2 + \frac{1}{2}f''^2(\mu_z)\sigma_z^4 + \mathcal{O}(3).$$

Case 2 (μ_z unknown, but ξ_0 known as a fixed effect approximation (this model is implied in *E. Grafarend* and *B. Schaffrin* 1983, p.470, $\xi_0 \neq \mu_z$)):
By *Taylor series expansion* we have $f(z) = f(\xi_0) + \frac{1}{1!}f'(\xi_0)(z - \xi_0) + \frac{1}{2!}f''(\xi_0)(z - \xi_0)^2 + \mathcal{O}(3)$ using

$$\xi_0 = \mu_z + (\xi_0 - \mu_z) \Rightarrow z - \xi_0 = z - \mu_z + (\mu_z - \xi_0)$$

we have

$$f(z) = f(\xi_0) + \frac{1}{1!}f'(\xi_0)(z - \mu_z) + \frac{1}{1!}f'(\xi_0)(z - \xi_0)$$
$$+\frac{1}{2!}f''(\xi_0)(z - \mu_z)^2 + \frac{1}{2!}f''(\xi_0)(z - \xi_0)^2$$
$$+f''(\xi_0)(z - \mu_z)(z - \xi_0) + \mathcal{O}(3)$$

and

$$E\{y\} = E\{f(z)\} = f(\xi_0) + f'(\xi_0)(\mu_z - \xi_0) + \frac{1}{2}f''(\xi_0)\sigma_z^2$$
$$+\frac{1}{2}f''(\xi_0)(\mu_z - \xi_0)^2 + \mathcal{O}(3)$$

leading to $E\{[y - E\{y\}][[y - E\{y\}]\}$ as

$$\begin{aligned}\sigma_z^2 &= f'^2(\xi_0)\sigma_z^2 + f'f''(\xi_0)E\{(z-\mu_z)^3\} + 2f'f''(\xi_0)\sigma_z^2(\mu_z-\xi_0)\\ &+\frac{1}{4}f''^2(\xi_0)E\{(z-\mu_z)^4\} + f''^2(\xi_0)E\{(z-\mu_z)^3\}(\mu_z-\xi_0)\\ &-\frac{1}{4}f''^2(\xi_0)\sigma_z^4 + f''^2(\xi_0)\sigma_z^2(\mu_z-\xi_0)^2 + \mathcal{O}(3)\end{aligned}$$

and with z being quasi-normally distributed, we have

$$\sigma_z^2 = f'^2(\xi_0)\sigma_z^2 + 2f'f''(\xi_0)\sigma_z^2(\mu_z-\xi_0) + \frac{1}{2}f''^2(\xi_0)\sigma_z^4 + f''^2(\xi_0)\sigma_z^2(\mu_z-\xi_0)^2 + \mathcal{O}(3)$$

with the *first* and *third* terms (on the right hand side) being the right hand side-terms of case 1 (cf. *E. Grafarend* and *B. Schaffrin* 1983, p.470).
Case 3 (μ_z unknown, but z_0 known as a random effect approximation):
By *Taylor series expansion* we have

$$\begin{aligned}f(z) &= f(\mu_z) + \frac{1}{1!}f'(\mu_z)(z-\mu_z) + \frac{1}{2!}f''(\mu_z)(z-\mu_z)^2 +\\ &+\frac{1}{3!}f'''(\mu_z)(z-\mu_z)^3 + \mathcal{O}(4)\end{aligned}$$

changing

$$z-\mu_z = z_0-\mu_z = z_0 - E\{z_0\} - (\mu_z - E\{z_0\})$$

and the initial bias

$$-(\mu_z - E\{z_0\}) = E\{z_0\} - \mu_z =: \beta_0$$

leads to

$$z-\mu_z = z_0 - E\{z_0\} + \beta_0.$$

Consider

$$(z-\mu_z)^2 = (z_0 - E\{z_0\})^2 + \beta_0^2 + 2(z_0 - E\{z_0\})\beta_0$$

we have

$$\begin{aligned}f(z) &= f(\mu_z) + \frac{1}{1!}f'(\mu_z)(z_0 - E\{z_0\}) + \frac{1}{1!}f'(\mu_z)\beta_0\\ &+\frac{1}{2!}f''(\mu_z)(z_0-E\{z_0\})^2 + \frac{1}{2!}f''(\mu_z)\beta_0^2 + f''(\mu_z)(z_0 - E\{z_0\})\beta_0 + \mathcal{O}(3)\\ E\{y\} &= f(\mu_z) + f'(\mu_z)\beta_0 + \frac{1}{2}f''(\mu_z)\sigma_{z_0}^2 + \frac{1}{2}f''(\mu_z)\beta_0^2 + \mathcal{O}(3)\end{aligned}$$

leading to $E\{[y-E\{y\}][[y-E\{y\}]\}$ as

$$\begin{aligned}\sigma_y^2 = {} & f'^2(\mu_z)\sigma_{z_0}^2 + f'f''(\mu_z)E\{(z_0 - E\{z_0\})^3\} \\ & +2f'f''(\mu_z)\sigma_{z_0}^2\beta_0 + \frac{1}{4}f''^2(\mu_z)E\{(z_0 - E\{z_0\})^4\} \\ & +f''^2(\mu_z)E\{(z_0 - E\{z_0\})^3\}\beta_0 + f''^2(\mu_z)\sigma_{z_0}^2\beta_0^2 \\ & +\frac{1}{4}f''^2(\mu_z)\sigma_{z_0}^4 - \frac{1}{2}f''^2(\mu_z)E\{(z_0 - E\{z_0\})^2\}\sigma_{z_0}^2 + \mathcal{O}(3)\end{aligned}$$

and with z_0 being quasi-normally distributed, we have

$$\sigma_y^2 = f'^2(\mu_z)\sigma_{z_0}^2 + 2f'f''(\mu_z)\sigma_{z_0}^2\beta_0 + \frac{1}{2}f''^2(\mu_z)\sigma_{z_0}^4 + f''^2(\mu_z)\sigma_{z_0}^2\beta_0^2 + \mathcal{O}(3)$$

with the *first* and *third* terms (on the right-hand side) being the right-hand side terms of case 1.

Example 8.2 Nonlinear vector valued error propagation with random effect models

In a *GeoInformation System* we ask for the quality of a nearly *rectangular planar surface element*. Four points $\{P_1, P_2, P_3, P_4\}$ of an element are assumed to have the coordinates $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ and form a 88 full variance-covariance matrix (central moments of order two) and moments of higher order. The planar surface element will be computed according the *Gauß trapezoidal*:

$$F = \sum_{i=1}^{4} \frac{y_i + y_{i+1}}{2}(x_i - x_{i+1})$$

with the *side condition* $x_5 = x_1, y_5 = y_1$. Note that within the *Error Propagation Law*

$$\frac{\partial^2 F}{\partial x \partial y} \neq 0$$

holds (Fig. 8.3).

First question

? What is the structure of the variance-covariance matrix of the four points if we assume statistical homogeneity and isotropy of the network (Taylor–Karman structure)?

Second question

! Approach the criterion matrix in terms of absolute coordinates. Interpolate the correlation function linear! (Tables 8.1 and 8.2)

Our example refers to the *Taylor–Karman structure* or the structure function introduced in Sect. 3-222.

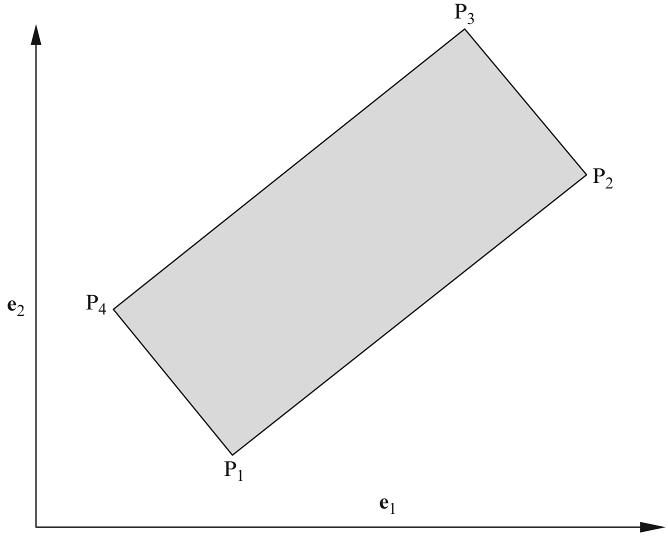


Fig. 8.3 Surface element of a building in the map

Table 8.1 Coordinates of a four-dimensional simplex

Point	x	y
P_1	100.00m	100.00m
P_2	110.00m	117.32m
P_3	101.34m	122.32m
P_4	91.34m	105.00m

Table 8.2 Longitudinal and lateral correlation functions *Σ_{m} and Σ_ℓ for a Taylor–Korman structured 4 point network*

$\lvert\mathbf{x}\rvert$	$\Sigma_m(\lvert\mathbf{x}\rvert)$	$\Sigma_l(\lvert\mathbf{x}\rvert)$
10m	0.700	0.450
20m	0.450	0.400
30m	0.415	0.238

:Solution:

The Gauß trapezoidal surface element has the size:

$$F = \frac{y_1 + y_2}{2}(x_1 - x_2) + \frac{y_2 + y_3}{2}(x_2 - x_3) + \frac{y_3 + y_4}{2}(x_3 - x_4) + \frac{y_4 + y_1}{2}(x_4 - x_1).$$

Once we apply the "*error propagation law*" we have to use.

$$\sigma_{\mathrm{F}}^2 = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' + \frac{1}{2}\mathbf{H}(\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}'.$$

In our case, $n = 1$ holds since we have only one function to be computed. In contrast, the *variance-covariance matrix* enjoys the format 88, while the *Jacobi matrix of first derivatives* is a 18 matrix and the *Hesse matrix of second derivatives* is a 164 matrix.

Table 8.3 Distances and meridian correlation function Σ_m and longitudinal correlation function Σ_ℓ

$p-q$	$\lvert\mathbf{x}_p-\mathbf{x}_q\rvert$	$\lvert\mathbf{x}_p-\mathbf{x}_q\rvert^2$	Σ_m	Σ_l	x_p-x_q	y_p-y_q
1-2	20.000	399.982	0.45	0.40	−10	−17.32
1-3	22.360	499.978	0.44	0.36	−1.34	−22.32
1-4	10.000	100.000	0.70	0.45	8.66	−5
2-2	10.000	100.000	0.70	0.45	8.66	−5
2-4	22.360	499.978	0.44	0.36	18.66	12.32
3-4	20.000	399.982	0.45	0.40	10	17.32

Table 8.4 Distance function versus $\Sigma_m(\mathbf{x})$, $\Sigma_\ell(\mathbf{x})$

$\lvert\mathbf{x}\rvert$	$\Sigma_{\mathrm{m}}(\mathbf{x})$	$\Sigma_\ell(\mathbf{x})$
10-20	$0.95-0.025\lvert\mathbf{x}\rvert$	$0.5-0.005\lvert\mathbf{x}\rvert$
20-30	$0.52-0.0035\lvert\mathbf{x}\rvert$	$0.724-0.0162\lvert\mathbf{x}\rvert$

Table 8.5 Taylor–Karman matrix for the case study

	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
x_1	1	0	0.438	–0.022	0.441	–0.005	0.512	0.108
y_1		1	–0.022	0.412	–0.005	0.361	0.108	0.638
x_2			1	0	0.512	0.108	0.381	–0.037
y_2				1	0.108	0.634	–0.037	0.417
x_3			symmetric		1	0	0.438	–0.022
y_3						1	–0.022	0.412
x_4							1	0
y_4								1

(i) The structure of the homogeneous and isotropic variance-covariance matrix is such that locally 22 variance-covariance matrices appear as unit matrices generating *local error circles of identical radius*.

(ii) The celebrated *Taylor–Karman matrix* for absolute coordinates is given by

$$\boldsymbol{\Sigma}_{ij}(\mathbf{x}_p,\mathbf{x}_q)=\Sigma_{\mathrm{m}}(|\mathbf{x}_p-\mathbf{x}_q|)\delta_{ij}+[\Sigma_\ell(|\mathbf{x}_p-\mathbf{x}_q|)-\Sigma_{\mathrm{m}}(|\mathbf{x}_p-\mathbf{x}_q|)]\frac{\Delta x_i\Delta x_j}{|\mathbf{x}_p-\mathbf{x}_q|^2}$$

subject to

$$\Delta x_1 := \Delta x = x_p - x_q,\ \Delta x_2 := \Delta y = y_p - y_q,\ i,j\in\{1,2\};\ p,q\in\{1,2,3,4\}.$$

By means of a *linear interpolation* we have derived the Taylor–Karman matrix by *Tables 8.3* and *8.4*.

Once we take care of $\sum_{\mathrm{m}}$ and $\sum_\ell$ as a function of the distance for gives values of tabulated distances we arrive at the Taylor–Karman correlation values of type *Table 8.5*.

Finally, we have computed the Jacobi matrix of first derivatives in Table 8.6 and the Hesse matrix of second derivatives in *Table 8.7*.

Table 8.6 Table of the Jacobi matrix

"*Jacobi matrix*"

$$\mathbf{J} = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial y_1}, \ldots, \frac{\partial F}{\partial x_4}, \frac{\partial F}{\partial y_4}\right]$$

$$\mathbf{J} = \frac{1}{2}[y_2 - y_4,\ x_4 - x_2,\ y_3 - y_1,\ x_1 - x_3,\ y_4 - y_2,\ x_2 - x_4, y_1 - y_3,\ x_3 - x_1]$$

$$\mathbf{J} = \tfrac{1}{2}[12.32,\ -18.66,\ 22.32,\ -1.34,\ -12.32,\ 18.66,\ -22.32,\ 1.34].$$

Note:

$$\left.\begin{array}{l} \dfrac{\partial F}{\partial x_i} = \dfrac{y_{i+1} - y_{i-1}}{2} \\ \dfrac{\partial F}{\partial y_i} = \dfrac{x_{i+1} - x_{i-1}}{2} \end{array}\right] \quad \begin{array}{l} x_0 = x_4 \\ y_0 = y_4 \\ x_5 = x_1 \\ y_5 = y_1. \end{array}$$

Table 8.7 Table Hesse matrix

"*Hesse matrix*"

$$\begin{aligned} \mathbf{H} &= \tfrac{\partial}{\partial \mathbf{x}'} \otimes \tfrac{\partial}{\partial \mathbf{x}'} F(\mathbf{x}) = \\ &= \frac{\partial}{\partial \mathbf{x}'} \otimes \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial y_1}, \ldots, \frac{\partial F}{\partial x_4}, \frac{\partial F}{\partial y_4}\right] \\ &= \left[\frac{\partial^2 F}{\partial x_1^2}, \frac{\partial^2 F}{\partial x_1 \partial y_1}, \ldots, \frac{\partial^2 F}{\partial x_1 \partial y_4}, \frac{\partial^2 F}{\partial y_1 \partial x_1}, \ldots, \frac{\partial^2 F}{\partial y_4 \partial x_4}, \frac{\partial^2 F}{\partial y^2}\right] \\ &= \left[\frac{\partial}{\partial x_1}\left(\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial y_4}\right), \ldots, \frac{\partial}{\partial y_4}\left(\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial y_4}\right)\right]. \end{aligned}$$

Note the detailed computation in *Table 8.8*.

Table 8.8 Second derivatives 0, +1/2, -1/2

:"*interims formulae: Hesse matrix*":

$$
\begin{aligned}
\frac{\partial^2 F}{\partial x_i \partial x_j} &= \frac{\partial^2 F}{\partial y_i \partial y_j} = 0 \quad \forall\ i, j = 1,2,3,4\\
\frac{\partial^2 F}{\partial x_i \partial y_i} &= \frac{\partial^2 F}{\partial y_i \partial x_i} = 0 \quad \forall\ i = 1,2,3,4\\
\frac{\partial^2 F}{\partial x_i \partial y_{i-1}} &= \frac{\partial^2 F}{\partial y_i \partial x_{i+1}} = -\frac{1}{2} \quad \forall\ i = 1,2,3,4\\
\frac{\partial^2 F}{\partial y_i \partial x_{i-1}} &= \frac{\partial^2 F}{\partial x_i \partial y_{i+1}} = \frac{1}{2} \quad \forall\ i = 1,2,3,4.
\end{aligned}
$$

Results

At first, we list the distances $\{P_1P_2, P_2P_3, P_3P_4, P_4P_1\}$ of the *trapezoidal finite* element by $|P_1P_2| = 20$ (for instance 20 m), $|P_2P_3| = 10$ (for instance 10 m), $|P_3P_4| = 20$ (for instance 20 m) and $|P_4P_1| = 10$ (for instance 10 m).

Second, we compute $\sigma_F^2(\textit{first term}) = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}'$ by

$$
\begin{aligned}
&\sigma_F^2(\textit{first term})\\
&= \frac{1}{2}[12.32,\ -18.66,\ 22.32,\ -1.34,\ -12.32,\ 18.62,\ -22.32,\ 1.34]\\
&\times \begin{bmatrix}
1 & 0 & 0.438 & -0.022 & 0.442 & -0.005 & 0.512 & 0.108\\
0 & 1 & -0.022 & 0.412 & -0.005 & 0.362 & 0.108 & 0.638\\
0.438 & -0.022 & 1 & 0 & 0.512 & 0.108 & 0.386 & -0.037\\
-0.022 & 0.412 & 0 & 1 & 0.108 & 0.638 & -0.037 & 0.418\\
0.442 & -0.005 & 0.512 & 0.108 & 1 & 0 & 0.438 & -0.022\\
-0.005 & 0.362 & 0.108 & 0.638 & 0 & 1 & -0.022 & 0.412\\
0.512 & 0.108 & 0.386 & -0.037 & 0.438 & -0.022 & 1 & 0\\
0.108 & 0.638 & -0.037 & 0.418 & -0.022 & 0.412 & 0 & 1
\end{bmatrix}\\
&\times \frac{1}{2}[12.32,\ -18.66,\ 22.32,\ -1.34,\ -12.32,\ 18.62,\ -22.32,\ 1.34]'\\
&= 334.7117.
\end{aligned}
$$

Third, we need to compute $\sigma_F^2(\textit{second term}) = \dfrac{1}{2}\mathbf{H}(\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}'$ by

$$
\sigma_F^2(\textit{second term}) = \frac{1}{2}\mathbf{H}(\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}' = 7.2222 \times 10^{-35}
$$

where

$$\mathbf{H} = \text{vec}\begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix},$$

and

$$\text{vec}\boldsymbol{\Sigma} = \text{vec}\begin{bmatrix} 1 & 0 & 0.438 & -0.022 & 0.442 & -0.005 & 0.512 & 0.108 \\ 0 & 1 & -0.022 & 0.412 & -0.005 & 0.362 & 0.108 & 0.638 \\ 0.438 & -0.022 & 1 & 0 & 0.512 & 0.108 & 0.386 & -0.037 \\ -0.022 & 0.412 & 0 & 1 & 0.108 & 0.638 & -0.037 & 0.418 \\ 0.442 & -0.005 & 0.512 & 0.108 & 1 & 0 & 0.438 & -0.022 \\ -0.005 & 0.362 & 0.108 & 0.638 & 0 & 1 & -0.022 & 0.412 \\ 0.512 & 0.108 & 0.386 & -0.037 & 0.438 & -0.022 & 1 & 0 \\ 0.108 & 0.638 & -0.037 & 0.418 & -0.022 & 0.412 & 0 & 1 \end{bmatrix}.$$

Finally, we get the variance of the planar surface element F

$$\sigma_F^2 = 334.7117 + 7.2222 \times 10^{-35} = 334.7117$$

i.e.

$$\sigma_F = \pm 18.2951(\text{m}^2).$$

Example 8.3 Nonlinear vector valued error propagation with random effect models

The *distance element* between P_1 and P_2 has the size:

$$F = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Once we apply the "*error propagation law*" we have to use.

$$\sigma_{\text{F}}^2 = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' + \frac{1}{2}\mathbf{H}(\text{vec}\boldsymbol{\Sigma})(\text{vec}\boldsymbol{\Sigma})'\mathbf{H}'.$$

Table 8.6: Table of the Jacobi matrix

"*Jacobi matrix*"

$$\mathbf{J} = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial y_1}, \ldots, \frac{\partial F}{\partial x_4}, \frac{\partial F}{\partial y_4}\right]$$

$$\mathbf{J} = \left[\frac{-(x_2 - x_1)}{F}, \ \frac{-(y_2 - y_1)}{F}, \ \frac{(x_2 - x_1)}{F}, \ \frac{(y_2 - y_1)}{F}, \ 0, \ 0, \ 0, \ 0\right]$$

$$\mathbf{J} = [-0.5, \ -0.866, \ 0.5, \ 0.866, \ 0, \ 0, \ 0, \ 0].$$

Table 8.7: Table Hesse matrix

"*Hesse matrix*"

$$\mathbf{H} = \frac{\partial}{\partial \mathbf{x}'} \otimes \frac{\partial}{\partial \mathbf{x}'} F(\mathbf{x})$$

$$= \frac{\partial}{\partial \mathbf{x}'} \otimes \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial y_1}, \ldots, \frac{\partial F}{\partial x_4}, \frac{\partial F}{\partial y_4}\right]$$

$$= \left[\frac{\partial^2 F}{\partial x_1^2}, \frac{\partial^2 F}{\partial x_1 \partial y_1}, \ldots, \frac{\partial^2 F}{\partial x_1 \partial y_4}, \frac{\partial^2 F}{\partial y_1 \partial x_1}, \ldots, \frac{\partial^2 F}{\partial y_4 \partial x_4}, \frac{\partial^2 F}{\partial y^2}\right]$$

$$= \left[\frac{\partial}{\partial x_1}\left(\frac{\partial F}{\partial x_1}, \ \ldots, \ \frac{\partial F}{\partial y_4}\right), \ \ldots, \ \frac{\partial}{\partial y_4}\left(\frac{\partial F}{\partial x_1}, \ \ldots, \ \frac{\partial F}{\partial y_4}\right)\right].$$

Note the detailed computation in *Table 8.8*.

Table 8.8: Second derivatives

: "interims formulae: Hesse matrix":

$$\frac{\partial}{\partial x_1}\left(\frac{\partial F}{\partial x_1}, \ \ldots, \ \frac{\partial F}{\partial y_4}\right) = \left[\frac{1}{F} - \frac{(x_2 - x_1)^2}{F^3}, \ -\frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \right.$$
$$\left. -\frac{1}{F} + \frac{(x_2^2 - x_1^2)}{F^3}, \ \frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \ 0, \ 0, \ 0, \ 0\right],$$

$$\frac{\partial}{\partial y_1}\left(\frac{\partial F}{\partial x_1}, \ \ldots, \ \frac{\partial F}{\partial y_4}\right) = \left[-\frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \ \frac{1}{F} - \frac{(y_2 - y_1)^2}{F^3}, \right.$$
$$\left. \frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \ -\frac{1}{F} + \frac{(y_2^2 - y_1^2)}{F^3}, \ 0, \ 0, \ 0, \ 0\right]$$

$$\frac{\partial}{\partial x_2}\left(\frac{\partial F}{\partial x_1}, \ \ldots, \ \frac{\partial F}{\partial y_4}\right) = \left[-\frac{1}{F} + \frac{(x_2 - x_1)^2}{F^3}, \ \frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \right.$$
$$\left. \frac{1}{F} - \frac{(x_2^2 - x_1^2)}{F^3}, \ -\frac{(x_2 - x_1)(y_2 - y_1)}{F^3}, \ 0, \ 0, \ 0, \ 0\right],$$

$$\frac{\partial}{\partial y_2}\left(\frac{\partial F}{\partial x_1},\ldots,\frac{\partial F}{\partial y_4}\right)=\left[\frac{(x_2-x_1)(y_2-y_1)}{F^3},\ -\frac{1}{F}+\frac{(y_2-y_1)^2}{F^3},\right.$$
$$\left.-\frac{(x_2-x_1)(y_2-y_1)}{F^3},\ \frac{1}{F}-\frac{(y_2^2-y_1^2)}{F^3},\ 0,\ 0,\ 0,\ 0\right],$$

$$\left.\begin{array}{l}\frac{\partial}{\partial x_i}\left(\frac{\partial F}{\partial x_1},\ldots,\frac{\partial F}{\partial y_4}\right)=[0,\ 0,\ 0,\ ,0,\ 0,\ 0,\ 0,\ 0]\\ \frac{\partial}{\partial y_i}\left(\frac{\partial F}{\partial x_1},\ldots,\frac{\partial F}{\partial y_4}\right)=[0,\ 0,\ 0,\ ,0,\ 0,\ 0,\ 0,\ 0]\end{array}\right],\ \forall\ i=3,4.$$

Results

At first, we list the distance $\{P_1P_2\}$ of the distance element by $|P_1P_2|$ =20 (for instance 20m).

Second, we compute $\sigma_F^2(\textit{first term})=\mathbf{J}\boldsymbol{\Sigma}\mathbf{J}'$ by

$$\begin{aligned}&\sigma_F^2(\textit{first term})\\&=[-0.5,\ -0.866,\ 0.5,\ 0.866,\ 0,\ 0,\ 0,\ 0]\\&\times\begin{bmatrix}1&0&0.438&-0.022&0.442&-0.005&0.512&0.108\\0&1&-0.022&0.412&-0.005&0.362&0.108&0.638\\0.438&-0.022&1&0&0.512&0.108&0.386&-0.037\\-0.022&0.412&0&1&0.108&0.638&-0.037&0.418\\0.442&-0.005&0.512&0.108&1&0&0.438&-0.022\\-0.005&0.362&0.108&0.638&0&1&-0.022&0.412\\0.512&0.108&0.386&-0.037&0.438&-0.022&1&0\\0.108&0.638&-0.037&0.418&-0.022&0.412&0&1\end{bmatrix}\\&\times[-0.5,\ -0.866,\ 0.5,\ 0.866,\ 0,\ 0,\ 0,\ 0]'\\&=1.2000.\end{aligned}$$

Third, we need to compute $\sigma_F^2(\textit{second term})=\frac{1}{2}\mathbf{H}(\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}'$ by

$$\sigma_F^2(\textit{second term})=\frac{1}{2}\mathbf{H}(\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}'=0.0015$$

where

$$\mathbf{H}=\mathrm{vec}\begin{bmatrix}0.0375&-0.0217&-0.0375&0.0217&0\ 0\ 0\ 0\\-0.0217&0.0125&0.0217&-0.0125&0\ 0\ 0\ 0\\-0.0375&0.0217&0.0375&-0.0217&0\ 0\ 0\ 0\\0.0217&-0.0125&-0.0217&0.0125&0\ 0\ 0\ 0\\0&0&0&0&0\ 0\ 0\ 0\\0&0&0&0&0\ 0\ 0\ 0\\0&0&0&0&0\ 0\ 0\ 0\\0&0&0&0&0\ 0\ 0\ 0\end{bmatrix},$$

and

$$
\operatorname{vec}\boldsymbol{\Sigma} = \operatorname{vec}\begin{bmatrix} 1 & 0 & 0.438 & -0.022 & 0.442 & -0.005 & 0.512 & 0.108 \\ 0 & 1 & -0.022 & 0.412 & -0.005 & 0.362 & 0.108 & 0.638 \\ 0.438 & -0.022 & 1 & 0 & 0.512 & 0.108 & 0.386 & -0.037 \\ -0.022 & 0.412 & 0 & 1 & 0.108 & 0.638 & -0.037 & 0.418 \\ 0.442 & -0.005 & 0.512 & 0.108 & 1 & 0 & 0.438 & -0.022 \\ -0.005 & 0.362 & 0.108 & 0.638 & 0 & 1 & -0.022 & 0.412 \\ 0.512 & 0.108 & 0.386 & -0.037 & 0.438 & -0.022 & 1 & 0 \\ 0.108 & 0.638 & -0.037 & 0.418 & -0.022 & 0.412 & 0 & 1 \end{bmatrix}.
$$

Finally, we get the variance of the distance element F

$$
\sigma_F^2 = 1.2000+0.0015 = 1.2015
$$

i.e.

$$
\sigma_F = \pm 1.0961(\mathrm{m}).
$$

Chapter 9
The Fifth Problem of Algebraic Regression: The System of Conditional Equations: Homogeneous and Inhomogeneous Equations: $\{\mathbf{By} = \mathbf{Bi} \text{ versus } -\mathbf{c} + \mathbf{By} = \mathbf{Bi}\}$

Conditional equations in its linear form are a standard topic in Geodetic Sciences. We mention for *example F.R. Helmert* (1907) and *H. Wolf* (1986). Here we start with *weighted Least Squares* ($\mathbf{G_y}$ - *LESS*) of inconsistent *homogeneous and inhomogeneous linear conditional equations*. Simple examples are (i) *the triplet of angular observations* and (ii) the *sum of planar triangles*. The stochastic approach is left to *Chapter 13*.

Here we shall outline two systems of poor conditional equations, namely *homogeneous* and *inhomogeneous inconsistent equations*. *First, Definition 9.1* gives us $\mathbf{G_y}$-LESS of

inconsistent homogeneous conditional equations

which we characterize as the *least squares solution* with respect to the $\mathbf{G_y}$-*seminorm* ($\mathbf{G_y}$ -*norm*) by means of *Lemma 9.2*, *Lemma 9.3* ($\mathbf{G_y}$-*norm*) and *Lemma 9.4* ($\mathbf{G_y}$-seminorm). *Second*, *Definition 9.5* specifies $\mathbf{G_y}$)-LESS of

inconsistent inhomogeneous conditional equations

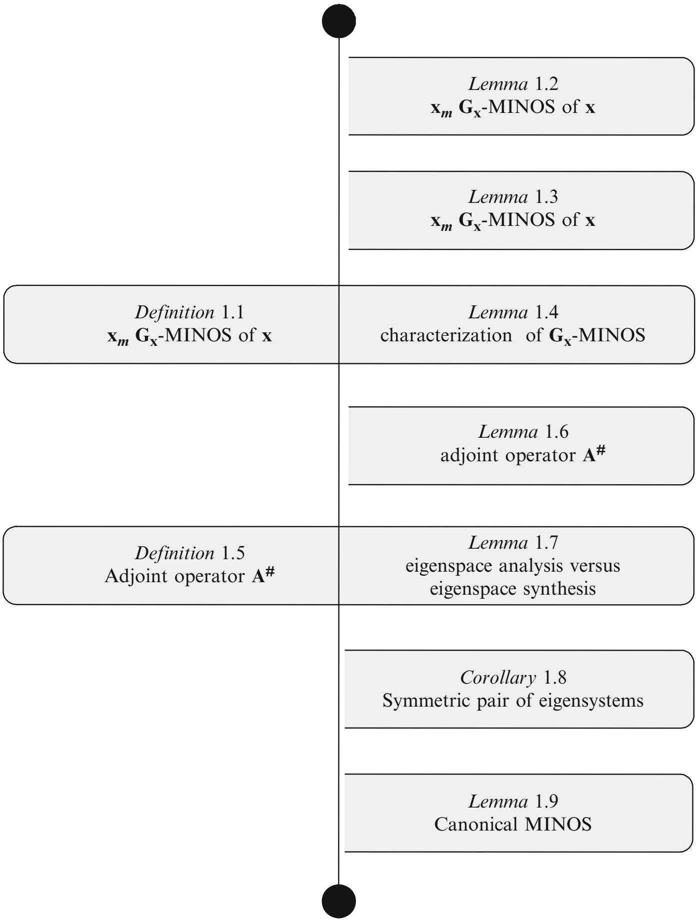
which alternatively characterize as the corresponding least squares solution with respect to the $\mathbf{G_y}$-seminorm by means of *Lemma 9.6*. *Third*, we come up with examples.

9-1 $\mathbf{G_y}$-LESS of a System of a Inconsistent Homogeneous Conditional Equations

Our point of departure is *Definition 9.1* by which we *define* $\mathbf{G_y}$-LESS of a *system of inconsistent homogeneous conditional equations*.

Definition 9.1. ($\mathbf{G_y}$-LESS of a system of inconsistent homogeneous conditional equations):

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2_9,
© Springer-Verlag Berlin Heidelberg 2012



The guideline of chapter one: definitions, lemmas and corollary

An n× 1 vector $\mathbf{i}_\ell$ of inconsistency is called $\mathbf{G_y}$-LESS (LEast Squares Solution with respect to the $\mathbf{G_y}$-seminorm) of the inconsistent system of linear conditional equations

$$\mathbf{Bi} = \mathbf{By}, \tag{9.1}$$

if in comparison to all other vectors $\mathbf{i} \in \mathbb{R}^n$ the inequality

$$||\mathbf{i}_\ell||^2_{\mathbf{G_y}} := \mathbf{i}'_\ell \mathbf{G_y} \mathbf{i}_\ell \leq \mathbf{i}' \mathbf{G_y} \mathbf{i} =: ||\mathbf{i}||^2_{\mathbf{G_y}} \tag{9.2}$$

holds in particular if the vector of inconsistency $\mathbf{i}_\ell$ has the least $\mathbf{G_y}$-seminorm. *Lemma 9.2* characterizes the *normal equations* for the least squares solution of the system of inconsistent homogeneous conditional equations with respect to the $\mathbf{G_y}$-seminorm.

Lemma 9.2. (least squares solutions of the system of *inconsistent homogeneous conditional equations* with respect to the $\mathbf{G_y}$-seminorm):

An $n \times 1$ vector $\mathbf{i}_\ell$ of the system of *inconsistent homogeneous conditional equations*

$$\mathbf{Bi} = \mathbf{By} \tag{9.3}$$

is $\mathbf{G_y}$-LESS if and only if the system of normal equations

$$\begin{bmatrix} \mathbf{G_y} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{i}_\ell \\ \lambda_\ell \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{B_y} \end{bmatrix} \tag{9.4}$$

with the $q\times 1$ vector λ_ℓ of "*Lagrange multipliers*" is fulfilled.

Proof. $\mathbf{G_y}$-LESS of $\mathbf{Bi} = \mathbf{By}$ is constructed by means of the *Lagrangean*

$$\mathcal{L}(\mathbf{i}, \lambda) := \mathbf{i}'\mathbf{G_y}\mathbf{i} + 2\lambda'(\mathbf{Bi} - \mathbf{By}) = \min_{\mathbf{i}, \lambda}.$$

The *first derivatives*

$$\frac{\partial \mathcal{L}}{\partial \mathbf{i}}(\mathbf{i}_\ell, \lambda_\ell) = 2(\mathbf{G_y}\mathbf{i}_\ell + \mathbf{B}'\lambda_\ell) = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda}(\mathbf{i}_\ell, \lambda_\ell) = 2(\mathbf{B}\mathbf{i}_\ell - \mathbf{By}) = 0$$

constitute the *necessary conditions*. (The theory of vector-valued derivatives is presented in Appendix B). The *second derivatives*

$$\frac{\partial \mathcal{L}}{\partial \mathbf{i} \partial \mathbf{i}'}(\mathbf{i}_\ell, \lambda_\ell) = 2\mathbf{G_y} \geq 0$$

build up due to the positive semidefiniteness of the matrix $\mathbf{G_y}$ the *sufficiency condition* for the minimum. The normal equations (9.4) are derived from the two equations of first derivatives, namely

$$\begin{bmatrix} \mathbf{G_y} & \mathbf{B}' \\ \mathbf{B} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{i}_\ell \\ \lambda_\ell \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{By} \end{bmatrix}.$$

Lemma 9.3 is a short review of the system of *inconsistent homogeneous conditional equations with respect to the* $\mathbf{G_y}$*-norm, Lemma 9.4* alternatively *with respect to the* $\mathbf{G_y}$-seminorm.

Lemma 9.3. (least squares solution of the system of *inconsistent homogeneous conditional equations* with respect to the $\mathbf{G_y}$-norm):

An n × 1 vector $\mathbf{i}_\ell$ of the system of *inconsistent homogeneous conditional equations* $\mathbf{Bi} = \mathbf{By}$ is the *least squares solution* with respect to the $\mathbf{G_y}$-*norm* if and only if it solves the normal equations

$$\mathbf{G_y i}_\ell = \mathbf{B}'(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By}. \tag{9.5}$$

The solution

$$\mathbf{i}_\ell = \mathbf{G_y^{-1}B}'(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By} \tag{9.6}$$

is unique. The "*goodness of fit*" of $\mathbf{G_y}$-LESS is

$$||\mathbf{i}_\ell||^2_{\mathbf{G_y}} = \mathbf{i}'_\ell \mathbf{G_y i}_\ell = \mathbf{y}'\mathbf{B}'(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By}. \tag{9.7}$$

Proof. A basis of the proof could be *C.R. Rao's Pandora Box*, the theory of *inverse partitioned matrices* (Appendix A: Fact: Inverse Partitioned Matrix /IPM/ of a symmetric matrix). Due to the rank identity $rk\mathbf{G_y} = n$, the *normal equations* (9.4) can be faster solved by *Gauss elimination*.

$$\mathbf{G_y i}_\ell + \mathbf{B}'\lambda_\ell = 0$$

$$\mathbf{Bi}_\ell = \mathbf{By}.$$

Multiply the *first* normal equation by $\mathbf{BG_y^{-1}}$ and substitute the *second* normal equation for $\mathbf{Bi}_\ell$.

$$\left.\begin{array}{r} \mathbf{BG_y^{-1}G_y i}_\ell = \mathbf{Bi}_\ell = -\mathbf{BG_y^{-1}B}'\lambda_\ell \\ \mathbf{Bi}_\ell = \mathbf{By} \end{array}\right] \Rightarrow$$

$$-\mathbf{BG_y^{-1}B}'\lambda_\ell = \mathbf{By}$$

$$\lambda_\ell = -(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By}.$$

Finally we substitute the "*Lagrange multiplier*" λ_ℓ back to the first normal equation in order to prove

$$\mathbf{G_y i}_\ell + \mathbf{B}'\lambda_\ell = \mathbf{G_y i}_\ell - \mathbf{B}'(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By} = 0$$

$$\mathbf{i}_\ell = \mathbf{G_y^{-1}B}'(\mathbf{BG_y^{-1}B}')^{-1}\mathbf{By}.$$

We switch immediately to *Lemma 9.4*.

Lemma 9.4. (least squares solution of the system of *inconsistent homogeneous conditional equations* with respect to the $\mathbf{G_y}$-seminorm):

An n × 1 vector $\mathbf{i}_\ell$ of the system of *inconsistent homogeneous conditional equations* $\mathbf{Bi} = \mathbf{By}$ is the *least squares solution* with respect to the $\mathbf{G_y}$-*seminorm* if the *compatibility condition*

$$\mathcal{R}(\mathbf{B}') \subset \mathcal{R}(\mathbf{G_y}) \tag{9.8}$$

is fulfilled, and solves the system of normal equations

$$\mathbf{G_y i}_\ell = \mathbf{B}'(\mathbf{B G_y^{-1} B}')^{-1}\mathbf{By}, \tag{9.9}$$

which is independent of the choice of the g-inverse $\mathbf{G_y^-}$.

9-2 Solving a System of Inconsistent Inhomogeneous Conditional Equations

The text point of departure is *Definition 9.5*, a definition of $\mathbf{G_y}$-LESS of a *system of inconsistent inhomogeneous condition equations*.

Definition 9.5. ($\mathbf{G_y}$-LESS of a system of inconsistent inhomogeneous conditional equations):

An $n\times 1$ vector $\mathbf{i}_\ell$ of inconsistency is called $\mathbf{G_y}$-LESS (LEast Squares Solution with respect to the $\mathbf{G_y}$-seminorm) of the *inconsistent system of inhomogeneous conditional equations*

$$-\mathbf{c} + \mathbf{By} = \mathbf{B\,i} \tag{9.10}$$

(*the minus sign is conventional*),

if in comparison to all other vectors $\mathbf{i} \in \mathbb{R}^n$ the inequality

$$||\mathbf{i}_\ell||^2 := \mathbf{i}'_\ell \mathbf{G_y i}_\ell \le \mathbf{i}'\mathbf{G_y i} =: ||\mathbf{i}||^2_{\mathbf{G_y}} \tag{9.11}$$

holds, in particular if the vector of inconsistency $\mathbf{i}_\ell$ has the least $\mathbf{G_y}$-seminorm.

Lemma 9.6 characterizes the *normal equations* for the least squares solution of the system of inconsistent *inhomogeneous conditional equations* with respect to the $\mathbf{G_y}$-seminorm.

Lemma 9.6. (least squares solution of the system of inconsistent inhomogeneous conditional equations with respect to the $\mathbf{G_y}$-seminorm):

An $n \times 1$ vector $\mathbf{i}_\ell$ of the system of inconsistent homogeneous conditional equations

$$\mathbf{Bi} = \mathbf{By} - \mathbf{c} = \mathbf{B}(\mathbf{y} - \mathbf{d}) \tag{9.12}$$

is $\mathbf{G_y}$-LESS if and only if the system of normal equations

$$\begin{bmatrix} \mathbf{G_y} & \mathbf{B}' \\ \mathbf{B} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{i}_\ell \\ \lambda_\ell \end{bmatrix} = \begin{bmatrix} \mathbf{O} \\ \mathbf{By} - \mathbf{c} \end{bmatrix}, \tag{9.13}$$

with the $q\times 1$ vector λ of *Lagrange multipliers* is fulfilled. $\mathbf{i}_\ell$ exists surely if

$$\mathcal{R}(\mathbf{B}') \subset \mathcal{R}(\mathbf{G_y}) \tag{9.14}$$

and it solves the normal equations

$$\mathbf{G_y i}_\ell = \mathbf{B}'(\mathbf{B G_y^{-1} B}')^{-1}(\mathbf{By} - \mathbf{c}), \tag{9.15}$$

which is independent of the choice of the g-inverse $\mathbf{G_y^-}$. $\mathbf{i}_\ell$ is unique if the matrix $\mathbf{G_y}$ is *regular* and in consequence *positive definite*.

9-3 Examples

Our two examples relate to the *triangular condition*, the so-called *zero misclosure*, within a triangular network, and the condition that the *sum* within a flat triangle accounts to 180^o.

(i) The *first example*: triplet of angular observations
We assume that *three observations* of height differences within the triangle $P_\alpha P_\beta P_\gamma$ sum up to zero. The condition of *holonomic heights* says

$$h_{\alpha\beta} + h_{\beta\gamma} + h_{\gamma\alpha} = 0,$$

namely

$$\mathbf{B} := [1,\ 1,\ 1],\ \mathbf{y} := \begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix},\ \ \mathbf{i} := \begin{bmatrix} i_{\alpha\beta} \\ i_{\beta\gamma} \\ i_{\gamma\alpha} \end{bmatrix}.$$

The *normal equations* of the inconsistent condition read for the case $\mathbf{G_y} = \mathbf{I}_3$:

$$\mathbf{i}_\ell = \mathbf{B}'(\mathbf{BB}')^{-1}\mathbf{By},$$

$$\mathbf{B}'(\mathbf{BB}')\mathbf{B} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$(i_{\alpha\beta})_\ell = (i_{\beta\gamma})_\ell = (i_{\gamma\alpha})_\ell = \frac{1}{3}(\underline{h}_{\alpha\beta} + \underline{h}_{\beta\gamma} + \underline{h}_{\gamma\alpha})$$

(ii) The *second example*: sum of planar triangles
Alternatively, we assume: three angles which form a planar triangle of *sum* to

$$\alpha + \beta + \gamma = 180^\circ$$

namely

The *normal equations* of the inconsistent condition equation read in our case $\mathbf{G_y} = \mathbf{I}_3$:

$$\mathbf{i}_\ell = \mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1}(\mathbf{B}\mathbf{y} - \mathbf{c}),$$

$$\mathbf{B}'(\mathbf{B}\mathbf{B}')^{-1} = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{B}\mathbf{y} - \mathbf{c} = \underline{\alpha} + \underline{\beta} + \underline{\gamma} - 180°,$$

$$(\mathbf{i}_\alpha)_\ell = (\mathbf{i}_\beta)_\ell = (\mathbf{i}_\gamma)_\ell = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}(\underline{\alpha} + \underline{\beta} + \underline{\gamma} - 180°).$$

Chapter 10
The Fifth Problem of Probabilistic Regression

Setup of BLUUE for the moments of first order (Kolmogorov–Wiener prediction)

We define the fifth *problem of probabilistic regression* by the *inhomogeneous general linear Gauss-Markov model* including fixed effects as well as random effect, namely by **A**ξ + **C**E{**z**} + **y** = E{**y**} together with variance-covariance matrices Σ_z and Σ_y being unknown as well as ξ, E{**z**}, and E{**y**}, **y**. It is the standard model of *Kolmogorov-Wiener Prediction in its general form*. By *Theorem 10.3* we estimate ξ and E{**z**} by ($\hat{\xi}$, $E\{\hat{\mathbf{z}}\}\Sigma$y - *Best Linear Uniformly Unbiased Estimation* ($\hat{\xi}$, $E\{\hat{\mathbf{z}}\}\Sigma\mathbf{y}$ – BLUUE)) following *C.R. Rao and J. Kleffe* (1988 pp. 161 – 180) in section one. Section two is an attempt to construct an explicit representation of the *error budget* in the *general Gauss-Markov model with mixed effects*. Restrictive is the assumption D{**y**} = Vσ^2 (one parameter) and D{**z**} = Zσ^2 (one parameter) and the correlations (V,Z) to be known! For such a model, we will be able by *Theorem 10.5* to construct a *homogeneous quadratic estimate* of σ^2. We leave the question open whether such a model is realistic.

Instead we will give an *example* from *linear collocation* with ξ, E{**z**}, E{**y**}, Σ_z, Σ_y to be unknown, *but* y&$$$; to be known. We depart from analyzing a *height network observed at two epochs*. At the initial epoch three height differences have been observed. From the first epoch to the second epoch *we assume height differences which change linear in time,* for instance caused by an *Earthquake*. We apply a height varying model

$$h_{\alpha\beta}(\tau) = h_{\alpha\beta}(O) + h_{\alpha\beta}(O)\tau + O(\tau^2)$$

where τ denotes the time difference *from* the first epoch *to* the second epoch.

Section four will be a list of comments about *linear and nonlinear prediction of type Kolmogorov-Wiener,* an extensive list of *geodetic examples* from hybrid concepts of *vertical deflections, gravity gradients* and *gravity values* and *specific trend model*. An *extensive list of references is given*. Finally, we comment on

(*i*) best inhomogeneous linear prediction,
(*ii*) best homogeneous linear prediction
(*iiii*) best homogeneous linear unbiased prediction by the dispersion ordering D3 ≤ D2 ≤ D1

When we apply instead of Kolmogorov-Wiener linear prediction for absolute quantities, we arrive at the more general concept of *Krige prediction* when we use *differences,* for instance

$$\|\mathbf{y}_{P_1} - \mathbf{y}_{P_2}\|^2 = E\left\{\mathbf{y}_{P_1} - \mathbf{y}_{P_2} - (E\{\mathbf{y}_{P_1}\} - E\{\mathbf{y}_{P_2}\})\right\}^2$$

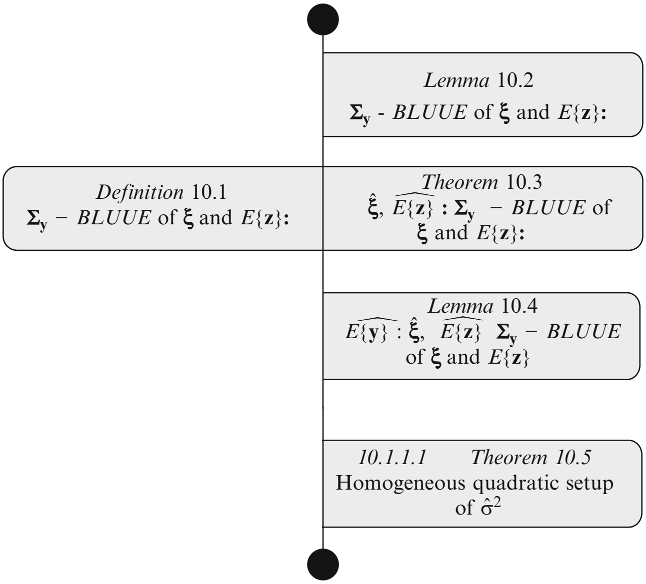
E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_10,
© Springer-Verlag Berlin Heidelberg 2012

For a *weakly relative translational invariant stochastic process* establishing the *Kolmogorov's structure function*. Alternatively, we are able to analyze *higher order variance-covariance function,* for instance designed by *E. Grafarend (1984)*.

Fast track reading: Read only *Theorems* 10.3 and 10.5.

"*Prediction company's chance of success is not zero, but close to it.*"–Eugene Fama

"*The best way to predict the future is to invent it.*"–Alan Kay



The inhomogeneous general linear Gauss–Markov model with fixed effects and random effects will be presented first. We review the special *Kolmogorov–Wiener* model and extend it by the proper stochastic model of type BIQUUE given by *Theorem* 10.5.

The extensive example for the general *linear Gauss–Markov model with fixed effects* and *random effects* concentrates on a height network observed at two epochs. At the first epoch we assume *three measured height differences*. Inbetween the first and the second epoch we assume height differences which change *linear in time*, for instance as a result of an earthquake we have found the height difference model

$$h_{\alpha\beta}(\tau) = h_{\alpha\beta}(0) + h^{\bullet}_{\alpha\beta}\tau + \mathcal{O}(\tau^2).$$

Namely, indicates the time interval from the first epoch to the second epoch relative to the height difference $h^{\bullet}_{\alpha\beta}$. Unknown are

- The fixed effects $h_{\alpha\beta}$
- The expected values of stochastic effects of type height difference velocities $h_{\alpha\beta}$

given the *singular dispersion matrix of height differences*. Alternative *estimation and prediction* producers of

- type $(\mathbf{V} + \mathbf{CZC}')$-BLUUE for the unknown *fixed parameter vector* ξ *of height differences of initial epoch and*
- the *expectation data* $E\{\mathbf{z}\}$ of stochastic height difference velocities z, and
- of type $(\mathbf{V} + \mathbf{CZC}')$-BLUUE for the expectation data $E\{\mathbf{y}\}$ of height difference measurements y,
- of type $\tilde{\mathbf{e}}_{\mathbf{y}}$ of the empirical error vector,
- as well as of type $(\mathbf{V} + \mathbf{CZC}')$-BLUUP of the *stochastic vector* $\tilde{\mathbf{z}}$ of *height difference velocities*.

For the unknown variance component σ^2 of height difference observations we review estimates of type BIQUUE. At the end, we intend to generalize the concept of *estimation and prediction* of fixed and random effects by a short historical remark.

10-1 Inhomogeneous General Linear Gauss–Markov Model (Fixed Effects and Random Effects)

Here we focus on the general inhomogeneous *linear Gauss–Markov model* including *fixed effects and random effects*. By means of *Definition* 10.1 we review $\mathbf{\Sigma}_{\mathbf{y}}$-BLUUE of ξ and $E\{\mathbf{z}\}$ followed by the related *Lemma* 10.2, *Theorem* 10.3 and *Lemma* 10.4.

Box 10.1. Inhomogeneous general linear Gauss–Markov model (fixed effects and random effects)

$$\mathbf{A}\xi + \mathbf{C}E\{\mathbf{z}\} + \gamma = E\{\bar{\mathbf{y}}\} \tag{10.1}$$

$$\mathbf{C}\{\bar{\mathbf{y}}, \mathbf{z}\} = 0$$

$$\mathbf{\Sigma}_{\mathbf{z}} := D\{\mathbf{z}\},\ \mathbf{\Sigma}_{\bar{\mathbf{y}}} := D\{\bar{\mathbf{y}}\}$$

$$C\{\bar{\mathbf{y}}, \mathbf{z}\} = 0$$

$$\xi, E\{\mathbf{z}\}, E\{\bar{\mathbf{y}}\}, \mathbf{\Sigma}_{\mathbf{z}}, \mathbf{\Sigma}_{\bar{\mathbf{y}}}\ \textit{unknown}$$

$$\gamma\ \textit{unknown}$$

$$E\{\bar{\mathbf{y}}\} - \gamma \in ([\mathbf{A}, \mathbf{C}]) \tag{10.2}$$

The $n \times 1$ stochastic vector $\bar{\mathbf{y}}$ of observations is transformed by means of $\bar{\mathbf{y}} - \gamma =: \mathbf{y}$ to the new $n \times 1$ stochastic vector $\mathbf{y}$ of reduced observations which is characterized by second order statistics, in particular by the *first moments* $E\{\mathbf{y}\}$ and by the *central second moments* $D\{\mathbf{y}\}$.

Definition 10.1. ($\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of ξ and $E\{\mathbf{z}\}$):

The partitioned vector $\zeta = \mathbf{L}\mathbf{y} + \kappa$, namely

$$\begin{bmatrix} \hat{\xi} \\ \hat{\eta} \end{bmatrix} = \begin{bmatrix} \hat{\xi} \\ \widehat{E\{\mathbf{z}\}} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix} \mathbf{y} + \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}$$

is called $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of ζ (**B**est **L**inear **U**niformly **U**nbiased **E**stimation with respect to $\boldsymbol{\Sigma}_{\mathbf{y}}$- norm) in (10.1) if (1st) $\hat{\zeta}$ is uniformly unbiased in the sense of

$$E\{\hat{\zeta}\} = E\{\mathbf{L}\mathbf{y} + \kappa\} = \zeta \textit{ for all } \zeta \in \mathbb{R}^{m+l} \tag{10.3}$$

or

$$\begin{aligned} &E\{\hat{\xi}\} = E\{\mathbf{L}_1\mathbf{y} + \kappa_1\} = \xi \textit{ for all } \xi \in \mathbb{R}^m \\ &E\{\hat{\eta}\} = E\{\widehat{E\{\mathbf{z}\}}\} = E\{\mathbf{L}_2\mathbf{y} + \kappa_2\} = \eta = E\{\mathbf{z}\} \textit{ for all } \eta \in \mathbb{R}^l \end{aligned} \tag{10.4}$$

and (2nd) in comparison to all other linear uniformly unbiased estimation $\hat{\zeta}$ has *minimum variance*.

$$\mathrm{tr}D\{\hat{\zeta}\} := E\{(\hat{\zeta} - \zeta)'(\hat{\zeta} - \zeta)\} = \mathrm{tr}\mathbf{L}\Sigma_{\mathbf{y}}\mathbf{L}' = ||\mathbf{L}'||^2_{\boldsymbol{\Sigma}_{\mathbf{y}}} = \min_{\mathbf{L}} \tag{10.5}$$

or

$$\begin{aligned} &\mathrm{tr}D\,\{\hat{\eta}\} := \mathrm{tr}D\{\widehat{E\,\{\mathbf{z}\}}\} := E\,\left\{(\hat{\eta} - \eta)'(\hat{\eta} - \eta)\right\} \\ &= E\{(\widehat{E\,\{\mathbf{z}\}} - E\{\mathbf{z}\})'(\widehat{E\,\{\mathbf{z}\}} - E\{\mathbf{z}\}\,\{\mathbf{z}\})\} = \mathrm{tr}\mathbf{L}_2\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{L}'_2 = ||\mathbf{L}'||^2_{\boldsymbol{\Sigma}_{\mathbf{y}}} = \min_{\mathbf{L}_2}. \end{aligned} \tag{10.6}$$

We shall specify $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of ξ and $E\{\mathbf{z}\}$ by means of $\kappa_1 = 0, \kappa_2 = 0$ and writing the residual normal equations by means of "*Lagrange multipliers*".

Lemma 10.2. (*$\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE* of ξ and $E\{\mathbf{z}\}$):

An $(m + l) \times 1$ vector $[\hat{\xi}',\ \hat{\eta}']' = [\hat{\xi}',\ \widehat{E\{\mathbf{z}\}}']' = [\mathbf{L}'_1, \mathbf{L}'_2]'\,\mathbf{y} + [\kappa'_1, \kappa'_2]'$ is $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of $[\xi',\ \widehat{E\{\mathbf{z}\}}']'$ in (10.1), if and only if

$$\kappa_1 = 0,\ \kappa_2 = 0$$

hold and the matrices $\mathbf{L}_1$ and $\mathbf{L}_2$ fulfill the system of normal equations

$$\begin{bmatrix} \mathbf{\Sigma_y} & \mathbf{A} & \mathbf{C} \\ \mathbf{A}' & 0 & 0 \\ \mathbf{C}' & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{L}'_1 & \mathbf{L}'_2 \\ \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \mathbf{I}_m & 0 \\ 0 & \mathbf{I}_l \end{bmatrix} \tag{10.7}$$

or

$$\begin{aligned} \mathbf{\Sigma_y L}'_1 + \mathbf{A}\Lambda_{11} + \mathbf{C}\Lambda_{21} = \mathbf{0},\ \ \mathbf{\Sigma_y L}'_2 + \mathbf{A}\Lambda_{12} + \mathbf{C}\Lambda_{22} = \mathbf{0} \\ \mathbf{A}'\mathbf{L}'_1 = \mathbf{I}_m,\ \ \mathbf{A}'\mathbf{L}'_2 = \mathbf{0} \\ \mathbf{C}'\mathbf{L}'_1 = \mathbf{0},\ \ \mathbf{C}'\mathbf{L}'_2 = \mathbf{I}_l \end{aligned} \tag{10.8}$$

with suitable matrices Λ_{11}, Λ_{12}, Λ_{21} and Λ_{22} of "*Lagrange multipliers*". *Theorem 10.3* specifies the solution of the special normal equations by means of (10.9) relative to the specific "*Schur complements*" (10.10)–(10.13).

Theorem 10.3. ($\hat{\xi}$, $\widehat{E\{\mathbf{z}\}}\mathbf{\Sigma_y}$-BLUUE of ξ and $E\{\mathbf{z}\}$):

Let $[\hat{\xi}', \widehat{E\{\mathbf{z}\}'}]'$ be $\mathbf{\Sigma_y}$-BLUUE of the $[\xi', E\{\mathbf{z}\}']'$ in the mixed Gauss–Markov model (10.1). Then the equivalent representations of the solution of the normal equations (10.7)

$$\hat{\zeta} := \begin{bmatrix} \hat{\xi} \\ \hat{\eta} \end{bmatrix} := \begin{bmatrix} \hat{\xi} \\ \widehat{E\{\mathbf{z}\}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A} & \mathbf{A}'\Sigma_{y}^{-1}\mathbf{C} \\ \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A} & \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{A}' \\ \mathbf{C}' \end{bmatrix} \mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y} \tag{10.9}$$

$$\hat{\xi} = \{\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}\}^{-1} \times \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I_n} - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\,\mathbf{y}$$

$$\begin{aligned} \hat{\eta} = \widehat{E\{\mathbf{z}\}} &= \left\{\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}\right\}^{-1} \\ &\times \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma\mathbf{y}^{-1}]\,\mathbf{y} \end{aligned}$$

$$\hat{\xi} = S_{\mathbf{A}}^{-1}\mathbf{s_A}$$

$$\hat{\eta} := \widehat{E\{\mathbf{z}\}} = \mathbf{S}_{\mathbf{C}}^{-1}\mathbf{s_C}$$

$$\hat{\xi} = (\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \widehat{E\{\mathbf{z}\}})$$

$$\hat{\eta} := \widehat{E\{\mathbf{z}\}} = (\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \mathbf{A}\hat{\xi})$$

are completed by the dispersion matrices and the covariance matrices.

$$D\left\{\hat{\zeta}\right\} := D\left\{\begin{bmatrix} \hat{\xi} \\ \hat{\eta} \end{bmatrix}\right\} := D\left\{\begin{bmatrix} \hat{\xi} \\ \widehat{E\{\mathbf{z}\}} \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A} & \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} \\ \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A} & \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C} \end{bmatrix}^{-1} =: \mathbf{\Sigma}_{\hat{\zeta}}$$

$$
\begin{aligned}
D\left\{\hat{\xi}\right\} &= \left\{\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}\right\}^{-1} =: \boldsymbol{\Sigma}_{\hat{\xi}} \\
\mathbf{C}\{\hat{\xi}, \hat{\eta}\} &= \mathbf{C}\{\hat{\xi}, \widehat{E\{\mathbf{z}\}}\} \\
&= -\left\{\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}\right\}^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1} \\
&= -(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C}\left\{\mathbf{C}\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}\right\}^{-1} \\
D\{\hat{\eta}\} &:= D\{\widehat{E\{\mathbf{z}\}}\} = \{\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}\}^{-1} =: \boldsymbol{\Sigma}_{\hat{\eta}} \\
\mathbf{C}\{\hat{\xi}, \mathbf{z}\} &= \mathbf{0} \\
\mathbf{C}\{\hat{\eta}, \mathbf{z}\} &:= \mathbf{C}\{\widehat{E\{\mathbf{z}\}}, \mathbf{z}\} = \mathbf{0},
\end{aligned}
$$

where the "*Schur complements*" are defined by

$$\mathbf{S}_{\mathbf{A}} := \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}, \tag{10.10}$$

$$\mathbf{s}_{\mathbf{A}} := \mathbf{A}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{y}, \tag{10.11}$$

$$\mathbf{S}_{\mathbf{C}} := \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}, \tag{10.12}$$

$$\mathbf{s}_{\mathbf{C}} := \mathbf{C}'\Sigma_{\mathbf{y}}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{y} \tag{10.13}$$

Our final result (10.14)–(10.23) summarizes (i) the two forms (10.14) and (10.15) of estimating $\widehat{E\{\mathbf{y}\}}$ and $D\{\widehat{E\{\mathbf{y}\}}\}$ as derived covariance matrices, (ii) the empirical error vector $\tilde{\mathbf{e}}_y$ and the related variance-covariance matrices (10.19)–(10.21) and (iii) the dispersion matrices $D\mathbf{y}$ by means of (10.22)–(10.23).

Lemma 10.4. ($\widehat{E\{\mathbf{y}\}}$: $\hat{\xi}$, $\widehat{E\{\mathbf{z}\}}$ $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of ξ and $E\{\mathbf{z}\}$):

(i) With respect to the mixed Gauss–Markov model (10.1) $\boldsymbol{\Sigma}_{\mathbf{y}}$-BLUUE of the $E\{\mathbf{y}\} = \mathbf{A}\xi + \mathbf{C}E\{\mathbf{z}\}$ is given by

$$
\begin{aligned}
\widehat{E\{\mathbf{y}\}} &= \mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}} \\
&= \mathbf{A}\mathbf{S}_{\mathbf{A}}^{-1}\mathbf{s}_{\mathbf{A}} + \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{S}_{\mathbf{A}}^{-1}\mathbf{s}_{\mathbf{A}})
\end{aligned} \tag{10.14}
$$

or

$$
\begin{aligned}
\widehat{E\{\mathbf{y}\}} &= \mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}} \\
&= \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{S}_{\mathbf{C}}^{-1}\mathbf{s}_{\mathbf{C}}) + \mathbf{C}\mathbf{S}_{\mathbf{C}}^{-1}\mathbf{s}_{\mathbf{C}}
\end{aligned} \tag{10.15}
$$

with the corresponding *dispersion matrices*

$$
\begin{aligned}
D\{\widehat{E\{\mathbf{y}\}}\} &= D\{\mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}}\} \\
&= \mathbf{A}D\{\hat{\xi}\}\mathbf{A}' + \mathbf{A}\mathrm{cov}\{\hat{\xi}, \widehat{E\{\mathbf{z}\}}\}\mathbf{C}' + \mathbf{C}\mathrm{cov}\{\hat{\xi}, \widehat{E\{\mathbf{z}\}}\}\mathbf{A}' + \mathbf{C}D\{\widehat{E\{\mathbf{z}\}}\}\mathbf{C}'
\end{aligned}
$$

$$
\begin{aligned}
D\{\widehat{E\{\mathbf{y}\}}\} &= D\{\mathbf{A}\hat{\hat{\xi}} + \mathbf{C}\widehat{E\{\mathbf{z}\}}\} \\
&= \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}' + [\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}\mathbf{S}_{\mathbf{A}}^{-1} \\
&\quad \mathbf{A}'[\mathbf{I}_n - \Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}']
\end{aligned}
$$

$$
\begin{aligned}
D\{\widehat{E\{\mathbf{y}\}}\} &= D\{\mathbf{A}\hat{\hat{\xi}} + \mathbf{C}\widehat{E\{\mathbf{z}\}}\} \\
&= \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}' + [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1} \\
&\quad \mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}\mathbf{S}_{\mathbf{C}}^{-1}\mathbf{C}'[\mathbf{I}_n - \Sigma_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'], \\
&\quad \mathbf{e}_{\mathbf{y}}
\end{aligned}
$$

where $\mathbf{S}_{\mathbf{A}}, \mathbf{s}_{\mathbf{A}}, \mathbf{S}_{\mathbf{C}}, \mathbf{s}_{\mathbf{C}}$ are "*Schur complements*" (10.10), (10.11), (10.12) and (10.13).

The covariance matrix of $\widehat{E\{\mathbf{y}\}}$ and $\mathbf{z}$ amounts to

$$\operatorname{cov}\{\widehat{E\{\mathbf{y}\}},\ \mathbf{z}\} = \mathbf{C}\{\mathbf{A}\hat{\hat{\xi}} + \mathbf{C}\widehat{E\{\mathbf{z}\}},\ \mathbf{z}\} = \mathbf{0}. \tag{10.16}$$

(ii) If the "error vector" is empirically determined by means of the residual vector $\tilde{\mathbf{e}}_{\mathbf{y}} = \mathbf{y} - \widehat{E\{y\}}$ we gain the various representations of type

$$\tilde{\mathbf{e}}_{\mathbf{y}} = [\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}](\mathbf{y} - \mathbf{A}\mathbf{S}_{\mathbf{A}}^{-1}\mathbf{s}_{\mathbf{A}}) \tag{10.17}$$

or

$$\tilde{\mathbf{e}}_{\mathbf{y}} = [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}](\mathbf{y} - \mathbf{C}\mathbf{S}_{\mathbf{C}}^{-1}\mathbf{s}_{\mathbf{C}}) \tag{10.18}$$

with the corresponding *dispersion matrices*

$$
\begin{aligned}
D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} &= \boldsymbol{\Sigma}_{\mathbf{y}} - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}' \\
&\quad - [\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{A}\mathbf{S}_{\mathbf{A}}^{-1}\mathbf{A}'[\mathbf{I}_n - \Sigma_{\mathbf{y}}^{-1}\mathbf{C}(\mathbf{C}'\Sigma_{\mathbf{y}}^{-1}\mathbf{C})^{-1}\mathbf{C}']
\end{aligned} \tag{10.19}
$$

or

$$
\begin{aligned}
D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} &= \boldsymbol{\Sigma}_{\mathbf{y}} - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}' \\
&\quad - [\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}]\mathbf{C}\mathbf{S}_{\mathbf{C}}^{-1}\mathbf{C}'[\mathbf{I}_n - \Sigma_{\mathbf{y}}^{-1}\mathbf{A}(\mathbf{A}'\Sigma_{\mathbf{y}}^{-1}\mathbf{A})^{-1}\mathbf{A}'],
\end{aligned} \tag{10.20}
$$

where $\mathbf{S}_{\mathbf{A}}, \mathbf{s}_{\mathbf{A}}, \mathbf{S}_{\mathbf{C}}, \mathbf{s}_{\mathbf{C}}$ are "*Schur complements*" (10.10), (10.11), (10.12) and (10.13). $\tilde{\mathbf{e}}_{\mathbf{y}}$ and $\mathbf{z}$ are uncorrelated because of

$$\mathbf{C}\{\tilde{\mathbf{e}}_{\mathbf{y}},\ \mathbf{z}\} = \mathbf{0}. \tag{10.21}$$

(iii) The *dispersion matrices* of the observation vector is given by

$$D\{\mathbf{y}\} = D\{\mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}} + \tilde{\mathbf{e}}_\mathbf{y}\} = D\{\mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}$$
$$D\{\mathbf{y}\} = D\{\tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} + D\{\tilde{\mathbf{e}}_\mathbf{y}\}. \tag{10.22}$$

$\tilde{\mathbf{e}}_\mathbf{y}$ and $\widehat{E\{\mathbf{y}\}}$ are uncorrelated since

$$\mathbf{C}\{\tilde{\mathbf{e}}_\mathbf{y},\ \widehat{E\{\mathbf{y}\}}\} = \mathbf{C}\{\tilde{\mathbf{e}}_\mathbf{y},\ \mathbf{A}\hat{\xi} + \mathbf{C}\widehat{E\{\mathbf{z}\}}\} = \mathbf{C}\{\tilde{\mathbf{e}}_\mathbf{y}, \tilde{\mathbf{e}}_\mathbf{y} - \mathbf{e}_\mathbf{y}\} = \mathbf{0}\,. \tag{10.23}$$

10-2 Explicit Representations of Errors in the General Gauss–Markov Model with Mixed Effects

A collection of explicit representations of errors in the *general Gauss–Markov model with mixed effects* will be presented: ξ, $E\{\mathbf{z}\}$, $\bar{\mathbf{y}} - \gamma = \mathbf{y}$, $\mathbf{\Sigma_z}$, $\mathbf{\Sigma_y}$ will be assumed to be *unknown*, γ *known*. In addition, $\mathbf{C}\{\bar{\mathbf{y}}, \mathbf{z}\}$ will be assumed to *vanish*. The prediction of *random effects* will be summarized here. Note our *simple model*

$$\mathbf{A}\xi + \mathbf{C}E\{\mathbf{z}\} = E\{\mathbf{y}\},\ E\{\mathbf{y}\} \in \mathcal{R}([\mathbf{A}, \mathbf{C}]),\ \text{rk}[\mathbf{A}, \mathbf{C}] = m + \ell < n,$$

$$\begin{gathered} E\{\mathbf{z}\}\text{ unknown}, \mathbf{Z}\sigma^2 = D\{\mathbf{z}\}, \text{Z positive definite rk}\mathbf{Z} = s \le \ell, \\ \mathbf{V}\sigma^2 = D\{\mathbf{y} - \mathbf{Cz}\}, \mathbf{V}\text{ positive semidefinite} \\ \text{rk}\mathbf{V} = t \le n,\ \text{rk}[\mathbf{V}, \mathbf{CZ}] = n,\ \mathbf{C}\{\mathbf{z}, \mathbf{y} - \mathbf{Cz}\} = \mathbf{0}. \end{gathered}$$

A *homogeneous-quadratic ansatz* $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y}$ will be specified now.

Theorem 10.5. (homogeneous-quadratic setup of $\hat{\sigma}^2$):

(i) Let $\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y} = (\text{vec}\mathbf{M})'(\mathbf{y} \otimes \mathbf{y})$ be BIQUUE of σ^2 with respect to the model of the front desk.

Then

$$\begin{gathered} \hat{\sigma}^2 = (n - m - \ell)^{-1}[\mathbf{y}'\{\mathbf{I}_n - (\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{A}[\mathbf{A}'(\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'\} \\ (\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{y} - \mathbf{s}'_\mathbf{A}\mathbf{S}_\mathbf{C}^{-1}\mathbf{s}_\mathbf{C}] \end{gathered}$$

$$\begin{gathered} \hat{\sigma}^2 = (n - m - \ell)^{-1}[\mathbf{y}'\mathbf{Q}(\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{y} - \mathbf{s}'_\mathbf{A}\mathbf{S}_\mathbf{A}^{-1}\mathbf{s}_\mathbf{A}] \\ \mathbf{Q} := \mathbf{I}_n - (\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{C}[\mathbf{C}'(\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{C}]^{-1}\mathbf{C}' \end{gathered}$$

subject to

$$\begin{aligned} [\mathbf{S}_\mathbf{A}, \mathbf{s}_\mathbf{A}] &:= \mathbf{A}'(\mathbf{V} + \mathbf{CZC}')^{-1}\{\mathbf{I}_n - \mathbf{C}[\mathbf{C}'(\mathbf{V} + \mathbf{CZC}')^{-1}\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{V} + \mathbf{CZC}')^{-1}\} \\ [\mathbf{A}, \mathbf{y}] &= \mathbf{A}'\mathbf{Q}(\mathbf{V} + \mathbf{CZC}')^{-1}[\mathbf{A}, \mathbf{y}] \end{aligned}$$

and

$$[\mathbf{S_C}, \mathbf{s_C}] := \mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\{\mathbf{I}_n - \mathbf{A}[\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}[\mathbf{C}, \mathbf{y}],$$

where $\mathbf{S_A}$ and $\mathbf{S_C}$ are "*Schur complements*".
Alternately, we receive the empirical data based upon

$$\hat{\sigma}^2 = (n-m-\ell)^{-1}\mathbf{y}'(\mathbf{V}+\mathbf{CZC}')^{-1}\tilde{\mathbf{e}}_\mathbf{y}$$
$$= (n-m-\ell)^{-1}\tilde{\mathbf{e}}'_\mathbf{y}(\mathbf{V}+\mathbf{CZC}')^{-1}\tilde{\mathbf{e}}_\mathbf{y}$$

and the related variances

$$D\{\hat{\sigma}^2\} = 2(n-m-\ell)^{-1}\sigma^4 = 2(n-m-\ell)^{-1}(\sigma^2)^2$$

or replacing by the estimations

$$D\{\hat{\sigma}^2\} = 2(n-m-\ell)^{-1}(\hat{\sigma}^2)^2$$
$$= 2(n-m-\ell)^{-1}[\tilde{\mathbf{e}}'_\mathbf{y}(\mathbf{V}+\mathbf{CZC}')^{-1}\tilde{\mathbf{e}}_\mathbf{y}]^2.$$

(ii) If the cofactor matrix $\mathbf{V}$ is *positive definite*, we will find for the *simple representations* of type BIQUUE of σ^2 the *equivalent representations*

$$\hat{\sigma}^2 = (n-m-\ell)^{-1}\mathbf{y}'\{\mathbf{V}^{-1} - \mathbf{V}^{-1}[\mathbf{A}, \mathbf{C}]\begin{bmatrix} \mathbf{A}'\mathbf{V}^{-1}\mathbf{A} & \mathbf{A}'\mathbf{V}^{-1}\mathbf{C} \\ \mathbf{C}'\mathbf{V}^{-1}\mathbf{A} & \mathbf{C}'\mathbf{V}^{-1}\mathbf{C} \end{bmatrix}\begin{bmatrix} \mathbf{A}' \\ \mathbf{C}' \end{bmatrix}\mathbf{A}^{-1}\}\mathbf{y}$$

$$\hat{\sigma}^2 = (n-m-\ell)^{-1}(\mathbf{y}'\mathbf{Q}\mathbf{V}^{-1}\mathbf{y} - \mathbf{s}'_\mathbf{A}\mathbf{S}_\mathbf{A}^{-1}\mathbf{s}_\mathbf{A})$$

$$\hat{\sigma}^2 = (n-m-\ell)^{-1}\mathbf{y}'\mathbf{V}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}](\mathbf{y} - \mathbf{C}\mathbf{S}_\mathbf{C}^{-1}\mathbf{s}_\mathbf{C})$$

subject to the projection matrix

$$\mathbf{Q} = \mathbf{I}_n - \mathbf{V}^{-1}\mathbf{C}(\mathbf{C}'\mathbf{V}^{-1}\mathbf{C})^{-1}\mathbf{C}'$$

and

$$[\mathbf{S_A}, \mathbf{s_A}] := \mathbf{A}'\mathbf{V}^{-1}[\mathbf{I}_n - \mathbf{C}(\mathbf{C}'\mathbf{V}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{V}^{-1}][\mathbf{A}, \mathbf{y}] = \mathbf{A}'\mathbf{Q}\mathbf{V}^{-1}[\mathbf{A}, \mathbf{y}]$$

$$[\mathbf{S_C}, \mathbf{s_C}] := \{\mathbf{I}_\ell + \mathbf{C}'\mathbf{V}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}]\mathbf{CZ}\}^{-1}$$
$$\times\mathbf{C}'\mathbf{V}^{-1}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}][\mathbf{C}, \mathbf{y}].$$

Alternatively, we receive the empirical data based upon

$$\hat{\sigma}^2 = (n-m-\ell)^{-1}\mathbf{y}'\mathbf{V}^{-1}\tilde{\mathbf{e}}_\mathbf{y} = (n-m-\ell)^{-1}\tilde{\mathbf{e}}'_\mathbf{y}\mathbf{V}^{-1}\tilde{\mathbf{e}}_\mathbf{y}$$

and the related variances

$$D\{\hat{\sigma}^2\} = 2(n-m-\ell)^{-1}\sigma^4 = 2(n-m-\ell)^{-1}(\sigma^2)^2$$

$$\hat{D}\{\hat{\sigma}^2\} = 2(n-m-\ell)^{-1}(\hat{\sigma}^2)^2 = 2(n-m-\ell)^{-1}(\tilde{\mathbf{e}}'_{\mathbf{y}}\mathbf{V}^{-1}\tilde{\mathbf{e}}_{\mathbf{y}})^2.$$

The proofs are straight forward.

10-3 An Example for Collocation

Here we will focus on a special model with fixed effects and random effects, in particular with ξ, $E\{\mathbf{z}\}$, $E\{\bar{\mathbf{y}}\}$, $\boldsymbol{\Sigma}_{\mathbf{z}}$, $\boldsymbol{\Sigma}_{\mathbf{y}}$ *unknown*, but γ *known*.

We depart in analyzing *a height network observed at two epochs*. At the initial *epoch three height differences* have been observed. From the first epochs to the second epoch we assume height differences *which change linear in time*, for instance caused by an Earthquake. There is a *height varying model*

$$h_{\alpha\beta}(\tau) = h_{\alpha\beta}(0) + h^{\bullet}_{\alpha\beta}(0)\tau + \mathcal{O}(\tau^2),$$

where τ notes the time difference *from* the first epoch *to* the second epoch, related to the height difference. Unknown are the *fixed* height differences $h_{\alpha\beta}$ and the expected values of the *random* height difference velocities $h^{\bullet}_{\alpha\beta}$. Given is the singular dispersion matrix of height difference measurements. Alternative *estimation and prediction data* are of type $(\mathbf{V} + \mathbf{CZC}')$-BLUUE for the unknown parameter ξ of height difference at initial epoch and the expected data $E\{\mathbf{z}\}$ of *stochastic height difference velocities* $\mathbf{z}$, of type $(\mathbf{V} + \mathbf{CZC}')$-BLUUE of the expected data $E\{\mathbf{y}\}$ of height difference observations y, of type $\tilde{\mathbf{e}}_{\mathbf{y}}$ of the empirical error vector of observations and of type $(\mathbf{V} + \mathbf{CZC}')$-BLUUP for the stochastic vector $\tilde{\mathbf{z}}$ of height difference velocities. For the unknown variance component σ^2 of height difference observations we use estimates of type BIQUUE. In detail, our model assumptions are

epoch 1

$$E\left\{\begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix}\right\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \end{bmatrix}$$

epoch 2

$$E\left\{\begin{bmatrix} \underline{h}_{\alpha\beta} \\ \underline{h}_{\beta\gamma} \\ \underline{h}_{\gamma\alpha} \end{bmatrix}\right\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \tau & 0 \\ 0 & \tau \\ -\tau & \tau \end{bmatrix}\begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \\ E\{h^{\bullet}_{\alpha\beta}\} \\ E\{h^{\bullet}_{\beta\gamma}\} \end{bmatrix}$$

epoch 1 and 2

$$E\left\{\begin{bmatrix} \underset{\sim}{h}_{\alpha\beta} \\ \underset{\sim}{h}_{\beta\gamma} \\ \underset{\sim}{h}_{\gamma\alpha} \\ \underset{\sim}{k}_{\alpha\beta} \\ \underset{\sim}{k}_{\beta\gamma} \\ \underset{\sim}{k}_{\gamma\alpha} \end{bmatrix}\right\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \tau & 0 \\ 0 & \tau \\ -\tau & \tau \end{bmatrix} \begin{bmatrix} E\{h^{\bullet}_{\alpha\beta}\} \\ E\{h^{\bullet}_{\beta\gamma}\} \end{bmatrix}$$

$$\mathbf{y} := \begin{bmatrix} \underset{\sim}{h}_{\alpha\beta} \\ \underset{\sim}{h}_{\beta\gamma} \\ \underset{\sim}{h}_{\gamma\alpha} \\ \underset{\sim}{k}_{\alpha\beta} \\ \underset{\sim}{k}_{\beta\gamma} \\ \underset{\sim}{k}_{\gamma\alpha} \end{bmatrix}, \ \mathbf{A} := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} \ \xi := \begin{bmatrix} h_{\alpha\beta} \\ h_{\beta\gamma} \end{bmatrix}$$

$$\mathbf{C} := \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \tau & 0 \\ 0 & \tau \\ -\tau & \tau \end{bmatrix}, \quad E\{\mathbf{z}\} = \begin{bmatrix} E\{h^{\bullet}_{\alpha\beta}\} \\ E\{h^{\bullet}_{\beta\gamma}\} \end{bmatrix}$$

rank identities

$$rk\mathbf{A} = 2, rk\mathbf{C} = 2, rk[\mathbf{A}, \mathbf{C}] = m + l = 4.$$

The *singular dispersion* matrix $D\{\mathbf{y}\} = \mathbf{V}\sigma^2$ of the observation vector $\mathbf{y}$ and the *singular dispersion* matrix $D\{\mathbf{z}\} = \mathbf{Z}\sigma^2$ are determined in the following. We separate 3 *cases*.

(i) $rk\mathbf{V} = 6, rk\mathbf{Z} = 1$

$$\mathbf{V} = \mathbf{I}_6, \ \ \mathbf{Z} = \frac{1}{\tau^2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

(ii) $rk\mathbf{V} = 5, rk\mathbf{Z} = 2$

$$\mathbf{V} = \mathrm{Diag}(1,\ 1,\ 1,\ 1,\ 1,\ 0), \quad \mathbf{Z} = \frac{1}{\tau^2}\mathbf{I}_2, \ \ \mathrm{rk}(\mathbf{V}+\mathbf{CZC}') = 6$$

(iii) $rk\mathbf{V} = 4, rk\mathbf{Z} = 2$

$$\mathbf{V} = \mathrm{Diag}(1,\ 1,\ 1,\ 1,\ 0,\ 0), \quad \mathbf{Z} = \frac{1}{\tau^2}\mathbf{I}_2, \ \ \mathrm{rk}\,(\mathbf{V}+\mathbf{CZC}') = 6.$$

In order to be as simple as possible we use the time interval $\tau = 1$.

With the numerical values of matrix inversion and of "*Schur-complements*", e.g.
Table 10.1: $(\mathbf{V}+\mathbf{CZC}')^{-1}$
Table 10.2: $\{\mathbf{I}_n - \mathbf{A}[\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$
Table 10.3: $\{\mathbf{I}_n - \mathbf{C}[\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$
Table 10.4: "*Schur-complements*" $\mathbf{S_A}$, $\mathbf{S_C}$
Table 10.5: vectors $\mathbf{s_A}$, $\mathbf{s_C}$
1st case: $\hat{\xi}$, $D\{\hat{\xi}\}$, $\widehat{E\{\mathbf{z}\}}$, $D\{\widehat{E\{\mathbf{y}\}}\}$

$$\hat{\xi} = \frac{1}{3}\begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \end{bmatrix}\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2y_1 + y_2 - y_3 \\ y_1 + 2y_2 + y_3 \end{bmatrix},$$

$$D\{\hat{\xi}\} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

$$\widehat{E\{\mathbf{z}\}} = \begin{bmatrix} -2 & -1 & 1 & 2 & 1 & -1 \\ -1 & -2 & -1 & 1 & 2 & 1 \end{bmatrix}\mathbf{y} = \begin{bmatrix} -2y_1 - y_2 + y_3 + 2y_4 + y_5 - y_6 \\ -y_1 - 2y_2 - y_3 + y_4 + 2y_5 + y_6 \end{bmatrix},$$

$$D\{\widehat{E\{\mathbf{z}\}}\} = \frac{\sigma^2}{3}\begin{bmatrix} 7 & 5 \\ 5 & 7 \end{bmatrix},$$

2nd case: $\hat{\xi}$, $D\{\hat{\xi}\}$, $\widehat{E\{\mathbf{z}\}}$, $D\{\widehat{E\{\mathbf{z}\}}\}$

$$\hat{\xi} = \frac{1}{3}\begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \end{bmatrix}\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2y_1 + y_2 - y_3 \\ y_1 + 2y_2 + y_3 \end{bmatrix},$$

$$D\{\hat{\xi}\} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

$$\widehat{E\{\mathbf{z}\}} = \begin{bmatrix} -4 & -2 & 2 & 3 & 3 & -3 \\ -2 & -4 & -2 & 3 & 3 & 3 \end{bmatrix}\mathbf{y},$$

$$D\{\widehat{E\{\mathbf{z}\}}\} = \frac{\sigma^2}{6}\begin{bmatrix} 13 & 5 \\ 5 & 13 \end{bmatrix},$$

3rd case: $\hat{\xi}$, $D\{\hat{\xi}\}$, $\widehat{E\{\mathbf{z}\}}$, $D\{\widehat{E\{\mathbf{z}\}}\}$

$$\hat{\xi} = \frac{1}{3}\begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \end{bmatrix}\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2y_1 + y_2 - y_3 \\ y_1 + 2y_2 + y_3 \end{bmatrix},$$

$$D\{\hat{\xi}\} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

$$\widehat{E\{\mathbf{z}\}} = \begin{bmatrix} -2 & -1 & 1 & 3 & 0 & 0 \\ -1 & -2 & -1 & 0 & 3 & 0 \end{bmatrix}\frac{1}{3}\begin{bmatrix} -2y_1 - y_2 + y_3 + 3y_4 \\ -y_1 - 2y_2 - y_3 + 3y_5 \end{bmatrix},$$

Table 10.1 Matrix inverse $(\mathbf{V}+\mathbf{CZC}')^{-1}$ for a mixed Gauss–Markov model with fixed and random effects

	$\mathbf{V}+\mathbf{CZC}'$	$(\mathbf{V}+\mathbf{CZC}')^{-1}$
1st case	$\begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&2&1&0\\ 0&0&0&1&2&0\\ 0&0&0&0&0&1 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 3&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&3&0&0&0\\ 0&0&0&2&-1&0\\ 0&0&0&-1&2&0\\ 0&0&0&0&0&3 \end{bmatrix}$
2nd case	$\begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&2&0&-1\\ 0&0&0&0&2&1\\ 0&0&0&-1&1&2 \end{bmatrix}$	$\frac{1}{4}\begin{bmatrix} 4&0&0&0&0&0\\ 0&4&0&0&0&0\\ 0&0&4&0&0&0\\ 0&0&0&3&-1&2\\ 0&0&0&-1&3&-2\\ 0&0&0&2&-2&4 \end{bmatrix}$
3rd case	$\begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&-1\\ 0&0&0&0&1&1\\ 0&0&0&-1&1&3 \end{bmatrix}$	$\begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&2&-1&1\\ 0&0&0&-1&2&-1\\ 0&0&0&1&-1&1 \end{bmatrix}$

Table 10.2 Matrices $\{\mathbf{I}_n+[\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$ for a mixed Gauss–Markov model with fixed and random effects

	$\{\mathbf{I}_n-\mathbf{A}[\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$
1st case	$\frac{1}{24}\begin{bmatrix} 13&-7&4&-5&-1&4\\ -7&13&-4&-1&-5&-4\\ 4&-4&16&4&-4&-8\\ -11&-7&4&19&-1&4\\ -7&-11&-4&-1&19&-4\\ 4&-4&-8&4&-4&16 \end{bmatrix}$
2nd case	$\frac{1}{24}\begin{bmatrix} 13&-5&6&-4&-4&3\\ -5&13&-6&-4&-4&-3\\ 6&-6&12&0&0&-6\\ -11&-5&6&20&-4&3\\ -5&-11&-6&-4&20&-3\\ 6&-6&-12&0&0&18 \end{bmatrix}$

(Continued)

Table 10.2 Continued

3rd case	$\frac{1}{8}\begin{bmatrix} 5 & -1 & 2 & -3 & -1 & 0 \\ -1 & 5 & -2 & -1 & -3 & 0 \\ 2 & -2 & 4 & 2 & -2 & 0 \\ -3 & -1 & 2 & 5 & -1 & 0 \\ -1 & -3 & -2 & -1 & 5 & 0 \\ 2 & -2 & -4 & 2 & -2 & 8 \end{bmatrix}$

Table 10.3 Matrices $\{\mathbf{I}_n - \mathbf{C}[\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{C}']^{-1}\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$ for a mixed Gauss–Markov model with fixed and random effects

	$\{\mathbf{I}_n - \mathbf{C}[\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{V}+\mathbf{CZC}')^{-1}\}$
1st case	$\frac{1}{3}\begin{bmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
2nd case	$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
3rd case	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}$

Table 10.4 "*Schur-complements*" $\mathbf{S_A}$, $\mathbf{S_C}$, three cases, for a mixed Gauss–Markov model with fixed and random effects

	$\mathbf{S_A}$ (10.10)	$\mathbf{S_A}^{-1}$	$\mathbf{S_C}$ (10.12)	$\mathbf{S_C}^{-1}$
1st case	$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$	$\frac{1}{8}\begin{bmatrix} 7 & -5 \\ -5 & 7 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 7 & 5 \\ 5 & 7 \end{bmatrix}$
2nd case	$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$	$\frac{1}{24}\begin{bmatrix} 13 & -5 \\ -5 & 13 \end{bmatrix}$	$\frac{1}{6}\begin{bmatrix} 13 & 5 \\ 5 & 13 \end{bmatrix}$
3rd case	$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$	$\frac{1}{8}\begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$	$\frac{1}{3}\begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$

Table 10.5 Vectors $\mathbf{s_A}$ and $\mathbf{s_C}$, three cases, for a mixed Gauss–Markov model with fixed and random effects

	$\mathbf{s_A}$ (10.11)	$\mathbf{s_C}$ (10.13)
1st case	$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \mathbf{y}$	$\frac{1}{8}\begin{bmatrix} -9 & 3 & 12 & 9 & -3 & -12 \\ 3 & -9 & -12 & -3 & 9 & 12 \end{bmatrix} \mathbf{y}$
2nd case	$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \mathbf{y}$	$\frac{1}{24}\begin{bmatrix} -7 & -1 & 6 & 4 & 4 & -9 \\ -1 & -7 & -6 & 4 & 4 & -9 \end{bmatrix} \mathbf{y}$
3rd case	$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \mathbf{y}$	$\frac{1}{8}\begin{bmatrix} -3 & -1 & 2 & 5 & -1 & 0 \\ -1 & -3 & -2 & -1 & 5 & 0 \end{bmatrix} \mathbf{y}$

Table 10.6 *Numerical values, Case 1*

1st case: $\widehat{E\{\mathbf{y}\}}$, $D\{\widehat{E\{\mathbf{y}\}}\}$, $\tilde{\mathbf{e}}_y$, $D\{\tilde{\mathbf{e}}_{y_1}\}$

$$\widehat{E\{\mathbf{y}\}} = \begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & -1 & 1 & 2 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 2y_1 + y_2 - y_3 \\ y_1 + 2y_2 + y_3 \\ -y_1 + 2y_2 + 2y_3 \\ 2y_4 + y_5 - y_6 \\ y_4 + 2y_5 - y_6 \\ -y_4 + y_5 + 2y_6 \end{bmatrix}$$

$$D\{\widehat{E\{\mathbf{y}\}}\} = \frac{\sigma^2}{3} \begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & -1 \\ 0 & 0 & 0 & 4 & 5 & 1 \\ 0 & 0 & 0 & -1 & 1 & 2 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_y = \frac{1}{12} \begin{bmatrix} -1 & -5 & 8 & 5 & 1 & -4 \\ -5 & -1 & -8 & 1 & 5 & 4 \\ 8 & -8 & -4 & -4 & 4 & 8 \\ 11 & 7 & -4 & -7 & -11 & 8 \\ 7 & 11 & 4 & -11 & -7 & -8 \\ -4 & 4 & 8 & 8 & -8 & -4 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_y\} = \frac{\sigma^2}{3} \begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}.$$

First case:
$$\hat{\xi},\ D\{\hat{\xi}\},\ \widehat{E\{\mathbf{z}\}},\ D\{\widehat{E\{\mathbf{z}\}}\}$$

Second case:
$$\hat{\xi},\ D\{\hat{\xi}\},\ \widehat{E\{\mathbf{z}\}},\ D\{\widehat{E\{\mathbf{z}\}}\}$$

Third case:
$$\hat{\xi},\ D\{\hat{\xi}\},\ \widehat{E\{\mathbf{z}\}},\ D\{\widehat{E\{\mathbf{z}\}}\}.$$

Here are the results of computing

$$\widehat{E\{\mathbf{y}\}},\ D\{\widehat{E\{\mathbf{z}\}}\},\ \tilde{\mathbf{e}}_{\mathbf{y}} \text{ and } D\{\tilde{\mathbf{e}}_{\mathbf{y}}\},$$

ordered as case 1, case 2, and case 3 (Tables 10.6–10.10).

Table 10.7 *Numerical values, Case 2*

2nd case: $\widehat{E\{\mathbf{y}\}}$, $D\{\widehat{E\{\mathbf{y}\}}\}$, $\tilde{\mathbf{e}}_y$, $D\{\tilde{\mathbf{e}}_{y_1}\}$

$$\widehat{E\{\mathbf{y}\}} = \frac{1}{6}\begin{bmatrix} 4 & 2 & -2 & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & 0 & 0 \\ -2 & 2 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 3 & -3 \\ 0 & 0 & 0 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

$$D\{\widehat{E\{\mathbf{y}\}}\} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & -1 \\ 0 & 0 & 0 & 4 & 5 & 1 \\ 0 & 0 & 0 & -1 & 1 & 2 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \frac{1}{6}\begin{bmatrix} 2 & -2 & 2 & 0 & 0 & 0 \\ -2 & 2 & -2 & 0 & 0 & 0 \\ 2 & -2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -3 & 3 \\ 0 & 0 & 0 & -3 & 3 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{y} = \begin{bmatrix} 2y_1 - 2y_2 + 2y_3 \\ -2y_1 + 2y_2 - 2y_3 \\ 2y_1 - 2y_2 + 2y_3 \\ 3y_4 - 3y_5 + 3y_6 \\ 3y_4 - 3y_5 + 3y_6 \\ -3y_4 + 3y_5 - 3y_6 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \frac{\sigma^2}{6}\begin{bmatrix} 2 & -2 & 2 & 0 & 0 & 0 \\ -2 & 2 & -2 & 0 & 0 & 0 \\ 2 & -2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -3 & 0 \\ 0 & 0 & 0 & -3 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

1st case: $\widehat{E\{\mathbf{y}\}}$, $D\{\widehat{E\{\mathbf{y}\}}\}$, $\tilde{\mathbf{e}}_{\mathbf{y}}$, $D\{\tilde{\mathbf{e}}_{\mathbf{y}}\}$

$$\widehat{E\{\mathbf{y}\}} = \begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & -1 & 1 & 2 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 2y_1 + y_2 - y_3 \\ y_1 + 2y_2 + y_3 \\ -y_1 + 2y_2 + 2y_3 \\ 2y_4 + y_5 - y_6 \\ y_4 + 2y_5 - y_6 \\ -y_4 + y_5 + 2y_6 \end{bmatrix}$$

$$D\{\widehat{E\{\mathbf{y}\}}\} = \frac{\sigma^2}{3} \begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 4 & -1 \\ 0 & 0 & 0 & 4 & 5 & 1 \\ 0 & 0 & 0 & -1 & 1 & 2 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \frac{1}{12} \begin{bmatrix} -1 & -5 & 8 & 5 & 1 & -4 \\ -5 & -1 & -8 & 1 & 5 & 4 \\ 8 & -8 & -4 & -4 & 4 & 8 \\ 11 & 7 & -4 & -7 & -11 & 8 \\ 7 & 11 & 4 & -11 & -7 & -8 \\ -4 & 4 & 8 & 8 & -8 & -4 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \frac{\sigma^2}{3} \begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix}.$$

Table 10.8 *Numerical values, Case 3*

3rd case: $\widehat{E\{\mathbf{y}\}}$, $D\{\widehat{E\{\mathbf{y}\}}\}$, $\tilde{\mathbf{e}}_{\mathbf{y}}$, $D\{\tilde{\mathbf{e}}_{\mathbf{y}}\}$

$$\widehat{E\{\mathbf{y}\}} = \frac{1}{3} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 1 & 0 & 0 & 0 \\ -1 & -2 & -1 & 0 & 3 & 0 \\ 1 & -1 & 0 & -3 & 3 & 0 \end{bmatrix}$$

$$D\{\widehat{E\{\mathbf{y}\}}\} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & 1 & -1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ -1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & -3 \\ 0 & 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & -3 & 3 & 6 \end{bmatrix}$$

$$\tilde{\mathbf{e}}_{\mathbf{y}} = \frac{1}{3}\begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{y} = \begin{bmatrix} y_1 - y_2 + y_3 \\ -y_1 + 2y_2 - y_3 \\ y_1 - y_2 + y_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$D\{\tilde{\mathbf{e}}_{\mathbf{y}}\} = \frac{\sigma^2}{3}\begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Table 10.9 Data of type $\tilde{\mathbf{z}}$, $D\{\tilde{\mathbf{z}}\}$ for 3 cases

1st case:

$$\tilde{\mathbf{z}} = \frac{1}{3}\begin{bmatrix} -2 & -1 & 1 & 2 & 1 & -1 \\ -1 & -2 & -1 & 1 & 2 & 2 \end{bmatrix}\mathbf{y},$$

$$D\{\tilde{\mathbf{z}}\} = \frac{\sigma}{3}\begin{bmatrix} 7 & 5 \\ 5 & 7 \end{bmatrix},$$

2nd case:

$$\tilde{\mathbf{z}} = \frac{1}{3}\begin{bmatrix} -4 & -2 & 2 & 3 & 3 & -3 \\ -2 & -4 & -2 & 3 & 3 & 3 \end{bmatrix}\mathbf{y},$$

$$D\{\tilde{\mathbf{z}}\} = \frac{\sigma}{6}\begin{bmatrix} 13 & 5 \\ 5 & 13 \end{bmatrix},$$

3rd case:

$$\tilde{\mathbf{z}} = \frac{1}{3}\begin{bmatrix} -2 & -1 & 1 & 3 & 0 & 0 \\ -1 & -2 & -1 & 0 & 3 & 0 \end{bmatrix}\mathbf{y},$$

$$D\{\tilde{\mathbf{z}}\} = \frac{\sigma}{3}\begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix},$$

Table 10.10 Data of type $\hat{\sigma}^2$, $D\{\hat{\sigma}^2\}$, $\hat{D}\{\hat{\sigma}^2\}$ for 3 cases

1st case: $n = 6, m = 2, \ell = 2, n - m - \ell = 2$

$$\hat{\sigma}^2 = \frac{1}{12}\mathbf{y}'\begin{bmatrix} 7 & -1 & -2 & -5 & -1 & 4 \\ -1 & 7 & 2 & -1 & -5 & -4 \\ -2 & 2 & 10 & 4 & -4 & -8 \\ -5 & -1 & 4 & 7 & -1 & -2 \\ -1 & -5 & -4 & -1 & 7 & 2 \\ 4 & -4 & -8 & -2 & 2 & 10 \end{bmatrix}\mathbf{y},\ D\{\hat{\sigma}^2\} = \sigma^4,$$

$$\hat{D}\{\hat{\sigma}^2\} = \frac{1}{144}\{\mathbf{y}'\begin{bmatrix} 7 & -1 & -2 & -5 & -1 & 4 \\ -1 & 7 & 2 & -1 & -5 & -4 \\ -2 & 2 & 10 & 4 & -4 & -8 \\ -5 & -1 & 4 & 7 & -1 & -2 \\ -1 & -5 & -4 & -1 & 7 & 2 \\ 4 & -4 & -8 & -2 & 2 & 10 \end{bmatrix}\mathbf{y}\}^2,$$

2nd case: $n = 6, m = 2, \ell = 2, n - m - \ell = 2$

$$\hat{\sigma}^2 = \frac{1}{12}\mathbf{y}'\begin{bmatrix} 2 & -2 & 2 & 0 & 0 & 0 \\ -2 & 2 & -2 & 0 & 0 & 0 \\ 2 & -2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -3 & 3 \\ 0 & 0 & 0 & -3 & 3 & -3 \\ 0 & 0 & 0 & 3 & -3 & 3 \end{bmatrix}\mathbf{y},\ D\{\hat{\sigma}^2\} = \sigma^4,$$

$$\hat{D}\{\hat{\sigma}^2\} = \frac{1}{144}\{\mathbf{y}'\begin{bmatrix} 2 & -2 & 2 & 0 & 0 & 0 \\ -2 & 2 & -2 & 0 & 0 & 0 \\ 2 & -2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -3 & 3 \\ 0 & 0 & 0 & -3 & 3 & -3 \\ 0 & 0 & 0 & 3 & -3 & 3 \end{bmatrix}\mathbf{y}\}^2,$$

3rd case: $n = 6, m = 2, \ell = 2, n - m - \ell = 2$

$$\hat{\sigma}^2 = \frac{1}{6}\mathbf{y}'\begin{bmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\mathbf{y},\ D\{\hat{\sigma}^2\} = \sigma^4,$$

$$\hat{D}\{\hat{\sigma}^2\} = \frac{1}{144}\left\{\mathbf{y}'\begin{bmatrix} 1 & -1 & 1 & 0\,0\,0 \\ -1 & 1 & -1 & 0\,0\,0 \\ 1 & -1 & 1 & 0\,0\,0 \\ 0 & 0 & 0 & 0\,0\,0 \\ 0 & 0 & 0 & 0\,0\,0 \\ 0 & 0 & 0 & 0\,0\,0 \end{bmatrix}\mathbf{y}\right\}^2 .$$

Here is my journey's end.

10-4 Comments

(i) In their original contribution *A. N. Kolmogorov* (1941 a, b, c) and *N. Wiener* ("*yellow devil*", 1939) did not depart from our general setup of fixed effects and random effects. Instead they departed from the model

$$\text{“}y_{P_\alpha} = c_{P_\alpha} + \sum_{\beta=1}^{q} c_{P_\alpha Q_\beta} y_{Q_\beta} + \sum_{\beta,\gamma=1}^{q} c_{P_\alpha Q_\beta Q_\gamma} y_{Q_\beta} y_{Q_\gamma} + \mathcal{O}(y^3)''$$

of *nonlinear prediction*, e.g. *homogeneous linear prediction*

$$\begin{aligned} y_{P_1} &= y_{Q_1} c_{P_1 Q_1} + y_{Q_2} c_{P_1 Q_2} + \cdots + y_{Q_q} c_{P_1 Q_q} \\ y_{P_2} &= y_{Q_1} c_{P_2 Q_1} + y_{Q_2} c_{P_2 Q_2} + \cdots + y_{Q_q} c_{P_2 Q_q} \\ &\cdots \\ y_{P_{p-1}} &= y_{Q_1} c_{P_{p-1} Q_1} + y_{Q_2} c_{P_{p-1} Q_2} + \cdots + y_{Q_q} c_{P_{p-1} Q_q} \\ y_{P_p} &= y_{Q_1} c_{P_p Q_1} + y_{Q_2} c_{P_p Q_2} + \cdots + y_{Q_q} c_{P_p Q_q}. \end{aligned}$$

From given values of "*random effects*" $(y_{Q_1}, \ldots, y_{Q_p})$ other values of "*random effects*" $(y_{P_1}, \ldots, y_{P_p})$ have been predicted, namely under the assumption of "*equal correlation*" P of type

$$P_1 \bullet\!\!-\!\!\!-\!\!\bullet\, P_2 = Q_1 \bullet\!\!-\!\!\!-\!\!\bullet\, Q_2$$

$$\widehat{E\{y_P\}} = \mathbf{C}(\mathbf{C}\Sigma_{y_P}^{-1}\mathbf{C})^{-1}\mathbf{C}\Sigma_{y_P}^{-1} y_P$$

$$D\{\widehat{E\{y_P\}}\} = \mathbf{C}(\mathbf{C}\Sigma_{y_P}^{-1}\mathbf{C})^{-1}\mathbf{C}'$$

or

for all $y_P \in \{y_{P_1}, \ldots, y_{P_P}\}$

$$||y_P - \hat{y}_P|| := E\{(y_P - \hat{y}_P)\}^2 \leq \min,$$

$$E\{(y_P - \hat{y}_P)^2\} = E\{(y_{2i} - \sum_{Q_1}^{Q_q} y'_{1j} c_{ij})^2\} = \min(KW)$$

for homogeneous linear prediction.

"*ansatz*"

$$E\{y_P - E\{\hat{y}_P\}\} := 0,\ E\{(y_{P_1} - E\{\hat{y}_{P_1}\})(y_{P_2} - E\{\hat{y}_{P_2}\})\} = \mathrm{cov}(y_{P_1}, y_{P_2})$$

$$E\{(y_P - \hat{y}_P)^2\} = \mathrm{cov}(P, P) - \sum_{Q=Q_1}^{Q=Q_q} c_{pj}\,\mathrm{cov}(P, Q_j)$$

$$+ \sum_{Q_j} \sum_{Q_k} c_{pj} c_{pk}\,\mathrm{cov}(Q_j, Q_k) = \min$$

"*Kolmogorov–Wiener prediction*"

$$c_{pj}(KW) = \sum_{k=1}^{k=q} [\mathrm{cov}(Q_j, Q_k)]^{-1} \mathrm{cov}(Q_k, P)$$

$$\mathbf{E}\{(y_P - \hat{y}_P)^2 | KW\} = \mathrm{cov}(P, P) - \sum_{j=1}^{q} \sum_{k=1}^{q} \mathrm{cov}(P, Q_j)\mathrm{cov}(P, Q_k)[\mathrm{cov}(Q_j, Q_k)]^{-1}$$

constrained to

$$|Q_j - Q_k|$$

$$Q_j\text{—+——+——+——+——+—}Q_k$$

$$\mathrm{cov}(Q_j, Q_k) = \mathrm{cov}(Q_j - Q_k)$$

"$y_P - E\{y_P\}$ *is weakly translational invariant*"

KW prediction suffers from the effect that "a priori" we know only *one realization of the random function* $y(Q_1), \ldots, y(Q_q)$, for instance.

$$\widehat{\mathrm{cov}(\tau)} = \frac{1}{N_\delta} \sum_{|Q_j - Q_k|} (y_{Q_j} - E\{y_{Q_j}\})(y_{Q_k} - E\{y_{Q_k}\}).$$

Modified versions of the KW prediction exist if we work with random fields in the plane (*weak isotropy*) or on the plane (*rotational invariances*).

As a model of "*random effects*" we may write

$$\sum_{Q_\beta} c_{P_\alpha Q_\beta} E\{y_{Q_\beta}\} = E\{y_{P_\alpha}\} \sim CE\{z\} = E\{y\}.$$

(ii) The *first model* applies if we want to predict data of one type to predict data of the same type. Indeed, we have to generalize if we want to predict, for instance,

$$\left.\begin{array}{c}\textit{vertical deflections,}\\ \textit{gravity gradients,}\\ \textit{gravity values,}\end{array}\right] \textit{from gravity disturbances.}$$

The *second model* has to start from relating *one set of heterogeneous data to another set of heterogeneous data*. In the case we have to relate the various data sets to each other. An obvious alternative setup is

$$\sum_{Q_\beta} c_{1P_\alpha Q_\beta} E\{z_{1Q_\beta}\} + \sum_{R_\gamma} c_{2P_\alpha R_\gamma} E\{z_{2R_\gamma}\} + \cdots = E\{y_{P_\alpha}\}.$$

(iii) The *level of collocation* is reached if we include a trend model in addition to *Kolmogorov–Wiener prediction*, namely

$$E\{\mathbf{y}\} = \mathbf{A}\xi + \mathbf{C}E\{\mathbf{z}\},$$

the trend being represented by $A\xi$. The decomposition of "*trend*" and "*signal*" is well represented in *E. Grafarend* (1976), *E. Groten* (1970), *E. Groten and H. Moritz* (1964), *S. Heitz* (1967, 1968, 1969), *S. Heitz and C.C. Tscherning* (1972), *R.A. Hirvonen* (1956, 1962), *S. K. Jordan* (1972 a, b, c, 1973), *W.M. Kaula* (1959, 1963, 1966 a, b, c, 1971), *K.R. Koch* (1973 a, b), *K.R. Koch and S. Lauer* (1971), *L. Kubackove* (1973, 1974, 1975), S. Lauer (1971 a, b), *S.L. Lauritzen* (1972, 1973, 1975), *D. Lelgemann* (1972, 1974), P. Meissl (1970, 1971), in particular *H. Moritz* (1961, 1962, 1963 a, b, c, d, 1964, 1965, 1967, 1969, 1970 a, b, c, d, e, 1971, 1972, 1973 a, b, c, d, e, f, 1974 a, b, 1975), *H. Moritz and K.P. Schwarz* (1973), W. Mundt (1969), *P. Naicho* (1967, 1968), G. Obenson (1968, 1970), *A.M. Obuchow* (1947, 1954), *I. Parzen* (1960, 1963, 1972), *L.P. Pellinen* (1966, 1970), *V.S. Pugachev* (1962), *C.R. Rao* (1971, 1972, 1973 a, b), *R. Rupp* (1962, 1963, 1964 a, b, 1966 a, b, c, 1972, 1973 a, b, c, 1974, 1975), H. P. Robertson (1940), *M. Rosenblatt* (1959, 1966), *R. Rummel* (1975 a, b), *U. Schatz (*1970), I. Schoenberg (1942), W. Schwahn (1973, 1975), *K.P. Schwarz* (1972, 1974 a, b, c, 1975 a, b, c), *G. Seeber* (1972), *H.S. Shapiro* (1970), *L. Shaw et al* (1969), L. Sjoeberg (1975), *G. N. Smith* (1974), *F. Sobel* (1970), *G.F. Taylor* (1935, 1938), *C. Tscherning* (1972, a, b, 1973, 1974 a, b, 1975 a, b, c, d, e), *C. Tscherning and R.H Rapp* (1974), *V. Vyskocil* (1967, 1974 a, b, c), *P. Whittle* (1963 a, b), *N. Wiener* (1958, 1964), H. Wolf (1969, 1974), *E. Wong* (1969), *E. Wong and J.B. Thoma* (1961), *A.M. Yaglom* (1959, 1961).

(iv) An interesting comparison is the various solutions of type

- $\hat{y}_2 = \hat{l} + \hat{L}y_1$ (*best inhomogeneous linear prediction*)
- $\hat{y}_2 = \hat{L}y_1$ (*best homogeneous linear prediction*)
- $\hat{y}_2 = \hat{L}y_1$ (*best homogeneous linear unbiased prediction*)

dispersion identities

$$D_3 \leq D_2 \leq D_1.$$

(v) In spite of the effect that "*trend components*" and "*KW prediction*" may serve well the needs of an analyst, generalizations are obvious. For instance, in *Krige's prediction concept* it is postulated that only

$$||y_{p_1} - y_{p_2}||^2 =: E\{(y_{p_1} - y_{p_2} - (E\{y_{p_1}\} - E\{y_{p_2}\}))^2\}$$

is a *weakly relative translational invariant stochastic process. A.N. Kolmogorov* has called the weakly relative translational invariant random function

"structure function"

Alternatively, higher order variance-covariance functions have been proposed:

$$||y_1, y_2, y_3||^2 =: E\{(y_1 - 2y_2 + y_3) - (E\{y_1\} - 2E\{y_2\} + E\{y_3\})^2\}$$
$$||y_1, y_2, y_3, y_4||^2 =: E\{(y_1 - 3y_2 + 3y_3 - y_4) - (E\{y_1\} - 3E\{y_2\} + 3E\{y_3\} - E\{y_4\})^2\}$$

etc..

(vi) Another alternative has been the construction of higher order *absolute variance-covariance* functions of type

$$||(y_{Q_1} - E\{y_{Q_1}\})\,(y_{Q_2} - E\{y_{Q_2}\})\,(y_{Q_3} - E\{y_{Q_3}\})||$$
$$||(y_{Q_1} - E\{y_{Q_1}\})\,(...)\,(y_{Q_n} - E\{y_{Q_n}\})||$$

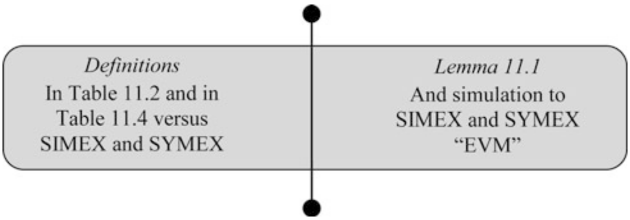
like in *E. Grafarend* (1984) derived from the characteristic functional, namely a series expression of higher order variance-covariance functions.

Chapter 11
The Sixth Problem of Probabilistic Regression - the random effect model – "errors-in-variable"

"*In difference to classical regression error-in-variables models here measurements occurs in the regressors. The naive use of regression estimators leads to severe bias in this situation. There are consistent estimators like the total least squares estimator (TLS) and the moment estimator (MME).*

J. Polzehl and S. Zwanzig
Department of Mathematics
Uppsala University, 2003 A.D."

Read Definition in *Tables* *11.1*, *11.2* and *11.3* and *Lemma* *11.1* as well as SIMEX/SYMEX in *Table 11.4*



In contrast to the conventional regression models, *error-in-variable models* (*E*rror-in-*V*ariables *M*odels: EVM) have characteristic measurement error or residuals in the *regressor*, with respect to such a model, those regression estimators lend to a *severe bias problem*. Here we only review linear and nonlinear error in variable models which are based on *Bayes Statistic*. Very often the EVM are called

Total Least Squares

when we use an extension of classical *Least Squares*.

There is a vast literature about EVMs. *Examples of Bayesian Statistics* for EVMs are *S.M. Berry et al* (2002), *R.J. Carrol et al* (1996i,ii), *W.A. Fuller* (1984) and *J. Pilz* (1994).

Notable algebraic contributions are *Z. Bai* and *J. Demmel* (1993), *A. Björek*, *P. Heggernes* and *P. Matstoms* (2000), *G.H. Golub, P.C. Hansen* and *D.P. O'Leary* (1999), *G.H. Golub* and *C. van Loan* (1980, 1996), *H. Guo* and *R.H. Renant* (2002), *P.C. Hansen* (1989,1992, 1994), *P.C. Hansen* and *D.O'Leary* (1993), *R.H. Horn*

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_11,
© Springer-Verlag Berlin Heidelberg 2012

Table 11.1 *Second order Statistics "Error-in-Variables"*

$$E\{\mathbf{y}\} = E\{\mathbf{X}\}\gamma$$

$$i, j \in \{1, \ldots, n\}, k, l \in \{1, \ldots, m\}$$

(i)	$E\{\mathbf{y}_i\} = \mu_i \sim E\{\mathbf{y}\} = \mu_y$	*First order moment of observation space* ***Y***
	$E\{\mathbf{y}_i - E\{\mathbf{y}_i\}\} = 0$	
(ii)	$E\{[\mathbf{y}_i - E\{\mathbf{y}_i\}][y_j - E\{\mathbf{y_j}\}] = D_{ij}$	*Second order moment of observation space* ***Y***
(iii)	$E\{\begin{bmatrix} \mathbf{x}_{i1} \\ \ldots \\ \mathbf{x}im \end{bmatrix}\} = E\{[\mathbf{x}_{ik}]\} = \mu_x$	*First order moments of random effects* $\mathbf{x}_{ik}$
	$E\{\mathbf{x}_{ik} - E\{\mathbf{x}_{ik}\}\} = 0 \forall i \in \{1, \ldots, n\}, k \in \{1, \ldots, m\}$	
	$\begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \mathbf{x}_{1m-1} & \mathbf{x}_{1m} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{2m-1} & \mathbf{x}_{2m} \\ \ldots & & & \\ \mathbf{x}_{n-11} & \mathbf{x}_{n-12} & \mathbf{x}_{n-1m-1} & \mathbf{x}_{n-1m} \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \mathbf{x}_{nm-1} & \mathbf{x}_{nm} \end{bmatrix} = \mathbf{X}$	
(iv)	$E\{[\mathbf{x}_{ik} - E\{\mathbf{x}_{ik}\}][x_{jl} - E\{\mathbf{y}_{jl}\}]\} := \sigma_{ijkl}$	*Second order moment of random effects* $\mathbf{X}_{ik}$
	$\begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \ldots & \ldots & \ldots \\ \sigma_{n1} & \ldots & \sigma_{nn} \end{bmatrix}$	*var-cov matrix of observation space* ***Y***
	$\begin{bmatrix} \sigma_{ij11} & \ldots & \sigma_{ij1m} \\ \ldots & \ldots & \ldots \\ \sigma_{ijm1} & \ldots & \sigma_{ijmn} \end{bmatrix}$	*var-cov matrix of random effects* $\mathbf{X}_{ik}$

model assumptions

(iv-1)	assumptions : $D_{ij} = \text{Diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$ "variance component model of observation space **Y**"
(iv-2)	assumptions : $D_{ij} \sim \sigma_{1j}$ "full symmetric matrix of observation space **Y**"
(iv-3)	assumptions: '*'simple model"* $D_{ij} \sim \delta_{ij}\sigma^2$
(iv-4)	assumptions: $D\{x_{ijkl}\} =: \sigma_{ijkl}$ "fourth order tensor" : $\sigma_{ijkl} = \sigma^2 \text{Diag}\{\tau_1, \ldots, \tau_m\}$
(iv-5)	assumptions:$\sigma_{ij}\sigma_{kl} = \sigma_{ijkl}$ *"decomposition"*
(iv-6)	assumptions:$\sigma_{ijkl} = \sigma_{ij}\sigma^2\delta_{kl}$ *"simple model"*
(iv-7)	assumptions:$\sigma_{ijkl} = \sigma^2\delta_{ij}\delta_{kl}$ *"super simple model"*

and *C.R. Johnson* (1985), *S. van Huffel* and *J. Vanderwalle* (1991), *B.N. Parlett* (1980), *S. van Huffel* and *H. Zha* (1993), *A.N. Tikhonov* (1963), *A.N. Tikhonov* and *V. Y. Arsenin* (1977), *R. A. Renaut* and *H. Guo* (2005).

Contribution on *straight-line-fit* were presented by *A. Acar et al* (2006), *M. Bipp* and *H. Krans* (199), *S. Kupferer* (2005), *J. Reinkind* (2008), *B. Schaffrin* (1983, 2002, 2008) and *B. Schaffrin* and *A. Wieser* (2008). Linear and quadratic constraints were used in *B. Schaffrin* and *Y.A. Fellus* (2009).

Table 11.2 *Explicit and implicit models*

(i) *Explicit model E*

$$E\{\mathbf{y}_i\} = E\{\mathbf{X}_i^T\}\gamma = \gamma^T\mathbf{X}_i \ \text{ for all } \gamma \in \tau := \{1, \ldots, m\}$$
$$\mathbf{y}_i = E\{\mathbf{y}_i\} + \mathbf{e}_y, \mathbf{X}_{ik} = \mathbf{E}\{\mathbf{X}_{ik}\} + E_{\mathbf{X}} \ \text{ for all } \mathbf{Y} \in \{1, \ldots, n\}, \mathbf{X} \in \{\mathbf{X}_{11}, \ldots, \mathbf{X}_{nm}\}$$

"$\mathbf{e}_y$ error vector, $\mathbf{E}_{\mathbf{X}}$ error matrix"

(ii) *Implicit model*

"We design our model that there no distinction between response variables **Y** and the predictor variables **X**"

Y are called response variables, **X** are called predictor variables.

Definitions: *(i)* $\mathbf{Z}_i^T := \{\mathbf{Y}_i, \mathbf{X}_i^T\}$
(ii) $\alpha^T := \{-1, \gamma_1, \ldots, \gamma_m\}$
(iii) $\eta_i^T := \{\mu_i, \mathbf{M}_i^T\}$

"Implicit model"

$$\boxed{\begin{gathered}\mathbf{O}_i := \{-1, \gamma_1, \ldots, \gamma_m\}[E\{\mathbf{y}_1\}, E\{\mathbf{X}_{\mathbf{i1}}\}, \ldots, E\{\mathbf{x}_{\mathbf{i1}}\}] \\ \mathbf{Z}_i := \eta_i + \mathbf{E}_{\mathbf{Z}}, \mathbf{Z}_i' := (\mathbf{Y}_i, \mathbf{X}_i'), \mathrm{var}\mathbf{E}_{\mathbf{Z}} := \mathbf{I}_{m+1}\sigma^2\end{gathered}}$$

References relating to *stochastic formulations* of *EVM*s are also very extensive. For instance, we mention to works of SIMEX models (*SIM*ulation *EX*trapolation estimators) by *R.J. Carroll et al* (1961i,ii), *R.J. Carroll et al* and *L.A. Stefanski* (1997), *J. R. Cook* and *L.A. Stefanski* (1987), *X. Lin* and *R.J. Carroll et al* (1999,2000), J. P. Holcomb (1999), *L.A. Stefanski* and *J. R. Cook* (1995), *N. Wang et al* (1998), in particular *S. Zwanzig* (1987,1989,1997,1998, 1999, 2000), *A. Kukash* and *S. Zwanzig*(2002), *J. Polzehl* and *S. Zwanzig* (2005), *S. Zwanzig*(2005), *J. Polzehl* and *S. Zwanzig* (2005) and *S. Zwanzig*(1998).

Finally, let us refer to Fig. 11.1 in order to illustrate the model "*errors-in-variables*".

We already mentioned them in contrast to standard regression models, EVMs have *"measurement errors"* in the *regressor*. The naive use of regression estimators leads to severe *bias problems*: The main idea of

SIMEX
(SIMulation EXtrapolation estimator)

is to eliminate the effect of *"measurement errors"* by *extrapolation*. Now samples with arbitrary larger variances are generated by added *simulated errors* to the original observed *regressor variables*. Here we follow to a large part the contribution of *J. Polzehl* and *S. Zwanzig* (2003) which summarizes a great number of contribution of *SIMEX*.

Table 11.3 *Inverse parameter transformation, naive estimator and Total Least Squares*

(11.31) $$O_i^* = [-E\{\mathbf{y}_i\}, \gamma_i E\{\mathbf{x}_{ij}\}, \ldots, \gamma_i E\{\mathbf{x}_{im}\}]/[-\gamma_\ell]$$
$$\text{for all } i = 1, \ldots, n, k = 1, \ldots, m$$

(11.32) $$\alpha^\ell := (-\alpha)/\gamma_\ell \quad \text{for } \ell = 1, \ldots, m, \alpha^0 := \alpha$$

"*inverse parameter transformation*"

(11.33) $$\gamma_k = -(\alpha_k^\ell)/\alpha_0^\ell \quad \text{for } k = 1, \ldots, m$$

(11.34) $$O_i^* =_i \xi_\ell - (\alpha_{(-\ell)}^\ell)\xi_{i(-\ell)}$$

"*explicit rewritten EVM*"

(11.35) $$\mathbb{Z}_\ell = (\alpha_{(-\ell)}^\ell \xi_{(-\ell)} + \mathbb{E}_\ell \quad \text{for } \ell = 0, 1, \ldots, m-1, m$$

"naive regression model"

(11.36) $$\mathbb{Z}_\ell = (\alpha_{(-\ell)}^\ell)'\mathbb{Z}_{-\ell} + \mathbb{E}_\ell$$

"The ℓth variable is considered as response and error in the other predictor variables are ignored"

(11.37) $$\hat{\alpha}_{(-\ell)}^\ell := (\mathbb{Z}'_{(-\ell)}\mathbb{Z}_{-\ell})^{-1}\mathbb{Z}'_{(-\ell)}\mathbb{Z}_\ell$$

(11.38) $$\alpha_\ell^\ell = -1$$

"solution of the naive estimator and its bias"

$$\min \ \alpha', \mathbb{Z}', \mathbb{Z}_\alpha$$
$$\alpha, \alpha_\ell = -1$$

"bias"

(11.30) $$E\{\hat{\alpha}_{(-\ell)}^\ell\} = E\{[(\xi'_{(-\ell)} + \mathbb{E}'_{(-\ell)})(\xi_{(-\ell)} + \mathbb{E}_{(-\ell)})]^{-1}(\xi'_{(-\ell)} + \mathbb{E}'_{(-\ell)})\}$$
$$\doteq (\xi'_{(-\ell)}\xi_{(-\ell)} + \sigma^2 I_m)^{-1}\xi'_{(-\ell)}\xi_{(-\ell)}(\alpha_{(-\ell)}^\ell$$

"criterion function for the ℓth naive regression"

(11.40) $$\hat{\gamma}^{\text{naive},\ell} = \underset{\gamma}{\arg\min} \frac{1}{\gamma_\ell^2}[-1, \gamma]\mathbb{Z}'\mathbb{Z}\begin{bmatrix} -1 \\ \gamma \end{bmatrix}$$

"Total Least Squares Estimator $\hat{\gamma}_{TLS}$, maximum likelihood estimator for a Gaussian error distribution"

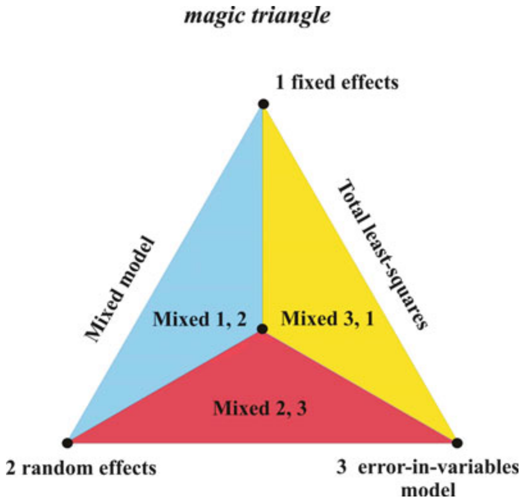
(11.41) $$\text{TLS: } \hat{\gamma}_{\text{TLS}} = \underset{\gamma}{\arg\min} \sum_{i=1}^n \min_{E\{\mathbf{x}_{ik}\} \in \mathbb{R}^m} (\|\mathbf{Y}_i - E\{\mathbf{x}'_{ik}\}\gamma\| - \|\mathbf{x}_{ik} - E\{\mathbf{x}_{ik}\}\|^2)$$

$$\text{TLS}(\gamma) = \frac{1}{1+\|\gamma\|^2}[-1, \gamma]\mathbb{Z}'\mathbb{Z}\begin{bmatrix} -1 \\ \gamma \end{bmatrix}$$

(11.42) $$\sum_{\ell=1}^m \text{Naive}_\ell(\gamma)^{-1} = \text{TLS}(\gamma)^{-1}$$

(11.43) $$\boxed{\sum_{\ell=1}^m \max_{\gamma} \text{Naive}_\ell(\gamma)^{-1} \geq \max_{\gamma} \text{TLS}(\gamma)^{-1}}$$

Fig. 11.1 Magic triangle



11-1 The Model of Error-in-Variables or Total Least Squares

Let us in more detail introduce the model definition of *Error-in-Variables*, also called *Total Least Squares*, in a statistical reference frame.

The *observation space* $\mathbf{Y}$ is generated by the *first order moments* $E\{\mathbf{y}_i\}$, the mean value operator, and the *central second order moments* $D\{\mathbf{y}_i\} =: D_{ij}$, the variance-covariance matrix in the observation space. *In contrast* the *random effects matrix-valued space* $\mathbf{X}$ is represented by the *first order, second order tensor* $E\{\mathbf{x}_y\}$ and by the *central second order, fourth order tensor* $D\{\mathbf{y}_{ik}\} =: \sigma_{ijkl}$. In modeling we present 7 sub-models, of variances and of covariances, of course, we assume that the observation space $\mathbf{Y}$ and the random effects space $\mathbf{X}_{ik}$ are *uncorrelated*.

$$\boxed{cov(\mathbf{y}_i, \mathbf{x}_{ik}) = 0}$$

We leave the formulation of Bayesian estimation for a model to the references. Instead we formulate equivalent models of type "implicit" and "explicit" here.

The implicit model underlines the symmetry structure of the *"error-in-variable models."*

Here we will proceed as following: *First* we will solve the nonlinear system of type *"error-in-variables"* by generalized least-squares. *Second* we use a *straight line fit* by technique of GLS ("generalized least squares"). Third we use the symmetry of EVM in order to introduce SYMEX for "*Total Least Squares*" and demonstrate its properties by a *simulated example*.

11-2 Algebraic Total Least Squares for the Nonlinear System of the Model "Error-in-Variables"

First, we define the random effect model of type "*errors-in-variables*" subject to the minimum condition $i_y' \mathbf{W}_y \, i_y + \mathrm{tr} \, \Im_{\mathbf{X}}' \mathbf{W}_{\mathbf{X}} \Im_{\mathbf{X}} = \min$. Second, we form the derivations, the *partial derivations* $i_y' \mathbf{W}_y \, i_y + \mathrm{tr} \, \Im_{\mathbf{X}}' \mathbf{W}_{\mathbf{X}} \Im_{\mathbf{X}} + 2\boldsymbol{\lambda}' \{y - \mathbf{X}\gamma + \Im_{\mathbf{X}}\gamma - i_y\}$, the *necessary conditions* for obtaining the minimum.

Definition 11.1. (the random effect model: "*errors-in-variables*")

The nonlinear model of type "*errors-in-variables*" is solved by "*total least squares*" based on the *risk function*

$$\mathbf{y} = E\{\mathbf{y}\} + \mathbf{e_y} \quad \sim \quad \mathbf{y} = \mathbf{y_0} + \mathbf{i_y} \tag{11.1}$$

$$\mathbf{X} = E\{\mathbf{X}\} + \mathbf{E_X} \quad \sim \quad \mathbf{X} = \mathbf{X_0} + \Im_{\mathbf{X}} \tag{11.2}$$

subject to

$$E\{\mathbf{y}\} \in \mathbb{R}^n \quad \sim \quad \mathbf{y}_0 \in \mathbb{R}^n \tag{11.3}$$

$$E\{\mathbf{X}\} \in \mathbb{R}^{n \times m} \quad \sim \quad \mathbf{X}_0 \in \mathbb{R}^{n \times m} \tag{11.4}$$

$$\mathrm{rk}\, E\{\mathbf{X}\} = m \sim \mathrm{rk}\, \mathbf{X}_0 = m$$

$$\text{and}\, n \geq m \tag{11.5}$$

$$\mathcal{L}_1 := \sum_{i_1, i_2} w_{i_i i_2} i_{i_1} i_{i_2} \qquad \text{and} \qquad \mathcal{L}_2 := \sum_{i_1, k_1, k_2} w_{k_1 k_2} i_{i_1 k_1} i_{i_1 k_2}$$

$$\mathcal{L}_1 := \mathbf{i'_y W i_y} \qquad \text{and} \qquad \mathcal{L}_2 := \mathrm{tr} \Im_{\mathbf{X}}' \mathbf{W_X} \Im_{\mathbf{X}}$$

$$\mathcal{L}_1 + \mathcal{L}_2 = \mathbf{i}'_y \mathbf{W_y i_y} + \mathrm{tr} \Im_{\mathbf{X}}' \mathbf{W_X} \Im_{\mathbf{X}} = \min_{\mathbf{y_0}, \mathbf{X_0}}$$

$$\mathcal{L} =: \; ||\mathbf{i_y}||^2_{\mathbf{W_y}} + ||\Im_{\mathbf{X}}||^2_{\mathbf{W_X}} \tag{11.6}$$

subject to

$$\mathbf{y} - \mathbf{X}_\gamma + \Im_{\mathbf{X}} \gamma - \mathbf{i_y} = 0. \tag{11.7}$$

The result of the minimization process is given by *Lemma 11.1*:

Lemma 11.1. (*error*-in-variables model, normal equations):

The *risk function* of the model "*errors-in-variables*" is minimal, if and only if

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\gamma} = -\mathbf{X}'\boldsymbol{\lambda}_\ell + \Im'_{\mathbf{X}}\lambda_\ell = \mathbf{0} \tag{11.8}$$

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\mathbf{i_y}} = \mathbf{W_y}\mathbf{i_y} - \lambda_\ell = \mathbf{0} \tag{11.9}$$

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\mathbf{I_X}} = \mathbf{W_X}\Im_{\mathbf{X}} + \gamma\lambda'_\ell \tag{11.10}$$

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\lambda} = \mathbf{y} - \mathbf{X_y} + \Im_{\mathbf{X}}\gamma - \mathbf{i_y} = \mathbf{0} \tag{11.11}$$

and

$$\det\left(\frac{\partial^2\mathcal{L}}{\partial\gamma_i\partial\gamma_j}\right) \tag{11.12}$$

Proof. First, we begin with the *modified risk function*

$$||\mathbf{i_y}||^2_{\mathbf{W_y}} + ||\Im_{\mathbf{X}}||^2 + 2\lambda'(\mathbf{y} - \mathbf{X}\gamma + \Im_{\mathbf{X}}\gamma - \mathbf{i_y}) = \min,$$

where the minimum condition is extended over

$$\mathbf{y},\ \mathbf{X},\ \mathbf{i_y},\ \Im_{\mathbf{X}},\ \lambda,$$

when λ denotes the *Lagrange parameter*.

$$\mathbf{i}'_{\mathbf{y}}\mathbf{W_y}\mathbf{i_y} + \mathrm{tr}\Im'_{\mathbf{X}}\mathbf{W_X}\Im_{\mathbf{X}} + 2\lambda'(\mathbf{y} - \mathbf{X}\gamma + \Im_{\mathbf{X}}\gamma - \mathbf{i_y}) = \min \tag{11.13}$$

if and only if

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\gamma} = -\mathbf{X}'\boldsymbol{\lambda}_\ell + \Im'_{\mathbf{X}}\lambda_\ell = \mathbf{0} \tag{11.14}$$

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial\mathbf{i_y}} = \mathbf{W_y}\mathbf{i_y} - \lambda_\ell \tag{11.15}$$

$$\frac{1}{2}\frac{\partial\mathcal{L}}{\partial(\mathrm{tr}\Im'_{\mathbf{X}}\mathbf{W_X}\Im_{\mathbf{X}})} = \mathbf{W_X}\Im_{\mathbf{X}} = \mathbf{0} \tag{11.16}$$

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \lambda_\ell} = \mathbf{y} - \mathbf{X}\gamma + \Im_{\mathbf{X}}\gamma - \mathbf{i}_{\mathbf{y}} = \mathbf{0} \tag{11.17}$$

and

$$\frac{\partial^2 \mathcal{L}}{\partial \gamma\, \partial \gamma'} \text{ positive semidefinite.} \tag{11.18}$$

The *first derivatives* guarantee the *necessity* of the solution, while the *second derivatives* being positive semidefinite assure the *sufficient condition*.

Second, we have the nonlinear equations, namely

$$(-\mathbf{X}' + \Im'_{\mathbf{X}})\lambda_\ell = 0\,(\textit{bilinear}) \tag{11.19}$$

$$\mathbf{W}_{\mathbf{y}}\mathbf{i}_{\mathbf{y}} - \lambda_\ell = 0\,(\textit{linear}) \tag{11.20}$$

$$\mathbf{W}_{\mathbf{X}}\Im_{\mathbf{X}} = 0\,(\textit{bilinear}) \tag{11.21}$$

$$\mathbf{y} - \mathbf{X}\gamma + \Im_{\mathbf{X}}\gamma - \mathbf{i}_{\mathbf{y}} = 0,\ (\textit{bilinear}) \tag{11.22}$$

which is a problem outside our orbit-of-interest. An example is given in the next chapter. Consult the literature list at the end of this chapter.

11-3 Example: The Straight Line Fit

Our example is based upon the "*straight line fit*"

$$\boxed{\mathbf{y} = a\mathbf{x} + b\mathbf{1}}$$

where $(\mathbf{x}, \mathbf{y})$ has been *measured*, in detail

$$E\{\mathbf{y}\} = \mathrm{a}E\{\mathbf{x}\} + \mathrm{b}\mathbf{1}$$

$$= [E\{\mathbf{x}\},\ \mathbf{1}]\begin{bmatrix} \mathrm{a} \\ \mathrm{b} \end{bmatrix}$$

or

$$\gamma_1 := a, \gamma_2 := b,\ \mathbf{x}\gamma_1 + \mathbf{1}\gamma_2 = [\mathbf{x},\ \mathbf{1}]\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

and

$$\boxed{\mathbf{y} - \mathbf{x}\gamma_1 - \mathbf{1}\gamma_2 + \mathbf{e}'_{\mathbf{x}}\gamma_1 - \mathbf{e}'_{\mathbf{y}} = 0.}$$

$(\gamma_1,\ \gamma_2)$ are the *two unknowns* in the *parameter space*. It has to be noted that the term $e_x\gamma_1$ includes *two coupled unknowns*, namely $\mathbf{e}_{\mathbf{x}}$ and γ_1.

Second, we formulate the *modified method of least squares*.

$$
\begin{aligned}
&\mathcal{L}(\gamma_1,\, \gamma_2,\, \mathbf{e_x},\, \mathbf{e_y},\, \lambda) \\
&= \mathbf{i}'\mathbf{W}\mathbf{i} + 2(\mathbf{y}' - \mathbf{x}'\gamma_1 - \mathbf{1}\gamma_2 + \mathbf{i}'_{\mathbf{x}}\gamma_1 - \mathbf{i}'_{\mathbf{y}})\lambda \\
&= \mathbf{i}'\mathbf{W}\mathbf{i} + 2\lambda'(\mathbf{y} - \mathbf{x}\gamma_1 - \mathbf{1}\gamma_2 + \mathbf{i}'_{\mathbf{x}}\gamma_1 - \mathbf{i}_{\mathbf{y}})
\end{aligned}
$$

or

$$
\begin{aligned}
&\mathbf{i}'_{\mathbf{y}}\mathbf{W_y}\mathbf{i_y} + \mathbf{i}'_{\mathbf{x}}\mathbf{W_x}\mathbf{i_x} \\
&\quad + 2(\mathbf{y}' - \mathbf{x}'\gamma_1 - \mathbf{1}'\gamma_2 + \mathbf{i}'_{\mathbf{x}}\gamma_1 - \mathbf{i}'_{\mathbf{y}})\lambda.
\end{aligned}
$$

Third, we present the *necessary* and *sufficient conditions* for obtaining the minimum of the *modified method of least squares*.

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \gamma_1} = -\mathbf{x}'\boldsymbol{\lambda}_\ell + \mathbf{i}'_{\mathbf{x}}\lambda_\ell = 0 \tag{11.23}$$

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \gamma_2} = -\mathbf{1}'^{\smile}_{\ell} = 0 \tag{11.24}$$

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \mathbf{i_y}} = \mathbf{W_y}\mathbf{i_y} - \lambda_\ell = 0 \tag{11.25}$$

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \mathbf{i_x}} = \mathbf{W_x}\mathbf{i_x} + \lambda_l\gamma_1 = 0 \tag{11.26}$$

$$\frac{1}{2}\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{y} - \mathbf{x}\gamma_1 - \mathbf{1}\gamma_2 + \mathbf{i_x}\gamma - \mathbf{i_y} = 0 \tag{11.27}$$

and

$$\det \begin{bmatrix} \partial^2\mathcal{L}/\partial^2\gamma_1 & \partial^2\mathcal{L}/\partial\gamma_1\partial\gamma_2 \\ \partial^2\mathcal{L}/\partial\gamma_1\partial\gamma_2 & \partial^2\mathcal{L}/\partial^2\gamma_2 \end{bmatrix} \geq 0. \tag{11.28}$$

Indeed, these conditions are *necessary* and *sufficient* for obtaining the minimum of the *modified method of least squares*.

By *Gauss elimination* we receive the results

$$(-\mathbf{x}' + \mathbf{i}'_x)\lambda_\ell = 0 \tag{11.29}$$

$$\lambda_1 + \cdots + \lambda_n = 0 \tag{11.30}$$

$$\mathbf{W_y}\mathbf{i_y} = \lambda_\ell \tag{11.31}$$

$$\mathbf{W_x}\mathbf{i_x} = -\lambda_\ell\gamma_1 \tag{11.32}$$

$$\mathbf{W_y}\mathbf{y} = \mathbf{W_y}\mathbf{x}\gamma_1 + \mathbf{W_y}\mathbf{1}\gamma_2 - \mathbf{W_y}\mathbf{i_x}\gamma_1 + \mathbf{W_y}\mathbf{i_y} \tag{11.33}$$

or

$$\begin{gathered}\mathbf{W_y y} = \mathbf{W_y x}\gamma_1 - \mathbf{W_y 1}\gamma_2 - (\mathbf{I_x} - \gamma_1^2)\lambda_\ell = 0 \\ \text{if } \mathbf{W_y} = \mathbf{W_x} = \mathbf{W}\end{gathered} \tag{11.34}$$

and

$$\mathbf{x}'\mathbf{W}\mathbf{y} - \mathbf{x}'\mathbf{W}\mathbf{x}\gamma_1 - \mathbf{x}'\mathbf{W}\mathbf{1}\gamma_2 - \mathbf{x}'(\mathbf{I_x} - \gamma_1^2)\lambda_\ell = 0 \tag{11.35}$$

$$\mathbf{y}'\mathbf{W}\mathbf{y} - \mathbf{y}'\mathbf{W}\mathbf{x}\gamma_1 - \mathbf{y}'\mathbf{W}\mathbf{1}\gamma_2 - \mathbf{y}'(\mathbf{I_x} - \gamma_1^2)\lambda_\ell = 0 \tag{11.36}$$

$$+\,\mathbf{x}'\boldsymbol{\lambda}_l = +\mathbf{i}'_\mathbf{x}\lambda_\ell \tag{11.37}$$

$$\lambda, +\ldots + \lambda_n = 0 \tag{11.38}$$

$$\Leftrightarrow$$

$$\begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\end{bmatrix}\mathbf{W_y} - \begin{bmatrix}\mathbf{y}'\\ \mathbf{x}'\end{bmatrix}\mathbf{W}\mathbf{x}\gamma_1 - \begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\end{bmatrix}\mathbf{W}\mathbf{1}\gamma_2 - \begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\end{bmatrix}(\mathbf{I_x} - \gamma_1^2)\lambda_\ell = 0 \tag{11.39}$$

subject to

$$\lambda_1 + \cdots + \lambda_n = 0, \tag{11.40}$$

$$\mathbf{x}'_\mathbf{1}\lambda_l = \mathbf{i}'_\mathbf{x}\lambda_l. \tag{11.41}$$

Let us iterate the solution.

$$\left[\begin{array}{cc|cc|c} 0 & 0 & 0 & 0 & (\mathbf{x}'_n - \mathbf{i}'_\mathbf{x}) \\ 0 & 0 & 0 & 0 & \mathbf{1}'_n \\ \hline 0 & 0 & \mathbf{W_x} & 0 & \gamma_1\mathbf{I}_n \\ 0 & 0 & 0 & \mathbf{W_y} & -\mathbf{I}_n \\ \hline \mathbf{x}_n & \mathbf{1}_n & -\gamma_1\mathbf{I}_n & \mathbf{I}_n & 0 \end{array}\right] \left[\begin{array}{c}\gamma_1\\ \gamma_2\\ \hline \mathbf{i_x}\\ \hline \mathbf{i_y}\\ \hline \lambda_\ell\end{array}\right] = \left[\begin{array}{c}0\\ 0\\ \hline 0\\ 0\\ \hline \mathbf{y}_n\end{array}\right].$$

We meet again the problem that the nonlinear terms γ_1 and $\mathbf{i_x}$ appear. Our *iteration* is based on the initial data

$$(i)\ \mathbf{W_x} = \mathbf{W_y} = \mathbf{W} = \mathbf{I}_n,\ (ii)\ (\mathbf{i_x})_0 = 0,\ (iii)\ (\gamma_1)_n = \gamma_1(\mathbf{y} - (\mathbf{i_y})_0 = \mathbf{x}(\gamma_1)_0 + \mathbf{1}_n(\gamma_2)_0)$$

in general

$$\left[\begin{array}{cc|cc|c} 0 & 0 & 0 & 0 & (\mathbf{x}' - \mathbf{i}'_\mathbf{x}) \\ 0 & 0 & 0 & 0 & \mathbf{1} \\ \hline 0 & 0 & \mathbf{W_x} & 0 & \gamma_{1(i)}\mathbf{I}_n \\ 0 & 0 & 0 & \mathbf{W_y} & -\mathbf{I}_n \\ \hline \mathbf{x}_n & \mathbf{1}_n & -\gamma_{1(i)}\mathbf{I}_n & \mathbf{I}_n & 0 \end{array}\right] \left[\begin{array}{c}\gamma_{1(i+1)}\\ \gamma_{2(i+1)}\\ \hline \mathbf{i}_{\mathbf{x}(i+1)}\\ \hline \mathbf{i}_{\mathbf{y}(i+1)}\\ \hline \lambda_{l(i+1)}\end{array}\right] = \left[\begin{array}{c}0\\ 0\\ \hline 0\\ 0\\ \hline \mathbf{y}_n\end{array}\right]$$

$$x_1 = \gamma_1,\ x_2 = \gamma_1,\ x_3 = \mathbf{i_x},\ x_4 = \mathbf{i_y},\ x_5 = \lambda_\ell.$$

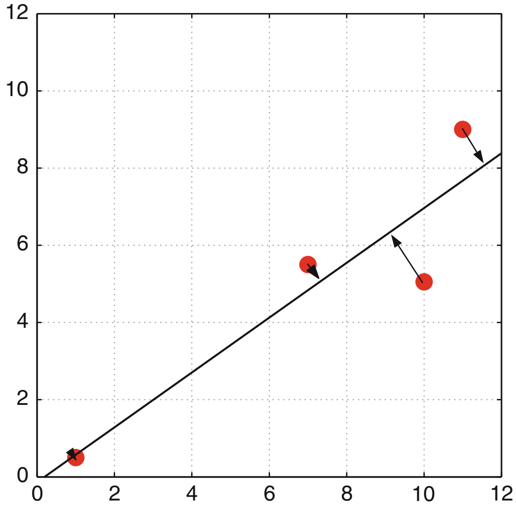


Fig. 11.2 General linear regression

The *five unknowns* have led us to the example within Fig. 11.2. Our solutions are collected as follows:

$$\begin{array}{llll} \gamma_1: & 0.752,6 & 0.866,3 & 0.781,6 \\ \gamma_2: & 0.152,0 & -1.201,6 & -0.244,6 \end{array}$$

$$\begin{array}{lll} \text{case :} & \text{case :} & \text{our general} \\ \sigma_{y(i)} = 0, \sigma_{X(i)} = 0 & & \text{result after} \\ \sigma_{X(i)} \neq 0, \sigma_{y(i)} \neq 0 & & \text{iteration} \end{array}$$

11-4 The Models SIMEX and SYMEX

(*R.J. Carroll, J.R. Cook, L.A.Stefanski, J. Polzehl* and *S. Zwanzig*)

Many authors take advantage of the symmetry of the *implicit model of Total Least Squares* or the simulation of *random condition equation* given before.

Using the alternative formulation of *type (11.26)* by dividing our model equation by minus γ_ℓ and taking care of the *inverse parameter transformation* (11.28). Equation (11.30) is a rewritten explicit model of EVM. We put emphasis on the *"naive regression model"* by means of (11.31), (11.29) and (11.33). Finally we estimate the bias *within naive regression within* (11.34). The related solution of type *"Total Least Squares"* which is maximum likelihood estimation is presented by (11.36) to (11.39). SIMEX and SYMEX is reviewed in special table, enriched by simulated examples and tables.

SIMEX versus SYMEX

Following *J.R. Cook* and *L.A. Stefanski*(1994), *L.A. Stefanski* and *J.R. Cook* (1995) on the one side and *J. Polzehl* and *S. Zwanzig* (2007) on the other side we compare *SIMEX* and *SYMEX*.

SIMEX	SYMEX
First step : Simulation For any values out of the set $\{\lambda_\alpha : \alpha = 0, \ldots, A\}$ starting with λ_0 we simulate samples with increased measurement variance $\mathbf{X}_{\beta ik}(\lambda) := \mathbf{X}_{ik} + \lambda^{1/2}\delta_{ik\beta}$ for all $\beta = 1, \ldots, n_\beta; i = 1, \ldots, n; k = 1, \ldots, K$ with independent identical distributed $\delta_{ik\beta} \sim \mathbf{N}(0, 1)$, replace the new observation matrix $\mathbb{Z}(\lambda)$ using $\mathbb{X}_{ik\beta}(\lambda)$ instead of $\mathbb{X}_{ik}$	*First step : Simulation* For values λ out of the set $\{\lambda_\alpha : \alpha = 0, \ldots, A\}$ starting with λ_0 we simulate samples with increased measurement variance as $\mathbf{Z}_{(-\ell)ik\beta}(\lambda) := \mathbf{Z}_{(-\ell)ik} + \lambda^{1/2}$ $\beta = 1, \ldots, n_\beta; i = 1, \ldots, n; k = 1, \ldots, K$ with independent identical distributed $\delta_{ik\beta} \sim \mathbf{N}(0, 1)$, for instance by adding independent error with variance λ in each component of the resulting matrix. Define the observation matrix $\mathbb{Z}(\lambda)$ as the matrix containing the value $\mathbb{Z}_{ik\beta}(\lambda)$, note that $\mathbb{Z}(\lambda_0)$ contains the n_β-times replicated the original sample.
Second step : Estimation Calculate the *naive regression estimate* $\hat{\gamma}(\lambda) = [\mathbf{Z}_{(-\ell)}(\lambda)'\mathbf{Z}_{(-\ell)}(\lambda)]^{-1}\mathbf{Y}$ for all values of λ	*Second step : Estimation* For all simulated samples, for instance values λ_α, calculate the *naive estimates* $\tilde{\alpha}(\lambda_\alpha)$ in our model by $\tilde{\alpha}^{\ell}_{(-\ell)}(\lambda) = [\mathbf{Z}'_{(-\ell)}(\lambda)\mathbf{Z}_{(-\ell)}(\lambda)]^{-1}\mathbf{Z}'_{(-\ell)}(\lambda)\mathbf{Z}_\ell$ subject to $\tilde{\alpha}^{\ell}_{\ell}(\lambda) = -1$

Third step : Estimation
Fit the model
$\gamma(\lambda,\mu) := (\mathbf{Z}'_{(-\ell)}\mathbf{Z}_{(-\ell)} + \lambda\mathbf{I}_m)^{-1}\mu$
with unknown parameter μ, used to describe the expectation of $\hat{\gamma}(\lambda)$ as a function of λ. Fit this model by *weighted least squares*.
$\hat{\mu} :=$ arg min

$$\sum_{\alpha=1}^{A} w_\alpha \|\hat{\gamma}(\lambda_\alpha) - \gamma(\lambda_\alpha,\mu)\|^2$$

subject to $w_0 \gg 1$ and $w_\alpha = 1$ for all $\alpha > 0$. A large value of w_0 can be used in order to force the model to fit perfectly for the informative estimate for instance $\lambda = 0$

Fourth step : Extrapolation
$\hat{\gamma}(\lambda,\mu)$ is used to *extrapolate* to the case of *zero measurement variance* using $\lambda = \sigma^2$. Such a result leads us to *the SIMEX estimate*
$\hat{\gamma}(\text{SIMEX}) = \gamma(-\sigma^2,\hat{\mu})$

Third step : Estimation
Fit the model
$\alpha^\ell_{(-\ell)}(\lambda,\mu) :=$
$(\mathbf{Z}'_{(-\ell)}\mathbf{Z}_{n(-\ell)} + \lambda\mathbf{I}_m)^{-1}\mu$
subject to
$\alpha^\ell_\ell(\lambda,\mu) := -1$
with parameter $\mu \in \mathbb{R}^m$. The estimate of μ is given by
$\hat{\mu}_\ell :=$ arg min

$$\sum_{\alpha=0}^{A} w_\alpha \|\hat{\alpha}^\ell(\lambda_\alpha) - \alpha^\ell(\lambda_\alpha,\mu)\|^2$$

subject to $w_0 \gg 1$ and $w_\alpha = 1$ for all $\alpha > 0$. Using our inverse transformation, we define
$\hat{\gamma}^k_\ell(\lambda,\hat{\mu}_\ell) := \alpha^k_\ell(\lambda,\hat{\mu}_\ell)/\alpha^0_\ell(\lambda,\hat{\mu}_\ell)$
for all $k = 1,\ldots,m$, λ arbitrary

Fourth step : Extrapolation
Determine an optimal value λ^* for *extrapolation* to the case of *zero measurement variance* of type

$$\lambda^* := \underset{\lambda \in (\lambda_{\min},0)}{\arg\min}$$

$$\sum_{\ell=0}^{m}\sum_{m=\ell+1}^{m} \|\hat{\gamma}^\ell(\lambda) - \hat{\gamma}^m(\lambda)\|^2$$

under the constraint

$$\lambda_{\min} = \min_\lambda\{\lambda \in \mathbb{R} :$$
$$\tfrac{1}{n}\mathbf{Z}'_{(-\ell)}\mathbf{Z}_{(-\ell)} + \lambda\mathbf{I}_m) > 0\}$$
for all $\ell = 0,\ldots,m$

Fifth step : SYMEX estimation
Estimate γ by

$$\hat{\gamma}(\text{SIMEX}) = \hat{\gamma}(\lambda^*) = \frac{1}{m+1}\sum_{\ell=0}^{m}\hat{\gamma}^{\ell}(\lambda^*)$$

The key idea of *SIMEX* as well as of *SYMEX* is to include the *extrapolation* generalized to the case of zero variance in using $\lambda = \sigma^2$ in the *extrapolation step*. The value $-\lambda := \sigma^2$ can be interpreted as an estimate of the measurement error variance, In addition, *J.Polzehl* and *S.Zwanzig* (2007) give a simple relation for *SYMEX* to calculate the naive estimator $\hat{\gamma}(\text{naive}), \mu, \lambda$ formulated in a *special theorem*. After all they conclude that *SYMEX*, in general, has *smaller root-mean-square errors* and *less outliers* when compared to *SIMEX*. This can be due to an approximation to the

optimal TLS estimate.

In all simulations *SYMEX* was not worse than *SIMEX*, namely with an *improvement being largest for larger values* γ or σ^2.

SIMULATION

All our simulations are of size 250 as presented by *J.Polzehl* and *S.Zwanzig* (2007). The grid of the quantities λ used $\lambda_\alpha = 0.2\alpha\sigma^2$ with $A = 10$. We generate samples of increased measurement variance with $n = 100$. The parameter space built on value γ and σ^2 are collected in Tables 11.3 up to 11.5.

Example 11.1. (one parameter $\gamma_1 = \gamma$)

$$E\{\mathbf{y}_i\} = \gamma E\{\mathbf{x}_i\}, \mathbf{Y}_i = E\{\mathbf{Y}_i\} + \mathbf{e}_y, \mathbf{X}_i = E\{\mathbf{X}_i\} + \mathbf{E}_\mathbf{X}$$

subject to

$$E\{\mathbf{e}_y\} = 0, E\{\mathbf{E}_\mathbf{X}\} = 0, D\{\mathbf{e}_y\} = \sigma^2, D\{\mathbf{E}_\mathbf{X}\} = \sigma^2$$

$$(i = 1, \ldots, n)$$

Design details

$$E\{\mathbf{X}_i\} \in \{-1, +1\}, \mathbf{P}(\mathbf{X}_i = -1) = \mathbf{P}\{\mathbf{X}_i = +1\} = 0.5$$

sample size: n=20

Example 11.2. (two parameter γ_1 and γ_2)

$$E\{\mathbf{y}_i\} = \gamma_1 E\{\mathbf{x}_{i1}\} + \gamma_2 E\{\mathbf{x}_{i2}\}, \mathbf{Y}_i = E\{\mathbf{Y}_i\} + \mathbf{e}_y, \mathbf{X}_{i1} = E\{\mathbf{X}_{i1}\} + \mathbf{E}_{\mathbf{X_{i1}}}$$

$$\mathbf{X}_{i2} = E\{\mathbf{X}_{i2}\} + \mathbf{E}_{\mathbf{X_{i2}}}$$

subject to

$$E\{\mathbf{e}_y\} = 0, E\{\mathbf{X}_{i1}\} = E\{\mathbf{X}_{i2}\} = 0, D\{\mathbf{e}_y\} = \sigma^2, D\{\mathbf{E}_{ik}\} = \delta_{ik}\sigma^2$$

for all $i \in \{1, \ldots, n\}, k \in \{1, 2\}$

Design details

$$E\{\mathbf{X}_{ik}\} \in \{-1, +1\}, \mathbf{P}\{\mathbf{X}_{ik} = -1\} = \mathbf{P}\{\mathbf{X}_{ik} = +1\} = 0.5$$

sample size: n = 25

Example 11.3. (two parameters γ_1 and γ_2)

We use the second example, but we vary the design towards a two point design in terms of $\mathbf{X}_{ik} \sim \mathbf{N}(0, \mathbf{I}_2)$

Table 11.5 Median absolute error, bias and standard deviation of estimates for simulation in Example 11.1 (m = 1)

σ	γ	MAE			bias			sd		
		Naive	SIMEX	SYMEX	Naive	SIMEX	SYMEX	Naive	SIMEX	SYMEX
0.125	2	0.0453	0.0411	0.0403	−0.025	0.0065	0.007	0.042	0.041	0.04
0.25	2	0.117	0.0847	0.0814	−0.11	0.012	0.011	0.088	0.087	0.082
0.5	2	0.389	0.19	0.173	−0.39	0.014	0.01	0.18	0.23	0.18
0.125	1	0.0285	0.027	0.0258	−0.015	−5.6e-05	0.0015	0.026	0.026	0.026
0.25	1	0.0702	0.0611	0.053	−0.059	−1.6e-05	−0.00091	0.055	0.055	0.051
0.5	1	0.195	0.137	0.116	−0.19	0.013	−0.0033	0.11	0.14	0.11
0.125	0.5	0.0211	0.0222	0.0214	−0.0058	0.0019	0.00029	0.02	0.02	0.02
0.25	0.5	0.0493	0.0467	0.0459	−0.025	0.0037	−0.0014	0.042	0.042	0.041
0.5	0.5	0.111	0.1	0.0965	−0.095	0.011	−0.0067	0.084	0.1	0.092

Table 11.6 Median absolute error of estimates for simulation in Example 11.2 (m = 2)

σ	γ_1				γ_2			
	true	Naive	SIMEX	SYMEX	true	Naive	SIMEX	SYMEX
0.125	2	0.059	0.0502	0.0481	2	0.0534	0.0534	0.0522
0.25	2	0.14	0.104	0.101	2	0.128	0.108	0.106
0.5	2	0.417	0.241	0.215	2	0.386	0.263	0.201
0.125	1	0.0382	0.0343	0.0319	1	0.0409	0.0353	0.0348
0.25	1	0.0761	0.0655	0.06	1	0.0719	0.0626	0.0573
0.5	1	0.216	0.147	0.137	1	0.185	0.155	0.125
0.125	0.5	0.0193	0.0211	0.021	0.5	0.0214	0.0224	0.0224
0.25	0.5	0.0443	0.0425	0.0408	0.5	0.0429	0.0453	0.0428
0.5	0.5	0.12	0.103	0.0991	0.5	0.0979	0.103	0.082
0.125	0.5	0.045	0.038	0.0369	0.5	0.0417	0.0406	0.0416
0.25	0.5	0.1343	0.0866	0.0763	0.5	0.0772	0.0854	0.0835
0.5	0.5	0.398	0.21	0.158	0.5	0.146	0.187	0.173

Table 11.7 Median absolute error of estimates for simulation in Example 11.2 (m = 2)

σ	γ_1				γ_2			
	true	Naive	SIMEX	SYMEX	true	Naive	SIMEX	SYMEX
0.125	2	0.0492	0.0411	0.0412	2	0.0399	0.0333	0.00316
0.25	2	0.132	0.0874	0.0847	2	0.108	0.0713	0.0627
0.5	2	0.402	0.195	0.178	2	0.372	0.168	0.137
0.125	1	0.027	0.0235	0.0227	1	0.0207	0.0194	0.0175
0.25	1	0.0688	0.0476	0.0467	1	0.0554	0.0403	0.0369
0.5	1	0.205	0.105	0.0994	1	0.189	0.0966	0.0753
0.125	0.5	0.0174	0.0158	0.0161	0.5	0.0143	0.0129	0.013
0.25	0.5	0.0403	0.0331	0.0314	0.5	0.0346	0.0275	0.0255
0.5	0.5	0.105	0.0711	0.0688	0.5	0.0983	0.0681	0.0567
0.125	0.5	0.0393	0.0289	0.0305	0.5	0.024	0.0239	0.024
0.25	0.5	0.123	0.0596	0.0579	0.5	0.0526	0.0499	0.0491
0.5	0.5	0.407	0.143	0.121	0.5	0.111	0.112	0.107

sample size: n = 25

Tables 11.4-11.4 provide the results of simulations in terms of median absolute error for the naive regression estimates. The SIMEX estimate, for instance using extrapolation with $\lambda = -\sigma^2$, and SYMEX estimate. In addition, estimated bias and standard deviation are given

J. Polzehl and *S. Zwanzig* (2003)

11-5 References

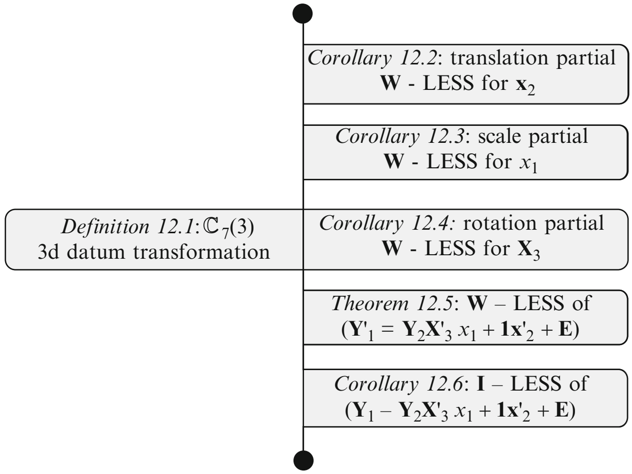
Please, contact the following references.

Abatzoglou, J.T. and Mendel, J.M. (1987), Abatzoglou, J.T. and Mendel, J.M. and Harada, G.A. (1991), Bajen, M.T., Puchal, T., Gonzales, A., Gringo, J.M., Castelao, A., Mora, J. and Comin, M. (1997), Berry, S.M., Carroll, R.J. and Ruppert, D. (2002), Bjrck, A. (1996), Bjrck, A. (1997), Bjrck, A., Elfving, T. and Strakos, Z. (1998), Bjrck, A., Heggernes, P. and Matstoms, P. (2000), Bojanczyk, A.W., Bront, R.P., Van Dooren, P., de Hoog, F.R. (1987), Bunch, J.R., Nielsen, C.P. and Sorensen, D.C. (1978), Bunke, H. and Bunke, O. (1989), Cap-derou, A., Douguet, D., Simislowski, T., Aurengo, A. and Zelter, M. (1997), Car-roll, R.J., Ruppert, D. and Stefanski, L.A. (1996), Carroll, R.J., Kschenhoff, H., Lombard, F. and Stefanski, L.A. (1996), Carroll, R.J. and Stefanski, L.A. (1997), Chandrasekuran, S. and Sayed, A.H. (1996), Chun, J. Kailath, T. and Lev-Ari, H. (1987), Cook, J.R. and Stefanski, L.A. (1994), Dembo, R.S., Eisenstat, S.C. and Steihaug, T. (1982), De Moor, B. (1994), Fuller, W.A. (1987), Golub, G.H. and Van Loan, C.F. (1980), Hansen, P.C. (1998), Higham, N.J. (1996), Holcomb, J. P. (1996), Humak, K.M.S. (1983), Kamm, J. and Nagy, J.G. (1998), Kailath, T. Kung, S. and More, M. (1979), Kailath, T. and Chun, J. (1994), Kailath, T. and Sayed, A.H.(1995), Kleming, J.S. and Goddard, B.A. (1974), Kung, S.Y., Arun, K.S. and Braskar Rao, D.V. (1983), Lemmerling, P., Van Huffel, S. and de Moor, B. (1997), Lemmerling, P., Dologlou, I. and Van Huffel, S. (1998), Lin, X. and Carroll, R.J. (1999), Lin, X. and Carroll, R.J. (2000), Mackens, W. and Voss, H. (1997), Mastronardi, N, Van Dooren, P. and Van Huffel, S., Mastronardi, N., Lemmerling, P. and Van Huffel, S. (2000), Nagy, J.G. (1993), Park, H. and El-den, L. (1997), Pedroso, J.J. (1996), Polzehl, J. and Zwanzig, S. (2003), Rosen, J.B., Park, H. and Glick, J. (1996), Stefanski, L.A. and Cook, J.R. (1995), Stew-art, M. and Van Dooren, P. (1997), Van Huffel, S. (1991), Van Huffel, S. and Vandewalle, J. (1991), Van Huffel, S., Decanniere, C., Chen, H. and Van Hecke, P.V. (1994), Van Huffel, S., Park, H. and Rosen, J.B. (1996), Van Huffel, S., Vandewalle, J., de Rao, M.C. and Willems, J.L.,Wang, N., Lin, X., Gutierrez, R.G. and Carroll, R.J. (1998), and Yang, T. (1996), Berry, S.M., Carroll, R.J. and D. Ruppert (2002), Bunke, H. and O. Bunke (1989), Carrol, R.J., Ruppert, D. and L.A. Stefanski (1996), Carroll, R.J., et al (1996), Carroll, R.J. and L.A. Stefanski (1997), Cook, J.R. and L.A. Stefanski, Fuller W.A (1987), Hollomb, J.R. (1999), Lin, X. and J. R. Cook (1999), Stefanski, L.A. and J.R. Cook (1995), Lin, X. and R.J. Carroll (2000), Wang et al. (1998)

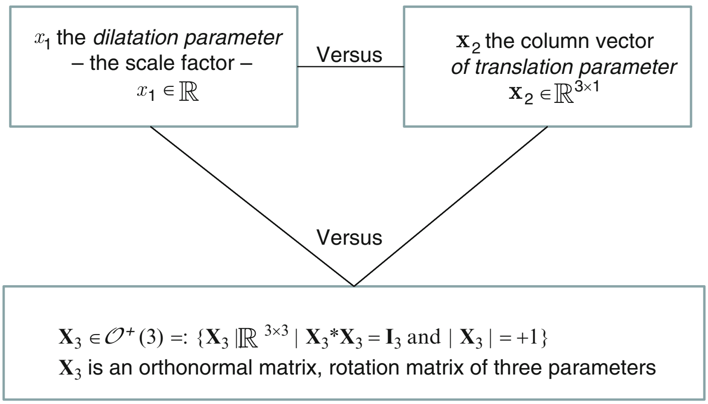
Chapter 12
The Nonlinear Problem of the 3d Datum Transformation and the Procrustes Algorithm

A special *nonlinear problem* is the *three-dimensional datum transformation* solved by the *Procrustes Algorithm*. A definition of the three-dimensional datum transformation with the *coupled unknowns* of type *dilatation unknown*, also called *scale factor*, *translation* and *rotation unknown* follows afterwards.

Fast track reading: Read *Definition 12.1*, *Corollary 12.2-12.4, 12.6* and *Lemma 12.5* and *Lemma 12.7*.



Let us specify the parameter space $\mathbb{X}$, namely



E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_12,
© Springer-Verlag Berlin Heidelberg 2012

which is built on the *scalar* x_1, the *vector* $\mathbf{x}_2$ and the *matrix* $\mathbf{X}_3$. In addition, by the matrices

$$\mathbf{Y}_1 := \begin{bmatrix} x_1 & x_2 & \dots & x_{n-1} & x_n \\ y_1 & y_2 & \dots & y_{n-1} & y_n \\ z_1 & z_2 & \dots & z_{n-1} & z_n \end{bmatrix} \in \mathbb{R}^{3\times n} \; and \; \mathbf{Y}_2 := \begin{bmatrix} X_1 & X_2 & \dots & X_{n-1} & X_n \\ Y_1 & Y_2 & \dots & Y_{n-1} & Y_n \\ Z_1 & Z_2 & \dots & Z_{n-1} & Z_n \end{bmatrix} \in \mathbb{R}^{3\times n}$$

we define a left and right *three-dimensional coordinate arrays* as an n-dimensional simplex of *observed data*. Our aim is

to determine the parameters of the three-dimensional datum transformation $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$ *out of a nonlinear transformation* (*conformal group* $\mathbb{C}_7(3)$). $\mathbf{x}_1 \in \mathbb{R}$ *stands for the dilatation parameter, also called scale factor*, $\mathbf{x}_2 \in \mathbb{R}^{3\times 1}$ denotes the column vector of translation parameters, and $\mathbf{X}_3 \in \mathcal{O}^+(3) =: \{\mathbf{X}_3 \in \mathbb{R}^{3\times 3} | \mathbf{X}_3'\mathbf{X}_3 = \mathbf{I}_3, |\mathbf{X}_3| = +1\}$ the orthonormal matrix, also called rotation matrix of three parameters.

The *key problem* is

how to determine the parameters for the unknowns of type $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$, namely scalar dilatation x_1, vector of translation and matrix of rotation, for instance by weighted least squares.

Example 12.1. (simplex of minimal dimension, $n = 4$ points tetrahedron):

$$\mathbf{Y}_1 := \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix}' \Leftrightarrow \begin{bmatrix} X_1 & X_2 & X_3 & X_4 \\ Y_1 & Y_2 & Y_3 & Y_4 \\ Z_1 & Z_2 & Z_3 & Z_4 \end{bmatrix}' =: \mathbf{Y}$$

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \\ X_4 & Y_4 & Z_4 \end{bmatrix} \mathbf{X}_3 x_1 + \mathbf{1x}_2' + \begin{bmatrix} \mathbf{e}_{11} & \mathbf{e}_{12} & \mathbf{e}_{13} \\ \mathbf{e}_{21} & \mathbf{e}_{22} & \mathbf{e}_{23} \\ \mathbf{e}_{31} & \mathbf{e}_{32} & \mathbf{e}_{33} \\ \mathbf{e}_{41} & \mathbf{e}_{42} & \mathbf{e}_{43} \end{bmatrix}.$$

Example 12.2. (W-LESS)

We depart from the setup of the pseudo-observation equation given in Example 12.1 (simplex of minimal dimension, $n = 4$ points, tetrahedron). For a diagonal weight $\mathbf{W} = \mathrm{Diag}(w_1, \dots, w_4) \in \mathbb{R}^{4\times 4}$ we compute the *Frobenius error matrix* $\mathbf{W}$-seminorm

$$||\mathbf{E}||^2_{\mathbf{W}} := \mathrm{tr}(\mathbf{E}'\mathbf{W}\mathbf{E}) = \mathrm{tr}\left(\begin{bmatrix} \mathbf{e}_{11} & \mathbf{e}_{21} & \mathbf{e}_{31} & \mathbf{e}_{41} \\ \mathbf{e}_{12} & \mathbf{e}_{22} & \mathbf{e}_{32} & \mathbf{e}_{42} \\ \mathbf{e}_{13} & \mathbf{e}_{23} & \mathbf{e}_{33} & \mathbf{e}_{43} \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & w_3 & 0 \\ 0 & 0 & 0 & w_4 \end{bmatrix} \times \begin{bmatrix} \mathbf{e}_{11} & \mathbf{e}_{12} & \mathbf{e}_{13} \\ \mathbf{e}_{21} & \mathbf{e}_{22} & \mathbf{e}_{23} \\ \mathbf{e}_{31} & \mathbf{e}_{32} & \mathbf{e}_{33} \\ \mathbf{e}_{41} & \mathbf{e}_{42} & \mathbf{e}_{43} \end{bmatrix} \right)$$

$$= \mathrm{tr}\left(\begin{bmatrix} \mathbf{e}_{11}w_1 & \mathbf{e}_{21}w_2 & \mathbf{e}_{31}w_3 & \mathbf{e}_{41}w_4 \\ \mathbf{e}_{12}w_1 & \mathbf{e}_{22}w_2 & \mathbf{e}_{32}w_3 & \mathbf{e}_{42}w_4 \\ \mathbf{e}_{13}w_1 & \mathbf{e}_{23}w_2 & \mathbf{e}_{33}w_3 & \mathbf{e}_{43}w_4 \end{bmatrix} \times \begin{bmatrix} \mathbf{e}_{11} & \mathbf{e}_{12} & \mathbf{e}_{13} \\ \mathbf{e}_{21} & \mathbf{e}_{22} & \mathbf{e}_{23} \\ \mathbf{e}_{31} & \mathbf{e}_{32} & \mathbf{e}_{33} \\ \mathbf{e}_{41} & \mathbf{e}_{42} & \mathbf{e}_{43} \end{bmatrix}\right)$$

$$\begin{aligned} &= w_1\mathbf{e}_{11}^2 + w_2\mathbf{e}_{21}^2 + w_3\mathbf{e}_{31}^2 + w_4\mathbf{e}_{41}^2 + w_1\mathbf{e}_{12}^2 + w_2\mathbf{e}_{22}^2 + \\ &+ w_3\mathbf{e}_{32}^2 + w_4\mathbf{e}_{42}^2 + w_1\mathbf{e}_{13}^2 + w_2\mathbf{e}_{23}^2 + w_3\mathbf{e}_{33}^2 + w_4\mathbf{e}_{43}^2 \,. \end{aligned}$$

Obviously, the coordinate errors $(\mathbf{e}_{11}, \mathbf{e}_{12}, \mathbf{e}_{13})$ have the same weight w_1, $(\mathbf{e}_{21}, \mathbf{e}_{22}, \mathbf{e}_{23})$ have the same weight w_2, $(\mathbf{e}_{31}, \mathbf{e}_{32}, \mathbf{e}_{33})$ have the same weight w_3, and finally $(\mathbf{e}_{41}, \mathbf{e}_{42}, \mathbf{e}_{43})$ have the same weight w_4. We may also say that the error weight is *pointwise isotropic*,

$$\text{weight } \mathbf{e}_{11} = \text{weight } \mathbf{e}_{12} = \text{weight } \mathbf{e}_{13} = w_1$$

etc. However, the error weight is *not homogeneous* since

$$w_1 = \text{weight } \mathbf{e}_{11} \neq \text{weight } \mathbf{e}_{21} = w_2.$$

Of course, an ideal *homogeneous and isotropic* weight distribution is guaranteed by the criterion $w_1 = w_2 = w_3 = w_4$.

12-1 The 3d Datum Transformation and the Procrustes Algorithm

First, we present **W**-LESS for our nonlinear adjustment problem for the unknowns of type *scalar, vector and special orthonormal matrix*. Second, we review the Procrustes Algorithm for the parameters $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$.

Definition 12.1. (nonlinear analysis for the three-dimensional datum transformation: the conformal group $\mathbb{C}_7(3)$):

The parameter array $\{x_{1\ell}, \mathbf{x}_{2\ell}, \mathbf{X}_{3\ell}\}$ is called **W**-LESS (LEast Squares Solution with respect to the **W** - Seminorm) of the inconsistent linear system of equations

$$\mathbf{Y}_2\mathbf{X}'_3 x_1 + \mathbf{1}\mathbf{x}'_2 + \mathbf{E} = \mathbf{Y}_1 \tag{12.1}$$

subject to

$$\mathbf{X}'_3\mathbf{X}_3 = \mathbf{I}_3, |\mathbf{X}_3| = +1 \tag{12.2}$$

of the field of parameters in comparison with alternative parameter arrays $\{x_{1\ell}, \mathbf{x}_{2\ell}, \mathbf{X}_{3\ell}\}$ fulfils the inequality equation

$$||\mathbf{Y}_1 - \mathbf{Y_2X'}_{3\ell} x_{1\ell} - \mathbf{1x'}_{2\ell}||^2_{\mathbf{W}}$$

$$:= \mathrm{tr}((\mathbf{Y}_1 - \mathbf{Y_2X'}_{3\ell} x_{1\ell} - \mathbf{1x'}_{2\ell})'\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2X'}_{3\ell} x_{1\ell} - \mathbf{1x'}_{2\ell})) \leq$$
$$=: \mathbf{tr}((\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1 - \mathbf{1x'}_2)'\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1 - \mathbf{1x'}_2))$$
$$=: ||\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1 - \mathbf{1x'}_2||^2_{\mathbf{W}} \tag{12.3}$$

in other words if

$$\mathbf{E}_\ell := \mathbf{Y}_1 - \mathbf{Y_2X'}_{3\ell} x_{1\ell} - \mathbf{1x'}_{2\ell} \tag{12.4}$$

has the least **W** - seminorm.

? How to compute the three unknowns $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$ by means of **W**-LESS ?

Here we will outline the computation of the parameter vector by means of partial **W**-LESS: At first, by means of **W**-LESS we determine $\mathbf{x}_{2\ell}$, secondly by means of **W**-LESS $x_{1\ell}$, followed by thirdly means of **W**-LESS $\mathbf{X}_3$. In total, we outline the *Procrustes Algorithm*.

Step one: $\mathbf{x}_2$

Corollary 12.2. *(partial* **W***-LESS for* $\mathbf{x}_{2\ell}$ *):*

A 3×1 vector $\mathbf{x}_{2\ell}$ is partial **W**-LESS of (12.1) subject to (12.2) if and only if $\mathbf{x}_{2\ell}$ fulfils the system of normal equations

$$\mathbf{1'W1x}_{2\ell} = (\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1)'\mathbf{W1}. \tag{12.5}$$

$\mathbf{x}_{2\ell}$ always exists and is represented by

$$\mathbf{x}_{2\ell} = (\mathbf{1'W1})^{-1}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X'}_3 x_1)'\mathbf{W1}. \tag{12.6}$$

For the special case $\mathbf{W} = \mathbf{I}_n$ the translated parameter vector $\mathbf{x}_{2\ell}$ is given by

$$\mathbf{x}_{2\ell} = \frac{1}{n}(\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1)'\mathbf{1}. \tag{12.7}$$

For the proof, we shall *first* minimize the *risk function*

$$(\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1 - \mathbf{1x'}_2)'(\mathbf{Y}_1 - \mathbf{Y_2X'}_3 x_1 - \mathbf{1x'}_2) = \min_{\mathbf{x}_2}$$

with respect to $\mathbf{x}_2$!

:Detailed Proof of Corollary 12.2:

W-LESS is constructed by the *unconstrained Lagrangean*

$$\begin{aligned}
L(x_1, \mathbf{x}_2, \mathbf{X}_3) &:= \frac{1}{2}||\mathbf{E}||^2_{\mathbf{W}} = ||\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1 - \mathbf{1}\mathbf{x}_2'||^2_{\mathbf{W}} \\
&= \frac{1}{2}\mathrm{tr}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1 - \mathbf{1}\mathbf{x}_2')'\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1 - \mathbf{1}\mathbf{x}_2') \\
&= \min_{x_1 \geq 0, \mathbf{x}_2 \in R^{3\times 1}, \mathbf{X}_3'\mathbf{X}_3 = \mathbf{I}_3} \\
\frac{\partial L}{\partial \mathbf{x}_2'}(\mathbf{x}_{2\ell}) &= (\mathbf{1}'\mathbf{W}\mathbf{1})\mathbf{x}_2 - (\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1)'\mathbf{W}\mathbf{1} = \mathbf{0}
\end{aligned}$$

constitutes the first necessary condition. Basics of the vector-valued differentials are found in *E. Grafarend and B. Schaffrin* (1993, pp. 439–451). As soon as we backward substitute the translational parameter $\mathbf{x}_{2\ell}$, we are led to the *centralized Lagrangean*

$$L(x_1, \mathbf{X}_3) = \frac{1}{2}\mathrm{tr}\{[\mathbf{Y}_1 - (\mathbf{Y_2}\mathbf{X}_3' x_1 + (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}'\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1))]'\mathbf{W}*$$

$$*[\mathbf{Y}_1 - (\mathbf{Y_2}\mathbf{X}_3' x_1 + (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}'\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1))]\}$$

$$L(x_1, \mathbf{X}_3) = \frac{1}{2}\mathrm{tr}\{[(\mathbf{I} - (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}')\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1)]'\mathbf{W}*$$

$$*[(\mathbf{I} - (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}')\mathbf{W}(\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1)]\}$$

$$C := \mathbf{I}_n - \frac{1}{2}\mathbf{1}\mathbf{1}'$$

being a definition of the *centering matrix*, namely for $\mathbf{W} = \mathbf{I}_n$

$$\boxed{C := \mathbf{I}_n - (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}'\mathbf{W}} \tag{12.8}$$

being in general symmetric. Substituting the *centering matrix* into the *reduced Lagrangean* $L(x_1, \mathbf{X}_3)$, we gain the *centralized Lagrangean*

$$\boxed{L(x_1, \mathbf{X}_3) = \frac{1}{2}\mathrm{tr}\{[\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1]'\mathbf{C}'\mathbf{W}\mathbf{C}[\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1]\}.} \tag{12.9}$$

Step two: x_1

Corollary 12.3. *(partial* **W**-*LESS for* $x_{1\ell}$*):*

A scalar $x_{1\ell}$ is partial **W**-LESS of (12.1) subject to (12.3) if and only if

$$\mathbf{x}_{1\ell} = \frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2} \tag{12.10}$$

holds. For the special case $\mathbf{W} = \mathbf{I}_n$ the real parameter is given by

$$\mathbf{x}_{1\ell} = \frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{C}\mathbf{Y}_2}. \tag{12.11}$$

The general condition is subject to

$$\mathbf{C} := \mathbf{I}_n - (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}'\mathbf{W}. \tag{12.12}$$

:Detailed Proof of *Corollary 12.3*:

For the proof we shall newly minimize the risk function

$$L(x_1, \mathbf{X}_3) = \frac{1}{2}\mathrm{tr}\{[\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1]'\mathbf{C}'\mathbf{W}\mathbf{C}[\mathbf{Y}_1 - \mathbf{Y_2}\mathbf{X}_3' x_1]\} = \min_{x_1}$$

subject to

$$\mathbf{X}_3'\mathbf{X}_3 = \mathbf{I}_3.$$

$$\frac{\partial L}{\partial x_1}(x_{1\ell}) = x_{1\ell}\mathrm{tr}\mathbf{X}_3\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3' - \mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3' = \mathbf{0}$$

constitutes the second necessary condition. Due to

$$\mathrm{tr}\mathbf{X}_3\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3' = \mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'\mathbf{X}_3 = \mathbf{Y_2}'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2$$

lead us to $x_{1\ell}$. While the forward computation of$(\partial L/\partial x_1)(x_{1\ell}) = 0$ enjoyed a representation of the optimal scale parameter $x_{1\ell}$, its backward substitution into the Lagrangean $L(x_1, \mathbf{X}_3)$ amounts to

$$L(\mathbf{X}_3) = \mathrm{tr}\left\{\left[\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_3'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}Y_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2}\right]\mathbf{C}'\mathbf{W}\mathbf{C}*\left[\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_3'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}Y_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2}\right]\right\}$$

$$\begin{aligned}
L(\mathbf{X}_3) = \frac{1}{2}\mathrm{tr}\Bigg\{&(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_1) - \mathrm{tr}(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3') * \frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2} - \\
&-\mathrm{tr}(\mathbf{X}_3\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_1)\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2} + \\
&+\mathrm{tr}(\mathbf{X}_3\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3')\frac{[\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3']^2}{[\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2]^2}
\end{aligned}$$

$$L(\mathbf{X}_3) = \frac{1}{2}\mathrm{tr}(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_1) - \frac{[\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3']^2}{[\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2]} + \frac{1}{2}\frac{[\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3']^2}{[\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2]}$$

$$\boxed{L(\mathbf{X}_3) = \frac{1}{2}\mathrm{tr}(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_1) - \frac{1}{2}\frac{[\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3']^2}{[\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2]} = \min_{\mathbf{X}_3'\mathbf{X}_3=\mathbf{I}_3}.} \tag{12.13}$$

Third, we are left with the proof for the *Corollary 12.4*, namely $\mathbf{X}_3$.

Step three: $\mathbf{X}_3$

Corollary 12.4. *(partial* **W***-LESS for* $\mathbf{X}_{3\ell}$ *):*

A 3×1 orthonormal matrix $\mathbf{X}_3$ is partial **W**-LESS of (12.1) subject to (12.3) if and only if

$$\mathbf{X}_{3\ell} = \mathbf{U}\mathbf{V}' \tag{12.14}$$

holds where $\mathbf{A} := \mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2 = \mathbf{U}\boldsymbol{\Sigma}_s\mathbf{V}'$ is a *singular value decomposition* with respect to a left orthonormal matrix $\mathbf{U},\mathbf{U}'\mathbf{U} = \mathbf{I}_3$, a right orthonormal matrix $\mathbf{V},\mathbf{V}\mathbf{V}' = \mathbf{I}_3$, and $\boldsymbol{\Sigma}_s = \mathrm{Diag}(\sigma_1,\sigma_2,\sigma_3)$ a diagonal matrix of singular values $(\sigma_1,\sigma_2,\sigma_3)$. The singular values are the canonical coordinates of the *right eigenspace* $(\mathbf{A}'\mathbf{A} - \boldsymbol{\Sigma}_s^2\mathbf{I})\mathbf{V} = 0$. The *left eigenspace* is based upon $\mathbf{U} = \mathbf{A}\mathbf{V}\boldsymbol{\Sigma}_s^{-1}$.

:Detailed Proof of *Corollary 12.4*:

The form $L(\mathbf{X}_3)$ subject to $\mathbf{X}_3'\mathbf{X}_3 = \mathbf{I}_3$ is minimal if

$$\mathrm{tr}(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_3') = \max_{x_1\geq 0,\mathbf{X}_3'\mathbf{X}_3=\mathbf{I}_3}.$$

Let $\mathbf{A} := \mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2 = \mathbf{U}\boldsymbol{\Sigma}V'$, a singular value decomposition with respect to a *left orthonormal matrix* $\mathbf{U},\mathbf{U}'\mathbf{U} = \mathbf{I}_3$, a right orthonormal matrix $\mathbf{V},\mathbf{V}\mathbf{V}' = \mathbf{I}_3$ and $\boldsymbol{\Sigma}_s = Diag(\sigma_1,\sigma_2,\sigma_3)$ a *diagonal matrix* of singular values $(\sigma_1,\sigma_2,\sigma_3)$. Then

$$\mathrm{tr}(\mathbf{A}\mathbf{X}_3') = \mathrm{tr}(\mathbf{U}\boldsymbol{\Sigma}_s\mathbf{V}'\mathbf{X}_3') = \mathrm{tr}(\boldsymbol{\Sigma}_s\mathbf{V}'\mathbf{X}_3\mathbf{U}) = \sum_{i=1}^{3}\sigma_i r_{ii} \leq \sum_{i=1}^{3}\sigma_i$$

holds, since

$$\boxed{\mathbf{R} = \mathbf{V}'\mathbf{X}_3'\mathbf{U} = [r_{ij}] \in \mathbb{R}^{3\times 3}} \tag{12.15}$$

is orthonormal with $||r_{ii}|| \leq 1$. The identity $tr(\mathbf{A}\mathbf{X}_3') = \sum_{i=1}^{3}\sigma_i$ applies, if

$$\mathbf{V}'\mathbf{X}_3'\mathbf{U} = \mathbf{I}_3, \text{i.e.}\mathbf{X}_3' = \mathbf{V}\mathbf{U}', \mathbf{X}_3 = \mathbf{U}\mathbf{V}',$$

namely, if $\mathrm{tr}(\mathbf{A}\mathbf{X}_3')$ is maximal

$$\boxed{\operatorname{tr}(\mathbf{A}\mathbf{X}_3') = \max_{\mathbf{X}_3'\mathbf{X}_3=\mathbf{I}_3} \Leftrightarrow \operatorname{tr}\mathbf{A}\mathbf{X}_3' = \sum_{i=1}^{3}\sigma_i \Leftrightarrow \mathbf{R} = \mathbf{V}'\mathbf{X}_3'\mathbf{U} = \mathbf{I}_3}. \tag{12.16}$$

An alternative proof of *Corollary 12.4* based on formal differentiation of *traces and determinants* has been given by *P.H. Schönemann* (1966) and *P.H. Schönemann and R.M. Carroll* (1970). Finally, we collect our sequential results in *Theorem 12.5* identifying the stationary point of **W**-LESS specialized for $\mathbf{W} = \mathbf{I}$ in *Corollary 12.5*. The highlight is the *Procrustes Algorithm* we review in *Table 12.1*.

Theorem 12.5. (W-LESS of $\mathbf{Y}_1' = \mathbf{Y}_2\mathbf{X}_3'x_1 + \mathbf{1}\mathbf{x}_2' + \mathbf{E}$):

(i) The parameter array $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$ is **W**-LESS if

$$x_{1\ell} = \frac{\operatorname{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{X}_{3\ell}'}{\operatorname{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2} \tag{12.17}$$

$$\mathbf{x}_{2\ell} = (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_{3\ell}'x_{1\ell})'\mathbf{W}\mathbf{1} \tag{12.18}$$

$$\mathbf{X}_3 = \mathbf{U}\mathbf{V}' \tag{12.19}$$

subject to the *singular value decomposition of the general* 3×3 *matrix*

$$\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2 = \mathbf{U}\operatorname{Diag}(\sigma_1, \sigma_2, \sigma_3)\mathbf{V}' \tag{12.20}$$

namely

$$[(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2)'(\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2) - \sigma_i\mathbf{I}]\mathrm{v}_i = \mathbf{0} \tag{12.21}$$

$$\mathbf{V} = [\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3], \mathbf{V}\mathbf{V}' = \mathbf{I}_3 \tag{12.22}$$

$$\mathbf{U} = \mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{V}\operatorname{Diag}(\sigma_1^{-1}, \sigma_2^{-1}, \sigma_3^{-1}), \tag{12.23}$$

$$\mathbf{U}'\mathbf{U} = \mathbf{I}_3 \tag{12.24}$$

and as well as the *centering matrix*

$$\mathbf{C} := \mathbf{I}_n - (\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}\mathbf{1}\mathbf{1}'\mathbf{W}. \tag{12.25}$$

(ii) The empirical error matrix of type **W**-LESS accounts for

$$\mathbf{E}_\ell = [\mathbf{I}_n - \mathbf{1}\mathbf{1}'\mathbf{W}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}]\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\operatorname{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\operatorname{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2}\right) \tag{12.26}$$

with the related *Frobenius matrix* **W**-*seminorm*

$$||\mathbf{E}_\ell||_{\mathbf{W}}^{\mathbf{2}} = \operatorname{tr}(\mathbf{E}_\ell'\mathbf{W}\mathbf{E}_\ell) = \operatorname{tr}\{\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\operatorname{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\operatorname{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2}\right)' *$$

$$*[\mathbf{I}_n - \mathbf{1}\mathbf{1}'\mathbf{W}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}]'\mathbf{W}[\mathbf{I}_n - \mathbf{1}\mathbf{1}'\mathbf{W}(\mathbf{1}'\mathbf{W}\mathbf{1})^{-1}]*$$

$$*\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}'\mathbf{W}\mathbf{C}\mathbf{Y}_2}\right)\Big\} \tag{12.27}$$

and the representative scalar measure of the error of type **W**-LESS

$$||\mathbf{E}_\ell||_{\mathbf{W}} = \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{W}\mathbf{E}_\ell)/3n}\,. \tag{12.28}$$

A special result is obtained if we specialize *Theorem 12.5* to the case $\mathbf{W} = \mathbf{I}_n$:

Corollary 12.6. *(I-LESS of* $\mathbf{Y}_1' = \mathbf{Y}_2\mathbf{X}_3'x_1 + \mathbf{1}\mathbf{x}_2' + \mathbf{E}$*):*

(i) The *parameter array* $\{x_1, \mathbf{x}_2, \mathbf{X}_3\}$ is **I**-LESS of $\mathbf{Y}_1' = \mathbf{Y}_2\mathbf{X}_3'x_1 + \mathbf{1}\mathbf{x}_2' + \mathbf{E}$ if

$$x_{1\ell} = \frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{X}_{3\ell}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2} \tag{12.29}$$

$$\mathbf{x}_{2\ell} = \frac{1}{n}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_{3\ell}'x_{1\ell})'\mathbf{1} \tag{12.30}$$

$$\mathbf{X}_{3\ell} = \mathbf{U}\mathbf{V}' \tag{12.31}$$

subject to the singular value decomposition of the general 3×3 matrix

$$\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2 = \mathbf{U}\mathrm{Diag}(\sigma_1, \sigma_2, \sigma_3)\mathbf{V}' \tag{12.32}$$

namely

$$[(\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2)'(\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2) - \sigma_i^2]\mathbf{I}\mathrm{v}_i = \mathbf{0}, \forall i \in \{1,2,3\}, \mathbf{V} = [\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3], \mathbf{V}\mathbf{V}' = \mathbf{I}_3 \tag{12.33}$$

$$\mathbf{U} = \mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathrm{Diag}(\sigma_1^{-1}, \sigma_2^{-1}, \sigma_3^{-1}), \mathbf{U}\mathbf{U}' = \mathbf{I}_3 \tag{12.34}$$

and as well as the *centering matrix*

$$\mathbf{C} := \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'. \tag{12.35}$$

(ii) *The empirical error matrix of type* **I**- *LESS accounts for*

$$\mathbf{E}_\ell = \left[\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right]\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}'\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2}\right) \tag{12.36}$$

with the related *Frobenius matrix* **W**-*seminorm*

$$||\mathbf{E}||_{\mathbf{I}}^{\mathbf{2}} = \mathrm{tr}(\mathbf{E}_\ell'\mathbf{E}_\ell) = \mathrm{tr}\{\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2}\right)' *$$

$$* \left[\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right]\left(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\frac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2}\right)\Bigg\} \tag{12.37}$$

and the representative scalar measure of the error of type I-LESS

$$||\mathbf{E}_\ell||_\mathbf{I} = \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{E}_\ell)/3n}. \tag{12.38}$$

In the proof of *Corollary 12.6* we only sketch the result that the matrix $\mathbf{I}_n - (1/n)\mathbf{1}\mathbf{1}'$ is idempotent:

$$\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = \mathbf{I}_n - \frac{2}{n}\mathbf{1}\mathbf{1}' + \frac{1}{n^2}(\mathbf{1}\mathbf{1}')^2$$

$$= \mathbf{I}_n - \frac{2}{n}\mathbf{1}\mathbf{1}' + \frac{1}{n^2}n\mathbf{1}\mathbf{1}' = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'.$$

As a summary of the various steps of *Corollary 12.2–12.4*, *12.5* and *Theorem 12.5*, *Table 12.1* presents us the celebrated *Procrustes Algorithm*, which is followed by one short and interesting Citation about "Procrustes".

Following *Table 12.1*, we present the celebrated *Procrustes Algorithm* which is a summary of the various steps of *Corollary 12.2–12.4, 12.5 and Theorem 12.5*.

Procrustes (the subduer), son of Poseidon, kept an inn benefiting from what he claimed to be a wonderful all-fitting bed. He lopped of excessive limbage from tall guests and either flattened short guests by hammering or stretched them by racking. The victim fitted the bed perfectly but, regrettably, died. To exclude the embarrassment of an initially exact-fitting guest, variants of the legend allow Procrustes two, different-sized beds. Ultimately, in a crackdown on robbers and monsters, the young Theseus fitted Procrustes to his own bed.

12-2 The Variance: Covariance Matrix of the Error Matrix E

By *Lemma 12.7* we review the variance–covariance matrix, namely the vector valued form of the transposed error matrix, as a function of

$$\boxed{\begin{array}{c}\boldsymbol{\Sigma}_{vec\mathbf{Y}_1}, \boldsymbol{\Sigma}_{vec\mathbf{Y}_2} \text{ and} \\ \text{the covariance matrix} \\ \boldsymbol{\Sigma}_{vec\mathbf{Y}_1', (\mathbf{I}_n \otimes x_1\mathbf{X}_3) vec\boldsymbol{\Sigma}_{\mathbf{Y}_2'}}.\end{array}}$$

Lemma 12.7. (Variance–covariance "error propagation"):

Let $vec\mathbf{E}'$ be the vector valued form of the transposed error matrix $\mathbf{E} := \mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_3'x_1 - \mathbf{1}\mathbf{x}_2'$. Then

Table 12.1 (Procrustes Algorithm):

Step 1: Read $\mathbf{Y}_1 = \begin{bmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix}$ and $\begin{bmatrix} X_1 & Y_1 & Z_1 \\ \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n \end{bmatrix} = \mathbf{Y}_2$

Step 2: Compute: $\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2 \text{subjectto} \mathbf{C} := \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$

Step 3: Compute: $SVD\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2 = \mathbf{U}\, Diag(\sigma_1, \sigma_2, \sigma_3)\mathbf{V}'$

3-1 $|(\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2)'(\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2) - \sigma_i^2\mathbf{I}| = \mathbf{0} \Rightarrow (\sigma_1, \sigma_2, \sigma_3)$

3-2 $((\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2)'(\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2) - \sigma_i^2\mathbf{I})\mathrm{v}_i = \mathbf{0}, \forall i \in \{1, 2, 3\}$

$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$right eigenvectors (right eigencolumns)

3-3 $\mathbf{U} = \mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{V}Diag(\sigma_1^{-1}, \sigma_2^{-1}, \sigma_3^{-1})$ left eigenvectors (left eigencolumns)

Step 4: Compute: $\mathbf{X}_{3\ell} = \mathbf{U}\mathbf{V}'$ rotation

Step 5: Compute: $x_{1\ell} = \dfrac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{X}_3'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2}$ (dilatation)

Step 6: Compute: $\mathbf{x}_{2\ell} = \dfrac{1}{n}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{X}_3'x_1)'\mathbf{1}$ (translation)

Step 7: Compute: $\mathbf{E}_\ell = \mathbf{C}(\mathbf{Y}_1 - \mathbf{Y}_2\mathbf{V}\mathbf{U}'\dfrac{\mathrm{tr}\mathbf{Y}_1'\mathbf{C}\mathbf{Y}_2\mathbf{V}\mathbf{U}'}{\mathrm{tr}\mathbf{Y}_2'\mathbf{C}\mathbf{Y}_2})$ (error matrix)

'optional control' $\mathbf{E}_\ell := \mathbf{Y}_1 - (\mathbf{Y}_2\mathbf{X}_3'x_{1\ell} + \mathbf{1}\mathbf{x}_{2\ell}')$

Step 8: Compute: $||\mathbf{E}_\ell||_\mathbf{I} := \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{E}_\ell)}$ (error matrix)

Step 9: Compute: $||\mathbf{E}_\ell||_\mathbf{I} := \sqrt{\mathrm{tr}(\mathbf{E}'_\ell\mathbf{E}_\ell)/3n}$ (mean error matrix)

$$\mathbf{\Sigma}_{vec\mathbf{E}'} = \mathbf{\Sigma}_{vec\mathbf{Y}_1'} + (\mathbf{I}_n \otimes x_1\mathbf{X}_3)\mathbf{\Sigma}_{vec\mathbf{Y}_2'}(\mathbf{I}_n \otimes x_1\mathbf{X}_3)' - 2\mathbf{\Sigma}_{vec\mathbf{Y}_1', (\mathbf{I}_n \otimes x_1\mathbf{X}_3)vec\mathbf{Y}_2'} \quad (12.39)$$

is the exact representation of the dispersion matrix (variance–covariance matrix) $\mathbf{\Sigma}_{vec\mathbf{E}'}$ of vec$\mathbf{E}'$ in terms of dispersion matrices (variance–covariance matrices) $\mathbf{\Sigma}_{vec\mathbf{Y}_1'}$ and $\mathbf{\Sigma}_{vec\mathbf{Y}_2'}$ of the two coordinates sets vec$\mathbf{Y}_1'$ and vec$\mathbf{Y}_2'$ as well as their covariance matrix

$$\mathbf{\Sigma}_{vec\mathbf{Y}_1', (\mathbf{I}_n \otimes \mathbf{X}_3)vec\mathbf{Y}_2'}.$$

The proof follows directly from "*error propagation*". Obviously the variance–covariance matrix of $\mathbf{\Sigma}_{vec\mathbf{E}'}$ can be decomposed in the variance–covariance matrix $\mathbf{\Sigma}_{vec\mathbf{Y}_1'}$, the product $(\mathbf{I}_n \otimes x_1\mathbf{X}_3)\mathbf{\Sigma}_{vec\mathbf{Y}_2'}(\mathbf{I}_n \otimes \mathbf{X}_3)'$ using prior information of x_1 and $\mathbf{X}_3$ and the covariance matrix $\mathbf{\Sigma}_{vec\mathbf{Y}_1', (\mathbf{I}_n \otimes x_1\mathbf{X}_3)vec\mathbf{Y}_2'}$ again using prior information of x_1 and $\mathbf{X}_3$.

12-21 Case Studies: The 3d Datum Transformation and the Procrustes Algorithm

By *Tables 12.1* and *12.2* we present two sets of coordinates, *first* for the local system A, *second* for the global system B, also called "*World Geodetic System 84*". The units are in meter. The results of

Table 12.2 Coordinates for system A (local system)

Station Name	$X(m)$	$Y(m)$	$Z(m)$	Positional error sphere
Solitude	4157222.543	664789.307	4774952.099	0.1433
Buoch Zeil	4149043.336	688836.443	4778632.188	0.1551
Hohenneuffen	4172803.511	690340.078	4758129.701	0.1503
Kuehlenberg	4177148.376	642997.635	4760764.800	0.1400
Ex Mergelaec	4137012.190	671808.029	4791128.215	0.1459
Ex Hof Asperg	4146292.729	666952.887	4783859.856	0.1469
Ex Kaisersbach	4138759.902	702670.738	4785552.196	0.1220

Table 12.3 Coordinates for system B (WGS 84)

Station Name	$X(m)$	$Y(m)$	$Z(m)$	Positional error sphere
Solitude	4157870.237	664818.678	4775416.524	0.0103
Buoch Zeil	4149691.049	688865.785	4779096.588	0.0038
Hohenneuffen	4173451.354	690369.375	4758594.075	0.0006
Kuehlenberg	4177796.064	643026.700	4761228.899	0.0114
Ex Mergelaec	4137659.549	671837.337	4791592.531	0.0068
Ex Hof Asperg	4146940.228	666982.151	4784324.099	0.0002
Ex Kaisersbach	4139407.506	702700.227	4786016.645	0.0041

I-*LESS, Procrustes Algorithm*

are listed in *Table 12.3*, especially

$$||\mathbf{E}_\ell||_\mathbf{I} := \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{E}_\ell)}, |||\mathbf{E}_\ell|||_\mathbf{I} := \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{E}_\ell)/3n}$$

and

W-*LESS, Procrustes Algorithm*

in *Table 12.4*, specially

$$||\mathbf{E}_\ell||_\mathbf{W} := \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{W}\mathbf{E}_\ell)}, |||\mathbf{E}_\ell|||_\mathbf{W} := \sqrt{\mathrm{tr}(\mathbf{E}_\ell'\mathbf{W}\mathbf{E}_\ell)/3n}$$

completed by *Table 12.5* of residuals from the *Linearized Least Squares* and by *Table 12.6* listing the *weight matrix*.

Discussion

By means of the *Procrustes Algorithm* which is based upon **W**-LESS with respect to Frobenius matrix W-norm we have succeeded to solve the normal equations of *Corollary 12.2 and 12.4* (necessary conditions) of the matrix-valued "error equations"

Table 12.4 Results of the **I**-LESS Procrustes transformation

	Values			
Rotation Matrix	0.999999999979023	$-4.33275933098276e-6$	4.81462518486797e-6	
$\mathbf{X}_3 \in \mathbb{R}^{3\times 3}$	$-4.8146461589238e-6$	0.999999999976693	$-4.84085332591588e-6$	
	4.33273602401529e-6	4.84087418647916e-6	0.999999999978896	
Translation	641.8804			
$\mathbf{x}_2 \in \mathbb{R}^{3\times 1}(m)$	68.6553			
	416.3982			
Scale $x_1 \in \mathbb{R}$	1.00000558251985			
Residual matrix $\mathbf{E}(m)$	Site	$X(m)$	$Y(m)$	$Z(m)$
	Solitude	0.0940	0.1351	0.1402
	Buoch Zeil	0.0588	-0.0497	0.0137
	Hohenneuffen	-0.0399	-0.0879	-0.0081
	Kuelenberg	0.0202	-0.0220	-0.0874
	Ex Mergelaec	-0.0919	0.0139	-0.0055
	Ex Hof Asperg	-0.0118	0.0065	-0.0546
	Ex Keisersbach	-0.0294	0.0041	0.0017
Error matrix norm *(m)*				
$\|\| E_l \|\|_{\mathrm{W}} := \sqrt{tr(\mathbf{E}_l^* \mathbf{E}_l)}$	0.2890			
Mean error matrix norm *(m)*				
$\|\| E_l \|\|_{\mathrm{W}} := \sqrt{tr(\mathbf{E}_l^* \mathbf{E}_l)/3n}$	0.0631			

Table 12.5 Results of the **W**-LESS Procrustes transformation

	Values			
Rotation Matrix	0.999999999979141	4.77975830372179e-6	$-4.34410139438235e-6$	
$\mathbf{X}_3 \in \mathbb{R}^{3\times 3}$	$-4.77977931759299e-6$	0.999999999976877	$-4.83729276438971e-6$	
	4.34407827309968e-6	4.83731352815542e-6	0.999999999978865	
Translation	641.8377			
$\mathbf{x}_2 \in \mathbb{R}^{3\times 1}(m)$	68.4743			
	416.2159			
Scale $x_1 \in \mathbb{R}$	1.00000561120732			
Residual matrix $\mathbf{E}(m)$	Site	$X(m)$	$Y(m)$	$Z(m)$
	Solitude	0.0948	0.1352	0.1407
	Buoch Zeil	0.0608	-0.0500	0.0143
	Hohenneuffen	-0.0388	-0.0891	-0.0072
	Kuelenberg	0.0195	-0.0219	-0.0868
	Ex Mergelaec	-0.0900	0.0144	-0.0052
	Ex Hof Asperg	-0.0105	0.0067	-0.0542
	Ex Keisersbach	-0.0266	0.0036	0.0022
Error matrix norm *(m)*				
$\|\| E_l \|\|_{\mathrm{W}} := \sqrt{tr(\mathbf{E}_l^* \mathbf{W} \mathbf{E}_l)}$	0.4268			
Mean error matrix norm *(m)*				
$\|\| E_l \|\|_{\mathrm{W}} := \sqrt{tr(\mathbf{E}_l^* \mathbf{W} \mathbf{E}_l)/3n}$	0.0930			

$$\boxed{\begin{gathered}\mathrm{vec}\mathbf{E}' = \mathrm{vec}\mathbf{Y}_1' - (\mathbf{I}_n \otimes x_1 \mathbf{X}_3)\mathrm{vec}\mathbf{Y}_2' - \mathrm{vec}\mathbf{x}_2 \mathbf{1}' \textit{ subject to} \\ \mathbf{X}_3' \mathbf{X}_3 = \mathbf{I}_3, |\mathbf{X}_3| = +1.\end{gathered}}$$

The scalar - valued unknown $x_1 \in \mathbb{R}$ represented *dilatation* (scale factor), the vector - valued unknown $x_2 \in \mathbb{R}^{3\times 1}$ the *translation vector*, and the matrix - valued

Table 12.6 Residuals from the linearized LS solution

Site	$X(m)$	$Y(m)$	$Z(m)$
Solitude	0.0940	0.1351	0.1402
Buoch Zeil	0.0588	−0.0497	0.0137
Hohenneuffen	−0.0399	−0.0879	−0.0081
Kuelenberg	0.0202	−0.0220	−0.0874
Ex Mergelaec	−0.0919	0.0139	−0.0055
Ex Hof Asperg	−0.0118	0.0065	−0.0546
Ex Keisersbach	−0.0294	0.0041	−0.0017

Table 12.7 Weight matrix

$$\mathbf{W} = \begin{bmatrix} 1.8110817 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.1843373 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.1145291 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.9918578 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.6288452 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2.1642460 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2.359370 \end{bmatrix},$$

unknown $\mathbf{X}_3 \in SO(3)$ the *orthonormal matrix*. The conditions of sufficiency, namely the *Hesse matrix* of second derivatives, of the Lagrangean $L(x_1, \mathbf{x}_2, \mathbf{X}_3)$ are not discussed here. They are given in the Procrustes references. In order to present you with a proper choice of the isotropic weight matrix $\mathbf{W}$, we introduced the corresponding "*random regression model*" (Table 12.7)

$$\mathbf{E}\{\text{vec}\mathbf{E}'\} = \mathbf{E}\{\text{vec}\mathbf{Y}_1'\} - (\mathbf{I}_n \otimes x_1\mathbf{X}_3)\mathbf{E}\{\text{vec}\mathbf{Y}_2'\} - \text{vec}\mathbf{x}_2\mathbf{1}' = \mathbf{0}$$

first moment identity,

$$\mathbf{D}\{\text{vec}\mathbf{E}'\} = \mathbf{D}\{\text{vec}\mathbf{Y}_1'\} - (\mathbf{I}_n \otimes x_1\mathbf{X}_3)\mathbf{D}\{\text{vec}\mathbf{Y}_2'\}(\mathbf{I}_n \otimes x_1\mathbf{X}_3)' - \\ -2\mathbf{C}\{\text{vec}\mathbf{Y}_1', (\mathbf{I}_n \otimes x_1\mathbf{X}_3)\text{vec}\mathbf{Y}_2'\},$$

second central moment identity.

12-3 References

Here is a list of important references: *Awange LJ* (1999), *Awange LJ* (2002), *Awange LJ, Grafarend E* (2001 a, b, c), *Awange LJ, Grafarend E* (2002), *Bernhardt T* (2000), *Bingham C, Chang T, Richards D* (1992), *Borg I, Groenen P* (1997), *Brokken FB* (1983), *Chang T, Ko DJ* (1995), *Chu MT, Driessel R* (1990), *Chu MT, Trendafilov NT* (1998), *Crosilla F* (1983a, b), *Dryden IL* (1998), *Francesco D, Mathien PP, Senechal D* (1997), *Golub GH* (1987), *Goodall C* (1991), *Gower JC* (1975), *Grafarend E and Awange LJ* (2000, 2003), *Grafarend E, Schaffrin B* (1993), *Grafarend E, Knickmeyer EH, Schaffrin B* (1982), *Green B* (1952), *Gulliksson M*(1995a, b), *Kent JT, Mardia KV* (1997), *Koch KR* (2001), *Krarup T* (1979), *Lenzmann E, Lenzmann L* (2001a, b), *Mardia K* (1978), *Mathar R* (1997), *Mathias R* (1993), *Mooijaart A, Commandeur JJF* (1990), *Preparata FP, Shamos MI* (1985), *Reinking J* (2001), *Schönemann PH* (1966),*Schönemann PH, Carroll RM* (1970), *Schottenloher M* (1997), *Ten Berge JMF* (1977), *Teunissen PJG* (1988), *Trefethen LN, Bau D* (1997) and *Voigt C* (1998).

Chapter 13
The Sixth Problem of Generalized Algebraic Regression

the system of conditional equations with unknowns - (Gauss–Helmert model)

C.F. Gauss *and* F.R. Helmert *introduced the generalized algebraic regression problem which can be identified as a system of conditional equations with unknowns. Here we present variance-covariance component estimation of Helmert type in the* Gauss–Helmert model.

Here we follow E.Grafarend: "Variance-covariance-component estimation of HELMERT type in the Gauss-Helmert model" in detail. In the first section we carefully define the Gauss-Helmert model of condition equations with unknown parameters in a linear model. The second section is E.Grafarend's model of variance-covariance-component estimation of HELMERT type within linear models of inhomogeous condition equations: $\mathbf{B}\,\varepsilon = \mathbf{By} - \mathbf{c}$, $\mathbf{By}$ not an element of $\mathrm{R}(\mathbf{A}) + \mathbf{c}$. In contrast, section three is E.Grafarend's model of variance-covariance-component estimation of HELMERT type within the linear model of inhomogeneous condition equations with unknown parameters, namely within the linear model model $\mathbf{Ax} + \varepsilon + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$, $\mathbf{By}$ not an element of $\mathrm{R}(\mathbf{A}) + \mathbf{c}$.

Finally we discuss in section four the block structure of the dispersion matrix D(y) and give two detailed examples for the variance-covariance estimation within the Gauss-Helmert model of Helmert type.

In order to improve an adjustment result in the *Gauss–Helmert* model (condition equations with unknowns: $\mathbf{Ax} + \mathbf{Be} = \mathbf{By} - \mathbf{c}$) the stepwise procedure of estimation of variance-covariance components of the observation vector y originating from *F.R. Helmert* for the Gauss–Markov model is developed. The basic result is the construction of a local Helmert-type inhomogeneous, invariant, quadratic and unbiased estimator of variance-covariance components. First results are presented for (*i*) a parameter matrix $\mathbf{A}$ of full column rank, (*ii*) a condition matrix $\mathbf{B}$ of full row rank and (*iii*) a positive-definite dispersion matrix of the observations. Two example are given.

F.R. Helmert (1924, pp. 258–267) introduced a stepwise procedure for the estimation of variance-covariance components in the *Gauss–Markov model of parameter adjustment* which will be extended here for *variance-covariance-component estimation* in the model of *condition adjustment* in the *general model of condition equations with unknown parameters* or the *Gauss–Helmert model* according to *H. Wolf* (1978). In contrast to other estimation techniques which are reviewed by *B. Schaffrin* (1983) the Helmert method has the following advantages:

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_13,
© Springer-Verlag Berlin Heidelberg 2012

In the *first step* unknowns and inconsistencies of the observations with respect to the model ("measurement errors") are determined by the *classical model of weighted least-squares* assuming a priori variances and covariances of the observations. In the *second step* the a priori values of the observational variances are corrected by solving the famous *Helmert system of linear equations*. In case that the corrected variance-covariance-component differ significantly from the a priori ones the estimation process can be started again, now with the new variances and covariances of the observations. The stepwise procedure can come to an end when a priori and a posteriori variance-covariance-component coincide, a situation we call *"reproducing"*.

There has been the argument the argument that *"variance-covariance-component estimation"* has no practical impact on the adjustment result. But referring to our own experience the contrary is true. From an adjustment of heterogeneously observed three-dimensional networks we found improvements of adjustment results through the stepwise Helmert procedure of more than a hundred percent.

E.Grafarend's contribution on "variance-covariance-component estimation of HELMERT type in the Gauss–Helmert model" (Zeitschrift für Vermessungswesen 109 (1984) 34–44), namely condition equations with unknowns, is organized in the following way: The first paragraph reviews the classical variance-covariance-component estimation of Helmert type of the linear Gauss–Markov model of parameter adjustment. A first extension for the model of linear inhomogeneous condition equations between the observations is given in paragraph two. As a notable result we mention the character of an inhomogeneous, quadratic unbiased of the generalized Helmert estimator for this model, in general. Paragraph three is devoted to variance-component of Helmert type in the linear Gauss–Helmert model, also called "inhomogeneous, inconsistent condition equations with unknown parameter". A local Helmert-type invariant, inhomogeneous, quadratic an unbiased estimator is constructed. Finally, the *fourth paragraph* is an introduction into the block structure of the variance-covariance matrix of the observations.

Our first results of variance-covariance-component estimation with respect to the model $\mathbf{Ax} + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$ are restricted to (*i*) a parameter matrix $\mathbf{A}$ of full column rank, (*ii*) a condition matrix B of full row rank and (*iii*) a positive-definite variance-covariance matrix of the observations. More general results will be given elsewhere. Here we have generalized the variance-covariance-component estimation of Helmert Type of Gauss–Markov model given by *E. Grafarend, A. Kleusberg* and *B. Schaffrin* (1980), first results for condition adjustment without parameter unknowns of *L. Sjöberg* (1983a,b) and four condition adjustment with unknowns of *C.G Person* (1982). Finally we mention a peculiar use of our variance-covariance-component estimation technique, the estimation of variance-covariance functions of *"signals"* and *"noise"* within *"collocation"*, e.g. in the version of *H. Wolf* (1977).

13-1 Variance-Covariance-Component Estimation in the Linear Model $\mathbf{Ax} + \boldsymbol{\varepsilon} = \mathbf{y}, \mathbf{y} \notin \mathcal{R}(\mathbf{A})$

According to E.Grafarend (1984): Variance-covariance-component estimation of Helmert type in the Gauss-Helmert model, Zeitschrift fuer Vermessungswesen 109 (1984) pp. 30–40 we follow his text: we assume that the system of linear equations $\mathbf{Ax} + \varepsilon = \mathbf{y}$ is formulated by (13.1) and (13.2). The $n \times m$ *fixed* matrix $\mathbf{A}$ of full column rank $r(\mathbf{A}) = m$ connects the $m \times 1$ *fixed* unknown vector $\mathbf{x}$ of parameters and the $n \times 1$ *stochastic* vector $\mathbf{y}$ of observations. The $n \times 1$ *stochastic* vector ε of *inconsistency* accounts for the fact that the observation vector y does not belong to the *range* $\mathcal{R}(\mathbf{A})$ of the matrix $\mathbf{A}$ spanned by its columns. ε is chosen in such a way that $\mathbf{y} - \varepsilon$ is an element of the column space of $\mathbf{A}$. Within second order statistics the linear model $\mathbf{Ax} + \varepsilon = \mathbf{y}$ is characterized by the first and second order moments of the vector y of observations.

$$E(\mathbf{y}) = \sum_{l=1}^{m} \mathbf{a}_i \mu_i \tag{13.1}$$

$$D(\mathbf{y}) = \sum_{j=1}^{l} \mathbf{C}_{jj}\sigma_j^2 + \sum_{\substack{j=1 \\ j<k}}^{l-1} \sum_{k=2}^{l} C_{jk}\sigma jk = \sum_{j=1}^{l(l+1)/2} \mathbf{C}_j \sigma_j \tag{13.2}$$

The first order design matrix $\mathbf{A} := [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ of order $o(\mathbf{A}) = n \times m$ and the symmetric *second order design* matrix blocks $\mathbf{C}_{jj}, \mathbf{C}_{jk}$ or order $n \times m$ or $\mathbf{C} := [\mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{22}, \ldots, \mathbf{C}_{l-1l}\mathbf{C}_{ll}]$ of order $o(\mathbf{C}) = n \times nl(l+1)/2$ such that $D(\mathbf{y})$ is positive definite are a priori given. For more details about block structure we refer to *Appendix*. The first moments μ_j, the l second moments σ_j^2 of type variance and the $l(l-1)/2$ second moments σ_{jk} of type covariance are unknown. They are collected in the $m \times 1$ vector $\mu := [\mu_1, \ldots, \mu_m]$, or $\mathbf{x}_1$ and in the $l(l+1)/2 \times 1$ vector $\sigma := [\sigma_1^2, \sigma_{12}, \sigma_2^2, \ldots, \sigma_{l,l-1}, \sigma_l^2]$ or $\mathbf{x} - 2$.

The idea of stepwise estimation of variance-covariance-component of F.R. Helmert is the following one:

The initial step

In the first step of (μ, σ) or $(\mathbf{x}_1, \mathbf{x}_2)$ estimation process we depart from an initial variance-covariance matrix Σ_0 of the dispersion matrix $\mathbf{D}(y) = \sigma$ of the observation vector $\mathbf{y}$, or equivalently, from an initial set σ_0 of variance-covariance-components. Using the initial dispersion matrix Σ_0, the unknown parameter vector μ or $\mathbf{x_1} = \mathbf{x_2}$ is estimated by *"weighted least-squares"*.

(i) Langarange-function

$$\mathbf{L}(\varepsilon) := \varepsilon' \Sigma_0^{-1} \varepsilon = (\mathbf{y} - \mathbf{Ax})' \Sigma_0^{-1} (\mathbf{y} - \mathbf{Ax}) = \min \tag{13.3}$$

$$\frac{\partial L}{\partial \varepsilon}(\varepsilon_l) = 2\Sigma_0^{-1}\varepsilon\varepsilon_l = 0 \tag{13.4}$$

$$\frac{\partial L}{\partial \mathbf{x}}(\mathbf{x}_l) = 2(\mathbf{A}'\Sigma_0^{-1}\mathbf{A}\mathbf{x}_1 - \mathbf{A}\Sigma_0^{-1}\mathbf{y}) = 0 \tag{13.5}$$

$$\frac{\partial^2 L}{\partial \varepsilon\varepsilon'}(\varepsilon_l) = 2\Sigma_0^{-1} \geq 0 \tag{13.6}$$

The first derivatives of the Lagrange-function constitute the *necessary* condition, the second derivatives the *sufficiency* condition. Since we have assumed a positive definite dispersion matrix Σ_0 (13.25) is always fulfilled. (Here it would have been sufficient to have started from a positive-semidefinite weight matrix, but this generalization will be treated elsewhere.) Thus we arrive at the classical least-square solution

$$\mathbf{x}_1 = (\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}\mathbf{y} \tag{13.7}$$

(ii) The invariance of the estimation

Let us make use of the estimation $\mathbf{x}_1$ in order to compute the stochastic vector ε_l of inconsistency.

$$\varepsilon_l = \mathbf{y} - \mathbf{A}\mathbf{x}_1 = [\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}A)^{-1}\mathbf{A}'\Sigma_0^{-1}]\mathbf{y} = \mathbf{D}_0\mathbf{y} \tag{13.8}$$

$$\mathbf{D}_0 := \mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1} \tag{13.9}$$

Lemma 13.1. Let $\varepsilon_l = \mathbf{D}_0 y$ be the weighted least-squares estimation of the stochastic vector ε of inconsistency. ε_l is invariant with respect to the transformation $\mathbf{y} \to \mathbf{y} - \mathbf{A}\mathbf{x} = \varepsilon$ such that

$$\varepsilon_l = \mathbf{D}_0\varepsilon \tag{13.10}$$

holds. $\mathbf{D}_0$ *is an idempotent matrix.*

For the proof replace $\mathbf{y}$ by $\mathbf{y} - \mathbf{A}\mathbf{x}$ within (13.10).

$$[\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}](\mathbf{y} - \mathbf{A}\mathbf{x})$$
$$[\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}]\mathbf{y} - [\mathbf{A} - [\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}]]\mathbf{x}$$
$$[\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}]\mathbf{y}\text{“q.e.d”}$$

Next we compute

$$\mathbf{D}_0^2 = [\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}][\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}]$$
$$\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}$$
$$+\mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}\mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}$$
$$\mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1} = \mathbf{D}_0$$

which is by definition the postulate of idempotence.

The underlying ideas of the variance-covariance-component estimation techniques of *Helmert type* are as following. At first we compute $E\{\varepsilon_l'\Sigma_0^{-1}\varepsilon_l\}$ which can be expressed via (13.10). $\varepsilon_l = \mathbf{D}_0\varepsilon$ in the terms of $E(\varepsilon\varepsilon') = \Sigma$. Secondly we partition Σ_0^{-1} and Σ into consistent blocks. Once we combine those procedure we gain a linear system of equations from which an unbiased estimation σ_H of variance-covariance-components, based on the Σ_0^{-1} weighted least-square estimation ε_l can be constructed. We give the steps of calculations in details, now.

$$E\{\varepsilon_l'\Sigma_0^{-1}\varepsilon_l\} = E\{\varepsilon'\mathbf{D}_0'\Sigma_0^{-1}\mathbf{D}_0\varepsilon\} = \mathrm{tr}D_0'\Sigma_0^{-1}\mathbf{D}_0(\varepsilon\varepsilon') = \mathrm{tr}\mathbf{D}_0'\Sigma_0^{-1}\mathbf{D}_0\Sigma \tag{13.11}$$

$$\Sigma = \sum_{j=1}^{l(l+1)/2} \mathbf{C}_j\sigma_j, \Sigma 0^{-1} = \sum_{j=1}^{l(l+1)/2} \mathbf{E}_j(\sigma_0) \tag{13.12}$$

$$\sum_{i=1}^{l(l+1)/2} E\{\varepsilon_l' E_i(\sigma_0)\varepsilon_l\} = \sum_{i=1}^{l(l+1)/2}\sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{D}_0'\mathrm{e}_i(\sigma_0)\mathbf{D}_0\mathbf{C}_j\sigma_j \tag{13.13}$$

$$E\{\varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l\} = \sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{D}_0'\mathbf{E}_i(\sigma_0)\mathbf{D}_0\mathbf{C}_j\sigma_j \qquad (i = 1,\dots,l(l+1)/2) \tag{13.14}$$

$$E(q_i) = \sum_{i=1}^{l(l+1)/2} H_{ij}\sigma_j \qquad (i = 1,\dots,l(l+1)/2) \tag{13.15}$$

$$q_i := \varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l = \mathbf{y}'\mathbf{D}_0'\mathbf{E}_i(\sigma_0)\mathbf{D}_0\mathbf{y} \qquad (i = 1,\dots,l(l+1)/2) \tag{13.16}$$

$$H_{ij} := \mathtt{tr}\mathbf{D}_0'\mathbf{E}_i(\sigma_0)\mathbf{D}_0\mathbf{C}_j \qquad (i, J = 1,\dots,l(l+1)/2) \tag{13.17}$$

$H_{ij} \sim \mathbf{H}$ is the famous $l(l+1)/2) \times l(l+1)/2)$ *Helmert matrix*. Let us assume $H_{ij} \sim \mathbf{H}$ is a regular matrix such that $\sigma = H^{-1}E\{q\}$. Now we choose

$$\sigma_H := \mathbf{H}^{-1}\mathbf{q} \tag{13.18}$$

as the *local* variance-covariance-component estimation of Helmert type, we can guarantee *unbiasedness*, $E(\sigma_H) = \sigma$, due to

$$E(\sigma_H) = H^{-1}Eq = \sigma$$

Theorem 13.1. Assume the Helmert matrix **H** to be regular. Then σ_H is unique and σ_0-*HIQUE* (σ_0-**H**elmert type **i**nvariant **q**uadratic **u**nbiased **e**stimation).

Uniqueness is obvious from (13.18). Invariance with respect to $\mathbf{y} \to \mathbf{y} - \mathbf{Ax} = \varepsilon$ is consequence of *Lemma 13.1* while the property of a quadratic unbiased estimation follows directly from construction principle. A more general theorem, especially based on *a singular* Helmert matrix, ha been given by *E. Grafarend*, *A Kleusberg* and *B. Schaffrin* (1980, pp. 165–167) to which we refer. We note that even if $\mathbf{H}$ is regular, σ_H is not σ_0-*BIQUE* (σ_0-best *IQUE*).

Step of higher order
It has been stated that the σ_0-Helmert type variance-covariance-component estimation σ_H is a *local* one. The function $\sigma_H(\sigma_0)$ illustrates this property. Conventionally, geodesists have designed algorithms which *iterate* the function $\sigma_H(\sigma_0)$ to the point $\sigma_H = \sigma_0$, the σ_0-reproducing variance-covariance-component estimation such that $\sigma_H = \mathbf{H}^{-1}(\sigma_0)\mathbf{q}(\sigma_0) = \sigma_0$. But it has been shown by *E. Grafarend* and *A. Kleusberg* (1980, pp. 133–135) that there are more than one reproducing points. We have therefore advocated to make a plot of the function $\sigma_H(\sigma_0)$ and $\sigma_H = \sigma_0$ in order to have a deeper insight into the local behavior of the estimation process. Anyway the higher order steps of iteration $\sigma_H(\sigma_0)$ will be stopped at some point $\sigma_H = \sigma_0$

13-2 Variance-Covariance-Component Estimation in the Linear Model $\mathbf{B}\boldsymbol{\varepsilon} = \mathbf{By} - \mathbf{c}$, $\mathbf{By} \notin \mathcal{R}(\mathbf{A}) + \mathbf{c}$

According to E.Grafarend (1984): Variance-covariance-component estimation of HELMERT type in the Gauss-Helmert model, Zeitschrift fuer Vermessungswesen 109 (1984) pp. 30–40 we follow his text: given the system of linear condition equation $\mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$: The $q \times n$ *fixed* matrix $\mathbf{B}$ of full row rank $r(\mathbf{B}) = q$ connects to $n \times 1$ *stochastic* vector ε of inconsistency and the $q \times 1$ *stochastic* vector $\mathbf{By} - \mathbf{c}$ where the $q \times 1$ fixed vector c of inhomogeneity accounts for the condition equations or constraints. The $n \times 1$ vector of ε of inconsistency is chosen in such a way that $\mathbf{By}$ does not belong to the range $\mathcal{R}(\mathbf{B})$ of the matrix $\mathbf{B}$ (spanned by its columns) and being translated by the vector c of inhomogeneity and that $\mathbf{B}E(\mathbf{y}) = \mathbf{c}$ holds. Within second order statistics the linear model $\mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$ is characterized by the first and second order moments of the vector $\mathbf{y}$ of observations.

$$E(\mathbf{y}) = \mathbf{y} - \varepsilon, \qquad E(\mathbf{By}) = \mathbf{B}E(\mathbf{y}) = \mathbf{c} \tag{13.19}$$

$$D(\mathbf{y}) = \sum_{j=1}^{l} \mathbf{C}_{jj}\sigma_j^2 + \sum_{\substack{j=1\\ j<k}}^{l-1}\sum_{k=2}^{j} \mathbf{C}_{jk}\sigma_{jk} = \sum_{j=1}^{l(l+1)/2} \mathbf{C}_j\sigma_j, \; D(\mathbf{By}) = \mathbf{B}D(\mathbf{y})\mathbf{B}' \tag{13.20}$$

The block structure of the dispersion matrix $D(\mathbf{y})$ is discussed in more detail in the first paragraph and in the *Appendix*. Here we develop the ides of stepwise estimation of variance-covariance-components of F.R. Helmert type with respect to the model of *condition adjustment* $\mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$.

The initial step

In the first step of the estimation process of $\{E(\mathbf{y}), D(\mathbf{y})\}$ we depart from an *initial* variance-covariance matrix Σ_0 of the dispersion matrix $D(\mathbf{y}) = \Sigma$ of the observation vector $\mathbf{y}$, or equivalently, from an initial set σ_0 of variance-covariance-components. Using the initial dispersion matrix Σ_0, the unknown vector ε of inconsistency are estimated by *"constrained least-squares"*:

(i) Lagrange-function

$$L(\varepsilon, \lambda) := \varepsilon' \sigma_0^{-1} \varepsilon + 2\lambda'(\mathbf{B}\varepsilon - \mathbf{B}\mathbf{y} + \mathbf{c}) \tag{13.21}$$

$$= [\mathbf{y} - E(\mathbf{y})]' \sigma_0^{-1} [\mathbf{y} - E(\mathbf{y})] + 2\lambda'[-\mathbf{B}\mathbf{E}(\mathbf{y}) + \mathbf{c} = \min_{\varepsilon, \lambda}$$

$$\frac{\partial L}{\partial \varepsilon}(\varepsilon_l, \lambda_l) = 2\sigma_0^{-1}\varepsilon_l + 2\mathbf{B}'\lambda_l = 0 \tag{13.22}$$

$$\frac{\partial L}{\partial \varepsilon}(\varepsilon_l, \lambda_l) = 2(\mathbf{B}\varepsilon_l - \mathbf{B}\mathbf{y} + \mathbf{c}) = 2\{\mathbf{B}[E(\mathbf{y})]_l + \mathbf{c}\} = 0 \tag{13.23}$$

$$\frac{\partial L}{\partial \varepsilon}\{[E(\mathbf{y})]_l, \lambda_l\} = -2\sigma_0^{-1}\{\mathbf{y} - [E(\mathbf{y})]_l\} - 2\mathbf{B}'\lambda_l = 0 \tag{13.24}$$

$$\frac{\partial^2 L}{\partial \varepsilon \partial \varepsilon'}(\varepsilon_l, \lambda_l) = 2\sigma_0^{-1} \geq 0 \tag{13.25}$$

The first derivatives of the Lagrange-function constitute the *necessary* conditions, the second derivatives the *sufficiency* condition. Since we assume a positive-definite dispersion matrix σ_0 in paragraph one, (13.25) is always fulfilled. (Here it would have been sufficient to have started from a positive-semidefinite weight matrix, but this generalization will be treated elsewhere.) Thus we arrive at the classical system of normal equations

$$\begin{bmatrix} \mathbf{\Sigma}_0^{-1} & \mathbf{B}' \\ \mathbf{B} & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_l \\ \lambda_l \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{B}\mathbf{y} - \mathbf{c} \end{bmatrix} \tag{13.26}$$

where the block matrix of normal equations is called the *fundamental matrix of constrained minimization*. (For more details see *S.L Campbell* and *C.D Meyer* (1979, pp. 63–70).)

For a $q \times n$ matrix $\mathbf{B}$ of full row rank, $r(\mathbf{B}) = q$, and for a positive-definite variance-covariance matrix σ_0 the vector ε_l of inconsistency and the vector λ_l of Lagrange-multipliers can be easily obtained from the system of normal equations.

$$\varepsilon_l = \mathbf{\Sigma}_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}(\mathbf{B}\mathbf{y} - \mathbf{c}), \qquad \lambda_l = -(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}(\mathbf{B}\mathbf{y} - \mathbf{c}) \tag{13.27}$$

$$[E(\mathbf{y})]_l = \mathbf{y} - \varepsilon_l = [I - \sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{B}]\mathbf{y} - \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{c} \tag{13.28}$$

(ii) The invariance of estimation

Let us discuss the "constrained least-squares" estimation ε_l of the vector ε of inconsistency. Once write

$$\varepsilon_l = \boldsymbol{\Sigma}_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}(\mathbf{By}-\mathbf{c}) = \mathbf{l}+\mathbf{L_0y} \tag{13.29}$$

$$[E(\mathbf{y})]_l = [I-\Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{B}]\mathbf{y}-\Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{c} = \mathbf{l}_0+\mathbf{M}_0\mathbf{y} \tag{13.30}$$

$$\mathbf{l}_0 := -\Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{c} \tag{13.31}$$

$$\mathbf{L}_0 := \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{B}, \qquad \mathbf{M}_0 := \mathbf{I}-\mathbf{L}_0 \tag{13.32}$$

It is obvious that (13.29), (13.30) represent linear, but inhomogeneous estimations.

Lemma 13.2. Let $\varepsilon_l = \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}(\mathbf{By}-\mathbf{c}) = \mathbf{l}+\mathbf{L_0y}$ be the constrained least-squares estimations of the stochastic vector ε of inconsistency. ε_l is invariant with respect to the transformation $\mathbf{By}-\mathbf{c} \to \mathbf{B}[\mathbf{y}-E(\mathbf{y})]$ such that

$$\varepsilon_l = \mathbf{L}_0\varepsilon \tag{13.33}$$

holds. $\mathbf{L}_0$ *is an idempotent matrix.*

The proof follows the lines of the one of *Lemma 13.1*.

(iii) Variance-covariance-component estimation of Helmert type

For a study of the underlying ideas of the variance-covariance-component estimation techniques of Helmert type we refer to paragraph 1 (iii). We only have to exchange $\mathbf{D}_0 = \mathbf{I}-\mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}$ by $\mathbf{L}_0 = \sigma_0\mathbf{B}'(\mathbf{B}\sigma_0\mathbf{B}')^{-1}\mathbf{B}$, e.g.

$$E\{\varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l\} = \sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{L}_0'\mathbf{E}_i(\sigma_0)\mathbf{L}_0\mathbf{C}_j\sigma_j \quad (i=1,\dots,l(l+1)/2 \tag{13.34}$$

$$E(q_i) = \sum_{j=1}^{l(l+1)/2} H_{ij}\sigma_j \quad (i=1,\dots,l(l+1)/2 \tag{13.35}$$

$$q_i := \varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l = (\mathbf{l}+\mathbf{L_0y})'\mathbf{E}_i(\sigma_0)(\mathbf{l}+\mathbf{L_0y}) \quad (i=1,\dots,l(l+1)/2 \tag{13.36}$$

$$H_{ij} := \mathrm{tr}\mathbf{L}_0'\mathbf{E}_i(\sigma_0)\mathbf{L}_0\mathbf{C}_j \quad (i,j=1,\dots,l(l+1)/2 \tag{13.37}$$

$H_{ij} \sim \mathbf{H}$ is the generalized $l(l+1)/2 \times l(l+1)/2$ *Helmert matrix*. Let us assume $H_{ij} \sim \mathbf{H}$ is a regular matrix such that $\sigma = \mathbf{H}^{-1}\mathbf{E}(\mathbf{q})$. Now we choose

$$\sigma_H := \mathbf{H}^{-1}\mathbf{q} \tag{13.38}$$

as the local variance-covariance-component estimation of *Helmert type*, we can guarantee unbiasedness, $E(\sigma_H) = \sigma$, due to $E(\sigma_H) = \mathbf{H}^{-1}\sigma\mathbf{E}(\mathbf{q}) = \sigma$.

Theorem 13.2. Assume the g-Helmert matrix H to be regular. Then σ_H is unique and σ_0-Helmert type invariant inhomogeneous quadratic unbiased estimation.

Uniqueness is obvious from (13.38). Invariance with respect to $\mathbf{By} - \mathbf{c} \to \mathbf{B}[\mathbf{y} - E(\mathbf{y})]$ is a consequence Lemmas 2 which the property of an inhomogeneous mixed linear-quadratic estimation follows directly form the construction principle. Note $\mathbf{c} = 0$ the σ_0-*Helmert type estimation* σ_H is σ_0-HIQUE. A more general theorem, especially based on a *singular g*-Helmert matrix, will be given elsewhere.

Steps of high order

The nature of a local estimation $\sigma_H(\sigma_0)$ leads again to an iteration scheme to a point $\sigma_H = \sigma_0$, the σ_0-*reproducing variance-covariance-component estimation* such that $\sigma_H = \mathbf{H}^{-1}(\sigma_0)\mathbf{q}(\sigma_0) = \sigma_0$. Again we advocate the computation of the complete graph $\sigma_H(\sigma_0)$ since more than one reproducing point might appear.

13-3 Variance-Covariance-Component Estimation in the Linear Model $\mathbf{Ax} + \boldsymbol{\varepsilon} + \mathbf{B}\boldsymbol{\varepsilon} = \mathbf{By} - \mathbf{c}$, $\mathbf{By} \notin \mathcal{R}(\mathbf{A}) + \mathbf{c}$

According to E.Grafarend (1984): Variance-covariance-component estimation of HELMERT type in the Gauss-Helmert model, Zeitschrift fuer Vermessungswesen 109 (1984) pp. 40–42 we follow his text: given the system of linear equations $\mathbf{Ax} + \varepsilon + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$: The stochastic $n \times 1$ vector ε of inconsistency pays attention to the fact that the $q \times 1$ vector $\mathbf{By}$ does not belong to the range, spanned by the column space of the $n \times m$ matrix $\mathbf{A}$ and translated by the $q \times 1$ vector c. The vector ε is chosen in such a way that the vector $\mathbf{By} - \mathbf{c}$ is an element of the column space of the matrix $\mathbf{A}$ translated by the vector $\mathbf{c}$. The fixed $q \times m$ matrix $\mathbf{A}$ is full of column rank, $r(\mathbf{A}) = m$, but the fixed $q \times n$ matrix $\mathbf{B}$ of full row rank $r(\mathbf{B}) = q$. Within second order statistics the linear model $\mathbf{Ax} + \varepsilon + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$ is characterized by the first and second order moments of the $n \times 1$ vector $\mathbf{y}$ of observation.

$$E\mathbf{y} = \mathbf{y} - \varepsilon \tag{13.39}$$

$$\mathbf{B}E(\mathbf{y}) = \sum_{i=1}^{m} \mathbf{a}_i \mu_i + \mathbf{c} = \mathbf{Ax} + \mathbf{c} \tag{13.40}$$

$$D(\mathbf{y}) = \sum_{j=1}^{l} \mathbf{C}_{jj}\sigma_j + \sum_{j=1}^{l-1}\sum_{k=2}^{l} \mathbf{C}_{jk}\sigma_{jk} = \sum_{j=1}^{l(l+1)/2} \mathbf{C}_j\sigma_j \tag{13.41}$$

$$D(\mathbf{By}) = \mathbf{B}D(\mathbf{y})\mathbf{B}' \tag{13.42}$$

The block structure of the dispersion matrix D($\mathbf{y}$) is discussed in more detail in Sect. 13-1 and in Sect. 13-4. Here we develop the idea of stepwise estimation of variance-covariance-component of *F.R. Helmert* type with respect to the linear model of *condition equations with unknown* $\mathbf{Ax} + \varepsilon + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$

$\boxed{\textit{The initial step}}$

In the first step of the estimation process of (μ, σ) or $(\mathbf{x_1}, \mathbf{x_2})$ we depart from initial variance-covariance matrix Σ_0 of the dispersion matrix $D(\mathbf{y}) = \Sigma$ of the observation vector $\mathbf{y}$, or, equivalently, from an initial set σ_0 of variance-covariance-components. Using the initial dispersion matrix Σ_0, the unknown vector $\mathbf{x}$ and ε are estimated by *"constrained least-squares"*:

(i) Lagrange-function

$$L(\varepsilon, \mathbf{x}, \lambda) := \varepsilon' \Sigma_0^{-1} \varepsilon + 2\lambda'(\mathbf{Ax} + \mathbf{B}\varepsilon - \mathbf{By} + \mathbf{c}) = \min_{\varepsilon, x, \lambda} \tag{13.43}$$

$$\frac{\partial L}{\partial \varepsilon}(\varepsilon_l, x_l, \lambda_l) = 2\Sigma_0^{-1}\varepsilon_l + 2\mathbf{B}'\lambda_l = 0$$

$$\frac{\partial L}{\partial \mathbf{x}}(\varepsilon_l, x_l, \lambda_l) = 2\mathbf{A}'\lambda_l = 0 \tag{13.44}$$

$$\frac{\partial L}{\partial \lambda}(\varepsilon_l, x_l, \lambda_l) = 2(\mathbf{Ax}_l + \mathbf{B}\varepsilon_l - \mathbf{By} + \mathbf{c}) = 0 \tag{13.45}$$

$$\frac{\partial' L}{\partial \varepsilon \partial \varepsilon'}(\varepsilon_l, x_l, \lambda_l) = 2\Sigma_0^{-1} \geq 0 \tag{13.46}$$

The first derivatives of the Lagrange-function constitute the *necessary* conditions, the second derivatives the *sufficiency* condition. Since we have assumed a positive-definite dispersion matrix Σ_0 in paragraph one, (13.46) is always fulfilled. (Here it would have been sufficient to have started from a positive-definite weight matrix, but this generalization will be treated elsewhere.) Thus we arrive at the classical system of normal equations.

$$\begin{bmatrix} \Sigma_0^{-1} & \mathbf{B}' & 0 \\ \mathbf{B} & 0 & \mathbf{A} \\ 0 & \mathbf{A}0 & \end{bmatrix} \begin{bmatrix} \varepsilon_l \\ \lambda_l \\ \mathbf{x}_l \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{By} - \mathbf{c} \\ 0 \end{bmatrix} \tag{13.47}$$

with the conventional block matrix of constrained minimization.

For a $q \times m$ matrix A of full column rank, $r(\mathbf{A}) = m$, a $q \times n$ matrix $\mathbf{B}$ of full rank, $r(\mathbf{B}) = q$ and a positive-semidefinite weight matrix Σ_0 the unknown vectors ε_l, λ_l and x_l can be easily obtained by successive reduction of the system of normal equations.

$$\begin{aligned}\varepsilon_l &= \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{\mathbf{I}-\mathbf{A}[\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\}(\mathbf{By}-\mathbf{c}) \quad (13.48)\\ &\quad \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{(\mathbf{By}-\mathbf{c})-\mathbf{Ax}_l\}\\ \lambda_l &= (\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{\mathbf{A}[\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}-\mathbf{I}\}(\mathbf{By}-\mathbf{c}) \quad (13.49)\\ \mathbf{x}_l &= [\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}(\mathbf{By}-\mathbf{c}) \quad (13.50)\end{aligned}$$

(ii) The variance of the estimation

Let us discuss the *"constrained least-squares"* estimation ε_l of the vector ε of inconsistency. Once write

$$\begin{aligned}\varepsilon_l &= \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{\mathbf{I}-\mathbf{A}[\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\}(\mathbf{By}-\mathbf{c}) \quad (13.51)\\ &= \mathbf{f}_0+\mathbf{F}_0\mathbf{y}\\ \mathbf{f}_0 &:= -\Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{\mathbf{I}-\mathbf{A}[\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\}\mathbf{c} \quad (13.52)\\ \mathbf{F}_0 &:= \Sigma_0\mathbf{B}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\{\mathbf{I}-\mathbf{A}[\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\mathbf{A}]^{-1}\mathbf{A}'(\mathbf{B}\Sigma_0\mathbf{B}')^{-1}\}\mathbf{B} \quad (13.53)\end{aligned}$$

it is obvious that (13.52) represents a linear, but inhomogeneous estimation.

Lemma 13.3. Let $\varepsilon_l = \mathbf{f}_0 + \mathbf{F}_0\mathbf{y}$ be the constrained least-squares estimation of the stochastic vector ε of inconsistency. ε_l is invariant with respect to the transformation $\mathbf{By}-\mathbf{c} \to \mathbf{Ax} + \mathbf{B}[\mathbf{y} - E(\mathbf{y})]$ such that

$$\varepsilon_l = \mathbf{F}_0\varepsilon \quad (13.54)$$

holds. $\mathbf{F}_0$ *is an idempotent matrix*

The proof follows the lines of *Lemma 13.1*.

(iii)Variance-covariance-components estimation of Helmert type

Again we refer for a motivation of the variance-covariance-component estimation technique of Helmert type to paragraph 1 (iii). We only have to exchange

$$\mathbf{D}_l = \mathbf{I} - \mathbf{A}(\mathbf{A}'\Sigma_0^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma_0^{-1}\text{by}\mathbf{F}_0 \quad (13.55)$$

in (13.53)

$$E\{\varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l\} = \sum_{j=1}^{l(l+1)/2} \mathrm{tr}\mathbf{F}_0'\mathbf{E}_i(\sigma_0)\mathbf{F}_0\mathbf{C}_j\sigma_j(i=1,\ldots,l(l+1)/2 \quad (13.56)$$

$$E(q_i) = \sum_{j=1}^{l(l+1)/2} H_{ij}\sigma_j(i=1,\ldots,l(l+1)/2 \quad (13.57)$$

$$q_i := \varepsilon_l'\mathbf{E}_i(\sigma_0)\varepsilon_l = (\mathbf{f_0}+\mathbf{F_0y})'\mathbf{E}_i(\sigma_0)(\mathbf{f_0}+\mathbf{F_0y})$$

$$(i = 1, \ldots, l(l+1)/2 \tag{13.58}$$

$$H_{ij} := \mathrm{tr}\mathbf{F}_0'\mathbf{E}_i(\sigma_0)\mathbf{F}_0\mathbf{C}_j(i, j = 1, \ldots, l(l+1)/2 \tag{13.59}$$

$H_{ij} \sim \mathbf{H}$ is the generalized $l(l+1)/2 \times l(l+1)/2$ *Helmert matrix*. Let us assume $H_{ij} \sim \mathbf{H}$ is a regular matrix such that $\sigma = \mathbf{H}^{-1}\mathbf{E}(\mathbf{q})$. Now we choose

$$\sigma_H := \mathbf{H}^{-1}\mathbf{q} \tag{13.60}$$

as the local variance-covariance-component estimation of *Helmert type*, we can guarantee unbiasedness, $E(\sigma_H) = \sigma$, due to $E(\sigma_H) = \mathbf{H}^{-1}\sigma\mathbf{E}(\mathbf{q}) = \sigma$.

Theorem 13.3. Assume the g-Helmert matrix $\mathbf{H}$ to be regular. Then σ_H is unique and σ_0-Helmert type invariant inhomogeneous quadratic unbiased estimation.

Uniqueness is obvious from (13.60). Invariance with respect to $\mathbf{By} - \mathbf{c} \rightarrow \mathbf{Ax} + \mathbf{B}[\mathbf{y} - E(\mathbf{y})]$ is a consequence of *Lemma 13.3* while the property of an inhomogeneous mixed linear-quadratic estimation follows directly from the construction principle, especially (13.52). Note that for $\mathbf{c} = 0$ the σ_0-Helmert type estimation σ_H is σ_0-HIQUE. A more general theorem, especially based on a *singular g*-Helmert matrix, will be given elsewhere.

Steps of high order

The local property of estimation $\sigma_H(\sigma_0)$ leads to the standard iteration problem. From the complete graph $\sigma_H(\sigma_0)$ we gain insight into the locations of the σ_0-reproducing variance-covariance-components estimation such that $\sigma_H = \mathbf{H}^{-1}(\boldsymbol{\sigma}_0)\mathbf{q}(\boldsymbol{\sigma}_0) = (\boldsymbol{\sigma}_0)$ In the applications of general linear model $\mathbf{Ax} + \varepsilon + \mathbf{B}\varepsilon = \mathbf{By} - \mathbf{c}$ a partitioned version appears:

$$\mathbf{A} := \begin{bmatrix} 0 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix}, \quad \mathbf{B} := \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ 0 \end{bmatrix}, \quad \mathbf{c} := \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix}, \quad \mathbf{c}_3 \in \mathcal{R}(\mathbf{A}_3) \tag{13.61}$$

(i) ***first partitioning***
inconsistent system of inhomogeneous condition equations between the observations only: $\mathbf{B}_1\varepsilon = \mathbf{B}_1\mathbf{y} - \mathbf{c}_1$

(ii) ***second partitioning***
inconsistent system of inhomogeneous condition equations with unknowns: $\mathbf{A}_2\mathbf{x}\mathbf{B}_2\varepsilon = \mathbf{B}_2\mathbf{y} - \mathbf{c}_2$

(iii) ***third partitioning***
consistent system of inhomogeneous condition equations between unknowns only: $\mathbf{A}_3 = \mathbf{c}_3, \quad \mathbf{c}_3 \in \mathcal{R}(\mathbf{A}_3)$

13-4 The Block Structure of Dispersion Matrix D{y}

According to E.Grafarend (1984): Variance-covariance-component estimation of HELMERT type in the Gauss-Helmert model, Zeitschrift fuer Vermessengswesen 109(1984) pp.42-44 at first we will review the block-partitioning of the dispersion matrix $\mathbf{D}\{\mathbf{y}\}$ of the observation vector y. An example from leveling and another one from distance observations will illustrate the technique of decomposition into blocks.

The variance-covariance matrix $\mathbf{D}\{\mathbf{y}\}$ of the stochastic vector y of observations is decomposed into

$$\mathbf{D}(\mathbf{y}) = \sum_{j=1}^{l} \mathbf{C}_{ii}\sigma_i^2 + \sum_{j=1}^{l-1}\sum_{k=1}^{l} \mathbf{C}_{jk}\sigma_{jk} = \begin{bmatrix} \sigma_1^2 Q_{11} & \sigma_{12}Q_{12} & \dots \sigma_{1l}Q_{1l} \\ \sigma_{12}Q'_{12} & \sigma_2^2 Q_{22} & \dots \sigma_{2l}Q_{2l} \\ \dots & & \\ \sigma_{1l}Q'_{1l} & \sigma_{2l}Q'_{2l} & \dots \mathbf{s}_l^2 Q_{ll} \end{bmatrix} \tag{13.62}$$

$$\begin{array}{l} \mathbf{C}_{jj} := \begin{bmatrix} 0 & \dots & 0 \\ \dots & \mathbf{Q}_{jj} & \dots \\ 0 & \dots & 0 \end{bmatrix} \text{row j} \quad (j = 1, \dots, l) \\ \qquad\qquad \text{column j} \end{array} \tag{13.63}$$

$$\begin{array}{l} \mathbf{C}_{jk} := \begin{bmatrix} 0 & \dots & & 0 \\ \dots & 0 & \mathbf{Q}_{jk} & \dots \\ & \mathbf{Q}'_{jk} & 0 & \\ 0 & \dots & & 0 \end{bmatrix} \text{row j} \quad (j, k = 1, \dots, l) \\ \qquad\qquad \text{column j} \end{array} \tag{13.64}$$

Example A1 (leveling)

Let us estimate the variance-components of forward and backward observations in a triangular leveling network, namely at two different epochs of observation. The 4×4 dispersion matrix $D(\mathbf{y}) = \Sigma(h_{12}, h_{23}, h_{21}, h_{32} = \Sigma$ of the observations h_{12}, h_{23} at epoch one and h_{21}, h_{32} at epoch two can be block-partitioned into

$$\Sigma = \left[\begin{array}{llll} \sigma_{11} + \sigma_{22} - 2\sigma_{12} & \sigma_{12} + \sigma_{23} - \sigma_{22} - \sigma_{13} & \sigma_2^1 + \sigma_1^2 - \sigma_1^1 - \sigma_2^2 & \sigma_2^2 + \sigma_1^3 - \sigma_1^2 - \sigma_2^3 \\ \sigma_{12} + \sigma_{23} - \sigma_{13} - \sigma_{22} & \sigma_{22} + \sigma_{33} - 2\sigma_{23} & \sigma_2^2 + \sigma_3^1 - \sigma_2^1 - \sigma_3^2 & \sigma_2^3 + \sigma_3^1 - \sigma_2^2 - \sigma_3^3 \\ \hline \sigma_1^2 + \sigma_2^1 - \sigma_1^1 - \sigma_2^2 & \sigma_2^2 + \sigma_3^1 - \sigma_2^1 - \sigma_2^3 & \sigma^{11} + \sigma^{22} - 2\sigma^{12} & \sigma^{12} + \sigma^{23} - \sigma_2 - \sigma^{13} \\ \sigma_1^3 + \sigma_2^2 - \sigma_1^2 - \sigma_2^3 & \sigma_2^3 + \sigma_3^2 - \sigma_2^2 - \sigma_3^3 & \sigma^{21} + \sigma^{32} - \sigma^{31} - \sigma^{22} & \sigma^{22} + \sigma^{33} - 2\sigma^{23} \end{array}\right] \tag{13.65}$$

where

$$\sigma_{ij} = E[\varepsilon_i(t_1)\varepsilon_j(t_1)] = \sigma_i(t_1)\sigma_j(t_1)\varrho_{ij} = \Phi_{ij}(t_1, t_1) \tag{13.66}$$

$$\sigma_i^j = E[\varepsilon_i(t_1)\varepsilon_j(t_2)] = \sigma_i(t_1)\sigma_j(t_2)\varrho_i^j = \Phi_{ij}(t_1, t_2) \tag{13.67}$$

$$\sigma^{ij} = E[\varepsilon_i(t_2)\varepsilon_j(t_2)] = \sigma_i(t_2)\sigma_j(t_2)\varrho^{ij} = \Phi_{ij}(t_2, t_2) \tag{13.68}$$

$$\tag{13.69}$$

We call $\sigma_i^i = \Phi_{ii}(t_1, t_2)$ *autovariance* and $\sigma_i^j = \Phi_{ij}(t_1, t_2), i \neq j$, *crossvariance.*

In a first design of second order let us assume that the absolute heights are uncorrelated, their variances to be identical within one epoch, but varying from epoch to epoch. Then the second order design matrix of observed height differences can be represented by

$$\Sigma = \left[\begin{array}{cc|cc} 2\sigma^2(t_1) & -\sigma^2(t_1) & 0 & \\ -\sigma^2(t_1) & 2\sigma^2(t_1) & & \\ \hline 0 & 0 & 2\sigma^2(t_2 1) & -\sigma^2(t_1) \\ 0 & 0 & -\sigma^2(t_1) & 2\sigma^2(t_2 1) \end{array}\right]$$

$$= \sigma_1^2 \left[\begin{array}{cc|cc} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] + \sigma_2^2 \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{array}\right]$$

$$= \sigma_1^2 \begin{bmatrix} \mathbf{Q}_{11} & 0 \\ 0 & 0 \end{bmatrix} + \sigma_2^2 \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{Q}_{22} \end{bmatrix} = \mathbf{C}_{11}(\sigma_1)^2 + \mathbf{C}_{22}(\sigma_2)^2$$

where $\sigma^2(t_1) = (\sigma_1)^2, \sigma^2(t_2) = (\sigma_2)^2$ are unknown variance-covariance-component of absolute heights at epoch t_1 and t_2.

$$\mathbf{Q}_{11} = \mathbf{Q}_{22} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \tag{13.70}$$

Here the open problem is to estimate $\sigma^2(t_1)$ and $\sigma^2(t_2)$

Example A2 (distance observations):

Three targets at distances 1, 2 and 4 km will be observed twice by distances. The two sets of observations lead to the conditions

$$E(y_1) = E(y_4), E(y_2) = E(y_5), E(y_3) = E(y_6)$$

or

$$\mathbf{B}\varepsilon = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \varepsilon = \mathbf{B}\mathbf{y} \tag{13.71}$$

Let us assume that the observations are uncorrelated, but the variances will be represented by

$$\sigma_{y_i}^2 = \sigma_1^2 + \sigma_i^{-2}\sigma_2^2 \tag{13.72}$$

where σ_1^2 is the distance-independent part, but σ_2^2 the distance-dependent part of the observational variance. Thus the 6×6 dispersion matrix $D(y) = \mathbf{\Sigma}(y_1, \ldots, y_6) = \mathbf{\Sigma}$ of observations $y_1, \ldots, y_6$ can be represented by

$$\Sigma = \left[\begin{array}{ccc|ccc} \sigma_1^2+\sigma_2^2 & & & 0 & 0 & 0 \\ & \sigma_1^2+\frac{1}{4}\sigma_2^2 & & 0 & 0 & 0 \\ & & \sigma_1^2+\frac{1}{16}\sigma_2^2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \sigma_1^2+\sigma_2^2 & & \\ 0 & 0 & 0 & & \sigma_1^2+\frac{1}{4}\sigma_2^2 & \\ 0 & 0 & 0 & & & \sigma_1^2+\frac{1}{16}\sigma_2^2 \end{array}\right] \tag{13.73}$$

$$= \sigma_1^2\mathbf{I} + \sigma_2^2 \mathrm{diag}\left(1, \frac{1}{4}, \frac{1}{16}, 1, \frac{1}{4}, \frac{1}{16}\right) \tag{13.74}$$

$$= \mathbf{C}_{11}(\sigma_1)^2 + \mathbf{C}_{22}(\sigma_2)^2$$

$$\mathbf{C}_{11} = \mathbf{I}, \mathbf{C}_{22} = \mathrm{diag}\left(1, \frac{1}{4}, \frac{1}{16}, 1, \frac{1}{4}, \frac{1}{16}\right) \tag{13.75}$$

Here the open problem is to estimate $(\sigma_1)^2$ and $(\sigma_2)^2$.

This introduction into conditional equation with *unknowns–the Gauss–Helmert model*–is based on our contribution in Zeit schrift für Vermessungswesen *109*(1984) 32–44, the first contribution which is based on *variance-covariance-component estimation*. The question is our model can be based on *Bayesian estimation* is still open. Model studies of our type we performed by *S.L Campbell* and *C.D. Meyer* (1979), *E. Grafarend* and *A. Kleusberg* (1980), *A. Kleusberg* and *B. Schaffrin*(1980), in particular *F.R. Helmert* (1924), *E. Grafarend* and *A. Kleusberg* (1980), *A. Kleusberg* and *E. Grafarend* (1981), *L.E. Sjöberg* (1983a,b) and in particular on *H. Wolf*(1978, 1988).

Chapter 14
Special Problems of Algebraic Regression and Stochastic Estimation

multivariate Gauss–Markov model, the n-way classification model, dynamical systems

Up to now, we have only considered an "*univariate Gauss–Markov model*". Its generalization towards a *multivariate Gauss–Markov model* will be given in *Sect. 14-1*. At first, we define a multivariate linear model by *Definition 14.1* by giving its first and second order moments. Its algebraic counterpart via multi-variate LESS is subject of *Definition 14.2*. *Lemma 14.3* characterizes the multi-variate LESS solution. Its multivariate Gauss–Markov counterpart is given by *Theorem 14.4*. In case we have constraints in addition, we define by *Definition 14.5* what we mean by "*multivariate Gauss–Markov model with constraints*". The complete solution by means of "*multivariate Gauss–Markov model with constraints*" is given by *Theorem 14.5*.

In contrast, by means of a MINOLESS solution we present the celebrated "*n-way classification model*". Examples are given for a *1-way classification model*, for a 2-way classification model *without interaction*, for a 2-way classification model with interaction with all numerical details for computing the *reflexive, symmetric generalized inverse* $(\mathbf{A}'\mathbf{A})^{-}_{rs}$. The higher classification with interaction is finally reviewed. We especially deal with the problem *how to compute a basis of unbiased estimable quantities from biased solutions*.

Finally, we take account of the fact that *in addition to observational models*, we have *dynamical system equations*. Additionally, we therefore review the *Kalman Filter* (*Kalman–Bucy Filter*). Two examples from tracking a satellite orbit and from statistical quality control are given. In detail, we define the stochastic process of type *ARMA* and *ARIMA*. A short introduction on "*dynamical system theory*" is presented. By two examples we illustrate the notions of "*a steerable state*" and of "*observability*". A careful review of the conditions "*steerability*" by *Lemma 14.7* and "*observability*" by *Lemma 14.8* is presented. Traditionally the state differential equation as well as the observational equation are solved by a typical Laplace transformation which we will review shortly. At the end, we focus on the *modern theory of dynamic nonlinear models* and comment on the *theory of chaotic behaviour* as its upto date counterpart.

14-1 The Multivariate Gauss–Markov Model: A Special Problem of Probabilistic Regression

Let us introduce the *multivariate Gauss–Markov model* as a *special problem of the probabilistic regression*. If for one matrix $\mathbf{A}$ of dimension $O(\mathbf{A}) = n \times m$ in a *Gauss–Markov model* instead of one vector of observations *several observation*

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_14,

© Springer-Verlag Berlin Heidelberg 2012

vectors $\mathbf{y}_i$ of dimension $O(\mathbf{y}_i) = n \times p$ with identical variance-covariance matrix Σ_{ij} are given and the fixed array of parameters ξ_i has to be determined, the model is referred to as a

Multivariate Gauss–Markov model.

The *standard Gauss–Markov model* is then called a *univariate Gauss–Markov model*. The analysis of variance-covariance is applied afterwards to a multivariate model if the effect of factors can be referred to not only by one characteristic of the phenomenon to be observed, but by several characteristics. Indeed this is the *multivariate analysis of variance-covariance*. For instance, *the effects of different regions on the effect of a species of animals are to be investigated, the weight of the animals can serve as one characteristic and the height of the animals as a second one*.

Multivariate models can also be setup, if *observations are repeated at different times, in order to record temporal changes of a phenomenon*. If measurements in order to detect temporal changes of manmade constructions *are repeated with identical variance-covariance matrices under the same observational program*, the matrix $\mathbf{A}$ of coefficients in the *Gauss–Markov model stays the same for each repetition* and each repeated measurement corresponds to one characteristic.

Definition 14.1. (multivariate Gauss–Markov model):

Let the matrix $\mathbf{A}$ of the order $n \times m$ be given, called the *first order design matrix*, let ξ_i denote the matrix of the order $m \times p$ of *fixed unknown parameters*, and let $\mathbf{y}_i$ be the matrix of the order $n \times p$ called the matrix of observations subject to $p \leq n$. Then we speak of a "*multivariate Gauss–Markov model*" if

$$E\{\mathbf{y}_i\} = \mathbf{A}\xi_i \quad \text{subject to} \quad \left[\begin{array}{l} O\{\mathbf{y}_\mathrm{i}\} = n \times p \\ O\{\mathbf{A}\} = n \times m \\ O\{\xi_i\} = m \times p \end{array}\right. \tag{14.1}$$

$$D\{\mathbf{y}_i, \mathbf{y}_j\} = \mathbf{I}_n \delta_{ij} \quad \text{subject } to \;\; O\{\delta_{ij}\} = p \times p \; and \;\; p.d. \tag{14.2}$$

for all $i,\ j\ \in \{1, \ldots, p\}$ apply for a second order statistics. Equivalent vector and matrix forms are

$$E\{\mathbf{Y}\} = \mathbf{A}\Xi \;\; and \;\; E\{\mathrm{vec}\mathbf{Y}\} = (\mathbf{I}_p \otimes \mathbf{A})\mathrm{vec}\,\Xi \tag{14.3}$$

$$D\{\mathrm{vec}\mathbf{Y}\} = \Sigma \otimes \mathbf{I}_n \;\; and \;\; d\{\mathrm{vec}\mathbf{Y}\} = d(\Sigma \otimes \mathbf{I}_n) \tag{14.4}$$

subject to

$$O\{\Sigma\} = p \times p,\;\; O\{\Sigma \otimes \mathbf{I}\} = np \times np,$$
$$O\{\mathrm{vec}\mathbf{Y}\} = np \times 1,\;\; O\{\mathbf{Y}\} = n \times p$$
$$O\{D\{\mathrm{vec}\mathbf{Y}\}\} = np \times np,\;\; O\{d(\mathrm{vec}\mathbf{Y})\} = np(np+1)/2.$$

The matrix $D\{\text{vec}\mathbf{Y}\}$ builds up the second order design matrix as the *Kronecker–Zehfuss product* Σ and $\mathbf{I}_n$.

In the *multivariate Gauss–Markov model both the matrices* ξ_i and σ_{ij} or Ξ and Σ are *unknown*. An algebraic equivalent of the multivariate linear model would read as given by *Definition 14.2*.

Definition 14.2. (multivariate linear model):

Let the matrix $\mathbf{A}$ of the order $n \times m$ be given, called *first order algebraic design matrix*, let $\mathbf{x}_i$ denote the matrix of the order of *fixed unknown parameter*, and $\mathbf{y}_i$ be the matrix of order $n \times p$ called the *matrix of observations* subject to $p \leq n$. Then we speak of an *algebraic multivariate linear model* if

$$\sum_{i=1}^{p} ||\mathbf{y}_i - \mathbf{A}\mathbf{x}_i||^2_{\mathbf{G_y}} = \min_{\mathbf{x}_i} \sim \ ||\text{vec}\mathbf{Y} - (\mathbf{I}_p \otimes \mathbf{A})\text{vec}\mathbf{X}||^2_{\mathbf{G}_{\text{vec}\mathbf{Y}}} = \min \tag{14.5}$$

establishing a $\mathbf{G_y}-$ or $\mathbf{G}_{vec\mathbf{Y}}$-weighted least squares solution of type *multivariate* LESS.

It is a standard solution of type *multivariate LESS* if

$$\mathbf{X} = (\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-}\mathbf{A}'\mathbf{G_y}\mathbf{Y} \left[\begin{array}{l} O\{\mathbf{X}\} = m \times p \\ O\{\mathbf{A}\} = n \times m \\ O\{\mathbf{Y}\} = n \times p\,, \end{array}\right. \tag{14.6}$$

which nicely demonstrates that the *multivariate LESS* solution is built on a *series of univariate LESS solutions*. If the matrix $\mathbf{A}$ is *regular* in the sense of $\text{rk}(\mathbf{A}'\mathbf{G_y}\mathbf{A}) = \text{rk}\mathbf{A} = m$, our *multivariate solution* reads

$$\mathbf{X} = (\mathbf{A}'\mathbf{G_y}\mathbf{A})^{-1}\mathbf{A}'\mathbf{G_y}\mathbf{Y} \tag{14.7}$$

excluding any rank deficiency caused by a *datum problem*. Such a result may be initiated by fixing a datum parameter of type *translation* (3 parameters at any epoch), *rotation* (3 parameters at any epoch) and *scale* (1 parameter at any epoch). These parameters make up the seven parameter conformal group $\mathbb{C}_7(3)$ at any epoch in a *three-dimensional Euclidian space* (pseudo-Euclidian space).

Lemma 14.3. (general multivariate linear model):

A general multivariate linear model is *multivariate LESS* if

$$\sum_{i,j=1}^{p,p} (\mathbf{y}_i - \mathbf{A}\mathbf{x}_i)\mathbf{G}^{ij}(\mathbf{y}_j - \mathbf{A}\mathbf{x}_j) = \min_{\mathbf{x}} \tag{14.8}$$

or

$$\sum_{\alpha=1}^{n}\sum_{\beta,\gamma}^{m,m}\sum_{i,j}^{p,p}(y_{\alpha i}-a_{\alpha\beta}x_{\beta i})\mathbf{G}^{ij}(y_{\alpha j}-a_{\alpha\gamma}x_{\gamma j}) = \min_{\mathbf{x}} \tag{14.9}$$

or

$$(\text{vec}\mathbf{Y}-(\mathbf{I}_p\otimes\mathbf{A})\text{vec}\mathbf{X})'(\mathbf{I}_n\otimes\mathbf{G}_{\mathbf{y}}(\text{vec}\mathbf{Y}-(\mathbf{I}_p\otimes\mathbf{A})\text{vec}\mathbf{X}) = \min_{\mathbf{x}}. \tag{14.10}$$

An array $\mathbf{X}$, $\dim\mathbf{X} = m\times p$ is multivariate LESS, if

$$\text{vec}\mathbf{X} = [(\mathbf{I}_p\otimes\mathbf{A})'(\mathbf{I}_n\otimes\mathbf{G}_{\mathbf{Y}})(\mathbf{I}_p\otimes\mathbf{A})]^{-1}(\mathbf{I}_p\otimes\mathbf{A})'(\mathbf{I}_n\otimes\mathbf{G}_{\mathbf{Y}})\text{vec}\mathbf{Y} \tag{14.11}$$

and

$$\text{rk}(\mathbf{I}_n\otimes\mathbf{G}_{\mathbf{Y}}) = np. \tag{14.12}$$

End of Lemma 14.3

Thanks to the *weight matrix* $\mathbf{G}^{ij}$ the *multivariate least squares solution* (14.3) differs from the special univariate model (14.7). The analogue to the general LESS model (14.8)–(14.10) of type *multivariate BLUUE* is given next.

Theorem 14.4. (multivariate Gauss–Markov model of type ξ_i, in particular $(\Sigma,\ \mathbf{I}_n)$-BLUUE):

A *multivariate Gauss–Markov model* is $(\Sigma,\ \mathbf{I}_n)$-BLUUE if the vector $\text{vec}\,\Xi$ of an array Ξ, $\dim\Xi = n\times p$, $\dim(\text{vec}\,\Xi) = np\times 1$ of unknowns is estimated by the matrix $\widehat{\xi}_i$, namely

$$\text{vec}\hat{\Xi} = [(\mathbf{I}_p\otimes\mathbf{A})'(\Sigma\otimes\mathbf{I}_n)^{-1}(\mathbf{I}_p\otimes\mathbf{A})]^{-1}(\mathbf{I}_p\otimes\mathbf{A})'(\Sigma\otimes\mathbf{I}_n)^{-1}\text{vec}\mathbf{Y} \tag{14.13}$$

subject to

$$\text{rk}(\Sigma\otimes\mathbf{I}_n)^{-1} = np. \tag{14.14}$$

$\Sigma\sim\sigma_{ij}$ denotes the variance-covariance matrix of multivariate effects $\mathbf{y}_{\alpha i}$ for all $\alpha = 1,\ldots,n$ and $i = 1,\ldots,p$. An *unbiased estimator* of the variance-covariance matrix of multivariate effects is

$$\boxed{\left[\begin{array}{l} i=j:\ \ \hat{\sigma}_i^2 = \dfrac{1}{n-q}(\mathbf{y}_i-\mathbf{A}\hat{\xi}_i)'(\mathbf{y}_i-\mathbf{A}\hat{\xi}_i) \\ i\neq j:\ \ \hat{\sigma}_{ij} = \dfrac{1}{n-q}(\mathbf{y}_i-\mathbf{A}\hat{\xi}_i)'(\mathbf{y}_j-\mathbf{A}\hat{\xi}_i) \end{array}\right.} \tag{14.15}$$

because of

$$\boxed{E\{(\mathbf{y}_i - \mathbf{A}\hat{\xi}_i)'(\mathbf{y}_j \mathbf{A}\hat{\xi}_j)\} = E\{\mathbf{y}'_i(I - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y}_j\} = \sigma_{ij}(n-q).} \tag{14.16}$$

A nice example is given in *K.R. Koch* (1988 pp. 281–286). For practical applications we need the *incomplete multivariate models* which do not allow a full rank matrix σ_{ij}. For instance, in the *standard multivariate model*, it is assumed that the matrix $\mathbf{A}$ of coefficients *has to be identical* for p vectors $\mathbf{y}_i$ and the *vectors* $\mathbf{y}_i$ *have to be completely given*.

If due to a change in the observational program in the case of repeated measurements or due to a loss of measurements, these assumptions *are not fulfilled*, an *incomplete multivariate model* results. If all the matrices of coefficients are different, but if p vectors $\mathbf{y}_i$ of observations agree with their dimension, the variance-covariance matrix Σ and the vectors ξ_i of first order parameters can be *iteratively estimated*. For example, if *the parameters of first order*, namely ξ_i, and the *parameters of second order*, namely σ_{ij}, the elements of the variance-covariance matrix, are unknown, we may use the *hybrid estimation of first and second order parameters* of type $\{\xi_i, \sigma_{ij}\}$ as outlined in Chap. 3, namely *Helmert type simultaneous estimation* of $\{\xi_i, \sigma_{ij}\}$ (B. Schaffrin 1983, p. 101).

An important generalization of the *standard multivariate Gauss–Markov model taking into account constraints*, for instance caused by rank definitions, e.g. the datum problem at r epochs, is the

multivariate Gauss–Markov model
with constraints

which we will treat at the end.

Definition 14.5. (multivariate Gauss–Markov model with constraints):

If in a multivariate model (14.1) and (14.2) the vectors ξ_i of parameters of first order are subject to constraints

$$\mathbf{H}\boldsymbol{\xi}_i = \mathbf{w}_i, \tag{14.17}$$

where $\mathbf{H}$ denotes the $r \times m$ matrix of known coefficients with the restriction

$$\mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A} = \mathbf{H}, \ \operatorname{rk}\mathbf{H} = r \leq m \tag{14.18}$$

and $\mathbf{w}_i$ known $r \times 1$ vectors, *then*

$$E\{\mathbf{y}_i\} = \mathbf{A}\xi_i, \tag{14.19}$$

$$D\{\mathbf{y}_i, \mathbf{y}_j\} = \mathbf{I}_n\sigma_{ij} \tag{14.20}$$

subject to

$$\mathbf{H}\boldsymbol{\xi}_i = \mathbf{w}_i \tag{14.21}$$

is called "*the multivariate Gauss–Markov model with linear homogeneous constraints*". If the p vectors $\mathbf{w}_i$ are collected in the $r \times p$ matrix $\mathbf{W}$, $\dim \mathbf{W} = r \times p$, the corresponding matrix model reads

$$E\{\mathbf{Y}\} = \mathbf{A}\Xi, \ D\{\mathrm{vec}\mathbf{Y}\} = \Sigma \otimes \mathbf{I}_n, \ \mathbf{H}\Xi = \mathbf{W} \tag{14.22}$$

subject to

$$\begin{aligned} &O\{\Sigma\} = p \times p, \ O\{\Sigma \otimes \mathbf{I}_n\} = np \times np, \\ &O\{\mathrm{vec}\mathbf{Y}\} = np \times 1, \ O\{\mathbf{Y}\} = n \times p \\ &O\{D\{\mathrm{vec}\mathbf{Y}\}\} = np \times np, \\ &O\{\mathbf{H}\} = r \times m, \ O\{\Xi\} = m \times p, \ O\{\mathbf{W}\} = r \times p. \end{aligned} \tag{14.23}$$

The *vector forms*

$$\boxed{\mathbf{E}\{\mathrm{vec}\mathbf{Y}\} = (\mathbf{I}_p \otimes \mathbf{A})\mathrm{vec}\Xi, \ \mathbf{D}\{\mathrm{vec}\mathbf{Y}\} = \Sigma \otimes \mathbf{I}_n, \ \mathrm{vec}\mathbf{W} = (\mathbf{I}_p \otimes \mathbf{H})\mathrm{vec}\Xi}$$

are equivalent to the *matrix forms*.

A key result is *Lemma 15.6* in which we solve for a given multivariate weight matrix $\mathbf{G}^{ij}$ – being *equivalent to* $(\Sigma \otimes \mathbf{I}_n)^{-1}$ – *a multivariate LESS problem.*

Theorem 14.6. (multivariate Gauss–Markov model with constraints):

A multivariate Gauss–Markov model with linear homogeneous constraints is $(\Sigma, \ \mathbf{I}_n)$-BLUUE if

$$\begin{aligned} \mathrm{vec}\hat{\Xi} = \ &(\mathbf{I}_p \otimes (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}')\mathrm{vec}\mathbf{Y} + \mathbf{Y}(\mathbf{I}_p \otimes (\mathbf{A}'\mathbf{A})\mathbf{H}'(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{H}')^{-1})\mathrm{vec}\mathbf{Y} \\ &-(\mathbf{I}_p \otimes (\mathbf{A}'\mathbf{A})^{-}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}')\mathrm{vec}\mathbf{Y} \end{aligned} \tag{14.24}$$

or

$$\hat{\Xi} = (\mathbf{A}'\mathbf{A})^{-}(\mathbf{A}'\mathbf{Y} + \mathbf{H}'(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{H}')^{-1}(\mathbf{W} - \mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{Y})) \tag{14.25}$$

An *unbiased estimation of the variance-covariance matrix* Σ is

$$\begin{aligned} \tilde{\Sigma} = &\frac{1}{n-m+r}\{(\mathbf{A}\hat{\Xi} - \mathbf{Y})'(\mathbf{A}\hat{\Xi} - \mathbf{Y}) \\ &+(\mathbf{H}\hat{\Xi} - \mathbf{W})'(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}')^{-1}(\mathbf{H}\hat{\Xi} - \mathbf{W})\}. \end{aligned} \tag{14.26}$$

End of Theorem 14.6

14-2 *n*-Way Classification Models

Another special model is called

n-way classification model.

We will define it and show how to solve its basic equations. Namely, we begin with the 1-way classification and to continue with the 2- and 3-way classification models. A specific feature of any classification model is the nature of the specific unknown vector which is either zero or one. The methods to solve the normal equation vary: *In one approach*, one assumption is that the unknown vector of zeros or ones is a *fixed effect*. The corresponding *normal equations* are solved by standard MINOLESS, weighted or not. *Alternatively*, one assumes that the parameter vector consists of *random effects*. *Methods of variance-covariance component estimation* are applied.

Here we only follow *a MINOLESS approach*, weighted or not. The interested reader of the alternative technique of *variance-covariance component estimation* is referred to *our Chap. 3* or to the literature, for instance *H. Ahrens and J. Laeuter* (1974) or *S.R. Searle* (1971), my favorite.

14-21 A First Example: 1-Way Classification

A *one-way classification model* is defined by

$$y_{ij} = E\{y_{ij}\} + e_{ij} = \mu + \alpha_i + e_{ij} \tag{14.27}$$

$$\mathbf{y}' := [\mathbf{y}'_1\, \mathbf{y}'_2 \ldots \mathbf{y}'_{p-1} \mathbf{y}'_p], \;\; \mathbf{x}' := [\mu\ \alpha_1\, \alpha_2 \ldots \alpha_{p-1}\, \alpha_p] \tag{14.28}$$

where the parameters μ and α_i are unknown. It is characteristic for the model that the coefficients of the unknowns are either one or zero. A MINOLESS (Minimum Norms LEast Squares Solution) for the unknown parameters μ, α_i is based on

$$||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathbf{I}} = \min_{\mathbf{x}} \text{ and } ||\mathbf{x}||^2_{\mathbf{I}} = \min_{\mathbf{x}}$$

we built around a *numerical example*.

Numerical example: 1-way classification

Here we will investigate data concerning the investment on consumer durables of people with different levels of education. Assuming that investment is measured by an index number, namely supposing that available data consist of values of this index for seven people: *Table 14.1* illustrates a very small example, but adequate for our purposes.

A suitable model for these data is

$$y_{ij} = \mu + \alpha_i + e_{ij}, \tag{14.29}$$

where y_{ij} is investment index of the jth person in the ith education level, μ is a general mean, α_i is the effect on investment of the ith level of education and e_{ij}

Table 14.1 (investment indices of seven people):

Level of education	Number of people	Indices	Total
1 (High School incomplete)	3	74, 68, 77	219
2 (High School graduate)	2	76, 80	156
3 (College graduate)	2	85,93	178
Total	7		553

is the random error term peculiar to y_{ij}. For the data of *Table 14.1* there are three educational levels and i takes the values $j = 1, 2, \ldots, n_i - 1,\ n_i$ where n_i is the number of observations in the ith educational level, in our case $n_1 = 3,\ n_2 = 2$ and $n_3 = 2$ in *Table 14.1*. Our model is the model for the 1-way classification. In general, the groupings such as educational levels are called classes and in our model y_{ij} as the response and levels of education as the classes, this is a model we can apply to many situations.

The normal equations arise from writing the data of *Table 14.1* in terms of our model equation.

$$\begin{bmatrix} 74 \\ 68 \\ \underline{77} \\ 76 \\ \underline{80} \\ 85 \\ 93 \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ \underline{y_{13}} \\ y_{21} \\ \underline{y_{22}} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 + e_{11} \\ \mu + \alpha_1 + e_{12} \\ \underline{\mu + \alpha_1 + e_{13}} \\ \mu + \alpha_2 + e_{21} \\ \underline{\mu + \alpha_2 + e_{22}} \\ \mu + \alpha_3 + e_{31} \\ \mu + \alpha_3 + e_{32} \end{bmatrix}, \ O(\mathbf{y}) = 7 \times 1$$

or

$$\begin{bmatrix} 74 \\ 68 \\ \underline{77} \\ 76 \\ \underline{80} \\ 85 \\ 93 \end{bmatrix} = \mathbf{y} = \begin{bmatrix} 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,0\,1\,0 \\ 1\,0\,1\,0 \\ 1\,0\,0\,1 \\ 1\,0\,0\,1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \underline{e_{13}} \\ e_{21} \\ \underline{e_{22}} \\ e_{31} \\ e_{32} \end{bmatrix} = \mathbf{Ax} + \mathbf{e_y}$$

and

$$\mathbf{A} = \begin{bmatrix} 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,1\,0\,0 \\ 1\,0\,1\,0 \\ 1\,0\,1\,0 \\ 1\,0\,0\,1 \\ 1\,0\,0\,1 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}, \ \ O\,(\mathbf{A}) = 7 \times 4, \ \ O(\mathbf{x}) = 4 \times 1$$

with $\mathbf{y}$ being the vector of observations and $\mathbf{e_y}$ the vector of corresponding error terms. As an inconsistent linear equation

$$\mathbf{y} - \mathbf{e_y} = \mathbf{Ax}, \quad O\{\mathbf{y}\} = 7 \times 1, \ O\{\mathbf{A}\} = 7 \times 4, \ O\{\mathbf{x}\} = 4 \times 1$$

we pose the key question:

?What is the rank of the design matrix A?

Most notable, the first column is $\mathbf{1}_n$ and the sum of the other three columns is also one, namely $\mathbf{c}_2 + \mathbf{c}_3 + \mathbf{c}_4 = \mathbf{1}_n$! Indeed, we have a proof for a linear dependence: $\mathbf{c}_1 = \mathbf{c}_2 + \mathbf{c}_3 + \mathbf{c}_4$. The rank $\text{rk}\mathbf{A} = 3$ is only three which differs from $O\{\mathbf{A}\} = 7 \times 4$. We have to build in this rank deficiency. For example, we could postulate the condition $x_4 = \alpha_3 = 0$ eliminating one component of the unknown vector. A more reasonable approach would be based on the computation of the

symmetric reflexive generalized inverse
such that

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}, \tag{14.30}$$

which would guarantee a *least squares minimum norm solution* or a $\mathbf{V}, \mathbf{S}$-BLUMBE solution (Best Linear $\mathbf{V}$-Norm Uniformly Minimum Bias $\mathbf{S}$-Norm Estimation) for $\mathbf{V} = \mathbf{I}, \mathbf{S} = \mathbf{I}$ and

$$\text{rk}\mathbf{A} = \text{rk}\mathbf{A}'\mathbf{A} = \text{rk}(\mathbf{A}'\mathbf{A})^-_{rs} = \text{rk}\mathbf{A}^+ \tag{14.31}$$

$\mathbf{A}'\mathbf{A}$ is a symmetric matrix $\Rightarrow$ $(\mathbf{A}'\mathbf{A})^-_{rs}$ is a symmetric matrix

or called

:*rank preserving identity*:
!*symmetry preserving identity*!

We intend to compute $\mathbf{x}_{lm}$ for our example (Table 14.2).

Summary

The general formulation of our *1-way classification problem* is generated by identifying the vector of *responses* as well as the vector of *parameters* (Table 14.3):

$$n = n_1 + n_2 + \ldots + n_p = \sum_{i=1}^{p} n_i \tag{14.32}$$

experimental design : number of observations	number of parameters :	rank of the design matrix :
$n = n_1 + n_2 + \ldots + n_p$	$1 + p$	$1 + (p - 1) = p$

(14.33)

Table 14.2 (1-way classification, example: normal equation):

$$\mathbf{A}'\mathbf{A} = \begin{bmatrix} 1&1&1&1&1&1&1 \\ 1&1&1&0&0&0&0 \\ 0&0&0&1&1&0&0 \\ 0&0&0&0&0&1&1 \end{bmatrix} \begin{bmatrix} 1&1&0&0 \\ 1&1&0&0 \\ 1&1&0&0 \\ 1&0&1&0 \\ 1&0&1&0 \\ 1&0&0&1 \\ 1&0&0&1 \end{bmatrix} = \begin{bmatrix} 7&3&2&2 \\ 3&3&0&0 \\ 2&0&2&0 \\ 2&0&0&2 \end{bmatrix}$$

$$\mathbf{A}'\mathbf{A} = \mathbf{DE}, \quad O\{\mathbf{D}\} = 4 \times 3, \quad O\{\mathbf{E}\} = 3 \times 4$$

$$\mathbf{D}'\mathbf{D} = \begin{bmatrix} 7&3&2 \\ 3&3&0 \\ 2&0&2 \\ 2&0&0 \end{bmatrix}, \ \mathbf{E}\,\textit{to be determined}:$$

$$\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{D}'\mathbf{DE} \ \Rightarrow \ (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{E}$$

$$\mathbf{D}'\mathbf{D} = \begin{bmatrix} 7&3&2&2 \\ 3&3&0&0 \\ 2&0&2&0 \end{bmatrix} \begin{bmatrix} 7&3&2 \\ 3&3&0 \\ 2&0&2 \\ 2&0&0 \end{bmatrix} = \begin{bmatrix} 66&30&18 \\ 30&18&6 \\ 18&6&8 \end{bmatrix}$$

$$\textit{:compute}\,(\mathbf{D}'\mathbf{D})^{-1}\ \textit{and}\ (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}':$$

$$\Rightarrow \ \mathbf{E} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A} = \begin{bmatrix} 1&0&0&1 \\ 0&1&0&-1 \\ 0&0&1&1 \end{bmatrix}$$

$$(\mathbf{A}'\mathbf{A})^{-}_{rs} = \mathbf{E}'(\mathbf{EE}')^{-1}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' =$$

$$= \begin{bmatrix} 0.0833 & 0 & 0.0417 & 0.0417 \\ 0 & 0.2500 & -0.1250 & -0.1250 \\ 0.0417 & -0.1250 & 0.3333 - 0.1667 & \\ 0.0417 & -0.1250 & -0.1667 & 0.3333 \end{bmatrix}$$

$$(\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}' =$$

$$\begin{bmatrix} 0.0833 & 0.0833 & 0.0833 & 0.1250 & 0.1250 & 0.1250 & 0.1250 \\ 0.2500 & 0.2500 & 0.2500 & -0.1250 & -0.1250 & -0.1250 & -0.1250 \\ -0.0833 & -0.0833 & -0.0833 & 0.3750 & 0.3750 & -0.1250 & -0.1250 \\ -0.0833 & -0.0833 & -0.0833 & -0.1250 & -0.1250 & 0.3750 & 0.3750 \end{bmatrix}$$

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}'\mathbf{y} = \begin{bmatrix} 60.0 \\ 13.0 \\ 18.0 \\ 29.0 \end{bmatrix}.$$

Table 14.3 (1-way classification):

$$\mathbf{y}' := [\mathbf{y}_{11}, \mathbf{y}_{12} \ldots \mathbf{y}_{1(n_1-1)}\mathbf{y}_{1n_1}|\mathbf{y}_{21}\mathbf{y}_{22} \ldots \mathbf{y}_{2(n_2-1)}\mathbf{y}_{2n_2}| \ldots |\mathbf{y}_{p1}\mathbf{y}_{p2} \ldots \mathbf{y}_{p(n_p-1)}\mathbf{y}_{(pn_p)}]$$

$$\mathbf{x}' := [\mu \;\; \alpha_1 \;\; \alpha_2 \;\; \ldots \alpha_{p-1} \;\; \alpha_p]$$

$$\mathbf{A} := \begin{bmatrix} 1 & 1 & 0 & 0 \\ \cdots & \cdots & & \cdots \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \cdots & \cdots & & \cdots \\ 1 & 0 & 1 & 0 \\ \cdots & \cdots & & \cdots \\ 1 & 0 & 0 & 1 \\ \cdots & \cdots & & \cdots \\ 1 & 0 & 0 & 1 \end{bmatrix}, \; O(\mathbf{A}) = n \times (p+1)$$

:MINOLESS:

$$||\mathbf{y} - \mathbf{A}\mathbf{x}||^2 = \min_{\mathbf{x}} \quad \text{and} \quad ||\mathbf{x}||^2 = \min \tag{14.34}$$

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}'\mathbf{y}. \tag{14.35}$$

14-22 A Second Example: 2-Way Classification Without Interaction

A *two-way classification model without interaction* is defined by

"MINOLESS"

$$y_{ijk} = E\{y_{ijk}\} + e_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \tag{14.36}$$

$$\mathbf{y}' := [\mathbf{y}'_{11}, \mathbf{y}'_{21}, \ldots, \mathbf{y}'_{p-1\,1}, \; \mathbf{y}'_{p1}, \mathbf{y}'_{12}, \mathbf{y}'_{22}, \ldots, \mathbf{y}'_{pq-1}, \mathbf{y}'_{pq}]$$

$$\mathbf{x}' = [\mu, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q] \tag{14.37}$$

$$||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathbf{I}} = \min_{\mathbf{x}} \quad \text{and} \quad ||\mathbf{x}||^2 = \min_{\mathbf{x}}. \tag{14.38}$$

Table 14.4 Level of factors

Level of the factor **B**	1	2	...	q
Level of factor **A**	n_{11}	n_{12}	...	n_{1q}
2	n_{21}	n_{22}	...	n_{2q}
...	...	...	...	...
p	n_{p1}	n_{p2}	...	n_{pq}

Table 14.5 Number of seconds beyond 3 minutes, taken to boil 2 quarts of water

		Make of Pan				Number of	
		A	B	C	total	observations	mean
	X	18	12	24	54	3	18
Brand of	Y	–	–	9	9	3	18
Stove	Z	3	–	15	18	3	18
	W	6	3	18	27	3	18
Total		27	15	66	108		
no. of. obs		3	2	4			
mean		9	$7\frac{1}{2}$	$16\frac{1}{2}$			

The *factor* **A** appears in p *levels* and the *factor* **B** in g *levels*. If n_{ij} denotes the number of observations under the influence of the i*th level* of the *factor* **A** and the jth level of the *factor* **B**, then the results of the experiment can be condensed in *Table 14.4*. If α_i and β_y denote the effects of the factors A and B, μ the mean of all observations, we receive

$$\mu+\alpha_i+\beta_j = E\{\mathbf{y}_{ijk}\} \text{ for all } i \in \{1, \dots, p\}, \; j \in \{1, \dots, q\}, \; k \in \{1, \dots, n_{ij}\} \tag{14.39}$$

as our model equation.

If $n_{ij} = 0$ for at least one pair$\{i, j\}$, then our experimental design is called incomplete. An experimental design for which n_{ij} is equal of all pairs $\{ij\}$, is said to be balanced. The data of Table 14.5 describe such a general model of $\mathbf{y}_{ijk}$ observations in the ith row (brand of stove) and jth column (make of the pan), μ is the mean, α_i is the effect of the ith row, β_j is the effect of the jth column, and e_{ijk} is the error term.

Outside the context of rows and columns α_i is equivalently the effect due to the ith level of the α factor and β_j is the effect due to the jth level of the β factor. In general, we have p levels of the α factor with $i = 1, \dots, p$ and q levels of the β factor with $j = 1, \dots, q$: in our example $p = 4$ and $q = 3$.
With balanced data every one of the pq cells in Table 14.5 would have one (or n) observations and $n \leq 1$ would be the only symbol needed to describe the number of observations in each cell. In our Table 14.5 some cells have zero observations and some have one. We therefore need n_{ij} as the number of observations in the ith row and jth column. Then all $n_{ij} = 0$ or 1, and the number of observations are the

values of

$$n_i = \sum_{j=1}^{q} n_{ij}, \; n_j = \sum_{i=1}^{p} n_{ij}, \; n = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}. \tag{14.40}$$

Corresponding totals and means of the observations are shown, too. For the observations in *Table 14.5* the linear equations of the model are given as follows,

$$\begin{bmatrix} 18 \\ 12 \\ \underline{24} \\ \underline{9} \\ 3 \\ \underline{15} \\ 6 \\ 3 \\ 18 \end{bmatrix} = \begin{bmatrix} y_{11} \\ y_{12} \\ \underline{y_{13}} \\ \underline{y_{23}} \\ y_{31} \\ \underline{y_{33}} \\ y_{41} \\ y_{42} \\ y_{43} \end{bmatrix} = \begin{bmatrix} 1\,1\cdot\cdot\cdot 1\cdot\cdot \\ 1\,1\cdot\cdot\cdot\cdot 1\cdot \\ 1\,1\cdot\cdot\cdot\cdot\cdot 1 \\ 1\cdot 1\cdot\cdot\cdot\cdot 1 \\ 1\cdot\cdot 1\cdot 1\cdot\cdot \\ 1\cdot\cdot 1\cdot\cdot\cdot 1 \\ 1\cdot\cdot\cdot 1\,1\cdot\cdot \\ 1\cdot\cdot\cdot 1\cdot 1\cdot \\ 1\cdot\cdot\cdot 1\cdot\cdot 1 \end{bmatrix} \begin{bmatrix} \underline{\mu} \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \underline{\alpha_4} \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{23} \\ e_{31} \\ e_{33} \\ e_{41} \\ e_{42} \\ e_{43} \end{bmatrix},$$

where dots represent zeros. In summary,

$$\begin{bmatrix} 18 \\ 12 \\ \underline{24} \\ \underline{9} \\ 3 \\ \underline{15} \\ 6 \\ 3 \\ 18 \end{bmatrix} = \mathbf{y} = \begin{bmatrix} 11000100 \\ 11000010 \\ 11000001 \\ 10100001 \\ 10010100 \\ 10010001 \\ 10001100 \\ 10001010 \\ 10001001 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \underline{e_{13}} \\ e_{23} \\ e_{31} \\ \underline{e_{33}} \\ e_{41} \\ e_{42} \\ e_{43} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1&1&0&0&0&1&0&0 \\ 1&1&0&0&0&0&1&0 \\ 1&1&0&0&0&0&1&1 \\ 1&0&1&0&0&0&0&1 \\ 1&0&0&1&0&1&0&0 \\ 1&0&0&1&0&0&0&1 \\ 1&0&0&0&1&1&0&0 \\ 1&0&0&0&1&0&1&0 \\ 1&0&0&0&1&0&0&1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \quad O(\mathbf{A}) = 9 \times 8, \quad O(\mathbf{x}) = 8 \times 1$$

with $\mathbf{y}$ being the vector of observations and $\mathbf{e_y}$ the vector of corresponding error terms. As an *inconsistent linear equation*

$$\mathbf{y} - \mathbf{e_y} = \mathbf{Ax},\ O\{\mathbf{y}\} = 9 \times 1,\ O\{\mathbf{A}\} = 9 \times 8,\ O\{\mathbf{x}\} = 8 \times 1$$

we pose the key question:

? What is the rank of the design matrix $\mathbf{A}$?

Most notable, the first column is $\mathbf{1}_n$ and the sum of the next 4 columns is also $\mathbf{1}_n$ as well as the sum of the remaining 3 columns is $\mathbf{1}_n$, too, namely $c_2+c_3+c_4+c_5 = \mathbf{1}_n$ and $c_6 + c_7 + c_8 = \mathbf{1}_n$. The rank $\mathrm{rk}\mathbf{A} = 1 + (p-1) + (q-1) = 1 + 3 + 2 = 6$ is only six which differs from $O\{\mathbf{A}\} = 9 \times 8$. *We have to take advantage of this rank deficiency*. For example, *we could postulate the condition* $\mathbf{x}_5 = 0$ and $\mathbf{x}_8 = 0$ eliminating two components of the unknown vector. A more reasonable approach would be based on the computation of the *symmetric reflexive generalized inverse*

such that

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}, \tag{14.41}$$

which would guarantee a least square minimum norm solution or a $\mathbf{I}$, $\mathbf{I}$-BLUMBE solution (Best Linear $\mathbf{I}$-Norm Uniformly Minimum Bias $\mathbf{I}$-Norm Estimation) and

$$\mathrm{rk}\mathbf{A} = \mathrm{rk}\mathbf{A}'\mathbf{A} = \mathrm{rk}(\mathbf{A}'\mathbf{A})^-_{rs} = \mathrm{rk}\mathbf{A}' \tag{14.42}$$

$\mathbf{A}'\mathbf{A}$ is a symmetric matrix $\Rightarrow$ $(\mathbf{A}'\mathbf{A})^-_{rs}$ is a symmetric matrix

or called

:rank preserving identity:
:*symmetry preserving identity* (Table 14.6):

Summary

The general formulation of our *2-way classification problem without interaction* is generated by identifying the vector of *responses* as well as the vector of *parameters* (Table 14.7).

subject to

$$c_2 + \ldots + c_p = \mathbf{1},\ c_{p+1} + \ldots + c_q = \mathbf{1}$$

$$n_i = \sum_{j=1}^{q} n_{ij},\ n_j = \sum_{i=1}^{p} n_{ij},\ n = \sum_{i=1,\,j=1}^{p,\,q} n_{ij} \tag{14.43}$$

Table 14.6 2-way classification without interaction, example: normal equation

$$
\mathbf{A}'\mathbf{A} = \begin{bmatrix} 1&1&1&1&1&1&1&1&1 \\ 1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&1&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1 \\ 1&0&0&0&1&0&1&0&0 \\ 0&1&0&0&0&0&0&1&0 \\ 0&0&1&1&0&1&0&0&1 \end{bmatrix}
\begin{array}{c} \begin{array}{cccccccc} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_1 & \beta_2 & \beta_3 \end{array} \\
\begin{bmatrix} 1&1&0&0&0&1&0&0 \\ 1&1&0&0&0&0&1&0 \\ 1&1&0&0&0&0&0&1 \\ 1&0&1&0&0&0&0&1 \\ 1&0&0&1&0&1&0&0 \\ 1&0&0&1&0&0&0&1 \\ 1&0&0&0&1&1&0&0 \\ 1&0&0&0&1&0&1&0 \\ 1&0&0&0&1&0&0&1 \end{bmatrix} \end{array}
$$

$$
= \begin{bmatrix} 9&3&1&2&3&3&2&4 \\ 3&3&0&0&0&1&1&1 \\ 1&0&1&0&0&0&0&1 \\ 2&0&0&2&0&1&0&1 \\ 3&0&0&0&3&1&1&1 \\ 3&1&0&1&1&3&0&0 \\ 2&1&0&0&1&0&2&0 \\ 4&1&1&1&1&0&0&4 \end{bmatrix}
$$

$$
\mathbf{A}'\mathbf{A} = \mathbf{DE}, \quad O\{\mathbf{D}\} = 8 \times 6, \quad O\{\mathbf{E}\} = 6 \times 8
$$

$$
\mathbf{D} = \begin{bmatrix} 9&3&1&2&3&2 \\ 3&3&0&0&1&1 \\ 1&0&1&0&0&0 \\ 2&0&0&2&1&0 \\ 3&0&0&0&1&1 \\ 3&1&0&1&3&0 \\ 2&1&0&0&0&2 \end{bmatrix}, \ \mathbf{E}\ \textit{to be determined}
$$

$$
\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{D}'\mathbf{DE} \Rightarrow (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{E}
$$

$$
\mathbf{D}'\mathbf{D} = \begin{bmatrix} 133&45&14&29&44&28 \\ 45&21&4&8&15&11 \\ 14&4&3&3&3&2 \\ 29&8&3&10&11&4 \\ 44&15&3&11&21&8 \\ 28&11&2&4&8&10 \end{bmatrix}
$$

compute $(\mathbf{D}'\mathbf{D})^{-1}$ and $(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$

$$\mathbf{E} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A}$$

$$= \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 1.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & -10.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & -1.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & -1.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & -1.0000 \end{bmatrix}$$

$$(\mathbf{A}'\mathbf{A})^{-}_{rs} = \mathbf{E}'(\mathbf{E}\mathbf{E}')\,(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$$

$$= \begin{bmatrix} 0.0665 & -0.0360 & 0.1219 & 0.0166 & -0.0360 & 0.0222 & 0.0748 & -0.0305 \\ -0.0360 & 0.3112 & -0.2327 & -0.0923 & -0.0222 & -0.0120 & -0.0822 & 0.0582 \\ 0.1219 & -0.2327 & 0.8068 & -0.2195 & -0.2327 & 0.1240 & 0.1371 & -0.1392 \\ 0.0166 & -0.0923 & -0.2195 & 0.4208 & -0.0923 & -0.0778 & 0.1020 & -0.0076 \\ -0.0360 & -0.0222 & -0.2327 & -0.0923 & 0.3112 & -0.0120 & -0.0822 & 0.0582 \\ 0.0222 & -0.0120 & 0.1240 & -0.0778 & -0.0120 & 0.2574 & -0.1417 & -0.0935 \\ 0.0748 & -0.0822 & 0.1371 & 0.1020 & -0.0822 & -0.1417 & 0.3758 & -0.1593 \\ -0.0305 & 0.0582 & -0.1392 & -0.0076 & 0.0582 & -0.0935 & -0.1593 & 0.2223 \end{bmatrix}$$

$$(\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}$$

$$= \begin{bmatrix} 0.0526 & 0.1053 & 0.0000 & 0.1579 & 0.1053 & 0.0526 & 0.0526 & 0.1053 & 0.0000 \\ 0.2632 & 0.1930 & 0.3333 & -0.2105 & -0.1404 & -0.0702 & -0.0702 & -0.1404 & 0.0000 \\ 0.0132 & 0.0263 & -0.2500 & 0.7895 & 0.0263 & -0.2368 & 0.0132 & 0.0263 & -0.2500 \\ 0.1535 & 0.0263 & -0.0833 & -0.2105 & 0.3596 & 0.4298 & -0.1535 & 0.0263 & -0.0833 \\ -0.0702 & -0.1404 & 0.0000 & -0.2105 & -0.1404 & -0.0702 & 0.2632 & 0.1930 & 0.3333 \\ 0.2675 & -0.1316 & -0.0833 & 0.0526 & 0.2018 & -0.1491 & 0.2675 & -0.1316 & -0.0833 \\ -0.1491 & 0.3684 & -0.1667 & 0.0526 & 0.0351 & 0.0175 & -0.1491 & 0.3684 & -0.1667 \\ -0.0658 & -0.1316 & 0.2500 & 0.0526 & -0.1316 & 0.1842 & -0.0658 & -0.1316 & 0.2500 \end{bmatrix}$$

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}'\mathbf{y} = \begin{bmatrix} 5.3684 \\ 10.8421 \\ -6.1579 \\ -1.1579 \\ 1.8421 \\ -0.2105 \\ -4.2105 \\ 9.7895 \end{bmatrix}.$$

Table 14.7 (2-way classification without interaction):

$$\mathbf{y}' := [\mathbf{y}'_{11},\ \ldots,\ \mathbf{y}'_{p1},\ \mathbf{y}'_{12},\ \ldots,\ \mathbf{y}'_{pq-1}\ \mathbf{y}'_{pq}]$$
$$\mathbf{x}' := [\mu,\ \alpha_1,\ \ldots,\ \alpha_p,\ \beta_1,\ \ldots,\ \beta_q]$$
$$\mathbf{A} := [\mathbf{1}_n,\ c_2,\ \ldots,\ c_p,\ c_{p+1},\ \ldots,\ c_q]$$

experimental design : number of observations	number of parameters :	rank of the design matrix :
$n = \sum_{i,j=1}^{p,q} n_{ij}$	$1+p+q$	$1+(p-1)+(q-1) = p+q-1$

(14.44)

:*MINOLESS*:

$$||\mathbf{y}-\mathbf{A}\mathbf{x}||^2 = \min_{\mathbf{x}} \quad \text{and} \quad ||\mathbf{x}||^2 = \min \tag{14.45}$$

$$\mathbf{x}_{lm} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}.$$

14-23 A Third Example: 2-Way Classification with Interaction

A two-way *classification model with interaction* is defined by

"*MINOLESS*"

$$y_{ijk} = E\{y_{ijk}\} + e_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk} \tag{14.46}$$

subject to

$$i \in \{1,\ldots,p\},\ j \in \{1,\ldots,q\}, k \in \{1,\ldots,n_{ij}\}$$
$$\mathbf{y}' := [\mathbf{y}'_{11},\ldots,\mathbf{y}'_{p-1},\mathbf{y}'_{p},\mathbf{y}'_{12}\mathbf{y}'_{22},\ldots\mathbf{y}'_{nq-1}\mathbf{y}'_{pq}]$$
$$\mathbf{x}' := [\mu,\ \alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q,(\alpha\beta)_{11},\ldots,(\alpha\beta)_{pq}] \tag{14.47}$$
$$||\mathbf{y}-\mathbf{A}\mathbf{x}||^2_{\mathrm{I}} = \min_{\mathbf{x}} \text{and} ||\mathbf{x}||^2_{\mathrm{I}} = \min_{\mathbf{x}}. \tag{14.48}$$

It was been in the second example on 2-way classification *without interaction* that the effects of different levels of the factors A and B were *additive*. An alternative model is a model in which the *additivity does not hold*: the observations *are not independent of each other*. Such a model is called a model with interaction between the factors whose effect $(\alpha\beta)_{ij}$ has to be reflected by means of

$$\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} = \mathrm{E}\{y_{ijk}\} \text{ for all } \begin{bmatrix} i \in \{1,\ldots,p\} \\ j \in \{1,\ldots,q\} \\ k \in \{1,\ldots,n_{ij}\} \end{bmatrix} \tag{14.49}$$

like our model equation. As an example we consider by means of *Table 14.8* a *plant breeder* carrying out a series of *experiments with three fertilizer* treatments on each of *four varieties of grain*. For each treatment-by-variety combination when he or she plants several $4' \times 4'$ plots. At harvest time she or he finds that many of the plots have been lost due to being wrongly ploughed up and all he or she is left with are the data of *Table 14.8*.

With four of the treatment-in-variety combinations there are no data at all, and with the others there are varying numbers of plots, ranging from 1 to 4 with a total of 18 plots in all. *Table 14.8* shows the yield of each plot, the total yields, the number of plots in each total and the corresponding mean, for each treatment-variety combination having data. Totals, numbers of observations and means are also shown for the three treatments, the four varieties and for all 18 plots. The symbols for the entries in the table, are also shown in terms of the model.

The equations of a suitable linear model for analyzing data of the nature of *Table 14.8* is for y_{ijk} as the kth observation in the ith treatment and jth variety. In our top table, μ is the mean, α_i is the effect of the ith treatment, β_j is the effect of the jth variety, $(\alpha\beta)_{ij}$ is the interaction effect for the ith treatment and the jth variety and ℓ_{ijk} is the error term. With balanced data every one of pq cells of our table would have n observations. In addition there would be pq levels of the $(\alpha\beta)$ factor, the interaction factor. However, with unbalanced data, when some cells have no observations they are only as many $(\alpha\beta)_{ij}$-levels in the data as there are non-empty cells. Let the number of such cells be s (s = 8 in Table 14.8). Then, if n_{ij}

Table 14.8 Weight of grain form $4' \times 4'$ trial plots

			Variety		
Treatment	1	2	3	4	Totals
1	8		12	7	
	13	-		11	
	9		–	–	
	30		12	18	60
	$y_{11}(n_{11})$		$y_{31}(n_{13})$	$y_{41}(n_{14})$	
2	6	12			
	12	14	-	-	
	18	26			
3	-	9	14	10	
		7	16	14	
				11	
		–	–	13	
		16	30	48	94
Total	48	42	42	66	198

is the number of observations in the (i, j)th cell of type "treatment i and variety j", s the number of cells in which $n_{ij} \neq 0$, in all other cases $n_{ij} > 0$. For these cells

$$y_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk}, \bar{y}_{ij} = y_{ij}/n_{ij} \tag{14.50}$$

is the *total yield* in the (i, j)th cell, and $\bar{y}_{ij}$ is the *corresponding mean*. Similarly,

$$\boxed{y = \sum_{i=1}^{p} y_i = \sum_{j=1}^{q} y_j = \sum_{i,j=1}^{p,q} y_{ij} = \sum_{i=1,j=1,k=1}^{p,q,n_{ij}} y_{ijk}} \tag{14.51}$$

is the total yield for all plots, the number of observations called "plots" therein being

$$\boxed{n = \sum_{i=1}^{p} n_i = \sum_{j=1}^{q} n_j = \sum_{i,j}^{p,q} n_{ij}.} \tag{14.52}$$

We shall continue with the corresponding *normal equations* being derived from the *observational equations*.

$$\begin{bmatrix} 8\\13\\9\\12\\7\\11\\6\\12\\12\\14\\9\\7\\14\\16\\10\\14\\11\\13 \end{bmatrix} = \begin{bmatrix} y_{111}\\y_{112}\\y_{113}\\y_{131}\\y_{141}\\y_{142}\\y_{211}\\y_{212}\\y_{221}\\y_{222}\\y_{321}\\y_{322}\\y_{331}\\y_{332}\\y_{341}\\y_{342}\\y_{343}\\y_{344} \end{bmatrix} = \begin{array}{c} \mu\,\alpha_1\;\alpha_2\;\alpha_3\,\beta_1\,\beta_2\;\beta_3\,\beta_4\,\overbrace{11\;13\;14\;21\;22\;31\;33\;34}^{(\alpha\beta)} \\ \begin{bmatrix}
1&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot\\
1&\cdot&1&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot\\
1&\cdot&1&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot\\
1&\cdot&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot\\
1&\cdot&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot\\
1&\cdot&\cdot&1&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot\\
1&\cdot&\cdot&1&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot&\cdot\\
1&\cdot&\cdot&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot\\
1&\cdot&\cdot&1&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1&\cdot\\
1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1\\
1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1\\
1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1\\
1&\cdot&\cdot&1&\cdot&\cdot&\cdot&1&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot&1
\end{bmatrix} \end{array} \cdot \begin{bmatrix} \underline{\mu}\\ \alpha_1\\ \alpha_2\\ \underline{\alpha_3}\\ \beta_1\\ \beta_2\\ \beta_3\\ \underline{\beta_4}\\ (\alpha\beta)_{11}\\ (\alpha\beta)_{13}\\ (\alpha\beta)_{14}\\ (\alpha\beta)_{21}\\ (\alpha\beta)_{22}\\ (\alpha\beta)_{32}\\ (\alpha\beta)_{33}\\ (\alpha\beta)_{34} \end{bmatrix} + \begin{bmatrix} e_{111}\\e_{112}\\e_{113}\\e_{131}\\e_{141}\\e_{142}\\e_{211}\\e_{212}\\e_{221}\\e_{222}\\e_{321}\\e_{322}\\e_{331}\\e_{332}\\e_{341}\\e_{342}\\e_{343}\\e_{344} \end{bmatrix}$$

where the *dots represent zeros*.

$$
\left[\begin{array}{cccccccccccccccc}
18 & 6 & 4 & 8 & 5 & 4 & 3 & 6 & 3 & 1 & 2 & 2 & 2 & 2 & 2 & 4 \\
\hline
6 & 6 & \cdot & \cdot & 3 & \cdot & 1 & 2 & 3 & 1 & 2 & \cdot & \cdot & \cdot & \cdot & \cdot \\
4 & \cdot & 4 & \cdot & 2 & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & 2 & \cdot & \cdot & \cdot \\
8 & \cdot & \cdot & 8 & \cdot & 2 & 2 & 4 & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & 2 & 4 \\
\hline
5 & 3 & 2 & \cdot & 5 & \cdot & \cdot & \cdot & 3 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot \\
4 & \cdot & 2 & 2 & \cdot & 4 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & 2 & \cdot & \cdot \\
3 & 1 & \cdot & 2 & \cdot & \cdot & 3 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 2 & \cdot \\
6 & 2 & \cdot & 4 & \cdot & \cdot & \cdot & 6 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & 4 \\
\hline
3 & 3 & \cdot & \cdot & 3 & \cdot & \cdot & \cdot & 3 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
2 & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot \\
2 & \cdot & 2 & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot \\
2 & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot & \cdot & \cdot \\
2 & \cdot & \cdot & 2 & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot & \cdot \\
2 & \cdot & \cdot & 2 & \cdot & \cdot & 2 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2 & \cdot \\
4 & \cdot & \cdot & 4 & \cdot & \cdot & \cdot & 4 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 4
\end{array}\right]
\left[\begin{array}{c}
\mu \\ \hline \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \hline \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \hline (\alpha\beta)_{11} \\ (\alpha\beta)_{13} \\ (\alpha\beta)_{14} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{32} \\ (\alpha\beta)_{33} \\ (\alpha\beta)_{34}
\end{array}\right]
=
\left[\begin{array}{c}
y \\ \hline y_1 \\ y_2 \\ y_3 \\ \hline y_{\cdot 1} \\ y_{\cdot 2} \\ y_{\cdot 3} \\ y_{\cdot 4} \\ \hline y_{11} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{32} \\ y_{33} \\ y_{34}
\end{array}\right]
=
\left[\begin{array}{c}
198 \\ \hline 60 \\ 44 \\ 94 \\ \hline 48 \\ 42 \\ 42 \\ 66 \\ \hline 30 \\ 12 \\ 18 \\ 18 \\ 26 \\ 16 \\ 30 \\ 48
\end{array}\right]
\sim \mathbf{A}'\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{y}.
$$

Now we again pose the key question:

?What is the rank of the design matrix $\mathbf{A}$?

The first column is $\mathbf{1}_n$ and the sum of other columns is $c_2 + c_3 + c_4 = \mathbf{1}_n$ and $c_5 + c_6 + c_7 + c_8 = \mathbf{1}_n$. How to handle the remaining sum $(\alpha\beta)\ldots$ of our incomplete model? Obviously, we experience $\mathrm{rk}[c_9, \ldots, c_{16}] = 8$, namely

$$\mathrm{rk}[(\alpha\beta)_{ij}] = 8 \text{ for } (\alpha\beta)_{ij} \in \{\gamma_{11}, \ldots, \gamma_{34}\}. \tag{14.53}$$

As a summary, we have computed $\mathrm{rk}(\mathbf{A}'\mathbf{A}) = 8$, a surprise for our special case. A more reasonable approach would be based on the computation of the *symmetric reflexive generalized inverse* such that

$$\mathbf{x}_{\ell m} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}, \tag{14.54}$$

which would assure a *minimum norm*, *least squares solution* or a $\mathbf{I}, \mathbf{I}$-BLUMBE solution (**B**est **L**inear **I**-**N**orm **U**niformly **M**inimum **B**ias **I**-Norm **E**stimation) and

$$\mathrm{rk}\mathbf{A} = \mathrm{rk}\mathbf{A}'\mathbf{A} = \mathrm{rk}(\mathbf{A}'\mathbf{A})^-_{rs} = \mathrm{rk}\mathbf{A}^+ \tag{14.55}$$

$\mathbf{A}'\mathbf{A}$ is a symmetric matrix $\Rightarrow$ $(\mathbf{A}'\mathbf{A})^-_{rs}$ is a symmetric matrix

or called

Table 14.9 (2-way classification with interaction):

$$\mathbf{y}' := [\mathbf{y}'_{11}, \ldots, \mathbf{y}'_{p1}, \mathbf{y}'_{12}, \ldots, \mathbf{y}'_{pq-1}, \mathbf{y}'_{pq}]$$
$$\mathbf{x}' := [\mu, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, (\alpha\beta)_{11}, \ldots, (\alpha\beta)_{pq}]$$
$$\mathbf{A} := [\mathbf{1}_n, c_1, \ldots, c_p, c_{p+1}, \ldots, c_q, c_{11}, \ldots, c_{pq}]$$

subject to

$$c_2 + \ldots + c_p = \mathbf{1}, c_{p+1} + \ldots + c_q = \mathbf{1}, \sum_{i,j=1}^{p,q} c_{ij} = (p-1)(q-1)$$

$$n = \sum_{i=1}^{p} n_i = \sum_{j=1}^{q} n_j = \sum_{i,j}^{p,q} n_{i,j}$$

:rank preserving identity:
!symmetry preserving identity!

Table 14.9 summarizes all the details of 2-way classification with interaction. In general, for *complete models* our table lists the general number of parameters and the rank of the design matrix which differs from our *incomplete design model.*

experimental design : number of observations	number of parameters :	rank of the design matrix :
$n = \sum_{i,j}^{p,q} n_{i,j}$	$1 + p + q + pq$	$1 + (p-1) + (q-1) + $ $+(p-1)(q-1)$ (14.56)

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 = \min_{\mathbf{x}} \text{ and } \|\mathbf{x}\|^2 = \min_{\mathbf{x}} \tag{14.57}$$

$$\mathbf{x}_{\ell m} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}. \tag{14.58}$$

For our key example we get from the *symmetric normal equation* $\mathbf{A}'\mathbf{A}\mathbf{x}_\ell = \mathbf{A}'\mathbf{y}$ the solution

$$\mathbf{x}_{\ell m} = (\mathbf{A}'\mathbf{A})^-_{rs}\mathbf{A}'\mathbf{y}$$
$$\textit{given}\, \mathbf{A}'\mathbf{A} \text{and} \mathbf{A}'\mathbf{y}$$
$$O\{\mathbf{A}'\mathbf{A}\} = 16 \times 16, O\{\mu, \ldots, (\alpha\beta)_{31}\} = 16 \times 1, O\{\mathbf{A}'\mathbf{y}\} = 16 \times 1$$
$$\mathbf{A}'\mathbf{A} = \mathbf{DE}, O\{\mathbf{D}\} = 18 \times 12, O\{\mathbf{E}\} = 12 \times 16$$

$$\mathbf{D} = \begin{bmatrix} 3 & 1 & 2 & 2 & 2 & 2 & 2 & 4 \\ 3 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 & 4 \\ 3 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 4 \\ 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

$$\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{D}'\mathbf{D}\mathbf{E} \Rightarrow (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A} = \mathbf{E}$$

$$\mathbf{E} = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A} =$$

$$\begin{bmatrix} 1.0000 & 1.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \\ 1.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0 & 0.0000 \\ 1.0000 & 0.0000 & 1.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 1.0000 & 0 & 1.0000 & 0.0000 & 0 & 1.0000 & 0.0000 \\ 1.0000 & 0 & 0 & 1.0000 & 0 & 1.0000 & 0 \\ 1.0000 & 0.0000 & 0.0000 & 1.0000 & 0 & 0.0000 & 1.0000 \\ 1.0000 & 0 & 0.0000 & 1.0000 & 0 & 0.0000 & 0 \end{bmatrix}$$

$$\mathbf{x}_{\ell m} = (\mathbf{A}'\mathbf{A})^{-}_{rs}\mathbf{A}'\mathbf{y} =$$
$$= [6.4602, 1.5543, 2.4425, 2.4634, 0.6943, 1.0579, 3.3540, 1.3540,$$
$$1.2912, 0.6315, -0.3685, -0.5969, 3.0394, -1.9815, 2.7224, 1.72245]'.$$

14-24 Higher Classifications with Interaction

If we generalize *1-way* and *2-way classifications with interactions* we arrive at a *higher classification* of type

$$\mu + \alpha_i + \beta_j + \gamma_k + \ldots +$$

$$+ (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\beta)_{jk} + \ldots + (\alpha\beta\gamma)_{ijk} + \ldots = E\{\mathbf{y}_{ijk\ldots\ell}\} \tag{14.59}$$

for all

$$i \in \{1,\ldots,p\}, j \in \{1,\ldots,q\}, k \in \{1,\ldots,r\},\ldots,\ell \in \{1,\ldots,n_{ijk},\ldots\}.$$

An alternative stochastic model assumes a fully occupied variance–covariance matrix of the observations, namely $\mathbf{D}\{\mathbf{y}\} \sim \mathbf{E}\{[\mathbf{y} - \mathbf{E}\{\mathbf{y}\}][\mathbf{y} - \mathbf{E}\{\mathbf{y}\}]'\} \sim \Sigma_{ij}$. Variance–covariance estimation techniques are to be applied. In addition, a mixed model for the effects are be applied, for instance of type

$$E\{\mathbf{y}\} = \mathbf{A}\xi + \mathbf{C}E\{\mathbf{z}\} \tag{14.60}$$

$$D\{\mathbf{y}\} = \mathbf{C}D\{\mathbf{z}\}\mathbf{C}'. \tag{14.61}$$

Here we conclude with a discussion of what is *unbiased estimable*:

Example (1-way classification):

If we depart from the model $E\{\mathbf{y}_{ij}\} = \mu + \alpha_i$ and note $\text{rk}\mathbf{A} = \text{rk}\mathbf{A}'\mathbf{A} = 1 + (p-1)$ for $i \in \{1,\ldots,p\}$, namely $\text{rk}\mathbf{A} = p$, we realize that the $1 + p$ *parameters* are *not* unbiased estimable: *The first column results from the sum of the other columns*. It is obvious the *difference* $\alpha_i - \alpha_1$ is *unbiased estimable*. This difference produces a *column matrix with full rank*.

Summary $\alpha_i - \alpha_1$ *quantities are unbiased estimable*

Example (2-way classification with interaction):

Our first statement relates to the unbiased estimability of the terms $\alpha_1,\ldots,\alpha_p$ and $\beta_1,\ldots,\beta_q$: obviously, the differences $\alpha_i - \alpha_1$ and $\beta_j - \beta_1$ for $i, j < 1$ are unbiased estimable. The first column is the sum of the other terms. For instance, the second column can be eliminated which is equivalent to estimating $\alpha_i - \alpha_1$ in order to obtain a design matrix of full column rank. The same effect can be seen with the other effect β_j for the properly chosen design matrix: $\beta_j - \beta_1$ for all $j > 1$ is unbiased estimable! If we add the pq effect $(\alpha\beta)_{ij}$ of interactions, only those interactions increase the rank of the de-sign matrix by one respectively, which refer to the differences $\alpha_i - \alpha_1$ and $\beta_j - \beta_1$, altogether $(p-1)(q-1)$ interactions. To the effect $(\alpha\beta)_{ij}$ of the interactions are estimable, $pq - (p-1)(q-1) = p + q - 1$ constants may be added, that is to the interactions

$$(\alpha\beta)_{i1}, (\alpha\beta)_{i2}, \ldots, (\alpha\beta)_{iq}$$

with

$$i \in \{1, \ldots, p\}$$

the *constants*

$$\Delta(\alpha\beta_{\cdot 1}), \Delta(\alpha\beta_{\cdot 2}), \ldots, \Delta(\alpha\beta_{\cdot q})$$

and to the interactions

$$(\alpha\beta)_{2j}, (\alpha\beta)_{3j}, \ldots, (\alpha\beta)_{pj}$$

with

$$j \in \{1, \ldots, q\}$$

the *constants*

$$\Delta(\alpha\beta_{2\cdot}), \Delta(\alpha\beta_{3\cdot}), \ldots, \Delta(\alpha\beta_{p\cdot}).$$

The constants $\Delta(\alpha\beta_{1\cdot})$ need not to be added which can be interpreted by $\Delta(\alpha\beta_{\cdot 1}) = \Delta(\alpha\beta_{1\cdot})$. A *numerical example* is

$$p = 2, q = 2$$

$$\mathbf{x}' = [\mu, \alpha_1, \alpha_{2,} \beta_1, \beta_2, (\alpha\beta)_{11}, (\alpha\beta)_{12}, (\alpha\beta)_{22}] \,.$$

Summary

$$\Delta\alpha = \alpha_2 - \alpha_1,$$

$$\Delta\beta = \beta_2 - \beta_1 \, for \, all \, i \in \{1, 2\}, j \in \{1, 2\} \tag{14.62}$$

as well as

$$\Delta(\alpha\beta_{\cdot 1}), \Delta(\alpha\beta_{\cdot 2}), \Delta(\alpha\beta_{2\cdot}) \tag{14.63}$$

are unbiased estimable.

At the end we review the number of parameters and the rank of the design matrix for a 3-way classification with interactions according to the following example.

$$\boxed{\text{3-way classification with interactions}}$$

experimental design : number of observations	*number of parameters :*	*rank of the design matrix :*
$n = \sum_{i,j,k=1}^{p,q,r} n_{ijk}$	$1 + p + q + r + \\ + pq + pr + qr + \\ + pqr$	$1 + (p-1) + (q-1) + (r-1) \\ +(p-1)(q-1)+ \\ +(p-1)(r-1) + (q-1)(r-1)+ \\ (p-1)(q-1)(r-1) = pqr$

(14.64)

14-3 Dynamical Systems

There are two essential items in the analysis of *dynamical systems: First*, there exists a "*linear or linearized observational equation* $\mathbf{y}(t) = \mathbf{Cz}(t)$" connecting a vector of *stochastic observations* $\mathbf{y}$ to a stochastic vector $\mathbf{z}$ of so called "*state variables*". *Second*, the other essential is the characteristic differential equation of type "$\mathbf{z}'(t) = F(t, \mathbf{z}(t))$", especially *linearizied* "$\mathbf{z}'(t) = \mathbf{Az}(t)$ ", which maps the *first derivative* of the "*state variable*" to the "*state variable*" its off. Both, $\mathbf{y}(t)$and$\mathbf{z}(t)$ are functions of a parameter, called "time t". The second equation describes the *time development of the dynamical system*. An alternative formulation of the dynamical system equation is "$\mathbf{z}(t) = A\mathbf{z}(t-1)$ ". Due to the *random nature* "of the two functions" $\mathbf{y}(t) = \mathbf{Cz}(t)$ and $\mathbf{z}' = \mathbf{Az}$ the complete equations read

$$E\{\mathbf{y}(t)\} = \mathbf{C}E\{\mathbf{z}\}\, and \left[\begin{array}{c} \mathbf{V}\{\mathbf{e}_\mathbf{y}(t_1), \mathbf{e}_\mathbf{y}(t_2)\} = \Sigma_\mathbf{y}(t_1, t_2) \\ D\{\mathbf{e}_\mathbf{y}(t)\} = \Sigma_\mathbf{y}(t), \end{array} \right. \tag{14.65}$$

$$E\{\mathbf{z}'(t)\} = \mathbf{A}E\{\mathbf{z}(t)\}\, and \left[\begin{array}{c} D\{e_\mathbf{z}(t)\} = \Sigma_\mathbf{z}(t), \\ \mathbf{V}\{\mathbf{e}_\mathbf{z}(t_1), \mathbf{e}_\mathbf{z}(t_2)\} = \Sigma_\mathbf{z}(t_1, t_2) \end{array} \right. \tag{14.66}$$

Here we only introduce "*the time invariant system equations*" characterized by $\mathbf{A}(t) = \mathbf{A}$. $\mathbf{z}'(t)$ abbreviates *the functional* $d\mathbf{z}(t)/dt$. There may be the case that the variance-covariance functions $\Sigma_\mathbf{y}(t)$ and $\Sigma_\mathbf{z}(t)$ do not change in time: $\Sigma_\mathbf{y}(t) = \Sigma_\mathbf{y}$, $\Sigma_\mathbf{z}(t) = \Sigma_\mathbf{z}$ equal a constant. Various models exist for the variance-covariance functions $\Sigma_\mathbf{y}(t_1, t_2)$ and $\Sigma_\mathbf{z}(t_1, t_2)$, e.q. linear functions as in the case of a Gauss process or a Brown process $\Sigma_\mathbf{y}(t_2 - t_1)$, $\Sigma_\mathbf{z}(t_2 - t_1)$. The analysis of dynamic system theory was initiated by *R.E. Kalman* (1960) and by *R.E. Kalman* and *R.S. Bucy* (1961): "KF" stand for "*Kalman filtering*".

Example 1 (tracking a satellite orbit):

Tracking a satellite's orbit around the Earth might be based on the *unknown state vector* $\mathbf{z}(t)$ being a function of the position and the speed of the satellite at time t with respect to a spherical coordinate system with origin at the mass center of the Earth. Position and speed of a satellite can be measured by GPS, for instance. If distances and accompanying angles are measured, they establish the observation $\mathbf{y}(t)$. The principles of space-time geometry, namely mapping $\mathbf{y}(t)$ into $\mathbf{z}(t)$, would be incorporated in the matrix $\mathbf{C}$ while $\mathbf{e}_\mathbf{y}(t)$ would reflect the *measurement errors* at the time instant t. The matrix $\mathbf{A}$ reflects the situation how position and speed change in time *according the physical lows* governing orbiting bodies, while e_z would allow for deviation from the lows owing to factors as *none uniformity of the Earth gravity field*.

Example 2 (statistical quality control):

Here the observation vector $\mathbf{y}(t)$ is a simple approximately normal transformation of the *number of derivatives* observed in a sample obtained at time t, while $y_1(t)$

and $y_2(t)$ represent respectively the refractive index of the process and the drift of the index. We have the *observation equation* and the *system equations*

$$y(t) = z_1(t) + e_y(t_1) \;\; and \;\; \begin{array}{c} z_1(t) = z_2(t) + e_{z1} \\ z_2(t) = z_2(t-1) + e_{z2}\,. \end{array}$$

In vector notation, *this system of equation* becomes

$$\mathbf{z}(t) = \mathbf{A}\mathbf{z}(t-1) + \mathbf{e_z}$$

namely

$$\mathbf{z}(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix},\; \mathbf{e_z} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} e_{\mathbf{z}_1} \\ e_{\mathbf{z}_2} \end{bmatrix},\; \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

do *not* change in time.

If we examine $y(t) - y(t-1)$ for this model, we observe that under the *assumption of constant variance*, namely $\mathbf{e_y}(t) = \mathbf{e_y}$ and $\mathbf{e_z}(t) = \mathbf{e_z}$, the *autocorrelation structure* of the difference is identical to that of an ARIMA (0,1,1) *process*. Although such a correspondence is sometimes easily discernible, we should in general *not* consider the two approaches to be equivalent.

A stochastic process is called an *ARMA process of the order* (p, q) if

$$\begin{gathered} \mathbf{z}(t) = a_1\mathbf{z}(t-1) + a_2\mathbf{z}(t-2) + \ldots + a_p\mathbf{z}(t-p) = \\ = b_0u(t) + b_1u(t-1) + b_2u(t-2) + \ldots + b_qu(t-q) \\ \text{forall} t \in \{1, \ldots, T\} \end{gathered} \tag{14.67}$$

also called a *mixed autoregressive/moving – average process*.

In practice, most time series are non-stationary. In order to fit a *stationary model*, it is necessary to remove *non-stationary sources of variation*. If the observed time series is non-stationary in the mean, then we can use the difference of the series. Differencing is widely used in all *scientific disciplines*. If $z(t),\; t \in \{1, \ldots, T\}$, is replaced by $\nabla^d z(t)$, then we have a model capable of describing certain types of *non-stationary signals*. Such a model is called an “integrated model” because the stationary model that *fits to the difference data has to the summed or “integrated”* to provide a model for *the original non-stationary data*. Writing

$$W(t) = \nabla^d z(t) = (1-B)^d z(t) \tag{14.68}$$

$$\text{for all } t \in \{1, \ldots, T\}$$

the *general autoregressive integrated moving-average* (ARIMA) process is of the form

$$\boxed{\text{ARIMA}}$$

$$W(t) = \alpha_1 W(t-1) + \ldots + \alpha_p W(t-p) + b_0 u(t) + \ldots + b_q u(t-q) \quad (14.69)$$

or

$$\Phi(B)\mathrm{W}(t) = \Gamma(B) \quad (14.70)$$

$$\Phi(B)(1-B)^d z(t) = \Gamma(B)u(t). \quad (14.71)$$

Thus we have an ARMA (p,q) model for $W(t), t \in \{1,\ldots,T\}$, while the model for $W(t)$ describing the dth differences for $z(t)$ is said to be an ARIMA process of order (p,d,q). *For our case*, ARIMA (0,1,1) *means a specific process* $p = 1$, $\nabla = 1$, $q = 1$. The model for $z(t)$, $t \in \{1,\ldots,T\}$, is *clearly non-stationary*, as the AR operators $\Phi(B)(1-q)^d$ has d roots on the unit circle since patting $B = 1$ makes the AR operator equal to zero. In practice, first differencing is often found to the adequate to make a series stationary, and accordingly the value of d *is often taken to be one*. Note the *random part* could be considered as an ARIMA (0,1,0) process.

It is a special problem of time-series analysis that the error variances are generally *not* known a priori. This can be dealt with by *guessing*, and *then updating them in an appropriate way*, or, *alternatively*, *by estimating then forming a set of data over a suitable fit period*.

In the state space modeling, the prime objective is to predict the *signal in the presence of noise*. In other words, we want to estimate the $m \times 1$ state vector $E\{\mathbf{z}(t)\}$ *which cannot usually be observed directly*. The *Kalman filter* provides a general method for doing this. It consists of a set of equations that allow us to *update the estimate* of $E\{\mathbf{z}(t)\}$ when a *new observation* becomes available. We will outline this updating procedure *with two stages*, called

- The prediction stage
- The updating stage.

Suppose we have observed a *univariate time series* up to time $(t-1)$, and that $E\{z(t-1)\}$ is "*the best estimator*" $E\{z(t-1)\}$ based on information up to this time. For instance, "*best*" is defined as an PLUUP *estimator*. Note that $z(t)$, $z(t-1)$ etc is a *random variable*. Further suppose that we have evaluated the $m \times m$ variance-covariance matrix of $E\{\widehat{z(t-1)}\}$ which we denote by $P\{t-1\}$. The *first stage* called the *prediction stage* is concerned with forecasting $E\{z(t)\}$ from data up to the time $(t-1)$, and we denote the resulting estimator in a obvious notation by $E\{\widehat{z(t)}|z(t-1)\}$. Considering the *state equations* where $D\{e_z(t-1)\}$ is still unknown at time $(t-1)$ the obvious estimator for $E\{z(t)\}$ is given by

$$E\{\mathbf{z}(t)|\hat{\mathbf{z}}(t-1)\} = G(t)E\{\widehat{\mathbf{z}(t-1)}\} \quad (14.72)$$

and the variance-covariance matrix

$$\mathbf{V}\{t|t-1\} = G(t)\mathbf{V}\{t-1\}G' + W\{t\} \quad (14.73)$$

called *prediction equations*. The last equations follows from the standard results on variance-covariance matrices for random vector variables. When the new observation at time t, namely when $y(t)$ has been observed the estimator for $E\{\mathbf{z}(t)\}$ can be modified to take *account* of the *extra information*. At time $(t-1)$, the best forecast of $\mathbf{y}(t)$ is given by $\mathbf{h}'\ E\{\mathbf{z}(t)|\hat{\mathbf{z}}(t-1)\}$ so that the *prediction error* is given by

$$\hat{e}_y(t) = y(t) - \mathbf{h}'(t)E\{\mathbf{z}(t)|\mathbf{z}(t-1)\}. \tag{14.74}$$

This quantity can be used to *update the estimate* of $E\{z(t)\}$ and its variance-covariance matrix.

$$E\{\hat{\mathbf{z}}(t)\} = E\{\mathbf{z}(t)|\hat{\mathbf{z}}(t-1)\} + K(t)\hat{\mathrm{e}}_{\mathrm{y}} \tag{14.75}$$

$$\mathbf{V}\{t\} = \mathbf{V}\{t|t-1\} - \mathbf{K}(t)\mathbf{h}'(t)\mathbf{V}\{t|t-1\} \tag{14.76}$$

$$\mathbf{V}\{t\} = \mathbf{V}\{t, t-1\}\mathbf{h}'(t)/[\mathbf{h}'(t)\mathbf{V}\{t|t-1\}\mathbf{h} + \sigma_n^2]. \tag{14.77}$$

$\mathbf{V}\{t\}$ is called the *gain matrix*. In the univariate case, $\mathbf{K}(t)$ is just a vector of size $(m-1)$. The previous equations constitute the second *updating stage* of the *Kalman filter*, thus they are called the *updating equations*.

A major practical advantage of the *Kalman filter* is that the calculations are *re-cursive* so that, although the current estimates are based on the whole past history of measurements, *there is no need for an ever expanding memory*. Rather the near estimate of the signal is *based solely on the previous estimate and the latest observations*. A *second advantage of the Kalman filter* is that it converges fairly quickly when there is a constant underlying model, but can also follow the movement of a system where the underlying model is evolving through time.

For special cases, there exist much simpler equations. An example is to consider the *random walk plus noise model* where the state vector $\mathbf{z}(t)$ consist of one state variable, the *current level* $\mu(t)$. It can be shown that the *Kalman filter* for this model in the steady state case for $t \to \infty$ reduces the simple recurrence relation

$$\hat{\mu}(t) = \hat{\mu}(t-1) + \alpha\hat{\mathrm{e}}(t),$$

where the smoothing constant α is a complicated function of the signal-to-noise ratio σ_w^2/σ_n^2. Our equation is *simple exponential smoothing*. When σ_w^2 tends to zero, $\mu(t)$ is a constant and we find that $\alpha \to 0$ would intuitively be expected, while as σ_w^2/σ_n^2 becomes large, then α approaches unity.

For a *multivariate time series* approach we may start from the *vector-valued equation* of type

$$E\{\mathbf{y}(t)\} = \mathbf{C}E\{\mathbf{z}\}, \tag{14.78}$$

where $\mathbf{C}$ is a known nonsingular $m \times m$ matrix.

By LESS we are able to predict

$$E\{\widehat{\mathbf{z}(t)}\} = \mathbf{C}^{-1}E\{\widehat{\mathbf{y}(t)}\}. \tag{14.79}$$

Once a model has been put into the *state-space form, the Kalman filter* can be used to *provide estimates of the signal*, and they in turn lead to algorithms for various other calculations, such as *making prediction* and handling missing values. For *instance, forecasts* may be obtained from the *state-space model* using the latest estimates of the state vector. Given data to time N, the best estimate of the state vector is written $E\{\widehat{\mathbf{z}(N)}\}$ and the h-step-ahead forecast is given by

$$\begin{aligned} E\{\widehat{\mathbf{y}(N)}\} &= \mathbf{h}'(N+\mathbf{h})E\{\widehat{\mathbf{z}(N+\mathbf{h})}\} \\ &= \mathbf{h}(N+\mathbf{h})G(N+\mathbf{h})G\{N+\mathbf{h}-1\}\ldots G(N+1)E\{\widehat{\mathbf{z}(N)}\} \end{aligned} \tag{14.80}$$

where we assume $\mathbf{h}\ (N+\mathbf{h})$ and future values of $G(t)$ are *known*. *Of course*, if $G(t)$ is a constant, say G, then we get

$$E\{\widehat{\mathbf{y}(N|\mathbf{h})}\} = \mathbf{h}'(N+\mathbf{h})G^{\mathbf{h}}E\{\widehat{\mathbf{z}(N)}\}. \tag{14.81}$$

If future values of $\mathbf{h}(t)\text{or}\mathbf{G}(t)$ are *not* known, then they must themselves be *forecasted* or *otherwise guessed*.

Up to this day a lot of research has be been done on nonlinear models in prediction theory relating to state-vectors and observational equations. There are excellent reviews, for instance by *P. H. Frances* (1988), *C. W. J. Granger and P. Newbold* (1986), *A. C. Harvey* (1993), *M. B. Priestley* (1981, 1988) and *H. Tong* (1990). *C. W. Granger and T. Teräsvirta* (1993) is a more advanced text.

In terms of the dynamical system theory we regularly meet the problem that the *observational equation is not of full column rank*. A state variable leads to a relation between the system input-output solution, especially a statement on how a system is developing in time. Very often it is reasonable to switch from a state variable, in one reference system *to another one with special properties*. Let $\mathbf{T}$ this time be a *similarity transformation*, namely described by a non-singular matrix of type

$$\mathbf{z} := \mathbf{T}\mathbf{z}^* \sim \mathbf{z}^* = \mathbf{T}^{-1}\mathbf{z} \tag{14.82}$$

$$\frac{d}{dt}\mathbf{z}^* = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}\mathbf{z}^* + \mathbf{T}^{-1}\mathbf{B}\mathbf{u}(t), \mathbf{z}_0^* = \mathbf{T}^{-1}\mathbf{z}_0 \tag{14.83}$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{T}\mathbf{z}^*(t) + \mathbf{D}\mathbf{u}(t). \tag{14.84}$$

The key question is now whether to the characteristic state equation there belongs a *transformation matrix* such that for a specific matrix $\mathbf{A}^*$ and $\mathbf{B}^*$ there exists an integer number r, $0 \leq r < n$, of the form

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11}^* & \mathbf{A}_{12}^* \\ 0 & \mathbf{A}_{22}^* \end{bmatrix}, \quad O\{\mathbf{A}^*\} = \begin{bmatrix} r \times r & r \times (n-r) \\ (n-1) \times r & (n-r) \times (n-r) \end{bmatrix}$$

$$\mathbf{B}^* = \begin{bmatrix} B_1^* \\ 0 \end{bmatrix}, \quad O\{\mathbf{B}^*\} = \begin{bmatrix} q \times r \\ q \times (n-r) \end{bmatrix}.$$

In this case the state equation separates in two distinct parts.

$$\frac{d}{dt}\mathbf{z}_1^*(t) = \mathbf{A}_{11}^*\mathbf{z}_1^* + \mathbf{A}_{12}^*\mathbf{z}_2^*(t) + B_1^* h(t), \mathbf{z}_1^*(0) = \mathbf{z}_{10}^* \tag{14.85}$$

$$\frac{d}{dt}\mathbf{z}_2^*(t) = \mathbf{A}_{22}^*\mathbf{z}_2^*(t), \;\; \mathbf{z}_2^*(0) = \mathbf{z}_{20}^*. \tag{14.86}$$

The last $n-r$ elements of $\mathbf{z}^*$ cannot be influenced in its time development. Influence is restricted to the *initial conditions* and to the *eigen dynamics of the partial system* 2 (characterized by the matrix $\mathbf{A}_{22}^*$). The state of the whole system cannot be influenced completely by the artificially given point of the state space. Accordingly, the *state differential equation* in terms of the matrix pair $(\mathbf{A}, \mathbf{B})$ is not steerable.

Example 3 (steerable state):

If we apply the dynamic matrix $\mathbf{A}$ and the *introductory matrix* of a *state model* of type

$$\mathbf{A} = \begin{bmatrix} -\frac{4}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{5}{3} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix},$$

we are led to the alternative matrices after using the *similarity transformation*

$$\mathbf{A}^* = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix}, \quad \mathbf{B}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

If the *initial state* is located along the z_1^*-axis, for instance $z_{20}^* = 0$, then the *state vector remains all times along this axis*. It is only possible to move this axis along a straight line "up and down".

In case that there exists *no similarity transformation* we call the state matrices $(\mathbf{A}, \mathbf{B})$ *steerable*. Steerability of a state differential equation may be tested by

Lemma 14.7. (Steerability):

The pair $(\mathbf{A}, \mathbf{B})$ is steerable if and only if

$$\mathrm{rk}[\mathbf{BAB}\ldots\mathbf{A}^{n-1}\mathbf{B}] = \mathrm{rk}\mathbf{F}(\mathbf{A}, \mathbf{B}) = n. \tag{14.87}$$

$\mathbf{F}(\mathbf{A}, \mathbf{B})$ is called matrix of steerability. If its rank $r < n$, then there exists a transformation T such that $\mathbf{A}^* = \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ and $\mathbf{B}^* = \mathbf{T}^{-1}\mathbf{B}$ has the form

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11}^* & \mathbf{A}_{12}^* \\ 0 & \mathbf{A}_{22}^* \end{bmatrix}, \mathbf{B}^* = \begin{bmatrix} \mathbf{B}_1^* \\ 0 \end{bmatrix} \tag{14.88}$$

and $(\mathbf{A}_{11}^*, \mathbf{B}_1^*)$ is *steerable*.

Alternatively we could search for a transformation matrix $\mathbf{T}$ such that transforms the *dynamic matrix* and the *exit matrix* of a state model to the form

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11}^* & 0 \\ \mathbf{A}_{21}^* & \mathbf{A}_{22}^* \end{bmatrix}, \quad O\{\mathbf{A}^*\} = \begin{bmatrix} r \times r & r \times (n-r) \\ (n-1) \times r & (n-r) \times (n-r) \end{bmatrix}$$

$$\mathbf{C}^* = [\mathbf{C}_1^*, 0], \quad O\{\mathbf{C}^*\} = [r \times p, (n-r) \times p]\,.$$

In this case the *state equation* and the *observational equations* read

$$\frac{d}{dt}\mathbf{z}_1^*(t) = \mathbf{A}_{11}^*\mathbf{z}_1^*(t) + \mathbf{B}_1^* u(t), \mathbf{z}_1^*(0) = \mathbf{z}_{10}^* \tag{14.89}$$

$$\frac{d}{dt}\mathbf{z}_2^*(t) = \mathbf{A}_{21}^*\mathbf{z}_1^*(t) + \mathbf{A}_{22}^*\mathbf{z}_2^*(t) + \mathbf{B}_2^* u(t), \;\; \mathbf{z}_2^*(0) = \mathbf{z}_{20}^* \tag{14.90}$$

$$\mathbf{y}(t) = \mathbf{C}_2^*\mathbf{z}_1^*(t) + Du(t). \tag{14.91}$$

The last $n-r$ elements of the vector $\mathbf{z}^*$ are not used in the *exit variable* $\mathbf{y}$. Since they *do not have an effect to* $\mathbf{z}_1^*$, the vector $\mathbf{g}$ contains *no information* of the component of the *state vector*. This state moves in the $n-r$ dimensional subspace of $\mathbb{R}^n$ *without any change in the exit variables*. Our model $(\mathbf{C}, \mathbf{A})$ is in this case called *non-observable*.

Example 4 (observability):

If the *exit matrix* and the *dynamic matrix* of a state model can be characterized by the matrices

$$\mathbf{C} = \begin{bmatrix} \frac{4}{5} & -\frac{2}{3} \end{bmatrix}, \; \mathbf{A} = \begin{bmatrix} 0 & -1 \\ 2 & -3 \end{bmatrix},$$

an application of the transformation matrix $\mathbf{T}$ leads to the matrices

$$\mathbf{C}^* = [1\;,\;\; 0]\,, \;\; \mathbf{A}^* = \begin{bmatrix} -1 & 0 \\ 1 & -2 \end{bmatrix}.$$

For an arbitrary motion of the state in the direction of the $\mathbf{z}_2^*$ axis *has no influence* on the *existing variable*. If there *does not exist a transformation* $\mathbf{T}$, we call the state vector *observable*.

A rank study helps again!

Lemma 14.8. (Observability test):

The pair $(\mathbf{C}, \mathbf{A})$ is observable if and only if

$$\mathrm{rk}\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{C}^{n-1} \end{bmatrix} = \mathrm{rk}\mathbf{G}(\mathbf{C}, \mathbf{A}) = n. \tag{14.92}$$

$\mathbf{G}(\mathbf{C}, \mathbf{A})$ is called observability matrix. If its rank $r < n$, then there exists a transformation matrix T such that $\mathbf{A}^* = \mathbf{T}^{-1}\mathbf{AT}$ and $C^* = C\mathbf{T}$ is of the form

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_{11}^* & \mathbf{0} \\ \mathbf{A}_{21}^* & \mathbf{A}_{22}^* \end{bmatrix}, \quad O\{\mathbf{A}^*\} = \begin{bmatrix} r \times r & r \times (n-r) \\ (n-1) \times r & (n-r) \times (n-r) \end{bmatrix} \tag{14.93}$$

$$\mathbf{C}^* = [\mathbf{C}_1^*, \mathbf{0}], \quad O\{\mathbf{C}^*\} = [r \times p, (n-r) \times p] \tag{14.94}$$

and C_1^*, A_{11}^* is *observable*.

With *Lemma 14.7* and *Lemma 14.8* we can only state whether a state model is *steerable or observable* or not, or *which dimension has a partial system* being classified as non-steerable and non-observable. In order to determine which part of a system is *non-steerable or non-observable* – which *eigen motion* is *not* excited or non-observable – we have to be able to construct proper transformation matrices $\mathbf{T}$. A tool is the PBH-test we do *not* analyze here.

Both the *state differential equation* as well as the *initial equation* we can *Laplace transform* easily. We only need the relations between the input, output and state variable via polynom matrices. If the initial conditions z_0 vanish, we get the *Laplace transformed* characteristical equations

$$(s\mathbf{I}_n - \mathbf{A})\mathbf{z}(s) = \mathbf{Bu}(s) \tag{14.95}$$

$$\mathbf{y}(s) = \mathbf{Cz}(s) + \mathbf{Du}(s). \tag{14.96}$$

For details we recommend to check the reference list. We only refer to *solving both the state differential equation as well as the initial* equation: Eliminating the state vector $\mathbf{z}(s)$ lead us to the *algebraic relation* between $\mathbf{u}(s)$and$\mathbf{y}(s)$:

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \tag{14.97}$$

or

$$\mathbf{G}(s) = \left[\mathbf{C}_1^* \mathbf{C}_2^*\right] \left(s \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & \mathbf{I}_{n-r} \end{bmatrix} - \begin{bmatrix} \mathbf{A}_{11}^* & \mathbf{A}_{12}^* \\ 0 & \mathbf{A}_{22}^* \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{B}_1^* \\ 0 \end{bmatrix} + \mathbf{D} \tag{14.98}$$

$$= \mathbf{C}_1^* (s\mathbf{I}_n - \mathbf{A}_{11}^*)^{-1} \mathbf{B}_1^* + \mathbf{D}.$$

Recently, the topic of chaos has attracted much attention. *Chaotic behavior* arises from certain types of *nonlinear models*, and a loose definition is apparently *random behavior* that is generated by a purely deterministic, nonlinear system.

Refer to the contributions of *K.S. Chan and H. Tong* (2001), *J. Gleick* (1987), *V. Isham* (1983), *H. Kants and I. Schreiber* (1997).

Chapter 15
Algebraic Solutions of Systems of Equations

Linear and Nonlinear Systems of Equations.

15-1 Introductory Remarks

Algebraic techniques of *Groebner bases* and *Multipolynomial resultants* are presented in this Chapter as efficient tools for solving explicitly systems of linear and nonlinear equations. Similar to the Gauss elimination technique applied to linear systems of equations, *Groebner bases* and *Multipolynomial resultants* are useful in eliminating several variables in multivariate systems of nonlinear equations in such a manner that the end product results into *univariate polynomial* equations whose roots are readily determined using existing functions such as the *roots* command in MATLAB.

The capability of *Groebner bases* and *Multipolynomial resultants* to solve explicitly nonlinear geodetic problems enables them to be used as the computational engine in the *Gauss–Jacobi combinatorial algorithm* proposed by *C.F. Gauss* (published posthumously, e.g., *Appendix D-3*) and *C.G.I. Jacobi* (1841) to solve the *nonlinear Gauss–Markov model*. By converting nonlinear geodetic observation equations into algebraic (polynomial) form, the *Gauss–Jacobi combinatorial algorithm* solves an overdetermined system of equations in two steps as follows:

1. In the first step, all the m combinations of minimal subsets of the observation equations are formed, and the n unknowns of each subset rigorously solved by means of Groebner bases or Multipolynomial resultants. The m solution sets are then represented as points in an n-dimensional space $\mathbb{R}^n$.
2. In a second step, the solution of the overdetermined Gauß-Markov-model is constructed as a weighted arithmetic mean of those m solution points, where the weights are obtained from the Error propagation law/variance-covariance propagation.

This chapter presents the theory of Groebner basis, multipolynomial resultants and the Gauss–Jacobi combinatorials methods, and is organized as follows. Section 15-2 provides the background to the algebraic and numerical methods used to derive exact solutions. In Sect. 15-3, the algebraic algorithms needed to solve the nonlinear system of equations are presented, and enriched with examples to help the reader understand them. The algorithms are finally applied successfully to three real-time geodetic problems in Sect. 15-4; transforming in a closed form geocentric coordinates to Gauss ellipsoidal coordinates (geodetic) through Minimum Distance Mapping (MDM); solution of GNSS pseudorange equations (using the GPS satellite

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2_15,
© Springer-Verlag Berlin Heidelberg 2012

vehicle as example); and to obtain the seven datum transformation parameters from two sets of coordinates. By means of *Groebner basis*, the *scale parameter* (in the seven parameters datum transformation problem) is shown to fulfill be a *univariate algebraic equation of fourth order (quartic polynomial)*, while the rotation parameters are functions in scale and the coordinates differences. What we present in this chapter is a nutshell of algebraic treatment of systems of polynomial equations. For a complete coverage, we refer to *J.L. Awange and E.W. Grafarend* (2005), and *J.L. Awange et al.* (2010).

15-2 Background to Algebraic Solutions

In Geodesy, Photogrammetry and Computer Vision, *nonlinear equations* are often encountered in several applications as they often relate the observations (measurements) to the unknown parameters to be determined. In case the number of observations n and the number of unknowns m are equal ($n = m$), the unknown parameters may be obtained by solving explicitly or in a closed form the system of equations relating observations to the unknown parameters. For example, *D. Cox et al.* (1998, pp. 28–32) illustrated that for systems of equations with exact solution, the system becomes vulnerable to small errors introduced during root findings and in case of extending the partial solution to the complete solution of the system, the errors may accumulate and thus become so large. If the partial solution is derived by iterative procedures, then the errors incurred during the root-finding may blow up during the extension of the partial solution to the complete solution (back substitution).

In some applications, symbolic rather than numerical solution are desired. In such cases, explicit procedures are usually employed. The resulting symbolic expressions often consists of univariate polynomials relating the unknown parameters (unknown variables) to the known variables (observations). By inserting known values into these univariate polynomials, numerical solutions are readily computed for the unknown variables. Advantages of explicit solutions are listed, e.g., by *E. L. Merritt* (1949) as; *provision of satisfaction to the users (Photogrammetrist and Mathematicians) of the methods*, *provision of data tools for checking the iterative methods*, *desired by Geodesist whose task of control densification does not favor iterative procedures*, *provision of solace* and *the requirement of explicit solutions rather than iterative by some applications*. In Geodesy for example, the *Minimum Distance Mapping* problem considered by *E. Grafarend* and *P. Lohse* (1991) relates a point on the Earth's topographical surface uniquely (one-to-one) to a point on the *International Reference Ellipsoid*. The solution of such an optimization problem requires that the equations be solved explicitly.

The draw back that was experienced with explicit solutions was that they were like rare jewels. The reason for this was partly because the methods required extensive computations for the results to be obtained, and partly because the resulting symbolic expressions were too large and required computers with large storage capacity. Until recently, the computers that were available could hardly handle large computations due to lack of faster Central Processing Unit (CPU), shortage of Random Access Memory (RAM) and limited hard disk space for storage. The other set back experienced by the explicit procedures was that some of the methods, especially those from algebraic fields, were formulated based on theoretical concepts that were hard to realize or comprehend without the help of computers. For a long time, these setbacks hampered progress of the explicit procedures. The advances made in computer technology in recent years, however, has changed the tides and led to improvements in explicit computational procedures which hitherto were difficult to achieve. Apart from the improvements in existing computational procedures, new computational techniques are continuously being added to the increasing list of computational methods with an aim of optimizing computational speed and effeiciency (see, e.g., Palancz and Awange 2012). In this category are the algebraic methods of *Groebner bases* and *Multipolynomial resultants*.

This Chapter presents algebraic computational techniques of *Groebner bases* and *Multipolynomial resultants* as suitable tools for solving explicitly nonlinear systems of observation equations that have been converted into algebraic (polynomial) forms in Geodesy and Geoinformatics. The algebraic techniques of *Groebner bases* and *Sylvester resultant* (resultant of two polynomials) for solving polynomial equations in Geodesy have been mentioned and examples of their applications to the two-dimensional case given in the work of *P. Lohse* (1994, pp. 36–39, 71–76). We will demonstrate using examples how the *Groebner bases* and *Multipolynomial resultants* (resultant of more than two polynomials) can be used to solve problems of three-dimensional nature inherent in Geodesy. *E. Grafarend* (1989) already suggested the use of *Groebner bases* approach in solving the perspective center problem in photogrammetry.

Other than revolutionizing computation procedures, the advances made in computer technology have also led to improvement in instrumentation for data acquisition as exemplified in the case of GPS positioning satellites. Since its inception as a positioning tool, the *Global Positioning System (GPS)* – referred to by *E. Grafarend* and *J. Shan* (1996) as the *Global Problem Solver* – has revolutionized geodetic positioning techniques and maintained its supremacy as a positioning tool. These improvements on instrumentation for data acquisition have led to improved data collection procedures and increase in accuracy. Lengthy geodetic procedures such as triangulation that required a lot of time, intervisibility between stations, and a large manpower are rapidly being replaced by satellite positioning techniques, which require shorter observation periods, no intervisibility requirement, weather independent and less manpower leading to optimization of time and money. Such improvements can be seen by the number of Global Navigation Satellite Systems (GNSS) being launched by various countries. These includes; Chinese's Compass/Beidou and Russian's GLONASS (see, e.g., Wellenhof 2008; Awange 2012). Whereas improved instrumentation is applauded, it comes a long with its own difficulties. One of the difficulties is that a lot of data is collected than required to determine unknown parameters leading to redundancies. In positioning with GPS for example, due to its constellation that offer a wider coverage, more than four satellites can be observed at any point of the earth. In the minimal case, only four satellites are required to fix the receiver position and the receiver clock bias assuming that the transmitter clock bias and the transmitter-receiver bias have been neglected for the *pseudo-range* type of observations. More than four satellites as is nowadays the case, therefore, leads to superfluous observations. In such cases, where $n > m$, the explicit solutions give way to optimization procedures such as *least squares solution* which work very well for *linear models* under specified conditions discussed in other chapters of this book.

In Geodesy and Geoinformatics, however, the observation equations are normally *nonlinear* thus requiring the use of *nonlinear Gauss–Markov model*, which is normally solved either by first *linearizing* the observation equations using Taylor series expansion to the second order terms about approximate values of the unknowns then applying linear model estimation procedures, or by using *iterative procedures* such as the *Gauss–Newton* approach. The *linearization approach* has the disadvantage that the linearized approximation of the nonlinear models may still suffer from nonlinearity and thus resulting in the estimates of such models being far from the real estimates of the *nonlinear models*. This can easily be checked by re-substituting the estimates from the *linearized model* into the original nonlinear model.

For the *iterative procedures,* the greatest undoing may be the requirement of the initial approximate values to start off the iterations, which may not be available for some applications. For simper models, the approximate initial values may be computed, for others however, the approximate values may be impossible to compute. Apart from the problem of getting the initial approximate values, there also exists the problem that poor choice of approximate values may lead to lack of convergence, or if the approximate values are far from the real solutions, then a large number of iterations may be required to get close to the real solutions thus rendering the whole procedure to be quite slow, especially where multiple roots are available. For other procedures, such as the *7-parameter datum transformation* that requires *linearization* and *iterative* methods, it is not feasible to take into account the stochasticity of both systems involved.

Clearly, a procedure for solving *nonlinear Gauss–Markov model* that can avoid the requirement of initial approximate starting values for *iteration* and *linearization* approaches, and also take into consideration the *stochasticity* of the systems involved is the desire of modern day geodesist and geoinformatist (see, e.g., *J. L. Awange* and *E. W. Grafarend* 2005; *J. L. Awange et al.* 2010). With this background, this chapter aims at answering the following questions:

- For geodetic, geoinformatics and Earth Science related problems requiring explicit solutions, can the algebraic tools of *Groebner bases* and *Multipolynomial resultants* that have found applications in other fields such as Robotics (for kinematic modelling of robots), Visions, Computer Aided Design (CAD), Engineering (offset surface construction in solid modelling), Computer Science (automated theorem proving) be used to solve *systems of nonlinear observation equations* of algebraic (polynomial) type?

- Is there any alternative for solving the *nonlinear Gauss–Markov model* without resorting to *linearization* or *iterative procedures* that require approximate starting values?

To answer the first question, the chapter uses the *Groebner bases* and *Multipolynomial resultants* to solve explicitly three geodetic problems of; *Minimum Distance Mapping*, *GPS four-point pseudo-ranging*, and the *7-parameter transformation problem* in Sect. 15-4. The answer to the second problem becomes clear once the first question has been answered. Should the algebraic techniques of *Groebner bases* or *Multipolynomial resultants* be successful in solving explicitly the selected geodetic problems, they are then used as the computational engine of the *combinatorial algorithm* that was first suggested by *C.F. Gauss* and later on by *C. G. I. Jacobi* (1841) and extended by *P. Werkmeister* (1920). We refer to this algorithm simply as the *Gauss–Jacobi combinatorial algorithm*. In attempting to answer the questions above, the objectives of this chapter are:

(1) To help you, the reader, understand the algebraic computational procedures of type *Groebner bases* and *Multipolynomial resultants* as applied to solve explicitly (in closed form) *geodetic nonlinear problems*. In this respect, the *Groebner bases* and *Multipolynomial resultants* techniques are used to solve explicitly (symbolically) geodetic problems of *GPS pseudo-ranging four-point P4P*, *Minimum Distance Mapping* and the *7-parameter transformation*. By converting the *nonlinear observation equations* of these selected geodetic problems into *algebraic* (*polynomial*), *Groebner bases* and *Multipolynomial resultants* techniques are used to eliminate several variables in a multivariate systems of nonlinear polynomial equations in such a manner that the end products from the initial systems of nonlinear observation equations result in *univariate polynomials*. The elimination procedure is similar to the Gauss elimination approach in linear systems.

(2) To introduce to you, the reader, an algebraic approach of Gauss–Jacobian combinatorial for solving overdetermined problems that are traditionally solved using least squares, which peg their operations on assumptions of linearity and prior information about the unknown values. From the principle of weighted arithmetic mean and using the *Gauss–Jacobi combinatorial lemma* (*C. G. I. Jacobi* 1841), an adjustment procedure that neither *linearizes* the *nonlinear observation equations* nor uses *iterative procedures* to solve the *nonlinear Gauss–Markov* model is developed. Linearization is permitted only for *nonlinear error propagation/variance-covariance propagation*. Such a procedure is only realized, thanks to the *univariate polynomial* generated by the algebraic computational procedures of type *Groebner bases* or *Multipolynomial resultants* as the computing engine for its *minimal combinatorial sets*.

15-3 Algebraic Methods for Solving Nonlinear Systems of Equations

In the present Section, we depart from the traditional iterative procedures for estimating the unknown fixed parameters of the *nonlinear Gauss–Markov model* and present a *combinatorial approach* that traces its roots back to the work of *C.F. Gauss*. *C.F. Gauss* first proposed the combinatorial approach using the products of squared distances (from unknown point to known points) and the square of the perpendicular distances from the sides of the error triangle to the unknown point as the weights, see, e.g., Appendix D3. According to *W.K. Nicholson* (1999, pp. 272–273), the motto in Gauss seal read *"pauca des matura"* meaning *few but ripe*. This belief led *C. F. Gauss* not to publish most of his important contributions. For instance, *W. K. Nicholson* (1999, pp. 272–273) writes "Although not all his results were recorded in the diary (many were set down only in letters to friends), several entries would have each given fame to their author if published. Gauss new about the quartenions before Hamilton. . . ". The combinatorial method, like many of his works, was later to be published after his death. Several years later, the method was to be developed further by *C. G. I. Jacobi* (1841) who used the square of the determinants as the weights in estimating the unknown parameters from the arithmetic mean. *P. Werkmeister* (1920) later established the relationship between the area of the error figure formed by the combinatorial approach and the standard error of the determined point. We will refer to this *combinatorial approach* as the *Gauss–Jacobi combinatorial algorithm* needed to solve an overdetermined system of equations algebraically.

While the solution of the *linear Gauss–Markov model* by **B***est* **L***inear* **U***niformly* **U***nbiased* **E***stimator* (BLUUE) is straight forward, the solution of the *nonlinear Gauss–Markov model* has no straight forward procedure owing to the *nonlinearity* of the *injective function* (or map function) that maps $\mathbb{R}^m$ to $\mathbb{R}^n$. The *Gauss–Jacobi combinatorial algorithm* is presented as a possible algebraic solution to the *nonlinear Gauss–Markov model.* In Sect. 15-31, we demonstrate how the procedure can be used to solve the *nonlinear Gauss–Markov model*.

15-31 Solution of Nonlinear Gauss–Markov Model

The *nonlinear Gauss–Markov model* is solved in two steps:

- Step 1: Combinatorial minimal subsets of observations are constructed and rigorously solved by means of the *Multipolynomial resultant* or *Groebner basis* (*J.L. Awange* and *E.W. Grafarend* 2005; *J. L. Awange et al.* 2010).
- Step 2: The combinatorial solution points of step 1 are reduced to their final adjusted values by means of an adjustment procedure where the **B**est **L**inear **U**niformly **U**nbiased **E**stimator (BLUUE) is used to estimate the vector of fixed

parameters within the linear Gauss–Markov model with the dispersion matrix of the real valued random vector of pseudo-observations from *Step 1* generated via the *nonlinear error propagation law* also known in this case as the *nonlinear variance-covariance propagation*.

15-311 Construction and Solution of the Combinatorial Subsets

Since $n > m$ we construct a minimal combinatorial subsets comprising m equations solvable in closed form using either *Groebner bases* or *Multipolynomial resultants* which we present in Sects. 15-312 and 15-313. We begin by the following elementary definitions:

Definition 15.1 (Permutation). Given a set S with elements $\{i, j, k\} \in S$, the arrangement obtained by placing $\{i, j, k\} \in S$ in some sequence is called *permutation*. If we choose any of the elements say i first, then each of the remaining elements j, k can be put in the second position, while the third position is occupied by the unused letter either j or k. For the set S, the following permutations can be made:

$$\left.\begin{array}{l} ijk \; ikj \; jik \\ jki \; kij \; kji \end{array}\right]. \tag{15.1}$$

From (15.1), there exist three ways of filling the first position, two ways of filling the second position, and one way of filling the third position. Thus the number of *permutations* is given by $3 \times 2 \times 1 = 6$. In general, for n different elements, the number of permutation is equal to $n \times \ldots \times 3 \times 2 \times 1 = n!$

Definition 15.2 (Combination). If for n elements only m elements are used for the permutation, then we have a *combination* of the mth order. If we follow the definition above, then the first position can be filled in n ways, the second in $n - 1$ ways and the mth in $n - (m - 1)$ ways. In (15.1) above, the combinations are identical and contain the same elements in different sequences. If the arrangement is to be neglected, then we have for n elements, a combination of mth order being given by

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n(n-1)\ldots(n-m+1)}{m \times \cdots \times 3 \times 2 \times 1}. \tag{15.2}$$

Given n nonlinear equations to be solved, we first form $\binom{n}{m}$ minimal combinatorial subsets each consisting of m elements (where m is the number of the unknown elements). Each minimal combinatorial subset is solved using the algebraic procedures discussed in Sects. 15-312 and 15-313. In geodesy, the number of elements n normally consist of the observations in the vector $\boldsymbol{y}$, while the number of elements m normally consist of the unknown fixed parameters in the vector $\boldsymbol{\xi}$. In the following Sections, we present two algebraic algorithms that are used to solve the minimal combinatorial subsets (15.2) in closed form. We first present the approach based on the *Groebner bases* and thereafter consider the *Multipolynomial resultants* approach.

15-312 Groebner Basis Method

Groebner basis is the greatest common divisors of a multivariate system of equations (*J.L. Awange* and *E.W. Grafarend* 2005; *J. L. Awange et al.* 2010). Its direct application is the elimination of variables in nonlinear systems of equations. Let us start by the problem of finding the greatest common divisors in Example 15.1:

Example 15.1. (Greatest common divisors (gcd)):

Given the numbers 12, 20, and 18, find their greatest common divisor. We proceed by writing the factors as

$$\left.\begin{array}{l} 12 = 2^2.3^1.5^0 \\ 20 = 2^2.3^0.5^1 \\ 18 = 2^1.3^2.5^0 \end{array}\right] \rightarrow 2^1.3^0.5^0 = 2, \tag{15.3}$$

leading to 2 as the greatest common divisor of 12, 20 and 18. Next, let us consider the case of univariate polynomials $f_1, f_2 \in k[x]$ in (15.4).

$$\left.\begin{array}{l} f_1 = 3x^4 - 3x^3 + 8x^2 + 2x - 5 \\ f_2 = 5x^4 - 4x^2 - 9x + 21 \end{array}\right] \rightarrow \textit{Euclidean algorithm} = f \in k[x]. \tag{15.4}$$

Equation (15.4) employs the Euclidean algorithm which obtains one univariate polynomial as the gcd of the two univariate polynomials f_1 and f_2. If on the other hand expressions in (15.4) were not univariate but multivariate, e.g., $g_1, g_2 \in k[x, y]$ as in (15.5), then one applies the Buchberger algorithm which then computes the Groebner bases. For example,

$$\left.\begin{array}{l} g_1 = xy + x - y - 1 \\ g_2 = xy - x - y + 1 \end{array}\right] \rightarrow \textit{Buchberger algorithm} = \textit{Groebner basis}. \tag{15.5}$$

Groebner basis therefore, is the greatest common divisors of a multivariate system of polynomial equations $\{g_1, g_2\}$.

As stated earlier, it is useful for eliminating variables in nonlinear systems of equations. Gauss elimination technique on the other hand is applicable for linear cases as shown in Example 15.2.

Example 15.2. (Gauss elimination technique)

Solve the linear system of equations

$$\begin{bmatrix} -x + y + 2z = 2 \\ 3x - y + z = 6 \\ -x + 3y + 4z = 4. \end{bmatrix} \tag{15.6}$$

The first step is to eliminate x in the second and third expressions of (15.6). This is achieved by multiplying the first expression by 3 and adding to the second expression to give the second expression of (15.7). The third expression of (15.7) is obtained by subtracting the first expression from the third expression in (15.6).

$$\begin{bmatrix} -x + y + 2z = 2 \\ 2y + 7z = 12 \\ 2y + 2z = 2. \end{bmatrix} \tag{15.7}$$

The second step is to eliminate y in the second and third expressions of (15.7). This is achieved by subtracting the second expression from the third expression in (15.7) to give (15.8).

$$\begin{bmatrix} -x + y + 2z = 2 \\ 2y + 7z = 12 \\ -5z = -10. \end{bmatrix} \tag{15.8}$$

The solution of $z = 2$ in (15.8) can now be substituted back into the second equation $2y + 7z = 12$ to give the value of $y = -1$, which together with the value of $z = 2$ are substituted into the first equation to give the value of $x = 1$ to complete the Gauss elimination technique.

In many applications however, equations relating unknown variables to the measured (observed) quantities are normally nonlinear and often consist of many variables (multivariate). In such cases, the Gauss elimination technique for the univariate polynomial equations employed in Example 15.2 gives way to Groebner basis as illustrated in Examples 15.3 and 15.4. In general, the Groebner basis algorithm reduces a system of multivariate polynomial equations. This is done by employing operations "*addition*" and "*multiplication*" on a polynomial ring (see, e.g., Awange and Grafarend 2005; Awange et al. 2010) to give more simplified expressions. Given a system of polynomial equations which are to be solved explicitly for unknowns, e.g., (15.5), Groebner basis algorithm is applied to reduce the set of polynomials into another set (e.g., from a system $F(x, y, z)$ to another system $G(x, y, z)$) of polynomials with suitable properties that allow solution. If $F(x, y, z)$ is a set of nonlinear system of polynomial equations, Groebner basis eliminates variables in a manner similar to Gauss elimination technique for linear cases to reduce it to $G(x, y, z)$. With Lexicographic ordering of the monomials (see *Definition* D1.2 in

Appendix D), one expression in $G(x, y, z)$ always turns out to be a univariate polynomial. Its roots are easily obtained using algebraic software of Matlab, Mathematica or Maple (see e.g., Appendix D3), and can be substituted in the other elements of the set $G(x, y, z)$ to obtain a complete solution which also satisfy the original set $F(x, y, z)$. Examples 15.3 and 15.4 elaborate on the application of Groebner basis.

Example 15.3. (Groebner basis computation)

Let us consider a simple example from Buchberger (2001). Consider a set $F(x, y) = \{f_1, f_2\}$ to have as its elements

$$\begin{bmatrix} f_1 = xy - 2y \\ f_2 = 2y^2 - x^2, \end{bmatrix} \tag{15.9}$$

where $\{f_1, f_2\} \in I$ are the generators of the Ideal I (see definition of Ideal on p. 537). We now seek a simplified set of generators of this Ideal using Buchberger algorithm. By employing operations "addition" and "multiplication", the Groebner basis algorithm (also called Buchberger algorithm) reduces the system of nonlinear equations (15.9) into another set G of F as

$$G := \{-2x^2 + x^3, -2y + xy, -x^2 + 2y^2\}. \tag{15.10}$$

In Mathematica software, using the lexicographic order $x > y$, i.e., x comes before y, the Groebner basis could simply be computed by entering the command

$$Groebner\ Basis[F, \{x, y\}]. \tag{15.11}$$

The set G in (15.10) contains one univariate polynomial $-2x^2+x^3$, which can easily be solved using roots command in Matlab for solutions $\{x = 0, x = 0, x = 2\}$ and substituted in any of the remaining elements of the set G to solve for y. The solutions of G, i.e., the roots $\{x = 0, x = 0, x = 2\}$) and those of y satisfy polynomials in F. This can be easily tested by substituting these solutions into (15.9) to give 0. Let us consider as a second example an optimization problem.

Example 15.4. (Minimum and maximization problem)

Find the minimum and maximum of $f(x, y, z) = x^3 + 2xyz - z^2$, such that $g(x, y, z) = x^2 + y^2 + z^2 - 1$. First, we obtain the partial derivatives of $f - L\,g = 0$ with respect to $\{x, y, z, L\}$, where L is the Lagrangean multiplier as

$$\frac{\partial f}{\partial\{x, y, z, L\}} := F = \begin{bmatrix} 3x^2 + 2yz - 2xL = 0 \\ 2xz - 2yL = 0 \\ 2xy - 2z - 2zL = 0 \\ x^2 + y^2 + z^2 - 1 = 0. \end{bmatrix} \tag{15.12}$$

Groebner basis is invoked in Mathematica by

$$Groebner\ Basis[\{F\}, \{x, y, L, z\}],$$

which leads to

$$G = \begin{bmatrix} L - 1.5x - 1.5yz - 43.7z^6 - 62.2z^4 - 17.5z^2 = 0 \\ x^2 + y^2 + z^2 - 1 = 0 \\ y^2z - 1.8z^5 + 2.8z^3 - z = 0 \\ z^7 - 1.5z^5 + 0.6z^3 - 0.04z = 0. \end{bmatrix} \tag{15.13}$$

The solution of z in (15.13) can then be substituted into the third equation $y^2z - 1.8z^5 + 2.8z^3 - z = 0$ to give the value of y. The obtained values of z and y are then substituted into the second equation to give the value of x, and thus complete the Groebner basis solution. Later in the chapter, we will introduce the *reduced Groebner basis* which can be used to obtain directly the last expression of (15.13), i.e., the univariate polynomial in z. The theory behind the operation of Groebner basis is however not so simple. In the remainder of this chapter, we will try to present in a simplified form the algorithm behind the computation of Groebner bases. The computation of Groebner basis is achieved by the capability to manipulate the polynomials to generate *Ideals* defined as

Definition 15.3 (Ideal). An Ideal is generated by a family of generators as consisting of the set of linear combinations of these generators with polynomial coefficients. Let $f_1, \ldots, f_s$ and $c_1, \ldots, c_s$ be polynomials in $k\,[x_1, \ldots, x_n]$, then

$$\langle f_1, \ldots, f_s \rangle = \sum_{i=1}^{s} c_i\, f_i. \tag{15.14}$$

In (15.14), $\langle f_1, \ldots, f_s \rangle$ is an Ideal and if a subset $I \subset k\,[x_1, \ldots, x_n]$ is an Ideal, it must satisfy the following conditions (Cox 1997, p. 29);

- $0 \in I$
- If $f, g \in I$, then $f + g \in I$ (i.e., I is an additive subgroup of the additive group of the field k)
- If $f \in I$ and $c \in k\,[x_1, \ldots, x_n]$, then $cf \in I$ (i.e., I is closed under multiplication ring element).

Example 15.5. (Ideal)

Consider equations expressed algebraically as

$$\begin{bmatrix} f_1 := (x_1 - x_0)^2 + (y_1 - y_0)^2 - d_1^2 \\ f_2 := (x_2 - x_0)^2 + (y_2 - y_0)^2 - d_2^2, \end{bmatrix} \tag{15.15}$$

where polynomials $\{f_1, f_2\}$ belong to the polynomial ring $\mathbb{R}[x_0, y_0]$. If the polynomials

$$\left[\begin{array}{l} c_1 := 4x_0 + 6 \\ c_2 := x_0 + y_0 \end{array}\right. \tag{15.16}$$

also belong to the same polynomial ring $\mathbb{R}[x_0, y_0]$, an Ideal is generated by a linear combination

$$I := \left[\begin{array}{l} \quad\langle f_1, f_2\rangle = c_1 f_1 + c_2 f_2 \\ = (4x_0 + 6) f_1 + (x_0 + y_0) f_2. \end{array}\right. \tag{15.17}$$

In this case, $\{f_1, f_2\}$ are said to be generators of the *Ideal* I. *Definition* (15.3) of an *Ideal* can be presented in terms of polynomial equations $f_1, \ldots, f_s \in k\,[x_1, \ldots, x_n]$. This is done by expressing the system of polynomial equations as

$$\left[\begin{array}{l} f_1 = 0 \\ f_2 = 0 \\ \quad . \\ \quad . \\ f_s = 0, \end{array}\right. \tag{15.18}$$

and using them to derive others by multiplying each individual equation f_i by another polynomial $c_i \in k\,[x_1, \ldots, x_n]$ and summing to get $c_1 f_1 + c_2 f_2 + \ldots + c_s f_s = 0$ (cf., 15.14). The *Ideal* $\langle f_1, \ldots, f_s\rangle$ thus consists of a system of equations $f_1 = f_2 = \ldots = f_s = 0$, thus indicating that if $f_1, \ldots, f_s \in k\,[x_1, \ldots, x_n]$, then $\langle f_1, \ldots, f_s\rangle$ is an *Ideal* generated by $f_1, \ldots, f_s$, i.e., being the *basis* of the *Ideal* I.

In this case, a collection of these *nonlinear algebraic equations* forming *Ideals* are referred to as the set of polynomials generating the *Ideal* and forms the elements of this *Ideal*. Perhaps a curious reader may begin to wonder why the term *Ideal* is used. To quench this curiosity we refer to Nicholson (1999, p. 220) and quote from Becker (1993, p. 59) who wrote:

> "On the origin of the term *Ideal*, the concept is attributed to *Dedekind* who introduced it as a set theoretical version of *Kummer's "Ideal number"* to circumvent the failure of unique factorization in certain natural extension of the domain $\mathbb{Z}$. The relevance of *Ideal* in the theory of *polynomial rings* was highlighted by *Hilbert Basis Theorem*. The systematic development of *Ideal* theory; in more general rings is largely due to *E. Noether*. In the older literature, the term "module" is sometimes used for *"Ideal"* (cf., Macaulay, 1916). The term "ring" seems to be due to *D. Hilbert*; *Kronecker* used the term "order" for ring".

Example 15.6. (Ideal)

Consider example (15.3) with polynomials in $\mathbb{R}\,[x, y]$. The *Ideal* $I = \langle xy - 2y, 2y^2 - x^2\rangle$. The generators of an Ideal can be computed using the *division algorithm* defined as

Definition 15.4 (Division algorithm). Fix a monomial order of polynomials say $x > y$ for polynomials $F = (h_1, \ldots, h_s)$. Then every $f \in k[x, y]$ can be written in the form $f = a_1 h_1 + a_2 h_2 + \ldots + a_s h_s + r$, where $a_i, r \in k[x, y]$ and either $r = 0$ or a linear combination with coefficients in k of monomials, none of which is divisible by any of $LT(f_1), \ldots, LT(f_s)$ (see Definition D1.5 in Appendix D for leading term LT).

Example 15.7. (Division algorithm in a univariate case)

Divide the polynomial $f = x^3 + 2x^2 + x + 5$ by $h = x^2 - 2$. We proceed as follows:

$$\left[\begin{array}{c} \underline{x+2} \\ x^2 - 2 \mid x^3 + 2x^2 + x + 5 \\ \underline{x^3 - 2x} \\ 2x^2 + 3x + 5 \\ \underline{2x^2 - 4} \\ 3x + 1, \end{array}\right. \tag{15.19}$$

implying $x^3 + 2x^2 + x + 5 = (x + 2)(x^2 - 2) + (3x + 1)$, with $a = (x + 2)$ and $r = (3x + 1)$. The *division algorithm* given in definition (15.4) fits well to the case of *univariate polynomials* as the remainder r can uniquely be determined. For *multivariate polynomials*, the remainder may not be uniquely determined as this depends on the order of the divisors. The division of the polynomial F by $\{f_1, f_2\}$ where f_1 comes before f_2 may not necessarily give the same remainder as the division of F by $\{f_2, f_1\}$ in whose case the order has been changed. This problem is overcome if we pass over to Groebner basis where the existence of every *Ideal* is assured by the Hilbert Basis Theorem (Cox 1997, pp. 47–61). The Hilbert Basis Theorem assures that every Ideal $I \subset k[x_1, \ldots, x_n]$ has a finite generating set, that is $I = \langle g_1, \ldots, g_s \rangle$ for some $\{g_1, \ldots, g_s\} \in I$. The finite generating set G in *Hilbert Basis Theorem* is what is known as a *basis*. Suppose every non-zero polynomial is written in decreasing order of its monomials:

$$\sum_{i=1}^{n} d_i x_i, \quad d_i \neq 0,\ x_i > x_{i+1}, \tag{15.20}$$

if we let the system of generators of the *Ideal* be in a set G, a polynomial f is reduced with respect to G if no leading monomial of an element of G (LM (G)) divides the leading monomial of f (LM(f)). The polynomial f is said to be *completely reduced* with respect to G if no monomials of f is divisible by the leading monomial of an element of G (Davenport 1988, pp. 96–97). The *basis* G, which completely reduces the polynomial f and uniquely determines the remainder r is also known as the Groebner basis and is defined as follows:

Definition 15.5 (Groebner basis). A system of generators G of an Ideal I is called a Groebner basis (with respect to the order $<$) if every reduction of $f \in I$ to a reduced polynomial (with respect to G) always gives zero as a remainder. This definition is a special case of a more general definition given as: Fix a monomial order and let $G = \{g_1, \ldots, g_t\} \subset k[x_1, \ldots, x_n]$. Given $f \in k[x_1, \ldots, x_n]$, then f reduces to zero Modulo G, written as

$$f \to_G 0, \tag{15.21}$$

if f can be written in the form (cf., 15.18 on p. 538)

$$f = a_1 g_1 + \ldots + a_t g_t, \tag{15.22}$$

such that whenever $a_i g_i \neq 0$, we have multideg(f)$\geq$ multideg($a_i g_i$) (see Definition D1.5 in Appendix D for leading term LT, LM and Multideg).

Following *Definition* 15.5, the reader can revisit Examples 15.3 and 15.4 which present the Groebner basis G of the original system F of equations.

Groebner basis has become a household name in algebraic manipulations and finds application in fields such as mathematics and engineering for solving partial differential equations, e.g.,(Lidl 1998, p. 432). It has also found use as a tool for discovering and proving theorems to solving systems of polynomial equations as elaborated in publications by Buchberger and Winkler (1998). Groebner basis also give a solution to the *Ideal* membership problem. By reducing a given polynomial f with respect to the Groebner basis G, f is said to be a member of the *Ideal* if zero remainder is obtained. This implies that if $G = \{g_1, \ldots, g_s\}$ is a Groebner basis of an *Ideal* $I \subset k[x_1, \ldots, x_n]$ and $f \in k[x_1, \ldots, x_n]$ a polynomial, $f \in I$ if and only if the remainder on division of f by G is zero. Groebner bases can also be used to show the equivalence of polynomial equations. Two sets of polynomial equations will generate the same *Ideal* if and only if their Groebner bases are equal with respect to any term ordering, e.g., the solutions of (15.10) satisfy those of (15.9). This property is important in that the solutions of the Groebner basis will satisfy the original system formed by the generating set of nonlinear equations. It implies that a system of polynomial equations $f_1(x_1, \ldots, x_n) = 0, \ldots, f_s(x_1, \ldots, x_n) = 0$ will have the same solutions with a system arising from any Groebner basis of $f_1, \ldots, f_s$ with respect to any term ordering. This is the main property of Groebner basis that is used to solve systems of polynomial equations as will be explained in the next section.

B. Buchberger algorithm

Given polynomials $g_1, \ldots, g_s \in I$, the *B. Buchberger algorithm* seeks to derive the standard generators or the Groebner basis of this *Ideal*. Systems of equations

$g_1 = 0, \ldots, g_s = 0$ to be solved in practise are normally formed by these same polynomials which here generating the *Ideal.* The *B. Buchberger algorithm* computes the *Groebner basis* by making use of pairs of polynomials from the original polynomials $g_1, \ldots, g_s \in I$ and computes the subtraction polynomial known as the *S-polynomial* explained in *D. Cox et al.* (1997; p. 81) as follows:

Definition 15.6 (*S-polynomial*). Let $f, g \in k[x_1, \ldots x_n]$ be two non-zero polynomials. If multideg $(f) = \alpha$ and multideg $(g) = \beta$, then let $\gamma = \gamma_1, \ldots, \gamma_n$, where $\gamma_i = \max\{\alpha_i, \beta_i\}$ for each i. x^γ is called the *Least Common Multiple* (LCM) of LM(f) and LM(g) expressed as $x^\gamma = LCM\{LM(f), LM(g)\}$. The *S-polynomial* of f and g is given as

$$S(f, g) = \frac{x^\gamma}{LT(f)} f - \frac{x^\gamma}{LT(g)} g. \tag{15.23}$$

The expression above gives S as a linear combination of the monomials $\frac{x^\gamma}{LT(f)}$, $\frac{x^\gamma}{LT(g)}$ with polynomial coefficients f and g and thus belongs to the *Ideal* generated by f and g.

Definition 15.7 (*Groebner basis* in terms of *S-polynomial*). A basis G is *Groebner basis* if and only if for every pair of polynomials f and g of G, $S(f, g)$ reduces to zero with respect to G. More generally a basis $G = \{g_1, \ldots, g_s\}$ for an *Ideal* I is a *Groebner basis* if and only if $S(f, g) \rightarrow_G 0, \quad i \neq j$.

The implication of *Definition 15.7* is the following: Given two polynomials $f, g \in G$ such that $LCM(LM(f), LM(g)) = LM(f).LM(g)$, the leading monomials of f and g are relatively prime leading to $S(f, g) \rightarrow_G 0$. The concept of prime integer is clearly documented in *K. Ireland* and *M. Rosen* (1990 pp. 1–17).

Example 15.8. (*S-Polynomial*):

Consider the two polynomials

$$\begin{bmatrix} g_1 = x_1^2 + 2a_{12}x_1x_2 + x_2^2 + a_{oo} \\ g_2 = x_2^2 + 2b_{23}x_2x_3 + x_3^2 + b_{oo}. \end{bmatrix} \tag{15.24}$$

The *S-polynomial* can then be computed as follows: First we choose a *lexicographic ordering* $\{x_1 > x_2 > x_3\}$ then

$$\left[\begin{array}{l} LM(g_1) = x_1^2,\ LM(g_2) = x_2^2, LT(g_1) = x_1^2,\ LT(g_2) = x_2^2 \\ LCM(LM(g_1), LM(g_2)) = x_1^2 x_2^2 \\ S(g_1, g_2) = \frac{x_1^2 x_2^2}{x_1^2}(x_1^2 + 2a_{12}x_1x_2 + x_2^2 + a_{oo}) - \frac{x_1^2 x_2^2}{x_2^2}(x_2^2 + 2b_{23}x_2x_3 + x_3^2 + b_{oo}) \\ = x_2^2 x_1^2 + 2a_{12}x_1x_2^3 + x_2^4 + a_{oo}x_2^2 - x_1^2x_2^2 - 2b_{23}x_1^2x_2x_3 - x_1^2x_3^2 - b_{oo}x_1^2) \\ = -b_{oo}x_1^2 - 2b_{23}x_1^2x_2x_3 - x_1^2x_3^2 + 2a_{12}x_1x_2^3 + x_2^4 + a_{oo}x_2^2 \end{array}\right. \tag{15.25}$$

Example 15.9. (*S-Polynomial*):

Consider two polynomial equations given as

$$\left[\begin{array}{l} g_3 = x_1^2 - 2a_0x_1 + x_2^2 - 2b_0x_2 + x_3^2 - 2c_0x_3 - x_4^2 + 2d_0x_4 + a_0^2 + b_0^2 + c_0^2 + d_0^2 \\ g_4 = x_1^2 - 2a_1x_1 + x_2^2 - 2b_1x_2 + x_3^2 - 2c_1x_3 - x_4^2 + 2d_1x_4 + a_1^2 + b_1^2 + c_1^2 + d_1^2. \end{array}\right. \tag{15.26}$$

By choosing the *lexicographic ordering* $\{x_1 > x_2 > x_3 > x_4\}$, the S-*polynomial* is computed as follows

$$\left[\begin{array}{l} LM(g_3) = x_1^2, LM(g_4) = x_1^2, LT(g_3) = x_1^2, LT(g_4) = x_1^2 \\ LCM(LM(g_3), LM(g_4)) = x_1^2 \\ S(g_3, g_4) = \frac{x_1^2}{x_1^2}(g_3) - \frac{x_1^2}{x_1^2}(g_4) = g_3 - g_4 \\ S(g_3, g_4) = 2(a_1 - a_o)x_1 + 2(b_1 - b_o)x_2 + 2(c_1 - c_o)x_3 + \\ +2(d_o - d_1)x_4 + a_o - a_1 + b_o - b_1 + c_o - c_1 + d_0 - d_1 \end{array}\right. \tag{15.27}$$

Example 15.10. (*S-Polynomial*):

As an additional example on the computation of S-polynomial, let us consider the minimum distance mapping problem. The last two polynomials equations are given as

$$\left[\begin{array}{l} g_5 = x_3 + b^2x_3x_4 - Z \\ g_6 = b^2x_1^2 + b^2x_2^2 + a^2x_3^2 - a^2b^2. \end{array}\right. \tag{15.28}$$

By choosing the *lexicographic ordering* $\{x_1 > x_2 > x_3 > x_4\}$, the S-polynomial is computed as follows

$$\left[\begin{array}{l} LM(g_5) = x_3, LM(g_6) = x_1^2, LT(g_5) = x_3, LT(g_6) = b^2x_1^2 \\ LCM(LM(g_5), LM(g_6)) = x_1^2x_3 \\ S(g_5, g_6) = \dfrac{x_1^2x_3}{x_3}(x_3 + b^2x_3x_4 - Z) - \dfrac{x_1^2x_3}{b^2x_1^2}(b^2x_1^2 + b^2x_2^2 + a^2x_3^2 - a^2b^2) \\ S(g_5, g_6) = -Zx_1^2 + b^2x_1^2x_3x_4 - x_2^2x_3 - \frac{a^2}{b^2}x_3^3 + a^2x_3 \end{array}\right. \tag{15.29}$$

Example 15.11. (computation of *Groebner basis* from the *S-polynomials*):

By means of an Example given by *J. H. Davenport et al.* (1988, pp. 101–102), we illustrate how the *B. Buchberger algorithm* works. Let us consider the *Ideal* generated by the polynomial equations

$$\left[\begin{array}{l} g_1 = x^3yz - xz^2 \\ g_2 = xy^2z - xyz \\ g_3 = x^2y^2 - z \end{array}\right. \tag{15.30}$$

with the *lexicographic ordering* $x > y > z$ adopted. The S-polynomials to be considered are $S(g_1, g_2)$, $S(g_2, g_3)$ and $S(g_1, g_3)$. We consider first $S(g_2, g_3)$ and show that the result is used to suppress g_1 upon which any pair $S(g_1, g_i)$ (e.g. $S(g_1, g_2)$ and $S(g_1, g_3)$) containing g_1 will not be considered. $LT(g_2) = xy^2z,\ LT(g_3) = x^2y^2$, then $LCM(g_2, g_3) = x^2y^2z$ respectively

$$\left[\begin{array}{l} S(g_2, g_3) = \dfrac{x^2y^2z}{xy^2z}g_2 - \dfrac{x^2y^2z}{x^2y^2}g_3 \\ = (x^2y^2z - x^2yz) - (x^2y^2z - z^2) \\ \qquad = -x^2yz + z^2 \end{array}\right. \tag{15.31}$$

We immediately note that the leading term of the resulting polynomial LT($S(g_2, g_3)$) is not divisible by any of the leading terms of the elements of G, and thus the remainder upon the division of $S(g_2, g_3)$ by the polynomials in G is not zero (i.e. when reduced with respect to G) thus G is *not* a *Groebner basis*. This resulting polynomial after the formation of the *S-polynomial* is denoted g_4 and its negative (to make calculations more reliable) added to the initial set of G leading to

$$\left[\begin{array}{l} g_1 = x^3yz - xz^2 \\ g_2 = xy^2z - xyz \\ g_3 = x^2y^2 - z \\ g_4 = x^2yz - z^2. \end{array}\right. \tag{15.32}$$

The *S-polynomials* to be considered are now $S(g_1, g_2)$, $S(g_1, g_3)$, $S(g_1, g_4)$, $S(g_2, g_4)$ and $S(g_3, g_4)$. In the set of G, one can write $g_1 = xg_4$ leading without any change in G to the suppression of g_1 leaving only $S(g_2, g_4)$ and $S(g_3, g_4)$ to be considered. Then

$$\left[\begin{array}{l} S(g_2, g_4) = xg_2 - yg_4 \\ \quad = -x^2yz + yz^2 \end{array}\right. \tag{15.33}$$

which can be reduced by adding g_4 to give $g_5 = yz^2 - z^2$, a non zero value thus the set G is *not* a *Groebner basis*. This value is added to the set of G to give

$$\left[\begin{array}{l} g_2 = xy^2z - xyz, \\ g_3 = x^2y^2 - z, \\ g_4 = x^2yz - z^2, \\ g_5 = yz^2 - z^2, \end{array}\right. \tag{15.34}$$

the *S-polynomials* to be considered are now $S(g_3, g_4)$, $S(g_2, g_5)$, $S(g_3, g_5)$ and $S(g_4, g_5)$. We then compute

$$\left[\begin{array}{l} S(g_3, g_4) = zg_3 - yg_4 \\ \quad = yz^2 - z^2 \end{array}\right. \tag{15.35}$$

which upon subtraction from g_5 reduces to zero. Further,

$$\left[\begin{array}{l} S(g_2, g_5) = zg_2 - xyg_5 \\ \quad = -xyz^2 + xyz^2 \\ \quad = 0 \end{array}\right. \tag{15.36}$$

and

$$\left[\begin{array}{l} S(g_4, g_5) = zg_4 - x^2yg_5 \\ \quad = x^2z^2 - z^3, \end{array}\right. \tag{15.37}$$

which is added to G as g_6 giving

$$\left[\begin{array}{l} g_2 = xy^2z - xyz, \\ g_3 = x^2y^2 - z, \\ g_4 = x^2yz - z^2, \\ g_5 = yz^2 - z^2, \\ g_6 = x^2y^2 - z^3, \end{array}\right. \tag{15.38}$$

The S polynomials to be considered are now $S(g_3, g_5)$, $S(g_2, g_6)$, $S(g_3, g_6)$, $S(g_4, g_6)$ and $S(g_5, g_6)$. We now illustrate that all this S-polynomials reduces to zero as follows

$$\begin{bmatrix} S(g_3, g_5) = z^2 g_3 - x^2 y g_5 = x^2 y z^2 - z^3 - z g_4 = 0 \\ S(g_2, g_6) = xz g_2 - y^2 g_6 = -x^2 y^2 z^2 + y^2 z^3 + y^2 g_4 = 0 \\ S(g_3, g_6) = z^2 g_3 - y^2 g_6 = y^2 z^3 - z^3 - (yz - z) g_5 = 0 \\ S(g_4, g_6) = z g_4 - y g_6 = y z^3 - z^3 - z g_5 = 0 \\ S(g_5, g_6) = x^2 g_5 - y g_6 = -x^2 z^2 + y z^3 + g_6 - z g_5 = 0. \end{bmatrix} \quad (15.39)$$

Thus (15.39) comprise the *Groebner basis* of the original set in (15.30).

The importance of the *S-polynomials* is that they lead to the cancelation of the leading terms of the polynomial pairs involved. In so doing the polynomial variables are systematically eliminated according to the polynomial ordering chosen. For example if the *lexicographic ordering* $x > y > z$ is chosen, x will be eliminated first, followed by y and the final expression may consist only of the variable z. *D. Cox et al.* (1998, p.15) has indicated the advantage of *lexicographic ordering* as being the ability to produce *Groebner basis* with systematic elimination of variables. *Graded lexicographic ordering* on the other hand has the advantage of minimizing the amount of computational space needed to produce the *Groebner basis*. The procedure is thus a *generalisation* of the *Gauss elimination procedure* for linear systems of equations. If we now put our system of polynomial equations to be solved in a set G, *S-pair* combinations can be formed from the set of G as explained in the *definitions* above. The *theorem*, known as the *Buchberger's S-pair polynomial criterion*, gives the criterion for deciding whether a given basis is a *Groebner basis* or not. It suffices to compute all the S-polynomials and check whether they reduce to zero. Should one of the polynomials not reduce to zero, then the basis fails to be a *Groebner basis*. Since the reduction is a linear combination of the elements of G, it can be added to the set G without changing the *Ideal* generated. *B. Buchberger* (1979) gives an *optimisation criterion* that reduces the number of the *S-polynomials* already considered in the algorithm. The criterion states that if there is an element p of G such that the leading monomial of p (LM(p)) divides the LCM($f, g \in G$), and if $S(f, p)$, $S(p, g)$ have already been considered, then there is no need of considering $S(f, g)$ as this reduces to zero.

The essential observation in using the *Groebner bases* to solve a system of polynomial equations is that the variety (simultaneous solution of system of polynomial equations) does not depend on the original system of the polynomials $F := \{f_1, \ldots, f_s\}$ but instead on the *Ideal* I generated by F. This therefore means that for the variety $V = V(I)$, one makes use of the special generating set (*Groebner basis*) instead of the actual system F. Since the *Ideal* is generated by F, the solution obtained by solving for the affine variety of this *Ideal* satisfies the original system F of equations. *B. Buchberger* (1970) proved that $V(I)$ is void, and thus giving a test as to whether the system of polynomial F can be solved, if and only if the computed *Groebner basis* of polynomial Ideal I has $\{1\}$ as its element. *B. Buchberger* (1970) further gives the criterion for deciding if $V(I)$ is finite. If the system has been proved to be solvable and finite then *F. Winkler* (1996, *theorem 8.4.4*, p.192) gives a theorem for deciding whether the system has finitely or infinitely many solutions. The *theorem* states that if G is a *Groebner basis*, then a solvable system of polynomial equations has finitely many solutions if and only if for every x_i, $1 \leq i \leq n$, there is a polynomial $g_i \in G$ such that $LM(g_i)$ is a pure power of x_i. The process of addition of the remainder after the reduction by the S*-polynomials* and thus expanding the generating set is shown by *B. Buchberger* (1970), *D. Cox et al.* (1997 p.88) and *J. H. Davenport et al.* (1988 p.101) to terminate.

The *B. Buchberger algorithm* (see Appendix D2), more or less a generalization of the *Gauss elimination* procedure, makes use of the subtraction polynomials known as the S*-polynomials* in *Definition 15.7* to eliminate the leading terms of a pair of polynomials. In so doing and if *lexicographic ordering* is chosen, the process end up with one of the computed S*-polynomials* being a univariate polynomial which can be solved and substituted back in the other S*-polynomials* using the *Extension Theorem* (*D. Cox et al.* 1998, pp. 25–26) to obtain the other variables. The *Groebner bases* approach adds to the treasures of methods that are used to solve *nonlinear algebraic systems of equations* in Geodesy, Photogrammetry, Machine Vision, Robotics and Surveying.

Having defined the term *Groebner basis* and illustrated how the *B. Buchberger algorithm* computes the *Groebner bases*, we briefly mention here how the *Groebner bases* can be computed using algebraic softwares of Mathematica and Maple. In Mathematica Versions 8, the *Groebner basis* command is executed by writing *In[1]:=GroebnerBasis[{polynomoials},{variables},{elims}]* (where In[1]:= is the mathematica prompt) which computes the *Groebner basis* for the *ideal* generated by the *polynomials* with respect to the *monomial order* specified by *monomial order options* with the *variables* specified as in the executable command giving the reduced *Groebner basis*. Without specifying the elims part that indicates the elimination order, one gets too many elements of the *Groebner basis* which may not be relevant. In Maple Version 5 the command is accessed by typing > *with (grobner);* (where > is the Maple prompt and the semicolon ends the Maple command). Once the *Groebner basis* package has been loaded, the execution command then becomes > *g basis (polynomials, variables, termorder)* which computes the *Groebner basis* for the *ideal* generated by the *polynomials* with respect to the *monomial ordering* specified by *termorder* and *variables* in the executable command. Following suggestions from *B. Buchberger* (1999), Mathematica software is applied to the examples presented in Sect. 15-4.

15-313 Multipolynomial Resultants Method

Whereas the resultant of two polynomials is well known and algorithms for computing it are well incorporated into computer algebra packages such as *Maple*, the *Multipolynomial resultant*, i.e. the resultant of more than two polynomials still remain an active area of research. In Geodesy, the use of the two polynomial resultant also known as the *Sylvester resultant* is exemplified in the work of *P. Lohse* (1994, pp. 72–76). This section therefore extends on the use of *Sylvester resultants* to resultants of more than two polynomials of multiple variables *(Multipolynomial resultant)*. *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010) illustrate how the tool can be exploited in Geodesy and Geoinformatics to solve *nonlinear system of equations*.

The necessity of *Multipolynomial resultant* method in Geodesy is due to the fact that many geodetic problems involve the solution of more than two polynomials of multiple variables. This is true since we are living in a three-dimensional world. We shall therefore understand the term *multipolynomial resultants* to mean resultants of more than two polynomials. We treat it as a tool besides the *Groebner bases* and perhaps more powerful to eliminate variables in solution of polynomial systems. Publications on the subject can be found in the works of *G. Salmon* (1876), *F. Macaulay* (1902, 1916), *A. L. Dixon* (1908), *C. Bajaj et al.* (1988), *J. F. Canny* (1988), *J. F. Canny et al.* (1989), *I. M. Gelfand et al.* (1990, 1994), *D. Manocha* (1992, 1993, 1994a,b,c), *D. Manocha* and *J.F. Canny* (1991, 1992, 1993), *G. Lyubeznik* (1995), *S. Krishna and D. Manocha* (1995), *J. Guckenheimer et al.*(1997), *B. Sturmfels* (1994, 1998) and *E. Cattani et al.* (1998). Text books on the subject have been written by *I. Gelfand et al.* (1994), by *D. Cox et al.* (1998, pp.71–122) and, more recently, by *J.L. Awange* and *E.W. Grafarend* (2005), and *J. L. Awange et al.* (2010) who provides interesting material.

In order to understand the *Multipolynomial resultants technique*, we first present the simple case; the resultant of two polynomial also known as the *Sylvester resultant*.

Resultant of two polynomials

Definition 15.8 (Homogeneous polynomial). If monomials of a polynomial p with non zero coefficients have the same *total degree*, the polynomial p is said to be *homogeneous*.

Example 15.12. (Homogeneous polynomial equation):

A homogeneous polynomial equation of total degree 2 is $p = x^2 + y^2 + z^2 + xy + xz + yz$ since the monomials $\{x, y, z, xy, xz, yz\}$ all have the sum of their powers (total degree) being 2.

To set the ball rolling, we examine next the resultant of two univariate polynomials $p, q \in k[x]$ of *positive degree* as

$$\left.\begin{array}{l} p = k_0 x^i + \ldots. + k_i,\ k_0 \neq 0,\ i > 0 \\ q = l_0 x^j + \ldots. + l_j,\ l_0 \neq 0,\ j > 0 \end{array}\right\} \tag{15.40}$$

the resultant of p and q, denoted Res(p, q), is the $(i + j) \times (i + j)$ determinant

$$Res\,(p, q) = det \begin{bmatrix} k_0 & k_1 & k_2 & . & . & . & k_i & 0 & 0 & 0 & 0 & 0 \\ 0 & k_0 & k_1 & k_2 & . & . & . & k_i & 0 & 0 & 0 & 0 \\ 0 & 0 & k_0 & k_1 & k_2 & . & . & . & k_i & 0 & 0 & 0 \\ 0 & 0 & 0 & k_0 & k_1 & k_2 & . & . & . & k_i & 0 & 0 \\ 0 & 0 & 0 & 0 & k_0 & k_1 & k_2 & . & . & . & k_i & 0 \\ 0 & 0 & 0 & 0 & 0 & k_0 & k_1 & k_2 & . & . & . & k_i \\ l_0 & l_1 & l_2 & . & . & . & l_j & 0 & 0 & 0 & 0 & 0 \\ 0 & l_0 & l_1 & l_2 & . & . & . & l_j & 0 & 0 & 0 & 0 \\ 0 & 0 & l_0 & l_1 & l_2 & . & . & . & l_j & 0 & 0 & 0 \\ 0 & 0 & 0 & l_0 & l_1 & l_2 & . & . & . & l_j & 0 & 0 \\ 0 & 0 & 0 & 0 & l_0 & l_1 & l_2 & . & . & . & l_j & 0 \\ 0 & 0 & 0 & 0 & 0 & l_0 & l_1 & l_2 & . & . & . & l_j \end{bmatrix} \quad (15.41)$$

where the coefficients of the first polynomial p of (15.40) occupies j rows while those of the second polynomial q occupies i rows. The empty spaces are occupied by zeros as shown above such that a square matrix is obtained. This resultant is also known as the *Sylvester resultant* and has the following properties (*B. Sturmfels* 1998, *D. Cox et al.* 1998, §3.5)

1. Res(p, q) is a polynomial in $k_0, \ldots, k_i,\ l_0, \ldots, l_j$ with integer coefficients
2. Res$(p, q) = 0$ if and only if $p(x)$ and $q(x)$ have a common factor in $Q[x]$
3. There exist a polynomial $r,\ s \in Q[x]$ such that $rp + sq = Res(p, q)$

Sylvester resultants can be used to solve two polynomial equations as shown in the example below

Example 15.13. (D. Cox et al.1998, p.72):

Consider the two equations

$$\left.\begin{array}{r} p := xy - 1 = 0 \\ q := x^2 + y^2 - 4 = 0. \end{array}\right\} \quad (15.42)$$

In order to eliminate one variable, we use the *hide variable technique* i.e. we consider one variable say y as a constant (of degree zero). We then have the *Sylvester resultant* from (15.41) as

$$Res\,(p, q, x) = det \begin{bmatrix} y & -1 & 0 \\ 0 & y & -1 \\ 1 & 0 & y^2 - 4 \end{bmatrix} = y^4 - 4y^2 + 1 \quad (15.43)$$

which can be readily solved for the variable y and substituted back in any of the equations in (15.42) to get the values of the other variable x. For two polynomials, the construction of resultant is relatively simpler and algorithms for the execution are incorporated in computer algebra algorithms. For the resultant of more than 2 polynomials of multiple variables, we turn to the *Multipolynomial resultants*.

Multipolynomial resultants

In defining the term *multipolynomial resultant*, *D. Manocha* (1994c) writes:

"Elimination theory, a branch of classical algebraic geometry, deals with conditions for common solutions of a system of polynomial equations. Its main result is the construction of a single resultant polynomial of n homogeneous polynomial equations in n unknowns, such that the vanishing of the resultant is a *necessary* and *sufficient* condition for the given system to have a non-trivial solution. We refer to this resultant as the *multipolynomial resultant* and use it in the algorithm presented in the paper".

We present here two approaches for the formation of the design matrix whose determinant we need; first the approach based on *F. Macaulay* (1902) formulation and then a more modern approach based on *B. Sturmfels* (1998) formulation.

Approach based on F. Macaulay *(1902) formulation:* With n polynomials, the construction of the matrix whose entries are the coefficients of the polynomials $f_1, \ldots, f_n$ can be done in five steps as illustrated in the following approach of *F. Macaulay* (1902):

Step 1: The given polynomials $f_1 = 0, \ldots, f_n = 0$ are considered to be homogeneous equations in the variables $x_1, \ldots, x_n$ and if not, they are homogenized. Let the degree of the polynomial f_i be d_i. The first step involves the determination of the *critical degree* given by *C. Bajaj et al.* (1988) as

$$d = 1 + \sum (d_i - 1). \tag{15.44}$$

Step 2: Once the *critical degree* has been established, the given monomials of the polynomial equations are multiplied with each other to generate a set X whose elements consists of monomials whose total degree equals the *critical degree*. Thus if we are given the polynomial equations $f_1 = 0, \ldots, f_n = 0$, then each monomial of $f_1$is multiplied by those of $f_2, \ldots, f_n$, those of f_2 are multiplied by those of $f_3, \ldots, f_n$ until those of f_{n-1} are multiplied by those of f_n. The set X of monomials generated in this form is

$$X^d = \{x^d \mid d = \alpha_1 + \alpha_2 + \cdots + \alpha_n\} \tag{15.45}$$

and the variable $x^d = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$.

Step 3: The set X containing the monomials each of total degree d is now partitioned according to the following criteria (*J. F. Canny* 1988, p. 54)

$$\begin{cases} X_1^d = & \{x^\alpha \in X^d \mid \alpha_1 \geq d_1\} \\ X_2^d = & \{x^\alpha \in X^d \mid \alpha_2 \geq d_2 \; and \; \alpha_1 < d_1\} \\ . \quad . & . \\ . \quad . & . \\ . \quad . & . \\ X_n^d = \{x^\alpha \in X^d \mid \alpha_n \geq d_n \; and \; \alpha_i < d_i, for\, i = 1, \ldots, n-1\}. \end{cases} \tag{15.46}$$

The resulting sets of X_i^d are disjoint and every element of X^d is contained in exactly one of them.

Step 4: From the resulting subsets $X_i^d \subset X^d$, a set of polynomials F_i which are homogeneous in n variables are defined as follows

$$F_i = \frac{X_i^d}{x_i^{d_i}} f_i \tag{15.47}$$

from which a *square matrix* $\boldsymbol{A}$ is now formed with the row elements being the *coefficients* of the monomials of the polynomials $F_i \;|_{i=1,\ldots,n}$ and the columns correspond to the N monomials of the set X^d. The formed square matrix is of the order $\binom{d+n-1}{d} \times \binom{d+n-1}{d}$ and is such that for a given polynomial F_i in (15.47), the row is made up of the symbolic coefficients of each polynomial. The square matrix $\boldsymbol{A}$ has a special property that the non trivial solution of the homogeneous equations F_i which also form the solution of the original equations f_i are in its null space. This implies that the matrix must be singular or its determinant, $det(\boldsymbol{A})$, must be zero. For the determinant to vanish therefore, the original equations f_i and their homogenized counterparts F_i must have the same non trivial solutions.

Step 5: After computing the determinant of the square matrix $\boldsymbol{A}$ above, *F. Macaulay* (1902) suggests the computation of *extraneous factor* in order to obtain the resultant. *D. Cox et al.* (1998, *Proposition* 4.6, p.99) explains the *extraneous factors* to be integer polynomials in the coefficients of $\bar{F}_0, \ldots, \bar{F}_{n-1}$, where $\bar{F}_i = F_i(x_0, \ldots, x_{n-1}, 0)$ and is related to the determinant via

$$determinant = Res(F_1, \ldots, F_n).Ext \tag{15.48}$$

with the determinant computed as in step 4, $Res(F_1, \ldots, F_n)$ being the *Multipolynomial resultant* and *Ext* the extraneous factor. This expression was established as early as 1902 by *F. Macaulay* (1902) and this procedure of *resultant* formulation named after him. *F. Macaulay* (1902) determines the extraneous factor from the sub-matrix of the $N \times N$ square matrix $\boldsymbol{A}$ and calls it a factor of minor obtained by deleting rows and columns of the $N \times N$ matrix $\boldsymbol{A}$. A monomial x^α of total degree d is said to be reduced if $x_i^{d_i}$ divides x^α for exactly one i. The *extraneous factor* is obtained by computing the determinant of the sub-matrix of the coefficient matrix $\boldsymbol{A}$ obtained by deleting rows and columns corresponding to reduced monomials x^α.

From the relationship of step 5, it suffices for our purpose to solve for the unknown variable hidden in the coefficients of the polynomials f_i by obtaining the determinant of the $N \times N$ square matrix $\boldsymbol{A}$ and equating it to zero neglecting the extraneous factor. This is because the extraneous factor is an integer polynomial and as such not related to the variable in the determinant of $\boldsymbol{A}$. The existence of the non-trivial solution provides the *necessary* and *sufficient* conditions for the vanishing of the determinant. It should be pointed out that there exists several procedures for computing the resultants as exemplified in the works of *D. Manocha* (1992, 1993, 1994a,b,c) who solves the *Multipolynomial resultants* using the eigenvalue-eigenvector approach, *J. F. Canny* (1988) who solves the resultant using the characteristic polynomial approach and *B. Sturmfels* (1994, 1998) who proposes a more compact approach for solving the resultants of a ternary quadric using the Jacobian determinant approach which we present below.

Approach based on B. Sturmfels *(1998, p.26) formulation:* Given three homogeneous equations of degree two as follows

$$\begin{aligned}
F_1 &:= a_{11}x^2 + a_{12}y^2 + a_{13}z^2 + a_{14}xy + a_{15}xz + a_{16}yz = 0\\
F_2 &:= a_{21}x^2 + a_{22}y^2 + a_{23}z^2 + a_{24}xy + a_{25}xz + a_{26}yz = 0\\
F_3 &:= a_{31}x^2 + a_{32}y^2 + a_{33}z^2 + a_{34}xy + a_{35}xz + a_{36}yz = 0
\end{aligned} \tag{15.49}$$

we compute the Jacobian determinants of (15.49) by

$$J = det\begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \\ \frac{\partial F_3}{\partial x} & \frac{\partial F_3}{\partial y} & \frac{\partial F_3}{\partial z} \end{bmatrix} \tag{15.50}$$

which is a cubic polynomial in the coefficients $\{x, y, z\}$. Since the determinant polynomial J in (15.50) is a cubic polynomial, its partial derivatives will be quadratic polynomials in variables $\{x, y, z\}$ and can be written in the form

$$
\begin{aligned}
\frac{\partial J}{\partial x} &:= b_{11}x^2 + b_{12}y^2 + b_{13}z^2 + b_{14}xy + b_{15}xz + b_{16}yz = 0 \\
\frac{\partial J}{\partial y} &:= b_{21}x^2 + b_{22}y^2 + b_{23}z^2 + b_{24}xy + b_{25}xz + b_{26}yz = 0 \\
\frac{\partial J}{\partial z} &:= b_{31}x^2 + b_{32}y^2 + b_{33}z^2 + b_{34}xy + b_{35}xz + b_{36}yz = 0.
\end{aligned}
\tag{15.51}
$$

The coefficients b_{ij} in (15.51) are quadratic polynomials in a_{ij} of equation (15.49). The final step in computing the resultant of the initial system (15.49) involves the computation of the determinant of a 6×6 matrix given by

$$
Res_{222}(F_1, F_2, F_3) = det \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{12} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & a_{36} \end{bmatrix}. \tag{15.52}
$$

The resultant (15.52) vanishes if and only if (15.49) have a common solution $\{x,\ y,\ z\}$, where $\{x,\ y,\ z\}$ are complex numbers or real numbers not all equal zero. In Sect. 15-4, we use the procedure to solve the GPS pseudo-range equations.

15-32 Adjustment of the combinatorial subsets

Once the *combinatorial minimal subsets* have been solved using either the *Groebner bases* or the *Multipolynomial resultant* approach, the resulting sets of solution are considered as pseudo-observations. For each combinatorial, the obtained minimal subset solutions considered as pseudo-observations are used as the approximate values to generate the dispersion matrix via the nonlinear error propagation law/variance-covariance propagation (*e.g. E. Grafarend* and *B. Schaffrin*, 1993, pp. 469–471) as follows:

From the nonlinear geodetic observation equations that have been converted into its algebraic (polynomial) form, the *combinatorial minimal subsets* will consist of polynomials $f_1, \ldots, f_m \in k[x_1, \ldots, x_m]$ with $\{x_1, \ldots, x_m\}$ being the unknown variables (fixed parameters) to be determined and $\{y_1, \ldots, y_n\}$ the known variables comprising the observations or pseudo-observations. We write the polynomials as

$$\left.\begin{array}{l} f_1 := g(x_1,\ldots,x_m,y_1,\ldots,y_n) = 0 \\ f_2 := g(x_1,\ldots,x_m,y_1,\ldots,y_n) = 0 \\ \qquad\qquad . \\ \qquad\qquad . \\ \qquad\qquad . \\ f_m := g(x_1,\ldots,x_m,y_1,\ldots,y_n) = 0 \end{array}\right\} \tag{15.53}$$

which is expressed in matrix form as

$$\boldsymbol{f} := \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0} \tag{15.54}$$

where the unknown variables $\{x_1,\ldots,x_m\}$ are placed in a vector $\boldsymbol{x}$ and the known variables $\{y_1,\ldots,y_n\}$ are placed in the vector $\boldsymbol{y}$. Next, we implement the error propagation from the observations (pseudo-observations) $\{y_1,\ldots,y_n\}$ to the parameters $\{x_1,\ldots,x_m\}$ that are to be explicitly determined namely characterized by the *first moments*, the expectation $E\{\boldsymbol{x}\} = \boldsymbol{\mu}_x$ and $E\{\boldsymbol{y}\} = \boldsymbol{\mu}_y$, as well as the *second moments*, the variance-covariance matrices/dispersion matrices $D\{\boldsymbol{x}\} = \boldsymbol{\Sigma}_x$ and $D\{\boldsymbol{y}\} = \boldsymbol{\Sigma}_y$. From *E. Grafarend* and *B. Schaffrin* (1993, pp. 470–471), we have up to nonlinear terms

$$D\{\boldsymbol{x}\} = \boldsymbol{J}_x^{-1}\boldsymbol{J}_y\boldsymbol{\Sigma}_y\boldsymbol{J}_y'(\boldsymbol{J}_x^{-1})' \tag{15.55}$$

with $\boldsymbol{J}_x, \boldsymbol{J}_y$ being the partial derivatives of (15.54) with respect to $\boldsymbol{x}$,$\boldsymbol{y}$ respectively at the Taylor points $(\boldsymbol{\mu}_x, \boldsymbol{\mu}_y)$. The approximate values of unknown parameters $\{x_1,\ldots,x_m\} \in \boldsymbol{x}$ appearing in the Jacobi matrices $\boldsymbol{J}_x, \boldsymbol{J}_y$ are obtained from *Groebner bases* or *Multipolynomial resultants* solution of the *nonlinear system of equations* (15.53).

Given $\boldsymbol{J}_i = \boldsymbol{J}_{x_i}^{-1}\boldsymbol{J}_{y_i}$ from the ith combination and $\boldsymbol{J}_j = \boldsymbol{J}_{x_j}^{-1}\boldsymbol{J}_{y_j}$ from the jth combinatorials, the correlation between the ith and jth combination is given by

$$\boldsymbol{\Sigma}_{ij} = \boldsymbol{J}_j\boldsymbol{\Sigma}_{y_j y_i}\boldsymbol{J}_i' \tag{15.56}$$

The sub-matrices variance-covariance matrix for the individual combinatorials $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \ldots, \boldsymbol{\Sigma}_k$ (where k is the number of combinations) obtained via (15.55) and the correlations between combinatorials obtained from (15.56) form the variance-covariance/dispersion matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} & . & . & . & \boldsymbol{\Sigma}_{1k} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 & . & . & . & \boldsymbol{\Sigma}_{2k} \\ . & & \boldsymbol{\Sigma}_3 & & & \\ . & & & . & & \\ . & & & & . & \\ \boldsymbol{\Sigma}_{k1} & . & . & . & & \boldsymbol{\Sigma}_k \end{bmatrix} \tag{15.57}$$

for the entire k combinations. The obtained dispersion matrix $\boldsymbol{\Sigma}$ is then used in the *linear Gauss–Markov model* to obtain the estimates $\hat{\boldsymbol{\xi}}$ of the unknown parameters

$\boldsymbol{\xi}$ with the combinatorial solution points in a polyhedron considered as pseudo-observations in the vector $\boldsymbol{y}$ of observations while the design matrix $\boldsymbol{A}$ comprises of integer values 1 being the coefficients of the unknowns as in (15.60). In order to understand the adjustment process, we consider the following example.

Example 15.14.

Consider that one has four observations in three unknowns in a nonlinear problem. Let the observations be given by $\{y_1, y_2, y_3, y_4\}$ leading to four combinations giving the solutions $z_I(y_1, y_2, y_3)$, $z_{II}(y_2, y_3, y_4)$, $z_{III}(y_1, y_3, y_4)$ and $z_{IV}(y_1, y_2, y_4)$. If the solutions are placed in a vector $z_J = [z_I \;\; z_{II} \;\; z_{III} \;\; z_{IV}]'$, the adjustment model is then defined as

$$E\{z_J\} = \boldsymbol{I}_{12\times3}\boldsymbol{\xi}_{3\times1},\ D\{z_J\}\mathit{from\ Variance/Covariance\ propagation}. \tag{15.58}$$

Let

$$\boldsymbol{\xi}^n = \boldsymbol{L}z_J \ \mathit{subject\ to}\ z_J := \begin{bmatrix} z_I \\ z_{II} \\ z_{III} \\ z_{IV} \end{bmatrix} \in \mathbb{R}^{12\times1} \tag{15.59}$$

such that the postulations $trD\{\boldsymbol{\xi}^n\} = min$ i.e. "*best*" and $E\{\boldsymbol{\xi}^n\} = \boldsymbol{\xi}\ \mathit{for\ all}\ \boldsymbol{\xi}^n \in \mathbb{R}^m$ i.e. "*uniformly unbiased*" holds. We then have from (15.57), (15.58) and (15.59) the result

$$\begin{aligned} \hat{\boldsymbol{\xi}} &= (\boldsymbol{I}'_{3\times12}\boldsymbol{\Sigma}_{z_J}\boldsymbol{I}_{12\times3})\boldsymbol{I}'_{3\times12}\boldsymbol{\Sigma}^{-1}_{z_J}z_J \\ \hat{\boldsymbol{L}} &= arg\{trD\{\boldsymbol{\xi}^n\} = tr\,\boldsymbol{L}\boldsymbol{\Sigma}_y\boldsymbol{L}' = min \mid UUE\} \end{aligned} \tag{15.60}$$

The dispersion matrix $D\{\hat{\boldsymbol{\xi}}\}$ of the estimates $\hat{\boldsymbol{\xi}}$ is obtained. The shift from arithmetic weighted mean to the use of *linear Gauss Markov model* is necessitated as we do not readily have the weights of the minimal combinatorial subsets but their dispersion which we obtain via *error propagation/Variance-Covariance propagation*. If we employ the equivalence *theorem* of *E. Grafarend* and *B. Schaffrin* (1993, pp. 339–341), an adjustment using linear Gauss markov model instead of weighted arithmetic mean is permissible.

Example 15.15. (error propagation for planar ranging problem):

Let us consider a simple case of the planar ranging problem. From an unknown point $P(X, Y) \in \mathbb{E}^2$, let distances S_1 and S_2 be measured to two known points $P_1(X_1, Y_1) \in \mathbb{E}^2$ and $P_2(X_2, Y_2) \in \mathbb{E}^2$ respectively. We have the distance equations expressed as

$$\begin{bmatrix} S_1^2 = (X_1 - X)^2 + (Y_1 - Y)^2 \\ S_2^2 = (X_2 - X)^2 + (Y_2 - Y)^2 \end{bmatrix} \tag{15.61}$$

which we express in algebraic form (15.53) as

$$\begin{bmatrix} f_1 := (X_1 - X)^2 + (Y_1 - Y)^2 - S_1^2 = 0 \\ \\ f_2 := (X_2 - X)^2 + (Y_2 - Y)^2 - S_2^2 = 0 \end{bmatrix}. \tag{15.62}$$

On taking total differential of (15.62) we have

$$\begin{bmatrix} df_1 := 2(X_1 - X)dX_1 - 2(X_1 - X)dX + 2(Y_1 - Y)dY_1 - \\ -2(Y_1 - Y)dY - 2S_1 dS_1 = 0 \\ \\ df_2 := 2(X_2 - X)dX_2 - 2(X_2 - X)dX + 2(Y_2 - Y)dY_2 - \\ -2(Y_2 - Y)dY - 2S_2 dS_2 = 0 \end{bmatrix} \tag{15.63}$$

which on arranging the differential vector of the unknown terms $\{X, Y\} = \{x_1, x_2\} \in \boldsymbol{x}$ on the left hand side and that of the known terms $\{X_1, Y_1, X_2, Y_2, S_1, S_2\} = \{y_1, y_2, y_3, y_4, y_5, y_6\} \in \boldsymbol{y}$ on the right hand side leads to

$$\boldsymbol{J}_x \begin{bmatrix} dX \\ dY \end{bmatrix} = \boldsymbol{J}_y \begin{bmatrix} dS_1 \\ dX_1 \\ dY_1 \\ dS_2 \\ dX_2 \\ dY_2 \end{bmatrix} \tag{15.64}$$

with

$$\boldsymbol{J}_x = \begin{bmatrix} \frac{\partial f_1}{\partial X} & \frac{\partial f_1}{\partial Y} \\ \\ \frac{\partial f_2}{\partial X} & \frac{\partial f_2}{\partial Y} \end{bmatrix} = \begin{bmatrix} -2(X_1 - X) & -2(Y_1 - Y) \\ -2(X_2 - X) & -2(Y_2 - Y) \end{bmatrix} \tag{15.65}$$

and

$$\boldsymbol{J}_y = \left[\begin{bmatrix} \frac{\partial f_1}{\partial S_1} & \frac{\partial f_1}{\partial X_1} & \frac{\partial f_1}{\partial Y_1} & 0 & 0 & 0 \\ \\ 0 & 0 & & \frac{\partial f_2}{\partial S_2} & \frac{\partial f_2}{\partial X_2} & \frac{\partial f_2}{\partial Y_2} \end{bmatrix} = \right.$$
$$\left. = \begin{bmatrix} 2S_1 & -2(X_1 - X) & -2(Y_1 - Y) & 0 & 0 & 0 \\ 0 & 0 & 0 & 2S_2 & -2(X_2 - X) & -2(Y_2 - Y) \end{bmatrix} \right] \tag{15.66}$$

If we consider that

$$D\{\boldsymbol{x}\} = \boldsymbol{\Sigma}_x = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{bmatrix} \tag{15.67}$$

and

$$D\{\boldsymbol{y}\} = \boldsymbol{\Sigma}_y = \begin{bmatrix} \sigma_{S_1}^2 & \sigma_{S_1X_1} & \sigma_{S_1Y_1} & \sigma_{S_1X_2} & \sigma_{S_1S_2} & \sigma_{S_1Y_2} \\ \sigma_{X_1S_1} & \sigma_{X_1}^2 & \sigma_{X_1Y_1} & \sigma_{X_1S_2} & \sigma_{X_1X_2} & \sigma_{X_1Y_2} \\ \sigma_{Y_1S_1} & \sigma_{Y_1X_1} & \sigma_{Y_1}^2 & \sigma_{Y_1S_2} & \sigma_{Y_1X_2} & \sigma_{Y_1Y_2} \\ \sigma_{S_2S_1} & \sigma_{S_2X_1} & \sigma_{S_2Y_1} & \sigma_{S_2}^2 & \sigma_{S_2X_2} & \sigma_{S_2Y_2} \\ \sigma_{X_2S_1} & \sigma_{X_2X_1} & \sigma_{X_2Y_1} & \sigma_{X_2S_2} & \sigma_{X_2}^2 & \sigma_{X_2Y_2} \\ \sigma_{Y_2S_1} & \sigma_{Y_2X_1} & \sigma_{Y_2Y_1} & \sigma_{Y_2S_2} & \sigma_{Y_2X_2} & \sigma_{Y_2}^2 \end{bmatrix} \tag{15.68}$$

we obtain with (15.64), (15.65) and (15.66) the dispersion (15.55) of the unknown variables $\{X, Y\} = \{x_1, x_2\} \in \boldsymbol{x}$. The unknown values of $\{X, Y\} = \{x_1, x_2\} \in \boldsymbol{x}$ appearing in the Jacobi matrices (15.65) and (15.66) are obtained from *Groebner bases* or *Multipolynomial resultants* solution of the *nonlinear system of equations* (15.61).

15-4 Examples

In this section, we present three examples from *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010) based on geodetic problems of *Minimum Distance Mapping*, *GPS pseudo-ranging four-point P4P*, and the *7-parameter transformation*.

Example 15.16. (Minimum distance mapping)

In order to relate a point P on the Earth's topographic surface to a point on the international reference ellipsoid $\mathbb{E}^2_{a,a,b}$, a bundle of half straight lines so called *projection lines* which depart from P and intersect $\mathbb{E}^2_{a,a,b}$ either not at all or in two points are used. There is *one projection line* which is at minimum distance relating P to p. Figure 15.1 is an illustration of such a minimum distance mapping. Let us formulate such an optimization problem by means of the Lagrangean $£(x_1, x_2, x_3, x_4)$ in Solution 15.1.

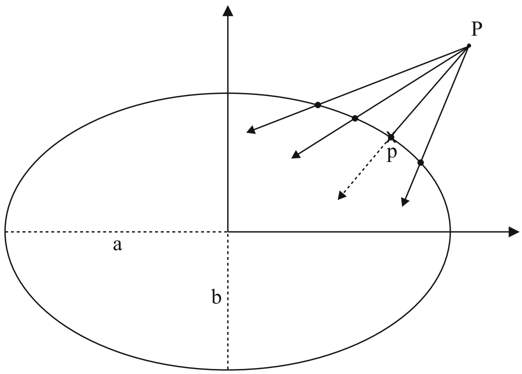


Fig. 15.1 Minimum distance mapping of a point P on the Earth's topographic surface to a point p on the *international reference ellipsoid* $\mathbb{E}^2_{a,a,b}$

Solution 15.1 (Constraint minimum distance mapping in terms of Cartesian coordinates).

$$\begin{aligned}
\pounds(x_1, x_2, x_3, x_4) &:= \frac{1}{2}\|\mathbf{X} - \mathbf{x}\|^2 + \frac{1}{2}x_4\left[b^2(x_1^2 + x_2^2) + ax_3^2 - a^2b^2\right] \\
&= \frac{1}{2}\{(X - x_1)^2 + (Y - x_2)^2 + (Z - x_3)^2 \\
&\quad + x_4\left[b^2(x_1^2 + x_2^2) + ax_3^2 - a^2b^2\right]\} \\
&= \underset{(x_1, x_2, x_3, x_4)}{min}
\end{aligned} \tag{15.69}$$

$$\mathbf{x} \in \mathbb{X} := \{\mathbf{x} \in \mathbb{R}^3 | \frac{x_1^2 + x_2^2}{a^2} + \frac{x_3^2}{b^2} = 1\} =: \mathbb{E}^2_{a,a,b} \tag{15.70}$$

In the first case, the *Euclidean distance* between points P and p in terms of *Cartesian coordinates* of $P(X, Y, Z)$ and of $p(x_1, x_2, x_3)$ is represented. The *Cartesian coordinates* (x_1, x_2, x_3) of the projection point P are unknown. The constraint that the point p is an element of the ellipsoid of revolution

$$\mathbb{E}^2_{a,a,b} := \{\mathbf{x} \in \mathbb{R}^3 | b^2(x_1^2 + x_2^2) + a^2x_2^3 - a^2b^2 = 0, \mathbb{R}^+ \ni a > b \in \mathbb{R}^+\}$$

is substituted into the Lagrangean by means of the Lagrange multiplier x_4, which is unknown too. $\{(x_1^{\wedge}, x_2^{\wedge}, x_3^{\wedge}, x_4^{\wedge}) = arg\{\pounds(x_1, x_2, x_3, x_4) = min\}$ is the argument of the minimum of the *constrained Lagrangean* $\pounds(x_1, x_2, x_3, x_4)$. The result of the minimization procedure is presented by Lemma 15.10. Equation (15.71) provides the *necessary conditions* to constitute an extremum: The normal equations are of bilinear type. *Products* of the unknowns for instance x_1x_4, x_2x_4, x_3x_4 and *squares* of the unknowns, for instance x_1^2, x_2^2, x_3^2 appear. Finally the matrix of second derivatives $\mathbf{H}_3$ in (15.73) which is *positive definite* constitutes the *sufficient condition* to obtain a minimum. Fortunately the matrix of second derivatives $\mathbf{H}_3$ is *diagonal*. Using (15.72i–15.72iv), together with (15.75) leads to (15.76), which are the eigenvalues of the Hesse matrix $\mathbf{H}_3$. These values are $\Lambda_1 = \Lambda_2 = X\backslash x_1^\wedge$, $\Lambda_3 = Z\backslash x_3^\wedge$ and must be positive.

Lemma 15.10 (Constrained minimum distance mapping).
The functional $\pounds(x_1, x_2, x_3, x_4)$ is minimal, if the conditions (15.71) and (15.73) hold.

$$\boxed{\frac{\partial \pounds}{\partial x_i}((x_1^\wedge, x_2^\wedge, x_3^\wedge, x_4^\wedge)) = 0 \quad \forall \quad \text{i=1,2,3,4.}} \tag{15.71}$$

On taking partial derivatives with respect to x_i, we have

$$\left[\begin{array}{ll} (i) & \dfrac{\partial \pounds}{\partial (x_1^\wedge)} = -(X - x_1^\wedge) + b^2 x_1^\wedge x_4^\wedge = 0 \\[2ex] (ii) & \dfrac{\partial \pounds}{\partial (x_2^\wedge)} = -(Y - x_2^\wedge) + b^2 x_2^\wedge x_4^\wedge = 0 \\[2ex] (iii) & \dfrac{\partial \pounds}{\partial (x_3^\wedge)} = -(Z - x_3^\wedge) + a^2 x_3^\wedge x_4^\wedge = 0 \\[2ex] (iv) & \dfrac{\partial \pounds}{\partial (x_4^\wedge)} = \dfrac{1}{2}[b^2(x_1^{\wedge 2} + x_2^{\wedge 2})] + a^2 x_3^{\wedge 2} - a^2 b^2 = 0 \end{array}\right. \tag{15.72}$$

$$\boxed{\frac{\partial^2 \pounds}{\partial x_i \partial x_j}(x_1^\wedge, x_2^\wedge, x_3^\wedge, x_4^\wedge)) > 0 \quad \forall \quad \text{i,j} \in \{1,2,3\}.} \tag{15.73}$$

$$\begin{aligned} \mathbf{H}_3 &:= \left[\frac{\partial^2 \pounds}{\partial x_i \partial x_j}(\mathbf{x}^\wedge)\right] \\ &= \begin{bmatrix} 1 + b^2 x_4^\wedge & 0 & \\ 0 & 1 + b^2 x_4^\wedge & 0 \\ 0 & 0 & 1 + a^2 x_4^\wedge \end{bmatrix} \quad \in \quad \mathbb{R}^{3\times 3} \end{aligned} \tag{15.74}$$

"*eigenvalues*"

$$|\mathbf{H}_3 - \Lambda \mathbf{I}_3| = 0 \quad \Longleftrightarrow \tag{15.75}$$

$$\left[\begin{array}{l} \Lambda_1 = \Lambda_2 := 1 + b^2 x_4^\wedge = \dfrac{X}{x_1^\wedge} = \dfrac{Y}{x_2^\wedge} \\ \Lambda_3 := 1 + a^2 x_4^\wedge = \dfrac{Z}{x_3^\wedge} \end{array}\right. \tag{15.76}$$

Without the various forward and backward reduction steps, we could automatically generate an equivalent algorithm for solving the normal equations (15.72i)–(15.72iv) in a closed form by means of Groebner basis approach.

Groebner basis computation of the normal equations (15.72i)–(15.72iv) leads to 14 elements presented in Solution 15.2 interpreted as follows: The *first expression* is a univariate polynomial of order four (quartic) in the Lagrange multiplier, i.e.,

$$\left[\begin{array}{l} \boxed{c_4 x_4^4 + c_3 x_4^3 + c_2 x_4^2 + c_1 x_4 + c_o = 0} \\ c_4 = a^6 b^6 \\ c_3 = (2a^6 b^4 + 2a^4 b^6) \\ c_2 = (a^6 b^2 + 4a^4 b^4 + a^2 b^6 - a^4 b^2 X^2 - a^4 b^2 Y^2 - a^2 b^4 Z^2) \\ c_1 = (2a^4 b^2 + 2a^2 b^4 - 2a^2 b^2 X^2 - 2a^2 b^2 Y^2 - 2a^2 b^2 Z^2) \\ c_o = (a^2 b^2 - b^2 X^2 - b^2 Y^2 - a^2 Z^2). \end{array}\right. \tag{15.77}$$

With the admissible values x_4 substituted in the linear equations (4),(8),(12) of the computed Groebner basis, i.e.,

$$\left[\begin{array}{l} (1 + a^2 x_4)x_3 - Z \\ (1 + b^2 x_4)x_2 - Y \\ (1 + b^2 x_4)x_1 - X, \end{array}\right. \tag{15.78}$$

the values $(x_1, x_2, x_3) = (x, y, z)$ are finally produced.

Solution 15.2 (Groebner basis MDM solution; see Awange et al. 2010 for further details and examples).

(1)

$$\left[\begin{array}{l} a^2 b^2 x_4^4 + (2a^6 b^4 + 2a^4 b^6)x_4^3 + (a^6 b^2 + 4a^4 b^4 + a^2 b^6 - a^4 b^2 X^2 - a^4 b^2 Y^2 - \\ a^2 b^4 Z^2)x_4^2 + +(2a^4 b^2 + 2a^2 b^4 - 2a^2 b^2 X^2 - 2a^2 b^2 Y^2 - 2a^2 b^2 Z^2)x_4 \\ +(a^2 b^2 - b^2 X^2 - b^2 Y^2 - a^2 Z^2). \end{array}\right.$$

(2)

$$\left[\begin{array}{l} (a^4 Z - 2a^2 b^2 Z + b^4 Z)x_3 - a^6 b^6 x_4^3 - (2a^6 b^4 + a^4 b^6)x_4^2 \\ -(a^6 b^2 + 2a^4 b^4 - a^4 b^2 X^2 - a^4 b^2 Y^2 - a^2 b^4 Z^2)x_4 \\ -a^2 b^4 + a^2 b^2 X^2 + a^2 b^2 Y^2 + 2a^2 b^2 Z^2 - b^4 Z^2. \end{array}\right.$$

(3)
$$\left[\begin{array}{l}(2b^2Z + b^4x_4Z - a^2Z)x_3 + a^4b^6x_4^3 + (2a^4b^4 + a^2b^6)x_4^2 \\ +(a^4b^2 + 2a^2b^4 - a^2b^2X^2 - a^2b^2Y^2 - b^4Z^2)x_4 \\ +a^2b^2 - b^2X^2 - b^2Y^2 - 2b^2Z^2.\end{array}\right.$$

(4)
$$(1 + a^2x_4)x_3 - Z$$

(5)
$$\left[\begin{array}{l}(a^4 - 2a^2b^2 + b^4)x_3^2 + (2a^2b^2Z - 2b^4Z)x_3 \\ -a^4b^6x_4^2 - 2a^4b^4x_4 - a^4b^2 + a^2b^2X^2 + a^2b^2Y^2 + b^4Z^2).\end{array}\right.$$

(6)
$$\left[\begin{array}{l}(2b^2 - a^2 + b^4x_4)x_3^2 - a^2Zx_3 + a^4b^6x_4^3 + (2a^4b^4 + 2a^2b^6)x_4^2 \\ + + (a^4b^2 + 4a^2b^4 - a^2b^2X^2 - a^2b^2Y^2 - b^4Z^2)x_4 \\ +2a^2b^2 - 2b^2X - 2bY^2 - 2b^2Z^2.\end{array}\right.$$

(7)
$$\left[\begin{array}{l}(X^2 + Y^2)x_2 + a^2b^4Yx_4^2 + Y(a^2b^2 - b^2x_3^2 - b^2Zx_3)x_4 \\ +Yx_3^2 - Y^3 - YZx_3 - YX^2.\end{array}\right.$$

(8)
$$(1 + b^2x_4)x_2 - Y$$

(9)
$$a^2x_3 - b^2x_3 + b^2Z)x_2 - a^2x_3Y$$

(10)
$$Yx_1 - Xx_2$$

(11) $$Xx_1 + a^2b^4x_4^2 + (a^2b^2 + b^2x_3^2 - b^2Zx_3)x_4 + x_3^2 - Zx_3 + Yx_2 - X^2 - Y^2.$$

(12)
$$(1 + b^2x_4)x_1 - X$$

(13)
$$(a^2x_3 - b^2x_3 + b^2Z)x_1 - a^2Xx_3$$

(14) $$x_1^2 + a^2b^4x_4^2 + (2a^2b^2 + b^2x_3^2 - b^2Zx_3)x_4 + 2x_3^2 - 2Zx_3 + x_2^2 - X^2 - Y^2.$$

Example 15.17. (GNSS positioning)

E. Grafarend and *J. Shan* (1996) have defined the *GPS pseudo-ranging four-point problem* ("pseudo 4P") as the problem of determining the four unknowns comprising the *three components of the receiver position* and the stationary receiver *range bias* from four observed pseudo-ranges to four satellite transmitter of given geocentric position. Geometrically, the four unknowns are obtained from the intersection of four spherical cones given by the pseudo-ranging equations. Several procedures have been put forward for obtaining closed form solution of the problem. Amongst the procedures include the vectorial approach evidenced in the works of *S. Bancroft* (1985)*, P. Singer et al.* (1993), *H. Lichtenegger* (1995) and *A. Kleusberg* (1994,1999). *E. Grafarend* and *J. Shan* (1996) propose two approaches.

One approach is based on the inversion of a 3×3 coefficient matrix of a linear system formed by differencing of the *nonlinear pseudo-ranging equations* in geocentric coordinates, while the other approach uses the coefficient matrix from the linear system to solve the same equations in barycentric coordinates. In this example we present both the approaches of *Groebner bases* and *Multipolynomial resultants* (*B. Sturmfel* 1998 approach) to solve the same problem. We demonstrate our algorithms by solving the *GPS Pseudo-ranging four-point problem* already solved by *A. Kleusberg* (1994) and *E. Grafarend* and *J. Shan* (1996). Both the *Groebner basis* and the *Multipolynomial resultant* solve the same linear equations as those of *E. Grafarend* and *J. Shan* (1996) and lead to identical results (see also *J. L. Awange* and *E. Grafarend* 2002). We start with the pseudo-ranging equations written algebraically as

$$\left[\begin{array}{l} (x_1-a_0)^2+(x_2-b_0)^2+(x_3-c_0)^2-(x_4-d_0)^2=0 \\ (x_1-a_1)^2+(x_2-b_1)^2+(x_3-c_1)^2-(x_4-d_1)^2=0 \\ (x_1-a_2)^2+(x_2-b_2)^2+(x_3-c_2)^2-(x_4-d_2)^2=0 \\ (x_1-a_3)^2+(x_2-b_3)^2+(x_3-c_3)^2-(x_4-d_3)^2=0 \\ \qquad\qquad \textit{where} \quad x_1,x_2,x_3,x_4 \in \\ \qquad\qquad (a_0,b_0,c_0)=(x^0,y^0,z^0) \sim P^0 \\ \qquad\qquad (a_1,b_1,c_1)=(x^1,y^1,z^1) \sim P^1 \\ \qquad\qquad (a_2,b_2,c_2)=(x^2,y^2,z^2) \sim P^2 \\ \qquad\qquad (a_3,b_3,c_3)=(x^3,y^3,z^3) \sim P^3 \end{array}\right. \tag{15.79}$$

with $\{P^0, P^1, P^2, P^3\}$ being the position of the four GPS satellites, their ranges to the stationary receiver at P given by $\{d_0, d_1, d_2, d_3\}$. The parameters

$$(\{a_0,b_0,c_0\},\{a_1,b_1,c_1\},\{a_2,b_2,c_2\},\{a_3,b_3,c_3\},\{d_0,d_1,d_2,d_3\})$$

are elements of the spherical cone that intersect at P to give the coordinates $\{x_1, x_2, x_3\}$ of the receiver and the stationary receiver range bias x_4. The equations above is expanded and arranged in the *lexicographic order* $\{x_1 > x_2 > x_3 > x_4\}$ and re-written with the linear terms on one side and the nonlinear terms on the other as (Awange and Grafarend 2005; Awange et al. 2010)

$$\begin{bmatrix} x_1^2 + x_2^2 + x_3^2 - x_4^2 = 2a_0x_1 + 2b_0x_2 + 2c_0x_3 - 2d_0x_4 + d_0^2 - a_0^2 - b_0^2 - c_0^2 \\ x_1^2 + x_2^2 + x_3^2 - x_4^2 = 2a_1x_1 + 2b_1x_2 + 2c_1x_3 - 2d_1x_4 + d_1^2 - a_1^2 - b_1^2 - c_1^2 \\ x_1^2 + x_2^2 + x_3^2 - x_4^2 = 2a_2x_1 + 2b_2x_2 + 2c_2x_3 - 2d_2x_4 + d_2^2 - a_2^2 - b_2^2 - c_2^2 \\ x_1^2 + x_2^2 + x_3^2 - x_4^2 = 2a_3x_1 + 2b_3x_2 + 2c_3x_3 - 2d_3x_4 + d_3^2 - a_3^2 - b_3^2 - c_3^2. \end{bmatrix} \tag{15.80}$$

On subtracting (15.80 iv) from (15.80 i), (15.80 ii), and (15.80 iii), to three equations which are linear with four unknowns are derived (see, e.g., Awange et al. 2010, p. 177). Treating the unknown variable x_4 as a constant, both *Groebner bases* and the *Multipolynomial resultant* techniques are applied to solve the linear system of equation for $x_1 = g(x_4), x_2 = g(x_4), x_3 = g(x_4)$, where $g(x_4)$ is a univariate polynomial (see Awange and Grafarend 2005 or Awange et al 2010, pp. 177–180 for further details).

Example 15.18. (7-parameter datum transformation)

Consider a case where coordinates have been given in two systems, A and B. For clarity purposes, let us assume the two coordinate systems to be e.g., photo image coordinates in system A and ground coordinates in system B. The ground coordinates $\{X_i, Y_i, Z_i | i, \ldots, n\}$ of the objects are obtained from, say, GPS measurements. Given the photo coordinates $\{a_i = x_i, b_i = y_i, c_i = -f | i, \ldots, n\}$ and their equivalent ground coordinates $\{X_i, Y_i, Z_i | i, \ldots, n\}$, the 7-parameter datum transformation problem concerns itself with determining;

(1) The scale parameter $x_1 \in \mathbb{R}$
(2) Three translation parameters $\mathbf{x}_2 \in \mathbb{R}^3$
(3) The rotation matrix $\mathbf{X}_3 \in \mathbb{R}^{3\times 3}$ comprising three rotation elements.

Once the listed unknowns have been determined, coordinates can subsequently be transformed from one system onto another. The nonlinear equations relating these unknowns and coordinates from both systems are given by

$$\begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix} = x_1 \mathbf{X}_3 \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + \mathbf{x}_2 \mid i = 1, 2, 3, \ldots, n, \tag{15.81}$$

subject to

$$\boxed{\mathbf{X}_3' \mathbf{X}_3 = \mathbf{I}_3}. \tag{15.82}$$

In (15.81), $\{a_i, b_i, c_i\}$ and $\{X_i, Y_i, Z_i\}$ are the coordinates of the same points, e.g., in both photo and ground coordinate systems respectively. The determination of the unknowns $x_1 \in \mathbb{R}$, $\mathbf{x}_2 \in \mathbb{R}^3$, $\mathbf{X}_3 \in \mathbb{R}^{3\times 3}$ require a minimum of *three points* in both systems whose coordinates are known. Owing to the nonlinearity of (15.81), the solutions have always been obtained using a least squares approach iteratively. With this approach, (15.81) is first linearized and some initial approximate starting values of the unknown parameters used. The procedure then iterates, each time improving on the solutions of the preceding iteration step. This is done until a convergence criteria is achieved.

Where the rotation angles are small e.g., in photogrammetry, the starting values of zeros are normally used. In other fields such as geodesy, the rotation angles are unfortunately not small enough to be initialized by zeros, thereby making the solution process somewhat difficult and cumbersome. Bad choices of initial starting values often lead to many iterations being required before the convergence criteria is achieved. In some cases, where the initial starting values are far from those of the unknown parameters, iteration processes may fail to converge. With these uncertainties in the initial starting values, the cumbersomeness of the linearization and iterations, procedures that would offer an exact solution to the 7-parameter datum transformation problem would be desirable. To answer this challenge, we propose algebraic approaches whose advantages over the approximate numerical methods have already been mentioned.

Apart from the computational difficulties associated with numerical procedures, the 7-parameter datum transformation problem poses another challenge to existing algorithms. This is, the incorporation of the variance-covariance (weight) matrices of the two systems involved. In practice, users have been forced to rely on iterative procedures and linearized least squares solutions which are incapable of incorporating the variance-covariance matrices of both systems in play. This challenge is addressed algebraically in *J.L. Awange and E.W. Grafarend* (2005) or *J.L. Awange et al.* (2010, Chap. 17).

15-5 Notes

The current known techniques for solving *nonlinear polynomial equations* can be classified into *symbolic*, *numeric* and *geometric* methods (*D. Manocha* 1994c). Symbolic methods, discussed in this chapter for solving closed form geodetic problems, apply the *Groebner bases* and the *Multipolynomial resultants techniques* to eliminate several variables in a multivariate system of equations in such a manner that the end product often consist of *univariate polynomials* whose roots can be determined by existing programs, however, such as the roots command in MATLAB. The current available programs however are efficient only for sets of low degree polynomial systems consisting of upto three to four polynomials due to the fact that computing the roots of the *univariate polynomials* can be ill conditioned for polynomials of degree greater than 14 or 15 (*D. Manocha* 1994c).

Elaborate literature on *Groebner bases* can be found in the works of *B. Buchberger* (1965, 1970), *J. H. Davenport et al.* (1988, pp. 95–103), *F. Winkler* (1996), *D. Cox et al.* (1997, pp.47-99), *H. M. Mller* (1998), *W. V. Vasconcelos*

(1998), *T. Becker* and *V. Weispfenning* (1993,1998), *B. Sturmfels* (1996), *G. Pistone* and *H. P. Wynn* (1996), *D. A. Cox* (1998) and *D. Cox et al.*(1998, pp. 1–50), while literature on *Multipolynomial resultants procedure* include the works of *G. Salmon* (1876), *F. Macaulay* (1902, 1921), *A. L. Dixon* (1908), *B. L. van Waerden* (1950), *C. Bajaj et al.* (1988), *J. F. Canny* (1988), *J. F. Canny et al*. (1989), *I. M. Gelfand et al.* (1990), *J. Weiss* (1993), *D. Manocha* (1992, 1993, 1994a,b,c, 1998), *D. Manocha* and *J. F. Canny* (1991, 1992, 1993), *I. M. Gelfand et al*. (1994), *G. Lyubeznik* (1995), *S. Krishna and D. Manocha* (1995), *J. Guckenheimer et al.*(1997), *B. Sturmfels* (1994, 1998), *E. Cattani et al.* (1998) and *D. Cox et al.* (1998, pp.71–122). Besides the *Groebner bases* and *resultant techniques*, there exists another approach for variable elimination developed by WU Wen Tsün (*W. T. Wu* 1984) using the ideas proposed by *J. F. Ritt* (1950). This approach is based on Ritts characteristic set construction and was successfully applied to automated geometric theorem by Wu. This algorithm is referred by *X. S. Gao* and *S. C. Chou* (1990) as the *Ritt-Wu's algorithm* (*D. Manocha* and *F. Canny* 1993). *C. L. Cheng* and *J. W. Van Ness* (1999) have presented polynomial measurement error models.

Numeric methods for solving polynomials can be grouped into *iterative* and *homotopy* methods. For *homotopy* we refer to *A. P. Morgan* (1992). Also in this category are geometric methods which have found application in curve and surface intersection whose convergence are however said to be slow (*D. Manocha* 1994c). In general, for low degree curve intersection, the algebraic methods have been found to be the fastest in practice. In Sect. 15-312 and 15-313 we present in a nut shell the theories of *Groebner bases* and *Multipolynomial resultants.*

The problem of *nonlinear adjustment* in Geodesy as in other fields continues to attract more attention from the modern day researchers as evidenced in the works of *R. Mautz* (2001) and *L. Guolin* (2000) who presents a procedure that tests using the F-distribution whether a *nonlinear model* can be linearized or not. The solution to the minimization problem of the *nonlinear Gauss–Markov model* unlike its linear counter part does not have a direct method for solving it and as such, always relies on the iterative procedures such as the Steepest-descent method, Newton's method and the Gauss-Newton's method discussed by *P. Teunissen* (1990). In particular, *P. Teunissen* (1990) recommends the Gauss-Newton's method as it exploits the structure of the objective function (sum of squares) that is to be minimized. *P. Teunissen* and *E. H. Knickmeyer* (1988) considers in a statistical way how the nonlinearity of a function manifests itself during the various stages of adjustment. *E. Grafarend* and *B. Schaffrin* (1989, 1991) while extending the work of *T. Krarup* (1982) on *nonlinear adjustment* with respect to geometric interpretation have presented the *necessary* and *sufficient* conditions for least squares adjustment of *nonlinear Gauss–Markov model* and provided the geometrical interpretation of these conditions.

Other geometrical approaches include the works of *G. Blaha* and *R. P. Besette* (1989) and *K. Borre* and *S. Lauritzen* (1983) while non geometrically treatment of nonlinear problems have been presented by *K. K. Kubik* (1967), *T. Saito* (1973), *H. J. Schek* and *P. Maier* (1976), *A. Pope* (1982) and *H. G. Bähr* (1988). A comprehensive review to the iterative procedures for solving the *nonlinear* *equations* is presented in the work of *P. Lohse* (1994). *M. Gullikson* and *I. Sderkvist* (1995) have developed algorithms for fitting surfaces which have been explicitly or implicitly defined to some measured points with negative weights being acceptable by the algorithm.

Our approach in the present chapter goes back to the work of *C. F. Gauss* (Published posthumously e.g. *Awange et al. 2010*) and *C. G. I. Jacobi* (1841). Within the framework of arithmetic mean, *C. F. Gauss* and *C. G. I. Jacobi* (1841) suggest that given n linear(ized) observation equations with m unknowns ($n > m$), σ combinations, each consisting of m equations be solved for the unknown elements and the weighted arithmetic mean be applied to get the final solution. Whereas *C.F. Gauss* proposes weighting by using the products of the square of the measured distances from the unknown point to known points and the distances of the side of the error triangles, *C. G. I. Jacobi* (1841) proposed the use of the square of the determinants as weights. In tracing the method of least squares to the arithmetic mean, *A. T. Hornoch* (1950) shows that the weighted arithmetic mean proposed by *C. G. I. Jacobi* (1841) leads to the least squares solution only if the weights used are the actual weights and the pseudo-observations formed by the combinatorial pairs are uncorrelated. Using a numerical example, *S. Wellisch* (1910, pp. 41–49) has shown the results of least squares solution to be identical to those of the *Gauss–Jacobi combinatorial algorithm* once proper weighting is applied.

P. Werkmeister (1920) illustrated that for planar resection case with three directional observations from the unknown point to three known stations, the area of the triangle (error figure) formed by the resulting three combinatorial coordinates of the new point is proportional to the determinant of the dispersion matrix of the coordinates of the new station. In this chapter, these concepts of the combinatorial *linear adjustment* were extended to *nonlinear adjustment*. Awange and Grafarend (2005) illustrated using a leveling network and planar ranging problem that the results of *Gauss–Jacobi combinatorial algorithm* are identical to those of *linear Gauss–Markov model* if the actual variance-covariance matrices are used.

To test the algebraic computational tools of *Groebner bases* and *Multipolynomial resultants*, geodetic problems of *Minimum Distance Mapping* and *GNSS pseudo-ranging four-point P4P*, and *7-parameter transformation* were solved. The procedures are, however, not limited only to these three examples but have been shown, e.g., in *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010) to be applicable to many problems in geodesy and geoinformatics. An example is the classical problem of resection. In general, the search towards the solution of the three-dimensional resection problem traces its origin to the work of a German mathematician *J. A. Grunert* (1841) whose publication appeared in the year 1841. *J. A. Grunert* (1841) solved the three-dimensional resection problem – what was then known as the *"Pothenot's"* problem – in a closed form by solving an algebraic equation of degree four. The problem had hitherto been solved by iterative means mainly in Photogrammetry and Computer Vision. Procedures developed later for solving the three-dimensional resection problem revolved around the improvements of the approach of *J. A. Grunert* (1841) with the aim of searching for the optimal means of distances determination. Whereas *J. A. Grunert* (1841) solves the problem by substitution approach in *three steps,* the more recent desire has been to solve the distance equations in lesser steps as exemplified in the works of *S. Finsterwalder* and *W. Scheufele* (1937), *E. L. Merritt* (1949), *M. A. Fischler* and *R. C. Bolles* (1981), *S. Linnainmaa et al.* (1988) and *E. Grafarend, P. Lohse* and *B. Schaffrin* (1989). Other research done on the subject of resection include the works of *F. J. Mller* (1925), *E. Grafarend* and *J. Kunz* (1965), *R. Horaud et al.* (1989), *P. Lohse* (1990), and *E. Grafarend* and *J. Shan* (1997a, 1997b). An extensive review of some of the procedures above are presented by *F. J. Müller* (1925) and *R. M. Haralick et al.* (1991, 1994).

R. M. Haralick et al. (1994) reviewed the performance of six direct solutions (*J. A. Grunert* 1841, *S. Finsterwalder* and *W. Scheufele* 1937, *E. L. Merritt* 1949, *M. A. Fischler* and *R. C. Bolles* 1981, *Linnainmaa et al.* 1988, and *E. Grafarend, P. Lohse* and *B. Schaffrin* 1989) with the aim of evaluating the numerical stability of the solutions and the effect of permutation within the various solutions. All the six approaches follow the outline adopted by *J. A. Grunert* (1841) with the major differences being the change in variables, the reduction of these variables, and the combination of different pairs of equations. The study revealed that the higher the order of the polynomials, the more complex the computations became and thus the less accurate the solutions were due numerical instabilities. Consequently, S. Finsterwalder's (SF) procedure which solves a third order polynomial is ranked first, J. A Grunert (JG), Fischler and Bolles (FB), and Grafarend et al. (GLS) solutions are ranked second, Linnainmaa et al. solution which generates an eighth order polynomial is ranked third. Though it does not solve the eight order polynomial, the complexity of the polynomial is still found to be higher than those of the other procedures. An amazing result is that of Merritt's procedure which is ranked last despite the fact that it is a fourth order polynomial and is similar to Grunert's approach except for the pairs of equations used. *R. M. Haralick et al.* (1994) attributes the poor performance of Merritt's procedure to the conversion procedure adopted by *E. L Merritt* (1949) in reducing the equations from forth to third order. For planar resection problem, solutions have been proposed e.g. by *D. Werner* (1913), *G. Brandsttter* (1974) and *J. van Mierlo* (1988).

Overdetermined planar resection problem has been treated graphically by *E. Hammer* (1896), *C. Runge* (1900), *P. Werkmeister* (1916) and *P. Werkmeister* (1920). *E. Gotthardt* (1940,1974) dealt with the overdetermined two-dimensional resection where more than four points were considered with the aim of studying the critical configuration that would yield a solution. The work was later to be extended by *K. Killian* (1990). A special case of an overdetermined two-dimensional resection has also been considered by *H. G. Bähr* (1991) who uses six known stations and proposes the measuring of three horizontal angles which are related to the two unknown coordinates by nonlinear equations. By adopting approximate coordinates of the unknown point, an iterative adjustment procedure is performed to get the improved two-dimensional coordinates of the unknown point. It should be noted that the procedure is based on the coordinate system of the six known stations. *K. Rinner* (1962) has also contributed to the problem of overdetermined two-dimensional resection.

In order to relate a point on the Earth's topographical surface uniquely (one-to-one) to a point on the *International Reference Ellipsoid, E. Grafarend* and *P. Lohse* (1991) have proposed the use of the *Minimum Distance Mapping (i.e., Example 15.16)*. Other procedures that have been proposed are either iterative, approximate "closed", closed or higher order equations. Iterative procedures include the works of *N. Bartelme* and *P. Meissl.* (1975), *W. Benning* (1987), *K. M. Borkowski* (1987, 1989), *N. Croceto* (1993), *A. Fitzgibbon et al.* (1999), *T. Fukushima* (1999), *W. Gander et al.* (1994), *B. Heck* (1987), *W. Heiskannen* and *H. Moritz* (1976), *R. Hirvonen* and *H. Moritz* (1963), *P. Loskowski* (1991), *K. C. Lin* and *J. Wang* (1995), *M. K. Paul* (1973), *L. E. Sjoeberg* (1999), *T. Soler* and *L. D. Hothem* (1989), *W. Torge* (1976), *T. Vincenty* (1978) and *R. J. You* (2000).

Approximate "closed" procedures include *B. R. Bowring* (1976, 1985), *A. Fotiou* (1998), *B. Hofman-Wellenhof et al.* (1992), *M. Pick* (1985) and *T. Vincenty* (1980), while closed procedures include *W. Benning* (1987), *H. Frhlich* and *H. H. Hansen* (1976), *E. Grafarend et al.* (1995) and *M. Heikkinen* (1982). Procedures that required the solution of higher order equations include *M. Lapaine* (1990), *M. J. Ozone* (1985), *P. Penev* (1978), *H. Snkel* (1976), and *P. Vaniceck and E. Krakiwski* (1982). In this chapter, the *Minimum Distance Mapping problem* is solved using the *Groebner bases* approach. A *univariate polynomial* of fourth order is obtained together with 13 other elements of the *Groebner basis*. The obtained *univariate polynomial* and the linear terms are compared to those of *E. Grafarend* and *P. Lohse* (1991). Other reference include *E. Grafarend* and *W. Keller* (1995) and *Mukherjee, K.* (1996)

The *GNSS four-point pseudo-ranging problem* presented in Example 15.17 is concerned with the determination of the coordinates of a stationary receiver together with its range bias. Several closed form procedures have been put forward for obtaining closed form solution of the problem. Amongst the procedures include the vectorial approach as evidenced in the works of *L. O. Krause* (1987), *J. S. Abel* and *J. W. Chaffee* (1991), *P. Singer et al.* (1993), *J. W. Chaffee* and *J. S. Abel* (1994), *H. Lichtenegger* (1995) and *A. Kleusberg* (1994, 1999). *E. Grafarend* and *J. Shan* (1996) propose two approaches; one approach is based on a closed form solution of the nonlinear pseudo-ranging equations in geocentric coordinates while the other approach solves the same equations in barycentric coordinates.

S. C. Han et al. (2001) have developed an algorithm for very accurate absolute positioning through Global Positioning System (GPS) satellite clock estimation while *S. Bancroft* (1985) provides an algebraic closed form solution of the overdetermined GPS pseudo-ranging observations. In this chapter we solved using *Groebner basis* and *Multipolynomial resultants GPS four-point pseudo-ranging problem*. For literature on three-dimensional positioning models, we refer to *E. W. Grafarend* (1975), *V. Ashkenazi* and *S, Grist* (1982), *G. W. Hein* (1982a, 1982b), *J. Zaiser* (1986), *F. W. O. Aduol* (1989), *U. Klein* (1997) and *S. M. Musyoka* 1999).

In *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010), two case studies; the conversion of the geocentric GPS points for the Baltic Sea level Project into the Gauss–Jacobi ellipsoidal coordinates and the determination of the seven datum transformation parameters are presented. Datum transformation models have been dealt with e.g. in *E. Grafarend, F. Okeke* (1998), *F. Krumm* and *F. Okeke* (1995), *E. Grafarend* and *F. Okeke* (1998) and *E. Grafarend* and *R. Syffus* (1998).

In summary, this chapter has demonstrated the power of the algebraic computational tools of *Groebner bases* and the *Multipolynomial resultants* in solving selected geodetic problems. In the case of the *minimal combinatorial set*, we have demonstrated how the problems of *minimum distance mapping* (Example 15.16) and the *pseudo-ranging four-point problem* (Example 15.17) could be solved in a closed form using either *Groebner bases* approach or the *Multipolynomial resultants* approach. We have succeeded in demonstrating that by converting the *nonlinear observation equations* of the selected geodetic problems above into algebraic (polynomials), the *multivariate* system of polynomial equations relating the unknown variables (indeterminate) to the known variables can be reduced into polynomial equations consisting of *a univariate polynomial* once *lexicographic monomial ordering* is specified. We have therefore managed to provide *symbolic solutions* to the problems of 7-parameter transformation, *minimum distance mapping* and the *GPS pseudo-ranging four-point P4P* by obtaining in each case *a univariate polynomial* that can readily be solved numerically once the observations are available. Although the algebraic techniques of *Groebner bases* and *Multipolynomial resultants* were applied to these selected geodetic problems, the tools can be used to solve explicitly any closed form problem in Geodesy. The only limitation may be the storage and computational speed of the computers when compounded with problems involving many variables. The ability of the algebraic tools (*Groebner bases* and the *Multipolynomial resultants*) to solve closed form solutions gave the *Gauss–Jacobi combinatorial algorithm* the required operational engine as evidenced in the solution of the *overdetermined GPS pseudo-ranging problem* in *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010) where the solutions are obtained without reverting to iterative or linearization procedures. The results of *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010) compared well with the solutions obtained using the *linearized Gauss–Markov model* giving legitimacy to the *Gauss–Jacobi combinatorial* procedure.

For the case studies (see e.g., *J.L. Awange* and *E.W. Grafarend* (2005) and *J. L. Awange et al.* (2010)), the *Groebner bases algorithm* successfully determines explicitly the 7-parameters of the datum transformation problem and shows the scale parameter to be represented by a *univariate polynomial* of *fourth order* while the rotation parameters are represented by linear functions comprising the coordinates of the two systems and the scale parameter. The admissible value of the scale is readily obtained by solving the *univariate polynomial* and restricting the scale to a positive real value $x_1 \in \mathbb{R}.^+$ This eliminates the negative components of the roots and the complex values. The admissible value $x_1 \in \mathbb{R}^+$ of the scale parameter is chosen and substituted in the linear functions that characterize the three elements of the *skew-symmetric matrix* $\boldsymbol{S}$ leading to the solution of the elements of the rotation matrix $\boldsymbol{X}_3$. The translation elements are then deduced from the transformation equation. The advantage of using the *Groebner bases algorithm* is the fact that there exists no requirement for prior knowledge of the approximate values of the 7-transformation parameters as is usually the case.

The *Groebner bases algorithm* managed to solve the *Minimum Distance Mapping problem* and in so doing, enabling the mapping of points from the *topographical surface* to the *International Reference Ellipsoid of Revolution*. The *univariate polynomial* obtained was identical to that obtained by Grafarend and Lohse (1991). This implies that the algebraic tools of *Groebner bases* and the *Multipolynomial resultants* can also be used to check the validity of existing closed form procedures in Geodesy.

The *Gauss–Jacobi combinatorial algorithm highlighted* one important fact while solving the overdetermined 7 parameter transformation problem; that the stochasticity of both systems involved can be taken into account. This has been the bottleneck of the problem of *7-datum transformation parameters*.

Appendix A
Tensor Algebra, Linear Algebra, Matrix Algebra, Multilinear Algebra

> The key word "*algebra*" is derived from the name and the work of the nineth-century scientist Mohammed ibn *Mûsâ al-Khowârizmi* who was born in what is now Uzbekistan and worked in Bagdad at the court of *Harun al-Rashid's son*. "*al-jabr*" appears in the title of his book *Kitab al-jabr wail muqabala* where he discusses symbolic methods for the solution of equations, K. Vogel: Mohammed Ibn Musa Alchwarizmi's Algorismus; Das forested Lehrbuch zum Rechnen mit indischen Ziffern, 1461, Otto Zeller Verlagsbuchhandlung, Aalen 1963). Accordingly what is an *algebra* and how is it tied to the notion of a *vector space*, a *tensor space*, respectively? By an *algebra* we mean *a set* $\mathbb{S}$ *of elements* and *a finite set* $\boldsymbol{M}$ *of operations*. Each *operation* $(\text{opera})_k$ is a single-valued function assigning to every finite ordered sequence $(x_1, \ldots, x_n)$ of $n = n(k)$ elements of $\mathbb{S}$ a *value* $(\text{opera})_k\ (x_1, \ldots, x_k) = x_l$ *in* $\mathbb{S}$. In particular for $(\text{opera})_k(x_1, x_2)$ the operation is called *binary*, for $(\text{opera})_k(x_1, x_2, x_3)$ ternary, in general for $(\text{opera})_k(x_1, \ldots, x_n)$ n-ary. For a given set of operation symbols $(\text{opera})_1$, $(\text{opera})_2, \ldots$, $(\text{opera})_k$ we define a *word*. In *linear algebra* the set $\boldsymbol{M}$ has basically two elements, namely two internal relations $(\text{opera})_1$ *worded* "addition" (including inverse addition: subtraction) and $(\text{opera})_2$ *worded* "multiplication" (including inverse multiplication: division). *Here* the elements of the set $\mathbb{S}$ are *vectors* over the field $\mathbb{R}$ of real numbers as long as we refer to *linear algebra*. In contrast, in *multilinear algebra* the elements of the set $\mathbb{S}$ are *tensors* over the field of *real numbers* $\mathbb{R}$. Only later *modules* as generalizations of vectors of linear algebra are introduced in which the "scalars" are allowed to be from *an arbitrary ring* rather than the field $\mathbb{R}$ of *real numbers*.

Let us assume that you as a potential reader are in some way familiar with the elementary notion of a threedimensional vector space $\mathbb{X}$ with elements called vectors $\mathbf{x} \in \mathbb{R}^3$, namely the intuitive space " we locally live in". Such an elementary vector space $\mathbb{X}$ is equipped with a metric to be referred to as *threedimensional Euclidean*.

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2,
© Springer-Verlag Berlin Heidelberg 2012

As a real threedimensional vector space we are going to give it a *linear and multilinear algebraic structure*. In the context of structure mathematics based upon

(a) `Order structure`
(b) `Topological structure`
(c) `Algebraic structure`

an algebra is constituted if at least two relations are established, namely one *internal* and one *external*. We start with *multilinear algebra*, in particular with the multilinearity of the tensor product before we go back to *linear algebra*, in particular to *Clifford algebra*.

A-1 Multilinear Functions and the Tensor Space $\mathbb{T}_q^p$

Let $\mathbb{X}$ be a finite dimensional linear space, e.g. a vector space over the field $\mathbb{R}$ of *real numbers*, in addition denote by $\mathbb{X}^*$ its dual space such that $n = \dim \mathbb{X} = \dim \mathbb{X}^*$. Complex, quaternion and octonian numbers $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ as well as *rings* will only be introduced later in the context. For $p, q \in \mathbb{Z}^+$ being an element of positive integer numbers we introduce

$$\mathbb{T}_q^p(\mathbb{X}, \mathbb{X}^*)$$

as the p-contravariant, q-covariant *tensor space* or *space of multilinear functions*

$$f : \mathbb{X}^* \times \ldots \times \mathbb{X}^* \times \mathbb{X} \times \ldots \times \mathbb{X} \longrightarrow \mathbb{R}^{p \dim \boldsymbol{X}^* \times q \dim \boldsymbol{X}}.$$

If we assume $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^p \in \mathbb{X}^*$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q \in \mathbb{X}$, then

$$\boxed{\boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q \in \mathbb{T}_q^p(\mathbb{X}^*, \mathbb{X})}$$

holds. *Multilinearity is understood as linearity in each variable*. "$\otimes$" identifies the tensor product, *the Cartesian product* of elements $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^p, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_q)$.

Example A.1: Bilinearity of the tensor product $\boldsymbol{x}_1 \otimes \boldsymbol{x}_2$
For every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{X}$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$ and $r \in \mathbb{R}$ bilinearity implies

$$\begin{aligned}
(\boldsymbol{x} + \boldsymbol{y}) \otimes \boldsymbol{x}_2 &= \boldsymbol{x} \otimes \boldsymbol{x}_2 + \boldsymbol{y} \otimes \boldsymbol{x}_2 && (\textit{internal left-linearity})\\
\boldsymbol{x}_1 \otimes (\boldsymbol{x} + \boldsymbol{y}) &= \boldsymbol{x}_1 \otimes \boldsymbol{x} + \boldsymbol{x}_1 \otimes \boldsymbol{y} && (\textit{internal right-linearity})\\
r\boldsymbol{x} \otimes \boldsymbol{x}_2 &= r(\boldsymbol{x} \otimes \boldsymbol{x}_2) && (\textit{external left-linearity})\\
\boldsymbol{x}_1 \otimes r\boldsymbol{y} &= r(\boldsymbol{x}_1 \otimes \boldsymbol{y}) && (\textit{external right-linearity}).
\end{aligned}$$

♣

The generalization of bilinearity of $\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \in \mathbb{T}_2^0$ to *multilinearity* of

$$\boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q \in \mathbb{T}_q^p$$

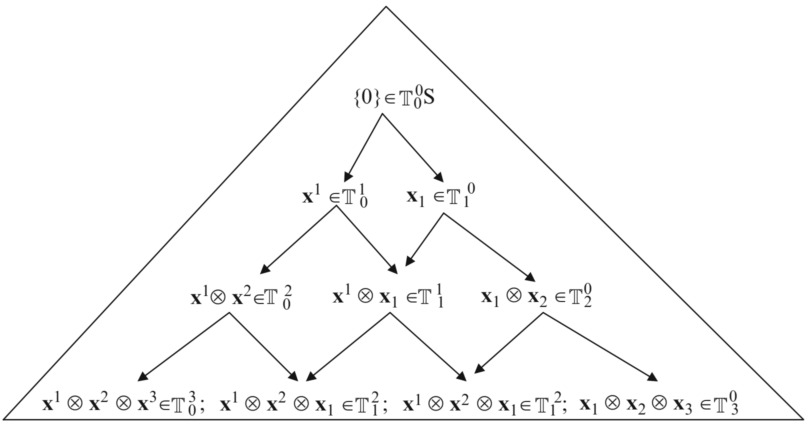
is obvious.

Definition A.1. (multilinearity of tensor space $\mathbb{T}_q^p$):

For every $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^p \in \mathbb{X}^*$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_q \in \mathbb{X}$ as well as $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{X}^*$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$ and $r \in \mathbb{R}$ multilinearity implies

$$\begin{aligned}
&(\boldsymbol{u} + \boldsymbol{v}) \otimes \boldsymbol{x}^2 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q \\
&\quad = \boldsymbol{u} \otimes \boldsymbol{x}^2 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q + \boldsymbol{v} \otimes \boldsymbol{x}^2 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \\
&\qquad \otimes \ldots \otimes \boldsymbol{x}_q \quad (\textit{internal left-linearity}) \\
&\boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes (\boldsymbol{x} + \boldsymbol{y}) \otimes \boldsymbol{x}_2 \otimes \ldots \boldsymbol{x}_q \\
&\quad = \boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x} \otimes \boldsymbol{x}_2 \otimes \ldots \otimes \boldsymbol{x}_q + \boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{y} \otimes \boldsymbol{x}_2 \\
&\qquad \otimes \ldots \otimes \boldsymbol{x}_q \quad (\textit{internal right-linearity}) \\
&r\boldsymbol{u} \otimes \boldsymbol{x}^2 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q \\
&\quad = r(u \otimes \boldsymbol{x}^2 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x}_1 \otimes \ldots \otimes \boldsymbol{x}_q) \quad (\textit{external left-linearity}) \\
&\boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes r\boldsymbol{x} \otimes \boldsymbol{x}_2 \otimes \ldots \otimes \boldsymbol{x}_q \\
&\quad = r(\boldsymbol{x}^1 \otimes \ldots \otimes \boldsymbol{x}^p \otimes \boldsymbol{x} \otimes \boldsymbol{x}_2 \otimes \ldots \otimes \boldsymbol{x}_q) \quad (\textit{external right-linearity}).
\end{aligned}$$

A possible way to visualize the different multilinear functions which span $\{\mathbb{T}_0^0, \mathbb{T}_0^1, \mathbb{T}_1^0, \mathbb{T}_0^2, \mathbb{T}_1^1, \mathbb{T}_2^0, \ldots, \mathbb{T}_q^p\}$ is to construct a *hierarchical diagram* or a special tree as follows.



When we learnt first about the tensor product symbolised by "$\otimes$" as well as its multilinearity we were left with the problem of developing an intuitive understanding of $\mathbf{x}^1 \otimes \mathbf{x}^2$, $\mathbf{x}^1 \otimes \mathbf{x}_1$ and higher order tensor products. Perhaps it is helpful to represent "the involved vectors" in a contravariant or in a covariant basis. For instances, $\mathbf{x}^1 = \mathbf{e}^1 x_1 + \mathbf{e}^2 x_2 + \mathbf{e}^3 x_3$ or $\mathbf{x}_1 = \mathbf{e}_1 x^1 + \mathbf{e}_2 x^2 + \mathbf{e}_3 x^3$ is a *left* representation of a three-dimensional vector in a *3-left basis* $\{\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3\}_l$ of contravariant type *or* in a *3-left basis* $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}_l$ of covariant type. Think in terms of $\mathbf{x}^1$ or $\mathbf{x}_1$ as a three-dimensional position vector with right coordinates $\{x_1, x_2, x_3\}$ or $\{x^1, x^2, x^3\}$, respectively. Since the intuitive algebras of vectors is *commutative* we may also represent the three-dimensional vector in a in a *3-right basis* $\{\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3\}_r$ of contravariant type *or* in a *3-right basis* $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}_r$ of covariant type such that $\mathbf{x}^1 = x_1 \mathbf{e}^1 + x_2 \mathbf{e}^2 + x_3 \mathbf{e}^3$ *or* $\mathbf{x}_1 = x^1 \mathbf{e}_1 + x^2 \mathbf{e}_2 + x^3 \mathbf{e}_3$ is a *right* representation of a three-dimensional position vector which coincides with its *left* representation thanks to *commutativity*. Further on, the tensor product $\mathbf{x} \otimes \mathbf{y}$ enjoys the left and right representations

$$\left(\mathbf{e}^1 x_1 + \mathbf{e}^2 x_2 + \mathbf{e}^3 x_3\right) \otimes \left(\mathbf{e}^1 y_1 + \mathbf{e}^2 y_2 + \mathbf{e}^3 y_3\right) = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{e}^i \otimes \mathbf{e}^j x_i y_j$$

and

$$\left(x_1 \mathbf{e}^1 + x_2 \mathbf{e}^2 + x_3 \mathbf{e}^3\right) \otimes \left(y_1 \mathbf{e}^1 + y_2 \mathbf{e}^2 + y_3 \mathbf{e}^3\right) = \sum_{i=1}^{3} \sum_{j=1}^{3} x_i y_j \mathbf{e}^i \otimes \mathbf{e}^j$$

which coincides again since we assumed a *commutative algebra* of vectors. The product of coordinates $(x_i y_j), i, j \in \{1, 2, 3\}$ is often called the *dyadic product*. Please do not miss the alternative covariant representation of the tensor product $\mathbf{x} \otimes \mathbf{y}$ which we introduced so far in the contravariant basis, namely

$$\left(\mathbf{e}_1 x^1 + \mathbf{e}_2 x^2 + \mathbf{e}_3 x^3\right) \otimes \left(\mathbf{e}_1 y^1 + \mathbf{e}_2 y^2 + \mathbf{e}_3 y^3\right) = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{e}_i \otimes \mathbf{e}_j x^i y^j$$

of left type and

$$\left(x^1 \mathbf{e}_1 + x^2 \mathbf{e}_2 + x^3 \mathbf{e}_3\right) \otimes \left(y^1 \mathbf{e}_1 + y^2 \mathbf{e}_2 + y^3 \mathbf{e}_3\right) = \sum_{i=1}^{3} \sum_{j=1}^{3} x^i y^j \mathbf{e}_i \otimes \mathbf{e}_j$$

of right type. In a similar way we produce

$$\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} = \sum_{i,j,k=1}^{3,3,3} \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}^k x_i y_j z_k = \sum_{i,j,k=1}^{3,3,3} x_i y_j z_k \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}^k$$

of contravariant type and

$$\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} = \sum_{i,j,k=1}^{3,3,3} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k x^i y^j z^k = \sum_{i,j,k=1}^{3,3,3} x^i y^j z^k \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k$$

of covariant type. "*Mixed covariant-contravariant*" representations of the tensor product $\mathbf{x}_1 \otimes \mathbf{y}^1$ are

$$\mathbf{x}_1 \otimes \mathbf{y}^1 = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{e}_i \otimes \mathbf{e}^j \mathbf{x}^i \mathbf{y}_j = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{x}^i \mathbf{y}_j \mathbf{e}_i \otimes \mathbf{e}^j$$

or

$$\mathbf{x}^1 \otimes \mathbf{y}_1 = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{e}^i \otimes \mathbf{e}_j \mathbf{x}_i \mathbf{y}^j = \sum_{i=1}^{3} \sum_{j=1}^{3} \mathbf{x}_i \mathbf{y}^j \mathbf{e}^i \otimes \mathbf{e}_j .$$

In addition, we have to explain the notion $\mathbf{x}^1 \in \mathbb{X}^*, \mathbf{x}_1 \in \mathbb{X}$: While the vector $\mathbf{x}_1$ is an element of the vector space $\mathbb{X}$, $\mathbf{x}^1$ is an element of its dual space. What is a dual space? Indeed the dual space $\mathbb{X}^*$ is the space of *linear functions* over the elements of $\mathbb{X}$. For instance, if the vector space $\mathbb{X}$ is equipped with *inner product*, namely $\langle \mathbf{e}_i | \mathbf{e}_j \rangle = g_{ij},\ i, j \in \{1, 2, 3\}$, with respect to the base vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ which span the vector space $\mathbb{X}$, then

$$\mathbf{e}^j = \sum_{i=1}^{3} \mathbf{e}_i g^{ij}$$

transforms the covariant base vectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ into the contravariant base vectors $\{\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3\} = \operatorname{span} \mathbb{X}^*$ by means of $[g^{ij}] = G^{-1}$, the inverse of the matrix $[g_{ij}] = G \in \mathbb{R}^{3\times 3}$. Similarly the coordinates g_{ij} of the metric tensor g are used for "*raising*" or "*lowering*" the indices of the coordinates x^i, x_j , respectively, for instance

$$x^i = \sum_{j=1}^{3} g^{ij} x_j , x_i = \sum_{j=1}^{3} g_{ij} x^j .$$

In a finite dimensional vector space, the power of a linear space $\mathbb{X}$ and its dual $\mathbb{X}^\star$ does not show up. In contrast, in an infinite dimensional vector space $\mathbb{X}$ the dual space $\mathbb{X}^\star$ is the space of *linear functionals* which play an important role in *functional* analysis. While through the tensor product "$\otimes$" which operated on vectors, e.g. $\mathbf{x} \otimes \mathbf{y}$, we constructed the p-contravariant, q-covariant tensor space or space of multilinear functions $\mathbb{T}^p_q(\mathbb{X}, \mathbb{X}^*)$, e.g. $\mathbb{T}^2_0, \mathbb{T}^0_2, \mathbb{T}^1_1$, we shall generalise the representation of the elements of $\mathbb{T}^p_q$ by means of

$$f = \sum_{i_1,\ldots,i_p=1}^{n=\dim \mathbb{X}^*} \mathbf{e}^{i_1} \otimes \ldots \otimes \mathbf{e}^{i_p} f_{i_1\ldots i_p} \in \mathbb{T}^p_0$$

$$f = \sum_{i_1,\ldots,i_q=1}^{n=\dim \mathbb{X}^*} \mathbf{e}_{i_1} \otimes \ldots \otimes \mathbf{e}_{i_q} f^{i_1 \ldots i_q} \in \mathbb{T}_q^0$$

$$f = \sum_{i_1,\ldots,i_p=1}^{n=\dim \mathbb{X}^*} \sum_{j_1,\ldots,j_q=1}^{n=\dim \mathbb{X}} \mathbf{e}^{i_1} \otimes \ldots \otimes \mathbf{e}^{i_p} \otimes \mathbf{e}_{j_1} \otimes \ldots \otimes \mathbf{e}_{j_q} f_{j_1,\ldots,j_q}^{i_1,\ldots,i_p} \in \mathbb{T}_q^p$$

for instance

$$f = \sum_{i,j=1}^{3} \mathbf{e}^i \otimes \mathbf{e}^j f_{ij} = \sum_{i,j=1}^{3} f_{ij} \mathbf{e}^i \otimes \mathbf{e}^j \in \mathbb{T}_0^2$$

$$f = \sum_{i,j=1}^{3} \mathbf{e}_i \otimes \mathbf{e}_j f^{ij} = \sum_{i,j=1}^{3} f^{ij} \mathbf{e}_i \otimes \mathbf{e}_j \in \mathbb{T}_2^0$$

$$f = \sum_{i,j=1}^{3} \mathbf{e}^i \otimes \mathbf{e}_j f_j^i = \sum_{i,j=1}^{3} f_j^i \mathbf{e}^i \otimes \mathbf{e}_j \in \mathbb{T}_1^1$$

We have to emphasize that the tensor coordinates $f_{i_1,\ldots,i_p}$, $f^{i_1,\ldots,i_q}$, $f_{j_1,\ldots,j_p}^{i_1,\ldots,i_p}$ are no longer of dyadic or product type. For instance, for

$$\boxed{\text{(2,0)-tensor : bilinear functions:}}$$

$$f = \sum_{i,j=1}^{n} \mathbf{e}^i \otimes \mathbf{e}^j f_{ij} = \sum_{i,j=1}^{n} f_{ij} \mathbf{e}^i \otimes \mathbf{e}^j \in \mathbb{T}_0^2$$

$$f_{ij} \neq f_i f_j$$

$$\boxed{\text{(2,1)-tensor : trilinear functions:}}$$

$$f = \sum_{i,j,k=1}^{n} \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}_k f_{ij}^k = \sum_{i,j=1}^{n} f_{ij}^k \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}_k \in \mathbb{T}_1^2$$

$$f_{ij}^k \neq f_i f_j f^k$$

$$\boxed{\text{(3,1)-tensor : ternary functions:}}$$

$$f = \sum_{i,j,k,l=1}^{n} \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}^k \otimes \mathbf{e}_l f_{ijk}^l = \sum_{i,j=1}^{n} f_{ijk}^l \mathbf{e}^i \otimes \mathbf{e}^j \otimes \mathbf{e}^k \otimes \mathbf{e}_l \in \mathbb{T}_1^3$$

$$f_{ijk}^l \neq f_i f_j f_k f^l$$

holds. Table A.1 is a list of (p, q)-tensors as they appear in various sciences. Of special importance is the decomposition of multilinear functions as elements of the space $\mathbb{T}_q^p$ into their symmetric, antisymetric and residual constituents we are going to outline.

Table A.1 Various examples of tensor spaces $\mathbb{T}_q^p$ ($p \,\&\, q$: rank of tensor) (2,0) tensor, tensor space $\mathbb{T}_0^2$

Metric tensor Gauss curvature tensor Ricci curvature tensor	Differential geometry
Gravity gradient tensor	Gravitation
Faraday tensor, Maxwell tensor Tensor of dielectric constant Tensor of permeability	Electromagnetic
Strain tensor, stress tensor	Continuum mechanics
Energy momentum tensor	Mechanics Electromagnetism General relativity
Second order multipole tensor	Gravitostatics Magnetostatics Electrostatics
Variance-covariance matrix	Mathematical statistics

Table A.2 Various examples of tensor spaces $\mathbb{T}_q^p$ ($p + q$: rank of tensor) (2,1) tensor, tensor space $\mathbb{T}_1^2$

Cartaan torsion tensor	Differential Deometry
Third order multipole tensor	Gravitostatics Magnetostatics Electrostatics
skewness tensor Third momentum tensor of probability distribution	Mathematical Statistics
Tensor of piezoelectric constant	Coupling of stress and electrostatic field

Table A.3 Various examples of tensor spaces $\mathbb{T}_q^p$ ($p+q$: rank of tensor) (3,1) tensor, (2,2) tensor, tensor space $\mathbb{T}_1^3$, $\mathbb{T}_2^2$

Riemann curvature tensor	Differential Geometry
Fourth order multipole tensor	Gravitostatics Magnetostatics Electrostatics
Hooke tensor	Stress-strain relation Constitutive equation Continuum mechanics Elasticity, viscosticity
Curtosis tensor Fourth moment tensor of a probability distribution	Mathematical Statistics

A-2 Decomposition of Multilinear Functions into Symmetric Multilinear Functions, Antisymmetric Multi-linear Functions and Residual Multilinear Functions $\boldsymbol{T T_q^p = \mathbb{S}_q^p \oplus \mathbf{A}_q^p \oplus \mathbb{R}_q^p}$

$\mathbb{T}_q^p$ as the space of multilinear functions follows the decomposition $\mathbb{T}_q^p = \mathbb{S}_q^p \oplus \mathbf{A}_q^p \oplus \mathbb{R}_q^p$ into the subspace $\mathbb{S}_q^p$ of *symmetric* multilinear functions, the subspace $\mathbf{A}_q^p$ of *antisymmetric* multilinear functions and the subspace $\mathbb{R}_q^p$ of *residual* multilinear functions:

Box A.2i. (Antisymmetry of the symbols $f_{i_1 \ldots i_p}$):

$$f_{ij} = -f_{ji}$$
$$f_{ijk} = -f_{ikj},\ f_{jki} = -f_{jik},\ f_{kij} = -f_{kji}$$
$$f_{ijkl} = -f_{ijlk},\ f_{jkli} = -f_{jkil},\ f_{klij} = -f_{klji},\ f_{kijk} = -f_{likj}$$

Box A.2ii. (Symmetry of the symbols $f_{i_1 \ldots i_p}$):

$$f_{ij} = f_{ji}$$
$$f_{ijk} = f_{ikj},\ f_{jki} = f_{jik},\ f_{kij} = f_{kji}$$
$$f_{ijkl} = f_{ijlk},\ f_{jkli} = f_{jkil},\ f_{klij} = f_{klji},\ f_{kijk} = f_{likj}$$

Box A.2iii. (The interior product of bases, span $\mathbb{S}^p$, $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 3$):

$$\mathbb{S}^1 : \frac{1}{1!}\boldsymbol{e}^i$$

$$\mathbb{S}^2 : \frac{1}{2!}(\boldsymbol{e}^i \otimes \boldsymbol{e}^j + \boldsymbol{e}^j \otimes \boldsymbol{e}^i)$$
$$=: \boldsymbol{e}^i \vee \boldsymbol{e}^j\,, \quad \boldsymbol{e}^i \vee \boldsymbol{e}^j = +\boldsymbol{e}^j \vee \boldsymbol{e}^i$$

$$\mathbb{S}^3 : \frac{1}{3!}(\boldsymbol{e}^i \otimes \boldsymbol{e}^j \otimes \boldsymbol{e}^k + \boldsymbol{e}^i \otimes \boldsymbol{e}^k \otimes \boldsymbol{e}^j + \boldsymbol{e}^j \otimes \boldsymbol{e}^k \otimes \boldsymbol{e}^i$$
$$+ \boldsymbol{e}^j \otimes \boldsymbol{e}^i \otimes \boldsymbol{e}^k + \boldsymbol{e}^k \otimes \boldsymbol{e}^i \otimes \boldsymbol{e}^j + \boldsymbol{e}^k \otimes \boldsymbol{e}^j \otimes \boldsymbol{e}^i)$$
$$=: \boldsymbol{e}^i \vee \boldsymbol{e}^j \vee \boldsymbol{e}^k$$

$$\boldsymbol{e}^i \vee \boldsymbol{e}^j \vee \boldsymbol{e}^k = \boldsymbol{e}^i \vee \boldsymbol{e}^k \vee \boldsymbol{e}^j = \boldsymbol{e}^k \vee \boldsymbol{e}^i \vee \boldsymbol{e}^j = \boldsymbol{e}^k \vee \boldsymbol{e}^j \vee \boldsymbol{e}^i = \boldsymbol{e}^j \vee \boldsymbol{e}^k \vee \boldsymbol{e}^i$$
$$= \boldsymbol{e}^j \vee \boldsymbol{e}^i \vee \boldsymbol{e}^k$$

Box A.2iv. (The exterior product of bases, span $\mathbf{A}^p$, $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 3$):

$$\mathbf{A}^1 : \frac{1}{1!}\boldsymbol{e}^i$$

$$\mathbf{A}^2 : \frac{1}{2!}(\boldsymbol{e}^i \otimes \boldsymbol{e}^j - \boldsymbol{e}^j \otimes \boldsymbol{e}^i)$$
$$=: \boldsymbol{e}^i \wedge \boldsymbol{e}^j\,, \quad \boldsymbol{e}^i \wedge \boldsymbol{e}^j = -\boldsymbol{e}^j \wedge \boldsymbol{e}^i$$

$$\mathbf{A}^3 : \frac{1}{3!}(\boldsymbol{e}^i \otimes \boldsymbol{e}^j \otimes \boldsymbol{e}^k - \boldsymbol{e}^i \otimes \boldsymbol{e}^k \otimes \boldsymbol{e}^j + \boldsymbol{e}^j \otimes \boldsymbol{e}^k \otimes \boldsymbol{e}^i$$
$$- \boldsymbol{e}^j \otimes \boldsymbol{e}^i \otimes \boldsymbol{e}^k + \boldsymbol{e}^k \otimes \boldsymbol{e}^i \otimes \boldsymbol{e}^j - \boldsymbol{e}^k \otimes \boldsymbol{e}^j \otimes \boldsymbol{e}^i)$$
$$=: \boldsymbol{e}^i \wedge \boldsymbol{e}^j \wedge \boldsymbol{e}^k$$

$$\boldsymbol{e}^i \wedge \boldsymbol{e}^j \wedge \boldsymbol{e}^k = -\boldsymbol{e}^i \wedge \boldsymbol{e}^k \wedge \boldsymbol{e}^j = +\boldsymbol{e}^k \wedge \boldsymbol{e}^i \wedge \boldsymbol{e}^j = -\boldsymbol{e}^k \wedge \boldsymbol{e}^j \wedge \boldsymbol{e}^i$$
$$= \boldsymbol{e}^j \wedge \boldsymbol{e}^k \wedge \boldsymbol{e}^i = -\boldsymbol{e}^j \wedge \boldsymbol{e}^i \wedge \boldsymbol{e}^k$$

Box A.2v. ($\mathbb{S}^p$, symmetric multilinear functions):

$$\mathbb{T}_0^1 \supset \mathbb{S}^1 \ni f = \left\{ \sum_{i=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i f_i \right\}$$

$$\mathbb{T}_0^2 \supset \mathbb{S}^2 \ni f = \left\{ \frac{1}{2!} \sum_{i,j=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \vee \boldsymbol{e}^j f_{ij} \right\}$$

$$= \left\{ \sum_{i \le j}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \vee \boldsymbol{e}^j f_{(ij)} | f_{(ij)} := \frac{1}{2!}(f_{ij} + f_{ji}) \right\}$$

$$\mathbb{T}_0^3 \supset \mathbb{S}^3 \ni f = \left\{ \frac{1}{3!} \sum_{i,j,k=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \vee \boldsymbol{e}^j \vee \boldsymbol{e}^k f_{ijk} \right\} = \left\{ \sum_{i<j<k}^{n} \boldsymbol{e}^i \vee \boldsymbol{e}^j \vee \boldsymbol{e}^k f_{(ijk)} | f_{(ijk)} \right.$$

$$\left. := \frac{1}{3!}(f_{ijk} + f_{ikj} + f_{jki} + f_{jik} + f_{kij} + f_{kji}) \right\}$$

$$\mathbb{T}_0^p \supset \mathbb{S}^p \ni f = \left\{ \frac{1}{p!} \sum_{i_1,i_2,\ldots,i_p=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^{i_1} \vee \boldsymbol{e}^{\boldsymbol{e}_2} \vee \ldots \vee \boldsymbol{e}^{i_p} f_{i_1 i_2 \ldots i_p} \right\}$$

$$= \left\{ \sum_{i_1 \le i_2 \le \ldots \le i_p}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^{i_1} \vee \boldsymbol{e}^{i_2} \vee \ldots \vee \boldsymbol{e}^{i_p} f_{(i_1 i_2 \ldots i_p)} | f_{(i_1 i_2 \ldots i_p)} \right.$$

$$\left. := \frac{1}{p!}(f_{i_1 \ldots i_{p-1} i_p} + f_{i_1 \ldots i_p i_{p-1}} + \ldots + f_{i_p \ldots i_1 i_2} + f_{i_p \ldots i_2 i_1}) \right\}.$$

Corollary: $\dim \mathbb{S}^p = \binom{n+p-1}{p}$, in particular if $n = p$, then $\dim \mathbb{S}^p = \binom{2p-1}{p}$.

Box A.2vi. ($\mathbf{A}^p$, antisymmetric multilinear functions):

$$\mathbb{T}_0^1 \supset \mathbf{A}^1 \ni f = \left\{ \sum_{i=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i f_i \right\}$$

$$\mathbb{T}_0^2 \supset \mathbf{A}^2 \ni f = \left\{ \frac{1}{2!} \sum_{i,j=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \wedge \boldsymbol{e}^j f_{ij} \right\}$$

$$= \left\{ \sum_{i<j}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \wedge \boldsymbol{e}^j f_{[ij]} | f_{[ij]} := \frac{1}{2!}(f_{ij} + f_{ji}) \right\}$$

$$\mathbb{T}_0^3 \supset \mathbf{A}^3 \ni f = \left\{ \frac{1}{3!} \sum_{i,j,k=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^i \wedge \boldsymbol{e}^j \wedge \boldsymbol{e}^k f_{ijk} \right\}$$

$$= \left\{ \sum_{i<j<k}^{n} \boldsymbol{e}^i \wedge \boldsymbol{e}^j \wedge \boldsymbol{e}^k f_{[ijk]} | f_{[ijk]} \right.$$

$$\left. := \frac{1}{3!}(f_{ijk} - f_{ikj} + f_{jki} - f_{jik} + f_{kij} - f_{kji}) \right\}$$

$$\mathbb{T}_0^p \supset \mathbf{A}^p \ni f = \left\{ \frac{1}{p!} \sum_{i_1,i_2,\ldots,i_p=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{\boldsymbol{e}_2} \wedge \ldots \wedge \boldsymbol{e}^{i_p} f_{i_1 i_2 \ldots i_p} \right\}$$

$$= \left\{ \sum_{i_1<i_2<\ldots<i_p}^{n=\dim \boldsymbol{X}^*} \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \ldots \wedge \boldsymbol{e}^{i_p} f_{[i_1 i_2 \ldots i_p]} | f_{[i_1 i_2 \ldots i_p]} \right.$$

$$\left. := \frac{1}{p!}(f_{i_1 \ldots i_{p-1} i_p} - f_{i_1 \ldots i_p i_{p-1}} + \ldots + f_{i_p \ldots i_1 i_2} - f_{i_p \ldots i_2 i_1}) \right\}.$$

Corollary: $\dim \mathbf{A}^p = n!/(p!(n-p)!) = \binom{n}{p}$ in particular if $n=p$, then $\dim \mathbf{A}^p = 1$.

Box A.2vii. ($\mathbf{A}_q$, antisymmetric multilinear functions, exterior product *Proposition*):

(a) For every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_q \in \mathbb{X}$ as well as $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$ and $r, s \in \mathbb{R}$ multilinearity implies

$$\boldsymbol{x}_1 \wedge \ldots \wedge \boldsymbol{x}_{i-1} \wedge (r\boldsymbol{x} + s\boldsymbol{y}) \wedge \boldsymbol{x}_{i+1} \wedge \ldots \wedge \boldsymbol{x}_q$$
$$= r(\boldsymbol{x}_1 \wedge \ldots \wedge \boldsymbol{x}_{i-1} \wedge \boldsymbol{x} \wedge x_{i+1} \wedge \ldots \wedge \boldsymbol{x}_q$$
$$+ s(\boldsymbol{x}_1 \wedge \ldots \wedge \boldsymbol{x}_{i-1} \wedge \boldsymbol{y} \wedge \boldsymbol{x}_{i+1} \wedge \ldots \wedge \boldsymbol{x}_q.$$

(b) For every permutation σ of $\{1, 2, \ldots, q\}$ we have

$$\boldsymbol{x}_{\sigma_1} \wedge \boldsymbol{x}_{\sigma_2} \wedge \ldots \wedge \boldsymbol{x}_{\sigma_q} = \operatorname{sign}(\sigma) \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \ldots \wedge \boldsymbol{x}_q .$$

(c) Let $\mathbf{A} \in \mathbf{A}_q(\mathbb{X})$, $\boldsymbol{B} \in \mathbf{A}_s(\mathbb{X})$; then

$$\boldsymbol{B} \wedge \mathbf{A} = (-1)^{qs} \mathbf{A} \wedge \boldsymbol{B} .$$

(d) For every q, $0 \leq q \leq n$, the tensor space $\mathbf{A}_q$ of antisymmetric multilinear functions has dimension

$$\dim \mathbf{A}_q = \binom{n}{q} = n!/(q!(n-q)!).$$

As detailed examples we like to decompose $\{\mathbb{T}_0^1, \mathbb{T}_1^0, \mathbb{T}_0^2, \mathbb{T}_2^0\}$ in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively, into symmetric and antisymmetric constituents.

Example A.2. ($\mathbb{T}_q^p = \mathbb{S}_q^p \oplus \mathbf{A}_q^p \oplus \mathbb{R}_q^p$, decomposition of multilinear functions into symmetric and antisymmetric constituents):

As a *first example* of the decomposition of multilinear functions (tensor space) into *symmetric* and *antisymmetric* constituents we consider a linear space $\mathbb{X}$ (vector space) of dimension $\dim \mathbb{X} = n = 2$. Its dual space $\mathbb{X}^*$, $\dim \mathbb{X}^* = \dim \mathbb{X} = n = 2$, is spanned by orthonormal contravariant base vectors $\{\boldsymbol{e}^1, \boldsymbol{e}^2\}$. Choose $q = 0$, $p = 1$ and $p = 2$.

$$\mathbb{X} = \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\} \quad \text{versus} \quad \mathbb{X}^* = \operatorname{span}\{\boldsymbol{e}^1, \boldsymbol{e}^2\}$$

$$\mathbb{T}_0^1 = \mathbf{A}^1 = \mathbb{S}^1 \ni f = \left\{ \sum_{i=1}^{2} \boldsymbol{e}^i f_i \right\} = \boldsymbol{e}^1 f_1 + \boldsymbol{e}^2 f_2 \in \mathbb{X}$$

$$\mathbb{T}_0^2 = \mathbf{A}^2 \oplus \mathbb{S}^2$$

$$\begin{aligned}
\mathbb{T}_0^2 \ni f &= \left\{ \sum_{i,j=1}^{2} \boldsymbol{e}^i \otimes \boldsymbol{e}^j f_{ij} \right\} \\
&= \boldsymbol{e}^1 \otimes \boldsymbol{e}^1 f_{11} + \boldsymbol{e}^1 \otimes \boldsymbol{e}^2 f_{12} + \boldsymbol{e}^2 \otimes \boldsymbol{e}^1 f_{21} + \boldsymbol{e}^2 \otimes \boldsymbol{e}^2 f_{22} \\
&= \boldsymbol{e}^1 \otimes \boldsymbol{e}^1 f_{11} + \boldsymbol{e}^2 \otimes \boldsymbol{e}^2 f_{22} + \frac{1}{2}(\boldsymbol{e}^1 \otimes \boldsymbol{e}^2 - \boldsymbol{e}^2 \otimes \boldsymbol{e}^1) f_{12} \\
&\quad + \frac{1}{2}(\boldsymbol{e}^1 \otimes \boldsymbol{e}^2 + \boldsymbol{e}^2 \otimes \boldsymbol{e}^1) f_{12} - \frac{1}{2}(\boldsymbol{e}^1 \otimes \boldsymbol{e}^2 - \boldsymbol{e}^2 \otimes \boldsymbol{e}^1) f_{21} \\
&\quad + \frac{1}{2}(\boldsymbol{e}^1 \otimes \boldsymbol{e}^2 + \boldsymbol{e}^2 \otimes \boldsymbol{e}^1) f_{21} \\
&= \boldsymbol{e}^1 \vee \boldsymbol{e}^1 f_{11} + \boldsymbol{e}^2 \vee \boldsymbol{e}^2 f_{22} + \boldsymbol{e}^1 \wedge \boldsymbol{e}^2 (f_{12} - f_{21})/2 + \boldsymbol{e}^1 \vee \boldsymbol{e}^2 (f_{12} + f_{21})/2
\end{aligned}$$

$$\dim \mathbb{S}^2 = 3, \quad \dim \mathbf{A}^2 = 1.$$

As a *second example* of the decomposition of multilinear functions (tensor space) into symmetric and antisymmetric constituents we consider a linear space $\mathbb{X}$ (vector space) of dimension $\dim \mathbb{X} = n = 3$, spanned by orthonormal contravariant base vectors $\{\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3\}$. Choose $p = 0, q = 1$ and 2.

$$\operatorname{span} \mathbb{X} = \{\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3\}$$

$$\mathbb{T}_1^0 = \mathbf{A}_1 = \mathbb{S}_1 \ni f = \left\{ \sum_{i=1}^{3} \boldsymbol{e}_i f^i \right\} = \boldsymbol{e}_1 f^1 + \boldsymbol{e}_2 f^2 + \boldsymbol{e}_3 f^3 \in \mathbb{X}$$

$$\mathbb{T}_2^0 = \mathbf{A}_2 \oplus \mathbb{S}_2$$

$$\begin{aligned}
\mathbb{T}_2^0 \ni f &= \left\{ \sum_{i,j=1}^{3} \boldsymbol{e}_i \otimes \boldsymbol{e}_j f^{ij} \right\} \\
&= \left\{ \sum_{i=1}^{3} \boldsymbol{e}_i \otimes \boldsymbol{e}_1 f^{i1} + \sum_{i=1}^{3} \boldsymbol{e}_i \otimes \boldsymbol{e}_2 f^{i2} + \sum_{i=1}^{3} \boldsymbol{e}_i \otimes \boldsymbol{e}_3 f^{i3} \right\} \\
&= \boldsymbol{e}_1 \otimes \boldsymbol{e}_1 f^{11} + \boldsymbol{e}_2 \otimes \boldsymbol{e}_1 f^{21} + \boldsymbol{e}_3 \otimes \boldsymbol{e}_1 f^{31} + \boldsymbol{e}_1 \otimes \boldsymbol{e}_2 f^{12} + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2 f^{22} \\
&\quad + \boldsymbol{e}_3 \otimes \boldsymbol{e}_2 f^{32} + \boldsymbol{e}_1 \otimes \boldsymbol{e}_3 f^{13} + \boldsymbol{e}_2 \otimes \boldsymbol{e}_3 f^{23} + \boldsymbol{e}_3 \otimes \boldsymbol{e}_3 f^{33} \\
&\quad + \frac{1}{2}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 - \boldsymbol{e}_2 \otimes \boldsymbol{e}_1) f^{12} + \frac{1}{2}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_1) f^{12} \\
&\quad - \frac{1}{2}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 - \boldsymbol{e}_2 \otimes \boldsymbol{e}_1) f^{21} + \frac{1}{2}(\boldsymbol{e}_1 \otimes \boldsymbol{e}_2 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_1) f^{12} \\
&\quad + \frac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_3 - \boldsymbol{e}_3 \otimes \boldsymbol{e}_2) f^{23} + \frac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_3 + \boldsymbol{e}_3 \otimes \boldsymbol{e}_2) f^{23} \\
&\quad - \frac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_3 - \boldsymbol{e}_3 \otimes \boldsymbol{e}_2) f^{32} + \frac{1}{2}(\boldsymbol{e}_2 \otimes \boldsymbol{e}_3 + \boldsymbol{e}_3 \otimes \boldsymbol{e}_2) f^{32} \\
&\quad + \frac{1}{2}(\boldsymbol{e}_3 \otimes \boldsymbol{e}_1 - \boldsymbol{e}_1 \otimes \boldsymbol{e}_3) f^{31} + \frac{1}{2}(\boldsymbol{e}_3 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_1 \otimes \boldsymbol{e}_3) f^{31} \\
&\quad - \frac{1}{2}(\boldsymbol{e}_3 \otimes \boldsymbol{e}_1 - \boldsymbol{e}_1 \otimes \boldsymbol{e}_3) f^{13} + \frac{1}{2}(\boldsymbol{e}_3 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_1 \otimes \boldsymbol{e}_3) f^{13} \qquad \clubsuit
\end{aligned}$$

Since the subspaces $\mathbb{S}_q^p$, $\mathbf{A}_q^p$ and $\mathbb{R}_q^p$ are independent, $\mathbb{S}_q^p \oplus \mathbf{A}_q^p \oplus \mathbb{R}_q^p$ denotes the *direct sum* of subspace $\mathbb{S}_q^p$, $\mathbf{A}_q^p$ and $\mathbb{R}_q^p$. Unfortunately $\mathbb{T}_q^p$ as the space of multilinear functions cannot be completely decomposed in the space of symmetric and antisymmetric multilinear functions: For instance, the dimension identities apply $\dim \mathbb{T}^p = n^p$, $\dim \mathbb{S}^p = \binom{n+p-1}{p}$, $\dim \mathbf{A}^p = \binom{n}{p}$ with respect of a vector space $\mathbb{X}$ of dimension $\dim \mathbb{X} = n$, such that $\dim \mathbb{R}^p = \dim \mathbb{T}^p - \dim \mathbb{S}^p - \dim \mathbf{A}^p = n^p - \binom{n+p+1}{p} - \binom{n}{p} < n^p$, in general. There is *one exception*, namely the (2,0) or (1,1) or (0,2) tensor space where the dimension of the subspace $\mathbb{R}_0^2$, $\mathbb{R}_1^1$ or $\mathbb{R}_2^0$ of *residual* multilinear functions *is zero*. An example is $\dim \mathbb{R}^2 = n^2 - \binom{n+1}{2} - \binom{n}{2} = n^2 - (n+1)n/2 - n(n-1)^2 = 0$.

A-3 Matrix Algebra, Array Algebra, Matrix Norm and Inner Product

Symmetry and antisymmetry of the symbols $f_{i_1 \ldots i_p}$ can be visualized by the trees of Box A.2i and A.2ii. With respect to the symbols of the *interior product* "$\vee$" and the *exterior product* "$\wedge$" ("*wedge product*") we are able to redefine symmetric and antisymmetric functions according to Box A.2iii–vi. Note the *isomorphism of tensor algebra* $\mathbb{T}^p_q$ and *array algebra*, namely of

(a) $[f_i] \in \mathbb{R}^n$ (onedimensional array, "column vector", $\dim[f_i] = n \times 1$)

(b) $[f_{ij}] \in \mathbb{R}^{n \times n}$ (twodimensional array, column-row array, "matrix", $\dim[f_{ij}] = n \times n$)

(c) $[f_{ijk}] \in \mathbb{R}^{n \times n \times n}$ (threedimensional array, "indexed matrix", $\dim[f_{ijk}] = n \times n \times n$)

etc. For the base space $\boldsymbol{x} \in \Omega \subset \mathbb{R}^3$ to be threedimensional *Euclidean* we had answered the question how to measure the length of a vector ("norm") and the angle between two vectors ("inner product"). The same question will finally been raised for tensors $t^p_q \in \mathbb{T}^p_q$. The answer is *constructively* based on the *vectorization* of the arrays $[f_{ij}]$, $[f_{ijk}]$, … $[f_{i_1 \ldots i_p}]$ by taking advantage of the symmetry-antisymmetry structure of the arrays and later on applying the *Euclidean* norm and the *Euclidean* inner product to the *vectorized array*.

For a 2-contravariant, 0-covariant tensor we shall outline the procedure.

(a) *Firstly* let $\boldsymbol{F} = [f_{ij}]$ be the quadratic matrix of dimension $\dim \boldsymbol{F} = n \times n$, an element of $\mathbb{T}^2_0$. Accordingly $\operatorname{vec} \boldsymbol{F}$ is the vector

$$\operatorname{vec} \boldsymbol{F} = \begin{bmatrix} f_{i_1} \\ f_{i_2} \\ \vdots \\ f_{i_{n-1}} \\ f_{i_n} \end{bmatrix}, \quad \dim \operatorname{vec} \boldsymbol{F} = n^2 \times 1$$

which is generated by stacking the elements of the matrix $\boldsymbol{F}$ *columnwise* in a vector. The *Euclidean* norm and the *Euclidean* inner product of $\operatorname{vec} \boldsymbol{F}$, $\operatorname{vec} \boldsymbol{G}$, respectively is

$$\boxed{\begin{aligned} \|\operatorname{vec} \boldsymbol{F}\|^2 &:= (\operatorname{vec} \boldsymbol{F})^T (\operatorname{vec} \boldsymbol{F}) &&= \operatorname{tr} \boldsymbol{F}^T \boldsymbol{F}, \\ \langle \operatorname{vec} F | \operatorname{vec} \boldsymbol{G} \rangle &:= (\operatorname{vec} \boldsymbol{F})^T \operatorname{vec} \boldsymbol{G} &&= \operatorname{tr} \boldsymbol{F}^T \boldsymbol{G}. \end{aligned}}$$

(b) *Secondly* let $\boldsymbol{F} = [f_{ij}] = [f_{ji}]$ be the *symmetric* matrix of dimension $\dim \boldsymbol{F} = n \times n$, an element of S^2. Accordingly $\operatorname{vech} \boldsymbol{F}$ (read "vector half") is the $n(n+1)/2 \times 1$ vector which is generated by stacking the elements *on and under* the main diagonal of the matrix $\boldsymbol{F}$ *columnwise* in a vector:

$$\boldsymbol{F} = [f_{ij}] = [f_{ji}] = \boldsymbol{F}^T \implies \operatorname{vech} \boldsymbol{F} := \begin{bmatrix} f_{11} \\ \cdot \\ f_{n1} \\ \hline f_{22} \\ \cdot \\ f_{2n} \\ \hline \cdot \\ \hline f_{nn} \end{bmatrix}, \ \dim \operatorname{vech} \boldsymbol{F} = n(n+1)/2$$

$$\operatorname{vech} \boldsymbol{F} = \boldsymbol{H} \operatorname{vec} \boldsymbol{F}, \quad \dim \boldsymbol{H} = n(n+1)/2 \times n^2.$$

The *Euclidean* norm and the *Euclidean* inner product of $\operatorname{vech} \boldsymbol{F}$, $\operatorname{vech} \boldsymbol{G}$, respectively is

$$\left.\begin{array}{l} \boldsymbol{F} = [f_{ij}] = [f_{ji}] = \boldsymbol{F}^T \\ \boldsymbol{G} = [g_{ij}] = [g_{ji}] = \boldsymbol{G}^T \end{array}\right] \implies$$

$$\boxed{\begin{aligned} \|\operatorname{vech} \boldsymbol{F}\|^2 &:= (\operatorname{vech} \boldsymbol{F})^T (\operatorname{vech} \boldsymbol{F}) \\ \langle \operatorname{vech} \boldsymbol{F} | \operatorname{vech} \boldsymbol{G} \rangle &:= (\operatorname{vech} F)^T (\operatorname{vech} \boldsymbol{G}). \end{aligned}}$$

(c) *Thirdly* let $\boldsymbol{F} = [f_{ij}] = -[f_{ji}]$ be the antisymmetric matrix of dimension $\dim F = n \times n$, an element of $\mathbf{A}^2$. Accordingly veck $\boldsymbol{F}$ (*read* "vector skew") is the $n(n-1)/2 \times 1$ vector which is generated by stacking the elements *under* the main diagonal of the matrix $\boldsymbol{F}$ columnwise in a vector:

$$\boldsymbol{F} = [f_{ji}] = -[f_{ji}] = -\boldsymbol{F}^T \implies \operatorname{veck} \boldsymbol{F} := \begin{bmatrix} f_{21} \\ \cdot \\ f_{n1} \\ \hline f_{32} \\ \cdot \\ f_{n2} \\ \hline \cdot \\ \hline f_{n-1n} \end{bmatrix}, \ \dim \operatorname{veck} \boldsymbol{F} = n(n-1)/2$$

$$\operatorname{veck} \boldsymbol{F} = \boldsymbol{K} \operatorname{vec} \boldsymbol{F}, \quad \dim \boldsymbol{K} = n(n-1)/2 \times n^2.$$

The *Euclidean* norm and the *Euclidean* inner product of $\operatorname{veck} \boldsymbol{F}$, $\operatorname{veck} \boldsymbol{G}$, respectively

$$\left.\begin{aligned} \boldsymbol{F} &= [f_{ij}] = -[f_{ji}] = -\boldsymbol{F}^T \\ \boldsymbol{G} &= [g_{ij}] = -[g_{ji}] = -\boldsymbol{G}^T \end{aligned}\right] \Longrightarrow$$

$$\boxed{\begin{aligned} \|\text{veck}\,\boldsymbol{F}\|^2 &:= (\text{veck}\,\boldsymbol{F})^T(\text{veck}\,\boldsymbol{F}) \\ \langle \text{veck}\,\boldsymbol{F} | \text{veck}\,\boldsymbol{G} \rangle &:= (\text{veck}\,\boldsymbol{F})^T(\text{veck}\,\boldsymbol{G}) \end{aligned}}\,.$$

You will find the forms (a) *vec F* and (b) *Vech F* in the standard book of (Harville, 1997, Chap. 16, 2001, Chap. 16) including many examples. Here we will present Example A.3 in computing the norm as well as the inner product of a 2-contravariant, 0-covariant tensor for the operator vec, vech and veck. The veck operator has been introduced by (Grafarend and Schaffrin, 1993, pp. 418–419).

Example A.3. (Norm and inner product of a 2-contravariant, 0-covariant tensor):

(a) $\mathbf{A} := \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & k \end{bmatrix}$, $\dim \mathbf{A} = 3 \times 3 \Longrightarrow \text{vec}\,\mathbf{A} = [a, b, c, d, e, f, g, h, k]^T$,

$\dim \text{vec}\,\mathbf{A} = 9 \times 1$

$\|\,\text{vec}\,\mathbf{A}\|^2 = (\text{vec}\,\mathbf{A})^T(\text{vec}\,\mathbf{A}) = \text{tr}\,\mathbf{A}^T\mathbf{A} = a^2 + \ldots + k^2$

(b) $\mathbf{A} := \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} = \mathbf{A}^T$, $\dim \mathbf{A} = 3 \times 3 \Longrightarrow \text{vech}\,\mathbf{A} = [a, b, c, d, e, f]^T$,

$\dim \text{vech}\,\mathbf{A} = 6 \times 1$

$\text{vech}\,\mathbf{A} = \boldsymbol{H}\,\text{vec}\,\mathbf{A}$

$$\forall\ \boldsymbol{H} := \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right], \quad \dim \boldsymbol{H} = 6 \times 9$$

$\|\text{vech}\,\mathbf{A}\|^2 = (\text{vech}\,\mathbf{A})^T(\text{vech}\,\mathbf{A}) = a^2 + b^2 + c^2 + d^2 + e^2 + f^2$
(Henderson and Searle (1978, p. 68–69))

(c) $\mathbf{A} := \begin{bmatrix} 0 & -a & -b & -c \\ a & 0 & -d & -e \\ b & d & 0 & -f \\ c & e & f & 0 \end{bmatrix} = -\mathbf{A}^T, \dim \mathbf{A} = 3 \times 3 \Longrightarrow$

$\text{veck}\,\mathbf{A} = [a, b, c, d, e, f]^T$, $\dim \text{veck}\,\mathbf{A} = 6 \times 1$

$\text{veck}\,\mathbf{A} = \boldsymbol{K}\,\text{vec}\,\mathbf{A}$

$$
\forall \;\; \boldsymbol{K} := \frac{1}{2}\left[\begin{array}{cccc|cccc|cccc|cccc}
0&1&0&0&-1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&-1&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&-1&0&0&0\\
\hline
0&0&0&0&0&0&1&0&0&-1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&0&0&0&-1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&-1&0
\end{array}\right],
$$

$\dim \boldsymbol{K} = 6 \times 16$

$\|\operatorname{veck} \mathbf{A}\|^2 = a^2 + b^2 + c^2 + d^2 + e^2 + f^2$ ♣

A-4 The Hodge Star Operator, Self Duality

In Chap. 3, we took advantage of the *Hodge star operator*, for instance for transforming an overdetermined system of linear equations (inconsistent system) into a system of condition equations. There we referred to multilinear operations "*join*" and"*meet*".
The most important operator of the algebra of antisymmetric multilinear functions is the "*Hodge star operator*" which we shall present finally. In addition, we shall bring to you the surprising special feature of *skew algebra* called "*selfduality*".
The algebra $\mathbf{A}_q^p$ of antisymmetric multilinear functions has been based on the exterior product "$\wedge$" ("wedge product"). There has been created a duality operator called the *Hodge star operator* $*$ which is a linear map of $\mathbf{A}^p \longrightarrow \mathbf{A}^{n-p}$ where $n = \dim \mathbb{X} = \dim \mathbb{X}^*$ denotes the dimension of the *base space* $\mathbb{X} \subset \mathbb{R}^3$. The basic idea of such a map of antisymmetric multilinear functions $f \in \mathbf{A}^p$ into antisymmetric linear functions $*f \in \mathbf{A}^{n-p}$ originates according to Box A.2vii from the following situation: The multilinear base of $\mathbf{A}^p$ is spanned by

$$\{1, \boldsymbol{e}^{i_1}, \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2}, \ldots \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \ldots \wedge \boldsymbol{e}^{i_p}\}$$

once we focus on $p = 0, 1, 2, \ldots n$, respectively. Obviously for any dimension number n and p-contravariant, q-covariant index of the skew tensor space $\mathbf{A}_q^p$ there is an *associated cobasis*, namely

$$n = 1\,, \quad p = 0, 1 \quad : \begin{cases} \textit{basis}: & \{1, \boldsymbol{e}^{i_1}\} \\ \textit{associated} & \\ \textit{cobasis}: & \{\boldsymbol{e}^{i_1}, 1\} \end{cases}$$

$$n = 2\,, \quad p = 0, 1, 2 : \begin{cases} \textit{basis}: & \{1, \boldsymbol{e}^{i_1}, \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2}\} \\ \textit{associated} & \\ \textit{cobasis}: & \{\boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2}, \boldsymbol{e}^{i_2}, 1\} \end{cases}$$

$$n = 3\,, \quad p = 0, 1, 2, 3: \begin{cases} \textit{basis}: & \{1, \boldsymbol{e}^{i_1}, \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3}, \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3}\} \\ \textit{associated} & \\ \textit{cobasis}: & \{\boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3}, \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3}, \boldsymbol{e}^{i_3}, 1\}. \end{cases}$$

in general, for arbitrary $n \in \mathbb{N}$, $p = 0, 1, \ldots, n-1, n$

$$\boxed{\begin{array}{l} \textit{basis}: \\ \{1, \boldsymbol{e}^{i_1}, \ldots, \boldsymbol{e}^{i_1} \wedge \ldots \wedge \boldsymbol{e}^{i_n}\} \\ \textit{associated} \quad \textit{cobasis}: \\ \{\boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \ldots \wedge \boldsymbol{e}^{i_{n-1}} \wedge \boldsymbol{e}^{i_n}, \boldsymbol{e}^{i_2} \wedge \ldots \boldsymbol{e}^{i_{n-1}} \wedge \boldsymbol{e}^{i_n}, \ldots, \boldsymbol{e}^{i_{n-1}} \wedge \boldsymbol{e}^{i_n}, \boldsymbol{e}^{i_n}, 1, \end{array}}$$

as long as we concentrate on p-contravariant $\mathbf{A}^p$ only. A similar set-up of basis-associated cobasis for q-covariant $\mathbf{A}_q$ and mixed $\mathbf{A}_q^p$ can be made. The linear map $\mathbf{A}^p \longrightarrow \mathbf{A}^{n-p}$, the *Hodge star operator*

$$\boxed{*(\boldsymbol{e}^{i_1} \wedge \ldots \wedge \boldsymbol{e}^{i_p}) := \frac{1}{(n-p)!} \epsilon^{i_1 \ldots i_p}_{i_{p+1} \ldots i_n} \boldsymbol{e}^{i_{p+1}} \wedge \ldots \wedge \boldsymbol{e}^{i_n}}$$

maps by means of the permutation symbol

$$\epsilon^{i_1 \ldots i_p}_{i_{p+1} \ldots i_n} := \left[\begin{array}{l} +1 \textit{ for an even permutation} \\ \qquad \textit{of } \{1, 2, \ldots, n-1, n\} \\ -1 \textit{ for an odd permutation} \\ \qquad \textit{of } \{1, 2, \ldots, n-1, n\} \\ \quad 0 \textit{ otherwise} \end{array} \right.$$

– sometimes called *Eddington's epsilons* – on orthonormal ("unimodular") base of $\mathbf{A}^p$ onto an orthonormal ("unimodular") base of $\mathbf{A}^{n-p}$.

For antisymmetric multilinear functions also called antisymmetric tensor-valued functions represented in an orthonormal ("unimodular") base the *Hodge star operator* is the following linear map

$$\mathbb{T}_0^p \supset \mathbf{A}^p \ni f = \left\{ \frac{1}{p!} \sum_{i_1, \ldots, i_p = 1}^{n = \dim \boldsymbol{X}^*} \boldsymbol{e}^{i_1} \wedge \ldots \wedge \boldsymbol{e}^{i_p} f_{i_1 \ldots i_p} \right\},$$

$$^*\mathbb{T}_0^p \supset \mathbf{A}^{n-p} \ni *f := \left\{ \frac{1}{(n-p)!} \sum_{i_{p+1}, \ldots, i_n}^{n = \dim \boldsymbol{X}^*} \sum_{i_1, \ldots, i_p}^{n = \dim \boldsymbol{X}} \frac{1}{p!} \epsilon^{i_1 \ldots i_p}_{i_{p+1} \ldots i_n} \boldsymbol{e}^{i_{p+1}} \wedge \ldots \wedge \boldsymbol{e}^{i_n} f_{i_1 \ldots i_p} \right\}.$$

As soon as the base space $\boldsymbol{x} \in \boldsymbol{\Omega} \subset \mathbb{R}^3$ is *not* covered by Cartesian coordinates, rather by *curvilinear coordinates*, its coordinates base

$$\{\boldsymbol{b}^1, \boldsymbol{b}^2, \boldsymbol{b}^3\} = \{dy^1, dy^2, dy^3\} \quad \textit{versus} \quad \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\} = \left\{\frac{\partial}{\partial y^1}, \frac{\partial}{\partial y^2}, \frac{\partial}{\partial y^3}\right\}$$

of contravariant *versus* covariant type is *neither* orthogonal *nor* normalized. It is for this reason that finally we present $*f$, the *Hodge star operator* of an antisymmetric multilinear function f, also called *the dual of* f, in a general coordinate base.

Definition A.2. (Hodge star operator, the dual of an antisymmetric multilinear function):

If an antisymmetric $(p,0)$ *multilinear function is an element of the skew algebra* $\mathbf{A}^p$ *with respect, to a general base* $\{\boldsymbol{b}^{i_1} \wedge \ldots \wedge \boldsymbol{b}^{i_p}\}$ *is given*

$$f = \left\{\frac{1}{p!} \sum_{i_1,\ldots,i_p=1}^{n=\dim \boldsymbol{X}^*} \boldsymbol{b}^{i_1} \wedge \ldots \wedge \boldsymbol{b}^{i_p} f_{i_1\ldots i_p}\right\}$$

then the *Hodge star operator*, the *dual* of f, can be uniquely represented by

(a) $$*f = \left\{\frac{1}{(n-p)!} \sum_{i_{p+1},\ldots,i_n}^{n=\dim \boldsymbol{X}^*} \sum_{i_1,\ldots i_p}^{n=\dim \boldsymbol{X}^*} \sum_{j_1,\ldots,j_p}^{n=\dim \boldsymbol{X}^*} \frac{1}{p!} \boldsymbol{b}^{i_{p+1}} \wedge \ldots \wedge \boldsymbol{b}^{i_p} \times \sqrt{g}\epsilon_{i_1\ldots i_p i_{p+1}\ldots i_n} g^{i_1 j_1} \ldots g^{i_p j_p} f_{j_1\ldots j_p}\right\}$$

(b) $$*f = \left\{\frac{1}{(n-p)!} \sum_{i_{p+1},\ldots,i_n}^{n=\dim \boldsymbol{X}^*} \sum_{i_1,\ldots,i_p}^{n=\dim \boldsymbol{X}^*} \frac{1}{p!} \boldsymbol{b}^{i_{p+1}} \wedge \ldots \wedge \boldsymbol{b}^{i_n} \sqrt{g}\epsilon_{i_1\ldots i_p i_{p+1}\ldots i_n} f^{i_1\ldots i_p}\right\}$$

(c) $$(*f)_{k_1\ldots k_{n-p}} = \left\{\sum_{i_1,\ldots i_p}^{n=\dim \boldsymbol{X}^*} \sqrt{g}\epsilon_{i_1\ldots i_p k_1\ldots k_{n-p}} f^{i_1\ldots i_p}\right\}$$

as an element of the *skew algebra* $\mathbf{A}^{n-p}$ in the general associated cobase $\{\boldsymbol{b}^{p+1} \wedge \ldots \wedge \boldsymbol{b}^{i_n}\}$ with respect to the base space $\boldsymbol{x} \in \mathbb{X} \supset \mathbb{R}^n$ of dimension $n = \dim \mathbb{X} = \dim \mathbb{X}^*$ and $[g^{kl}] = \boldsymbol{G}^{-1} = \operatorname{adj} \boldsymbol{G} / \det \boldsymbol{G}$, $\sqrt{g} = \sqrt{|g_{kl}|}$.

If we extend the algebra $\mathbf{A}^p$ of antisymmetric multilinear functions by $*1 = \boldsymbol{e}^1 \wedge \ldots \boldsymbol{e}^n \in \mathbf{A}^n$ and $*\boldsymbol{e}^1 \wedge \ldots \wedge \boldsymbol{e}^n = 1 \in \mathbf{A}^0 = \mathbb{R}$, respectively, let us collect some properties of $*f$, the *dual* of f.

Proposition A.1. *(***Hodge star operator***, the dual of an antisymmetric multilinear function):*

Let the linearly ordered base $\{\boldsymbol{e}^1, \ldots, \boldsymbol{e}^n\}$ *be orthonormal ("unimodular"). Then the Hodge star operator of an antisymmetric multilinear function* f*, the dual of* f*, with respect to* $\{\boldsymbol{e}^1, \ldots \boldsymbol{e}^n\}$ *satisfies the following:*

(a) $*$ *maps antisymmetric* p*-contravariant tensor-valued functions to antisymmetric* $(n-p)$*-contravariant tensor-valued functions:*
$* : \mathbf{A}^p \longrightarrow \mathbf{A}^{n-p}$

(b) $\begin{cases} *1 = \boldsymbol{e}^1 \wedge \ldots \wedge \boldsymbol{e}^n =: \mathrm{e} & \textit{for every } 1 \in \mathbf{A}^0, \ \mathrm{e} \in \mathbf{A} \\ *\mathrm{e} = 1 & \textit{for every } \ \mathrm{e} \in \mathbf{A}^n, \ 1 \in \mathbf{A}^p \end{cases}$

(c) $** f = (-1)^{p(n-p)} f$ *for every* $f \in \mathbf{A}^p$

(d) $f \wedge *f = \|f\|^2 \boldsymbol{e}^1 \wedge \ldots \wedge \boldsymbol{e}^n$ *with respect to the norm*

$$\|f\|^2 := \frac{1}{p!} \sum_{i_1,\ldots,i_p=1}^{n=\dim \boldsymbol{X}} f_{i_1\ldots i_p} f^{i_1\ldots i_p}.$$

Example A.4. (Hodge star operator $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 3$, span $\mathbb{X}^* = \{\boldsymbol{e}^1, \boldsymbol{e}^2, \boldsymbol{e}^3\}$, $\mathbf{A}^p \longrightarrow \mathbf{A}^{n-p}$):

$$n = 3\,, \ p = 0 : *\,1 = \boldsymbol{e}^1 \wedge \boldsymbol{e}^2 \wedge \boldsymbol{e}^3$$

$$n = 3\,, \ p = 1 : *\,\boldsymbol{e}^{i_1} = \frac{1}{2}\epsilon^{i_1}_{i_2 i_3} \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3}$$

$$\left[\begin{array}{l} *\boldsymbol{e}^1 = \frac{1}{2}(\boldsymbol{e}^2 \wedge \boldsymbol{e}^3 - \boldsymbol{e}^3 \wedge \boldsymbol{e}^2) = \boldsymbol{e}^2 \wedge \boldsymbol{e}^3 \\ *\boldsymbol{e}^2 = \frac{1}{2}(\boldsymbol{e}^3 \wedge \boldsymbol{e}^1 - \boldsymbol{e}^1 \wedge \boldsymbol{e}^3) = \boldsymbol{e}^3 \wedge \boldsymbol{e}^1 \\ *\boldsymbol{e}^3 = \frac{1}{2}(\boldsymbol{e}^1 \wedge \boldsymbol{e}^2 - \boldsymbol{e}^2 \wedge \boldsymbol{e}^1) = \boldsymbol{e}^1 \wedge \boldsymbol{e}^2 \end{array}\right.$$

$$n = 3\,, \ p = 2 : *\,\boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} = \epsilon^{i_1 i_2}_{i_3} \boldsymbol{e}^{i_3}$$

$$\left[\begin{array}{l} *\boldsymbol{e}^1 \wedge \boldsymbol{e}^2 = \boldsymbol{e}^3 \\ *\boldsymbol{e}^2 \wedge \boldsymbol{e}^3 = \boldsymbol{e}^1 \\ *\boldsymbol{e}^3 \wedge \boldsymbol{e}^1 = \boldsymbol{e}^2 \end{array}\right.$$

$$n = 3\,, \ p = 3 : *\,\boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3} = 1$$ ♣

Example A.5. (Hodge star operator of an antisymmetric tensor-valued function, $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 3$, $\mathbf{A}^p \longrightarrow \mathbf{A}^{n-p}$):

Throughout we apply the summation convention over repeated indices.

$$n = 3\,, \ p = 0 : \begin{cases} f & \text{“0-differential form”} \\ *f = f dx^1 \wedge dx^2 \wedge dx^3 & \text{“3-differential form''} \end{cases}$$

$$n = 3\,,\ p = 1 : \begin{cases} f &= dx^{i_1} f_{i_1} & \text{“1-differential form”} \\ *f &= \frac{1}{2}\epsilon^{i_1}{}_{i_2 i_3} dx^{i_2} \wedge dx^{i_3} f_{i_1} & \\ &= f_1 dx^2 \wedge dx^3 + f_2 dx^3 \wedge dx^1 & \\ &\quad + f_3 dx^1 \wedge dx^2 & \text{“2-differential form”} \end{cases}$$

$$n = 3\,,\ p = 2 : \begin{cases} f &= \frac{1}{2} dx^{i_1} \wedge dx^{i_2} f_{i_1 i_2} & \text{“2-differential form”} \\ *f &= \frac{1}{2}\epsilon^{i_1 i_2}{}_{i_3} dx^{i_3} f_{i_1 i_2} & \\ &= f_{23} dx^1 + f_{31} dx^2 + f_{12} dx^3 & \text{“1-differential form”} \end{cases}$$

$$n = 3\,,\ p = 3 : \begin{cases} f &= \frac{1}{6} dx^{i_1} \wedge dx^{i_2} \wedge dx^{i_3} f_{i_1 i_2 i_3} & \text{“3-differential form”} \\ *f &= \frac{1}{6}\epsilon^{i_1 i_2 i_3} f_{i_1 i_2 i_3} = f_{123} & \text{“0-differential form”} \end{cases} \quad \clubsuit$$

Example A.6. (Hodge star operator, $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 3$, “$\times$” product (cross product)):

By means of the *Hodge star operator* we are able to interpret the “$\times$” product (“*cross product*”) in *threedimensional* vector space. If the vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$, $\dim \mathbb{X} = 3$, presented in the orthonormal (“*unimodular*”) base $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ the following *equivalence* between $*\boldsymbol{x} \wedge \boldsymbol{y}$ and $\boldsymbol{x} \times \boldsymbol{y}$ holds:

$$\left.\begin{array}{lll} \boldsymbol{x} = \boldsymbol{e}_i \boldsymbol{x}^i\,, & \boldsymbol{y} = \boldsymbol{e}_j \boldsymbol{y}^j & (\textit{summation convention} \\ \boldsymbol{x} \in \mathbb{X}\,, & \boldsymbol{y} \in \mathbb{X} & \quad i, j \in \{1, 2, 3, \}) \end{array}\right] \Longrightarrow$$

$$\begin{aligned} \boldsymbol{x} \wedge \boldsymbol{y} &= \boldsymbol{e}_i \wedge \boldsymbol{e}_j \boldsymbol{x}^i \boldsymbol{y}^j \\ &= \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 (\boldsymbol{x}^1 \boldsymbol{y}^2 - \boldsymbol{x}^2 \boldsymbol{y}^1) + \boldsymbol{e}_2 \wedge \boldsymbol{e}_3 (\boldsymbol{x}^2 \boldsymbol{y}^3 - \boldsymbol{x}^3 \boldsymbol{y}^2) \\ &\quad + \boldsymbol{e}_3 \wedge \boldsymbol{e}_1 (\boldsymbol{x}^3 \boldsymbol{y}^1 - \boldsymbol{x}^1 \boldsymbol{y}^3) \\ \Longrightarrow *(\boldsymbol{x} \wedge \boldsymbol{y}) &= *(\boldsymbol{e}_1 \wedge \boldsymbol{e}_2)(\boldsymbol{x}^1 \boldsymbol{y}^2 - \boldsymbol{x}^2 \boldsymbol{y}^1) \\ &\quad + *(\boldsymbol{e}_2 \wedge \boldsymbol{e}_3)(\boldsymbol{x}^2 \boldsymbol{y}^3 - \boldsymbol{x}^3 \boldsymbol{y}^2) \\ &\quad + *(\boldsymbol{e}_3 \wedge \boldsymbol{e}_1)(\boldsymbol{x}^3 \boldsymbol{y}^1 - \boldsymbol{x}^1 \boldsymbol{y}^3) = \epsilon^k_{ij} \boldsymbol{e}_k \boldsymbol{x}^i \boldsymbol{y}^j \\ *(\boldsymbol{x} \wedge \boldsymbol{y}) &= \boldsymbol{e}_3 (\boldsymbol{x}^1 \boldsymbol{y}^2 - \boldsymbol{x}^2 \boldsymbol{y}^1) + \boldsymbol{e}_1 (\boldsymbol{x}^2 \boldsymbol{y}^3 - \boldsymbol{x}^3 \boldsymbol{y}_2) + \boldsymbol{e}_2 (\boldsymbol{x}^3 \boldsymbol{y}^1 - \boldsymbol{x}^1 \boldsymbol{y}^3) \\ &= \boldsymbol{e}_1 (\boldsymbol{x}^2 \boldsymbol{y}^3 - \boldsymbol{x}^3 \boldsymbol{y}^2) + \boldsymbol{e}_2 (\boldsymbol{x}^3 \boldsymbol{y}^1 - \boldsymbol{x}^1 \boldsymbol{y}^3) + \boldsymbol{e}_3 (\boldsymbol{x}^1 \boldsymbol{y}^2 - \boldsymbol{x}^2 \boldsymbol{y}^1) \end{aligned}$$

$$\left.\begin{array}{rl} \boldsymbol{x} \times \boldsymbol{y} &= \boldsymbol{e}_i \times \boldsymbol{e}_j \boldsymbol{x}^i \boldsymbol{x}^j \\ \boldsymbol{e}_i \times \boldsymbol{e}_j &:= \epsilon_{ij}{}^k \boldsymbol{e}_k \end{array}\right] \Longrightarrow \boxed{\boldsymbol{x} \times \boldsymbol{y} = *(\boldsymbol{x} \wedge \boldsymbol{y})} \qquad \clubsuit$$

Example A.7. (Hodge star operator, $n = \dim \mathbb{X} = \dim \mathbb{X}^* = 4, \mathbb{X} \in \mathbb{R}^4$, Minkwinski space, self-duality (Atiyah et al., 1978)):

By means of the Examples A-5–A-7 we like to make you familiar with (a) the Hodge star operator of an antisymmetric tensor-valued function over $\mathbb{R}^3$, (b) its equivalence to the "x" product ("*cross product*") and (c) *selfduality* in a *fourdimensional space*. Such a selfduality plays a key role in differential geometry and physics as being emphasized by Atiyah et al. (1978).

Historical Aside

Thus we have constructed an *anticommutative algebra* by implementing the "*exterior product*" "$\wedge$", also called "*wedge product*", initiated by *H. Grassmann* in "*Ausdehnungslehre*" (second version published in 1882). See also his collected works, *H. Grassmann* (1911). In addition the work by *G. Peano* (Calcolo geometrico secondo, l'Ausdehnungslehre di Grassmann, *Fratelli Bocca Editori*, Torino 1888) should be mentioned here. The historical development may be documented by the work of Forder (1960). A modern version of the "*wedge product*" is given by Berman (1961). In particular we mention the contribution by Barnabei et al. (1985) where by avoiding the notion of the dual $\mathbb{X}^*$ of a linear space $\mathbb{X}$ and based upon operations like *union*, *intersection*, and *complement* – i.e. known in *Boolean algebra* – have developed a *double algebra* with exterior products of *type one* ("*wedge product*", "*the join*") and of *type two* ("*the meet*"), namely "*to restore H. Grassmanns original ideas to full geometrical power*". The star operator "$*$" has been introduced by *W. V. D. Hodge*, being implemented into algebra in the work Hodge (1941) and Hodge and Pedoe (1968), pp. 232–309). Here the star operation has been called "*dual Grassmann coordinates*"; in addition "*intersections and joins*" have been introduced.

A-5 Linear Algebra

Multilinear algebra is built on *linear algebra* we are going into now. At first we give a careful definition of linear algebra which secondly we deepen by the diagrams "*Ass*", "*Uni*" and "*Comm*". The subalgebra "*ring with identity*" which is of central importance for solving polynomial equations by means of *Groebner bases*, the *Buchberger algorithm* and the *multipolynomial resultant* method is our third subject. Section 4 introduces the motion of division algebra and the non-associative algebra. *Fifthly*, we confront you with Lie algebra ("God is a lie Group"); in particular with *Witt algebra*. Section 6 compares *Lie algebra* and *Killing analysis*. Here we add some notes on the difficulties of a *composition algebra* in Sect. 7. Finally in Sect. 8

matrix algebra is presented again, but this time as a division algebra. As examples of a *division algebra* as well as composition algebra we introduce *complex algebra* (*Clifford algebra* $Cl(0, 1)$) in Sect. 9 and *quaternion algebra* in Sect. 10 (*Clifford algebra* $Cl(0, 2)$) which is followed by an interesting letter of *W. R. Hamilton* (16 October 1943) to his son reproduced in Sect. 11. *Octonian algebra* (*Clifford algebra* with respect to $\mathbb{H} \times \mathbb{H}$) in Sect. 12 is an example for a "non associative" algebra as well as a composition algebra. Of course, we have reserved "Sect. 13" for the fundamental *Hurwitz theorem* of composition algebra and the fundamental *Frobenius theorem* of division algebra.

A-51 Definition of a Linear Algebra

Up to now we have succeeded to introduce the *base space* $\mathbb{X}$ of vectors $\mathbf{x} \in \mathbb{X} = \mathbb{R}^{\mathbf{3}}$ equipped with a metric and specialized to be *threedimensional Euclidean*. We have extended the base space to a tensor space, namely from vector-valued functions to tensor-valuedfunctions on $\{\mathbb{R}^n, g_{ij}\}$. Now we proceed to give linear and multilinear functions *an algebraic structure*, namely by the definition of *two binary operations*, (*two internal relations*) $(\text{opera})_1 = \alpha$, $(\text{opera})_2 = \mu$ and *one binary operation* (*external relation*) $(\text{opera})_3 = \beta$.

Definition A.51. (linear algebra over the field of real numbers, linearity of vector space $\mathbb{X}$):

Let $\mathbb{R}$ *be the field of real numbers. A* linear algebra *over* $\mathbb{R}$ *or* $\mathbb{R}$*-algebra consists of a set* $\mathbb{X}$ *of objects, two internal relations (either* "additive" *or* "multiplicative"*) and one external relation*

$$(\text{opera})_1 =: \alpha : \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{X}$$
$$(\text{opera})_2 =: \beta : \mathbb{R} \times \mathbb{X} \longrightarrow \mathbb{X} \quad \text{or } \mathbb{X} \times \mathbb{R} \longrightarrow \mathbb{X}$$
$$(\text{opera})_3 =: \gamma : \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{X}.$$

$\boxed{1}$ *With respect to the internal relation* α *(*"join"*)* $\mathbb{X}$ *as a linear space is a* vector space *over* $\mathbb{R}$*, an* Abelian group *written* "additively" *or* "multiplicatively"*:*

$$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{X}$$

additively written Abelian group	*multiplicatively written Abelian group*
$\alpha(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} + \boldsymbol{y}$	$\alpha(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} \circ \boldsymbol{y}$
$(\boldsymbol{G1}+)$ $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$	$(\boldsymbol{G1}\circ)$ $(\boldsymbol{x} \circ \boldsymbol{y}) \circ \boldsymbol{z} = \boldsymbol{x} \circ (\boldsymbol{y} \circ \boldsymbol{z})$
(*additive associativity*)	(*multiplicative associativity*)

$(\boldsymbol{G2}+)$ $\boldsymbol{x} + \boldsymbol{0} = \boldsymbol{x}$

(*additive identity, neutral element*)

$(\boldsymbol{G2}\circ)$ $\boldsymbol{x} \circ 1 = \boldsymbol{x}$

(*multiplicative identity, neutral element*)

$(\boldsymbol{G3}+)$ $\boldsymbol{x} + (-\boldsymbol{x}) = \boldsymbol{0}$

(*additive inverse*)

$(\boldsymbol{G3}\circ)$ $\boldsymbol{x} \circ \boldsymbol{x}^{-1} = 1$

(*multiplicative inverse*)

$(\boldsymbol{G4}+)$ $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$

(*additive commutativity, Abelian axiom*)

$(\boldsymbol{G4}\circ)$ $\boldsymbol{x} \circ \boldsymbol{y} = \boldsymbol{y} \circ \boldsymbol{x}$

(*multiplicative commutativity, Abelian axiom*).

The triplet of axioms $\{(G1+), (G2+), (G3+)\}$ *or* $\{(G1\circ), (G2\circ), (G3\circ)\}$ *constitutes the set of group axioms.*

$\boxed{2}$ *With respect to the external relation* β *the following compatibility conditions are satisfied:*

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}, r, s \in \mathbb{R}$$

$$\beta(r, \boldsymbol{x}) =: r \times \boldsymbol{x}$$

$(\boldsymbol{D1}+)$
$$r \times (\boldsymbol{x} + \boldsymbol{y}) = (\boldsymbol{x} + \boldsymbol{y}) \times r = r \times \boldsymbol{x} + r \times \boldsymbol{y} = \boldsymbol{x} \times r + \boldsymbol{y} \times r$$
(1*st additive distributivity*)

$(\boldsymbol{D1}\circ)$
$$r \times (\boldsymbol{x} \circ \boldsymbol{y}) = (\boldsymbol{x} \circ \boldsymbol{y}) \times r = (r \times \boldsymbol{x}) \circ \boldsymbol{y} = \boldsymbol{x} \circ (\boldsymbol{y} \circ r)$$
(1*st multiplicative distributivity*)

$(\boldsymbol{D2}+)$
$$(r + s) \times \boldsymbol{x} = \boldsymbol{x} \times (r + s) = r \times \boldsymbol{x} + s \times \boldsymbol{x} = \boldsymbol{x} \times r + \boldsymbol{x} \times s$$
(2*nd additive distributivity*)

$(\boldsymbol{D2}\circ)$
$$(r \circ s) \times \boldsymbol{x} = \boldsymbol{x} \times (r \circ s) = r \times (s \times \boldsymbol{x}) = (\boldsymbol{x} \times r) \times s$$
(2*nd multiplicative distributivity*)

$$(\boldsymbol{D3}) \;\; 1 \times \boldsymbol{x} = \boldsymbol{x} \times 1 = \boldsymbol{x}$$

(*left and right identity*)

$\boxed{3}$ *With respect to the internal relation* γ *(“*meet*”) the following compatibility conditions are satisfied:*

$$\boldsymbol{x}, \boldsymbol{y}, z \in \mathbb{X}, r \in \mathbb{R}$$

$$\gamma(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} * \boldsymbol{y}$$

$$(\boldsymbol{G1}*) \;\; (\boldsymbol{x} * \boldsymbol{y}) * z = \boldsymbol{x} * (\boldsymbol{y} * z)$$

(*associativity w.r.t internal multiplication*)

$(\boldsymbol{D1} * +)$ $\boldsymbol{x} * (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} * \boldsymbol{y} + \boldsymbol{x} * \boldsymbol{z}$
$(\boldsymbol{x} + \boldsymbol{y}) * \boldsymbol{z} = \boldsymbol{x} * \boldsymbol{z} + \boldsymbol{y} * \boldsymbol{z}$
(*left and right additive distributivity w.r.t internal multiplication*)

$(\boldsymbol{D1} * \circ)$ $\boldsymbol{x} * (\boldsymbol{y} \circ \boldsymbol{z}) = (\boldsymbol{x} * \boldsymbol{y}) \circ \boldsymbol{z}$
$(\boldsymbol{x} \circ \boldsymbol{y}) * \boldsymbol{z} = \boldsymbol{x} \circ (\boldsymbol{y} * \boldsymbol{z})$
(*left and right multiplicative distributivity w.r.t internal multiplication*)

$(\boldsymbol{D2} * \times)$ $r \times (\boldsymbol{x} * \boldsymbol{y}) = (r \times \boldsymbol{x}) * \boldsymbol{y}$
$(\boldsymbol{x} * \boldsymbol{y}) \times r = \boldsymbol{x} * (\boldsymbol{y} r)$
(*left and right distributivity of internal and external multiplication*)

Please, in the following pay attention to the 3 axiom sets of the *linear algebra*. We specify now the three axiomatic sets called "*associativity*" (ass), "*unity*" (Uni) and "*commutativity*" (comm).

A-52 The Diagrams "Ass", "Uni" and "Comm"

Conventionally a linear algebra is minimally constituted by the triplet $(\mathbb{X}, \alpha, \beta)$ where $\mathbb{X}$ as a linear space is a vector space equipped with the linear maps $\alpha : \mathbb{X} \times \mathbb{X} \longrightarrow \mathbb{X}$ and $\beta : \mathbb{R} \times \mathbb{X} \longrightarrow \mathbb{X}$ satisfying the axioms (*Ass*) and (*Uni*) according to the following diagrams:

$\boxed{\text{(Ass):}}$

The square

$$\begin{tikzcd}
\mathbb{X}\times\mathbb{X}\times\mathbb{X} \arrow[r, "\alpha\times id"'] \arrow[d, "id\times\alpha"] & \mathbb{X}\times\mathbb{X} \arrow[d, "\alpha"] \\
\mathbb{X}\times\mathbb{X} \arrow[r, "\alpha"] & \mathbb{X}
\end{tikzcd}$$

Commutes.

$\boxed{\text{(Uni):}}$

The diagram

$$\begin{tikzcd}
\mathbb{R}\times\mathbb{X} \arrow[r, "\beta\times id"] \arrow[dr] & \mathbb{X}\times\mathbb{X} \arrow[d] & \mathbb{X}\times\mathbb{R} \arrow[l, "id\times\beta"'] \arrow[dl] \\
& \mathbb{X} &
\end{tikzcd}$$

Commutes.

$\boxed{\text{(Comm):}}$

The triangle

$$\begin{tikzcd}
\mathbb{X}\times\mathbb{X} \arrow[rr, "\tau_{\mathbb{X},\mathbb{X}}"] \arrow[dr, "\alpha"'] & & \mathbb{X}\times\mathbb{X} \arrow[dl, "\alpha"] \\
& \mathbb{X} &
\end{tikzcd}$$

Axiom (Ass) expresses the requirement that the multiplication α is *associative* whereas *Axiom (Uni)* means that the element $\beta(1)$ of $\mathbb{X}$ is a left as well as a right unit for α. The algebra $(\mathbb{X}, \alpha, \beta)$ is commutative if in addition it satisfies the axiom (Comm): commutes where $\tau_{\mathbb{X},\mathbb{X}}$ is the *flip* switching the factors: $\tau_{\mathbb{X},\mathbb{X}}(\boldsymbol{x} \circ \boldsymbol{y}) = \boldsymbol{y} \circ \boldsymbol{x}$. ♣

Indeed we have expressed a set of axioms both explicitly as well as in a diagrammatic approach which minimally constitute a linear algebra $(\mathbb{X}, \alpha, \beta)$. In addition, beside the first internal relation α called "*join*" we have experienced a second internal relation γ called "*meet*" which had to be made compatible with the other relations α and β, respectively. Actually the diagram for the *axiom (Dis)* is left as an *exercise*.
Obviously we have experienced the *words* "*addition*" and "*multiplication*" for various binary operations. Note that in the linear algebra isomorphic to the vector space as its *geometric counterpart* we have *not* specified the inner multiplication $\boldsymbol{\mu}(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{X}$. In a threedimensional *vector space* of *Euclidean* type

$$\gamma(\boldsymbol{x}, \boldsymbol{y}) =: *(\boldsymbol{x} \wedge \boldsymbol{y}) = \boldsymbol{x} \times \boldsymbol{y}$$

namely the star $*$ of the *exterior product* $\boldsymbol{x} \wedge \boldsymbol{y}$ or the "*cross product*" $\boldsymbol{x} \times \boldsymbol{y}$, for $\boldsymbol{x} \in \mathbb{R}^3$, $\boldsymbol{y} \in \mathbb{R}^3$ is an *example*. Sometimes

$$\gamma(\boldsymbol{x}, \boldsymbol{y}) =: [\boldsymbol{x}, \boldsymbol{y}]$$

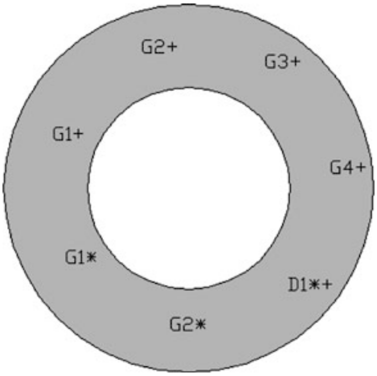
is written by rectangular brackets.

Historical Aside

Following a proposal of *L. Kronecker* ("Über die algebraisch auflösbaren Gleichungen (Abhandlung, 1929) the *axiom of commutativity* $(\boldsymbol{G}4+)$ or $(\boldsymbol{G}4\circ)$ is called after *N. H. Abel* (Memoire sur un classe particulière d'equations résolable algébraique, Crelle's J. reine angewandte Mathematik 4 (1828) 131–156 *Oeuvres* vol. 1, pp. 478–514, vol. 2, pp. 217–243, 329–331 edited by *S. Lie* and *L. Sylow*, Christiana 1881) who dealt with a particular class of equations of all degrees which are solvable by radicals, e.g. the cyclotomic equation $x^n - 1 = 0$. *N. H. Abel* has proved the following general theorem: If the roots of an equation are such that all roots can be expressed as rational functions of one of them, say x, and if any two of the roots, say $r_1 x$ and $r_2 x$ *where* r_1 *and* r_2 are *rational functions* are connected in such a way that $r_2 r_1 x = r_1 r_2 x$, then the equation can be solved by radicals. Refer $r_2 r_1 x = r_1 r_2 x$ to $(\boldsymbol{G}1)$.

A-53 Ringed Spaces: The Subalgebra "Ring with Identity"

In ($\boldsymbol{G}2\circ$) the *neutral element* 1 as well as in ($\boldsymbol{G}3\circ$) the *inverse element* has been multiplied from the *right*. Similarly *left multiplication* ($\boldsymbol{G}2\circ$) by the neutral element 1 as well as ($\boldsymbol{G}3\circ$) by the inverse element are defined. Indeed it can be shown that there exist exactly one neutral element which is both *left-neutral* and *right-neutral* as well as exactly one inverse element which is both *left-inverse* and *right-inverse*. A subalgebra is called a "*ring with identity*" if the following *seven* conditions hold:



A ring with identity ($\boldsymbol{G}3*$) is a *division ring* if every nonzero element of the ring has a *multiplicative inverse*. A *commutative ring* is a ring with *commutative multiplication* ($\boldsymbol{G}4*$). *Modules* are generalizations of the vector spaces of linear algebra in which the "scalars" are allowed to be from *an arbitrary ring*, rather than a field of real numbers. They will be discussed as soon as we introduce *superalgebras*. Now we take reference to

Lemma A.53. (anticommutativity): $\boldsymbol{x} \circ \boldsymbol{x} = \boldsymbol{0}$ for all $\boldsymbol{x} \in \mathbb{X} \Longleftrightarrow \boldsymbol{x} \circ \boldsymbol{y} = -\boldsymbol{y} \circ \boldsymbol{x}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{X}$. "$\circ$" is used in the notation "$\wedge$" accordingly.

Proof.

$$\text{“}\Longrightarrow\text{”}\ \boldsymbol{x} \circ \boldsymbol{y} + \boldsymbol{y} \circ \boldsymbol{x} = \boldsymbol{x} \circ \boldsymbol{x} + \boldsymbol{x} \circ \boldsymbol{y} + \boldsymbol{y} \circ \boldsymbol{x} + \boldsymbol{y} \circ \boldsymbol{y}$$

$$= \boldsymbol{x} \circ (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{y} \circ (\boldsymbol{x} + \boldsymbol{y}) = (\boldsymbol{x} + \boldsymbol{y}) \circ (\boldsymbol{x} + \boldsymbol{y}) = 0$$

$$\text{“}\Longleftarrow\text{”}\ \boldsymbol{x} = \boldsymbol{y} \Longrightarrow \boldsymbol{x} \circ \boldsymbol{x} = -\boldsymbol{x} \circ \boldsymbol{x} \Longrightarrow \boldsymbol{x} \circ \boldsymbol{x} = \boldsymbol{0}.$$

Later on we refer to the following algebras. ♣

A-54 Definition of a Division Algebra and Non-Associative Algebra

Definition A.23. (division algebra):

A $\mathbb{R}$-algebra is called *division algebra* over $\mathbb{R}$, if all non-null elements of $\mathbb{X}' := \mathbb{X} \setminus \{0\}$ form additionally a *group with respect to inner multiplication* $\mu =: \boldsymbol{x} \circ \boldsymbol{y}$, namely

$$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{X}\backslash\{0\}$$

$$(\boldsymbol{G}1\circ)\ (\boldsymbol{x} \circ \boldsymbol{y}) \circ \boldsymbol{z} = \boldsymbol{x} \circ (\boldsymbol{y} \circ \boldsymbol{z})$$
(associativity of inner multiplication)

$$(\boldsymbol{G}2\circ)\ \boldsymbol{x} \circ 1 = \boldsymbol{x}$$
(identity of inner multiplication)

$$(\boldsymbol{G}3\circ)\ \boldsymbol{x} \circ \boldsymbol{x}^{-1} = 1$$
(inverse of inner multiplication)

Definition A.24. (non-associative algebra):

A weaking of a $\mathbb{R}$-algebra is the *non-associative algebra* over $\mathbb{R}$, if the axioms $\boxed{1}$, $\boxed{2}$ and $\boxed{3}$ of linear algebra according to Definition A.21 hold *with the* exception of $(\boldsymbol{G}1\circ)$, that is the *associativity* of inner multiplication is cancelled.

A-55 Lie Algebra, Witt Algebra

"*perhaps god is a lie group*"

Many physicists believe that all modern physics is based on the operation called "*Lie algebra*". Indeed more than 1000 textbooks and papers are written on this most important subject. Indeed many physicists believe in "*perhaps god is a lie group*". Here, we can give a very short notice of "*Lie algebra*". We skip a note of comparison "*Lie algebra*" and its brother "*Killing analysis*".

Definition A.25. (*Lie* algebra):

A non-associative algebra is called Lie algebra *over* $\mathbb{R}$*, if the following operations with respect to inner multiplication* $\mu := \boldsymbol{x} \circ \boldsymbol{y}$ *hold:*

$$(\boldsymbol{L}1)\ \boldsymbol{x} \circ \boldsymbol{x} = 0$$

$$(\boldsymbol{L}2)\ (\boldsymbol{x} \circ \boldsymbol{y}) \circ \boldsymbol{z} + (\boldsymbol{y} \circ \boldsymbol{z}) \circ \boldsymbol{x} + (\boldsymbol{z} \circ \boldsymbol{x}) \circ \boldsymbol{y} = \mathbf{0}.$$
$$(\textit{Jacobi identity})$$

The examples of the *Lie algebra* are numerous. As a special *Lie algebra* we present the *Witt algebra* which is applied to *Laurent polynomials*.

Example A.8. (Witt algebra on the ring of Laurent polynomials (Chen, 1995)):

The *Witt algebra* W is the complex Lie algebra of polynomial fields on the unit circle $\mathbb{S}^1$. An element of W is a linear combination of the elements of the form $e^{in\phi}\frac{\partial}{\partial\phi}$, where ϕ is a real parameter, and the *Lie bracket* of W is given by

$$\left[e^{im\phi}\frac{\partial}{\partial\phi}, e^{in\phi}\frac{\partial}{\partial\phi}\right] = i(n-m)e^{i(m+n)}\frac{\partial}{\partial\phi}.$$

A-56 Definition of a Composition Algebra

Various algebras are generated by adding an additional structure to the minimal set of axioms of a linear algebra blocked by (a), (b) and (c). Later on we use such an additional structure for *matrix algebra*.

Definition A.26. (composition algebra):

A *non-associative algebra* with 1 as identity of inner multiplication is called *composition algebra* over $\mathbb{R}$, if there exists a regular quadratic form $Q : X \longrightarrow \mathbb{R}$ which is compatible with the corresponding operations that is the following operations hold:

(K1) $Q : X \longrightarrow \mathbb{R}$ *is a regular quadratic form*,

$$x, y, z \in \mathbb{R}, \quad r \in \mathbb{R}$$

$$Q(r \times x) = r^2 \times Q(\boldsymbol{x}) \ (\textit{quadratic form})$$

$$Q(\boldsymbol{x} + y + z) = Q(\boldsymbol{x} + y) + Q(\boldsymbol{x} + z) + Q(y + z) - Q(\boldsymbol{x}) - Q(y) - Q(z)$$

$$Q(r \times \boldsymbol{x} + y) - r \times Q(\boldsymbol{x} + y) = (r - 1) \times [r \times Q(\boldsymbol{x}) - Q(y)]$$

$$Q(\boldsymbol{x}) = 0 \Longleftrightarrow \boldsymbol{x} = 0 \ (\textit{regularity})$$

(K2) $Q(\boldsymbol{x} \wedge y) = Q(\boldsymbol{x}) \times Q(y)$ (*multiplicativity*)

$$Q(\mathbf{1}) = 1$$

The quadratic form introduced by Definition A.55 leads to the topological notion of scalar products, norm and metric we already used:

Lemma A.56. (scalar product, norm, metric):

In a composition algebra with a positive-definite quadratic form a *scalar product* ("inner product") is defined by the bilinear form $\langle \cdot | \cdot \rangle : X \times \mathbb{X}^* \longrightarrow \mathbb{R}$ with

$$\langle \boldsymbol{x} | \boldsymbol{y} \rangle := \frac{1}{2}[Q(\boldsymbol{x} + \boldsymbol{y}) - Q(\boldsymbol{x}) - Q(\boldsymbol{y})];$$

a *norm* is defined by $\| \cdot \| : X \longrightarrow \mathbb{R}$ with

$$\|\boldsymbol{x}\| := +[Q(\boldsymbol{x})]^{1/2}$$

and *metric* is defined by the bilinear form

$$\rho(\boldsymbol{x}, \boldsymbol{y}) := +[Q(\boldsymbol{x} - \boldsymbol{y})]^{1/2}.$$

Thus to the *algebraic structure* a *topological structure* is added, if in addition

$$(K3) \qquad Q(\boldsymbol{x}) \geq 0$$

for all $x \in X$ holds.

Proof.

$\boxed{\text{(i) scalar product}}$

$\langle \cdot | \cdot \rangle : X \times X \longrightarrow \mathbb{R}$ *is a scalar product since*

$$\begin{aligned}
(1) \qquad \langle \boldsymbol{x} | \boldsymbol{y} \rangle &= \frac{1}{2}[Q(\boldsymbol{x} + \boldsymbol{y}) - Q(\boldsymbol{x}) - Q(\boldsymbol{y})] \\
&= \frac{1}{2}[Q(\boldsymbol{y} + \boldsymbol{x}) - Q(\boldsymbol{y}) - Q(\boldsymbol{x})] = \langle \boldsymbol{y} | \boldsymbol{x} \rangle \ (\textit{symmetry}) \\
(2)\ \langle \boldsymbol{x} + \boldsymbol{y} | \boldsymbol{z} \rangle &= \frac{1}{2}[Q(\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}) - Q(\boldsymbol{x} + \boldsymbol{y}) - Q(\boldsymbol{z})] \\
&= \frac{1}{2}[Q(\boldsymbol{x} + \boldsymbol{z}) - Q(\boldsymbol{x}) - Q(\boldsymbol{z})] \\
&\quad + \frac{1}{2}[Q(\boldsymbol{y} + \boldsymbol{z}) - Q(\boldsymbol{y}) - Q(\boldsymbol{z})] \\
&= \langle \boldsymbol{x} | \boldsymbol{z} \rangle + \langle \boldsymbol{y} | \boldsymbol{z} \rangle \qquad (\textit{additivity})
\end{aligned}$$

$$(3)\ \langle r\boldsymbol{x}|\boldsymbol{y}\rangle = \frac{1}{2}[Q(r\boldsymbol{x}+\boldsymbol{y}) - Q(r\boldsymbol{x}) - Q(\boldsymbol{y})]$$
$$= \frac{1}{2}r[Q(\boldsymbol{x}+\boldsymbol{y}) - Q(\boldsymbol{x}) - Q(\boldsymbol{y})]$$
$$= r\cdot\langle\boldsymbol{x}|\boldsymbol{y}\rangle \qquad (homogenity)$$
$$(4)\quad \langle\boldsymbol{x}|\boldsymbol{x}\rangle = \frac{1}{2}[Q(\boldsymbol{x}+\boldsymbol{x}) - Q(\boldsymbol{x}) - Q(\boldsymbol{x})]$$
$$= \frac{1}{2}[Q(2\boldsymbol{x}) - 2\cdot Q(\boldsymbol{x})] = Q(\boldsymbol{x}) \geq 0 \ (positivity)$$

$\boxed{\text{(ii) norm}}$

$\|\cdot\| : X \longrightarrow \mathbb{R}$ *is a norm since*

$(\boldsymbol{N}1)\qquad \|\boldsymbol{x}\| = +[Q(\boldsymbol{x})]^{1/2} \geq 0$

(positivity) *and*

$$\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0}$$

$(\boldsymbol{N}2)\qquad \|r\boldsymbol{x}\| = +[Q(r\boldsymbol{x})]^{1/2} = +[r^2\cdot Q(\boldsymbol{x})]^{1/2} = |r|\times\|\boldsymbol{x}\|$
(homogenity)

$(\boldsymbol{N}3)\ \|\boldsymbol{x}+\boldsymbol{y}\| = +[Q(\boldsymbol{x}+\boldsymbol{y})]^{1/2} = +[Q(\boldsymbol{x}) + Q(\boldsymbol{y}) + 2\langle\boldsymbol{x}|\boldsymbol{y}\rangle]^{1/2}$

$$= +\Big[\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|\Big]^{1/2} \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|.$$

(*Cauchy-Schwarz' inequality*)*"triangle inequality"*)

$\boxed{\text{(iii) metric}}$

$\rho : X\times X \longrightarrow \mathbb{R}$ *is a metric since*

$(\boldsymbol{M}1)\ \rho(\boldsymbol{x},\boldsymbol{y}) = +[Q(\boldsymbol{x}-\boldsymbol{y})]^{1/2} = \|\boldsymbol{x}-\boldsymbol{y}\| \geq 0 \qquad (positivity)$

and

$$\rho(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0} \iff \boldsymbol{x}-\boldsymbol{y} = \boldsymbol{0} \iff \boldsymbol{x} = \boldsymbol{y}$$

$(\boldsymbol{M}2)\ \rho(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\| = |-1|\cdot\|\boldsymbol{y}-\boldsymbol{x}\| = \rho(\boldsymbol{y},\boldsymbol{x}) \quad (symmetry)$

$(\boldsymbol{M}3)\ \rho(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\| = \|(\boldsymbol{x}-z) + (z-\boldsymbol{y})\|$

$$\leq \|\boldsymbol{x}-z\| + \|z-\boldsymbol{x}\| = \rho(\boldsymbol{x},z) + \rho(z,\boldsymbol{y})$$

(triangle inequality) ♣

We treat in the next section a special case of a *division algebra*, namely *matrix algebra*.

A-6 Matrix Algebra Revisited, Generalized Inverses

It is known that every nonsingular matrix $\mathbf{A}$ has a unique inverse, usually denoted by $\mathbf{A}^{-1}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}A = \mathbf{I}$, where $\mathbf{I}$ is the unit matrix. A matrix has a inverse only if it is square, and even then only if it is nonsingular, or alternatively if its columns or rows are linearly independent. Over the many years past it was felt the need of some kind of *partial inverse* of a matrix that is *singular or even rectangularsingular matrix*. By the term generalized inverse of a given matrix $\mathbf{A}$ we shall mean a matrix $\mathbf{X}$ associated in some way with the matrix $\mathbf{A}$ that (a) exists for a class of matrices larger than the class of nonsingular matrices, (b) has some of the properties of the usual inverse, and (c) reduces to the usual inverse when $\mathbf{A}$ is non singular. Some others have used the term "*pseudoinvese*" rather than general inverse without specifying the chosen type of general inverse.
For a given matrix $\mathbf{A}$, the matrix equation $\mathbf{AXA} = \mathbf{A}$ alone characterizes those generalized inverses $\mathbf{X}$ that are of use in analyzing the solutions of the linear system $\mathbf{Ax} = \mathbf{b}$. For other purposes, other relationships play an essential role. Thus if we are concerned with least squares properties, the equation $\mathbf{AXA} = \mathbf{A}$ is not enough and must be supplemented by other relations. There results a more restricted class of generalized inverses. Here we are limited to generalized inverses *finite matrices*, but extension to infinite-dimensional space and to differential and integral operations are very well known.
E.H. Moore is attributed to have written one of the first papers on the topic of generalized inverses called by him the "*general reciprocal*" of any finite matrix,square or rectangular. His first publication on the subject was an abstract talk given at a meeting of the *American Mathematical Society* which appeared in 1920. Details of his talk were published only in 1935 (Moore, 1920), Moore (1935); (Moore and Nashed, 1974) after Moore's death! Little note was taken of Moore's discovery for 30 years after his first publication, during which time generalized inverses were taken for matrices by *C. L. Siegel* and for operators by Tseng (1936), Tseng (1949a)–Tseng (1949c), Tseng (1956), Murray and von Neumann (1936) and many others. A surprising revival of interest in general inverses appeared in 1950s when least squares properties were analyzed. The important properties were realized by Bjerhammar (1951a), Bjerhammar (1951b), (1968), one swedish colleague teaching LESS and geodetic applications of the subject. In 1955 *R. Penrose* extended *A. Bjerhammar's* results on linear systems and showed that *E.H. Moore* inverse satisfied four equations of type (a) $\mathbf{AXA} = \mathbf{A}$, (b)$\mathbf{XAX} = \mathbf{X}$, (c) $\mathbf{AX}^{'} = \mathbf{AX}$, and (d) $\mathbf{XA}^{'} = \mathbf{XA}$ nowadays called the axioms of the *Moore-Penrose inverse*, often denoted by $\mathbf{A}^{+}$.
There are excellent textbooks on matrix algebra and generalized inverses of matrices including many examples upto 500! We like to mention the excellent text

of Ben-Israel and Greville (1974), Bjerhammar (1973), Gere and Weaver (1965), Graybill (1963), Mardia et al. (1979), Nashed (1974), Rao and Mitra (1971), Rao and Rao (1998), Scarle (1982), Styan (1983). Here we will be unable to compete with the dense information given by these special texts.

We assume that the reader is familiar with the standard notion of a matrix as a rectangular array of numbers offered by any undergraduate text of applied mathematics. Familiarity of various notion, for instance the dimension of a matrix of type $n \times m$, multiplication and addition of two matrices of type (a) "*Cayley*" or simply the matrix product $\mathbf{C} := \mathbf{A}.\mathbf{B}$, (b)"*Kronecker-Zehfuss*" denoted by $\mathbf{C} := \mathbf{B} \otimes \mathbf{A}$, (c) "*Katri-Rao*" denoted by $\mathbf{C} := \mathbf{B} \odot \mathbf{A}$, and (d)"*Hadamard*" attributed by $\mathbf{K} := \mathbf{G} * \mathbf{H}$.

How are the various matrix product defined? Who defined the various algebras?

Matrix algebra is the special division algebra over the field of real numbers. There is a natural generalization in terms of complex numbers for instance.

`The three axioms`

$$\textit{Let } \mathbf{A} = [a_{ij}] \in \mathbb{R}^{n\times m} \textit{ be a rectangular matrix}$$

$\boxed{1}$ $\mathbf{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{nm}$, `first set of axiom` $\alpha(\mathbf{A}, \boldsymbol{B}) =: \mathbf{A} + \boldsymbol{B}$

$$\begin{aligned}
&(\boldsymbol{G}1+)\ (\mathbf{A} + \boldsymbol{B}) + \boldsymbol{C} &&= \mathbf{A} + (\boldsymbol{B} + \boldsymbol{C})\\
&(\boldsymbol{G}2+)\qquad \mathbf{A} + 0 &&= \mathbf{A}\\
&(\boldsymbol{G}3+)\qquad \mathbf{A} - \mathbf{A} &&= 0\\
&(\boldsymbol{G}4+)\qquad \mathbf{A} + \boldsymbol{B} &&= \boldsymbol{B} + \mathbf{A}
\end{aligned}$$

$\boxed{2}$ $\mathbf{A}, \boldsymbol{B} \in \mathbb{R}^{nm}$, `second set of axiom` $r, s \in \mathbb{R}$, $\beta(r, \mathbf{A}) =: r \times \mathbf{A}$

$$\begin{aligned}
&(\boldsymbol{D}1+)\ r \times (\mathbf{A} + \boldsymbol{B}) &&= r \times \mathbf{A} + r \times \boldsymbol{B}\\
&(\boldsymbol{D}2+)\ \ (r + s) \times \mathbf{A} &&= r \times \mathbf{A} + s\mathbf{A}\\
&(\boldsymbol{D}3)\qquad 1 \times \mathbf{A} &&= \mathbf{A}
\end{aligned}$$

$\boxed{3}$ "multiplication of matrices"

(a) "*Cayley-product*" (just "the matrix product")

$$\left.\begin{aligned}
\mathbf{A} &= [a_{ij}] \in \mathbb{R}^{n\times l}, \ \dim \mathbf{A} = n \times l\\
\boldsymbol{B} &= [b_{ij}] \in \mathbb{R}^{l\times m}, \ \dim \boldsymbol{B} = l \times m
\end{aligned}\right] \Longrightarrow \boldsymbol{C} := \mathbf{A} \cdot \boldsymbol{B} = [c_{ij}] \in \mathbb{R}^{nm},$$

$$\dim \boldsymbol{C} = n \times m \qquad c_{ij} := \sum_{k=1}^{l} a_{ik} b_{kl}.$$

The product was introduced by Cayley (1857); see also his Collected Works, vol. 2, 475–496. A historical perspective is given in Feldmann (1962).

(b) "*Kronecker-Zehfuß-product*"

$$\left.\begin{aligned} \mathbf{A} &= [a_{ij}] \in \mathbb{R}^{n\times m}, \quad \dim \mathbf{A} = n \times m \\ \boldsymbol{B} &= [b_{ij}] \in \mathbb{R}^{k\times l} \quad\ \dim \boldsymbol{B} = k \times l \end{aligned}\right]$$
$$\implies \boldsymbol{C} := \boldsymbol{B} \otimes \mathbf{A} = [c_{ij}] \in \mathbb{R}^{kn\times lm},$$
$$\dim \boldsymbol{C} = kn \times lm, \quad \boldsymbol{B} \otimes \mathbf{A} := [b_{ij}\mathbf{A}].$$

The product was early referenced to *L. Kronecker* by Mac Duffee (1946). The other reference is Zehfuss (1858). See also Henderson et al. (1983) and Horn and Johnson (1990). Reference is also made to Steeb (1991) and Graham (1981).

(c) "*Khatri-Rao-product*" (two rectangular matrices of identical column number)

$$\left.\begin{aligned} \mathbf{A} &= [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m] \in \mathbb{R}^{n\times m}, \quad \dim \mathbf{A} = n \times m \\ \boldsymbol{B} &= [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m] \in \mathbb{R}^{k\times m}, \quad \dim \boldsymbol{B} = k \times m \end{aligned}\right]$$
$$\implies \boldsymbol{C} := \boldsymbol{B} \odot \mathbf{A} := [\boldsymbol{b}_1 \otimes \boldsymbol{a}_1, \ldots, \boldsymbol{b}_m \otimes \boldsymbol{a}_m] \in \mathbb{R}^{kn\times m}$$
$$\dim \boldsymbol{C} = kn \times m.$$

"*two rectangular matrices of identical column numbers are multiplied.*"
The product was introduced by Khatri and Rao (1968).

(d) "*Hadamard product*" (two rectangular matrices of the same dimension, elementwise product)

$$\left.\begin{aligned} \boldsymbol{G} &= [g_{ij}] \in \mathbb{R}^{n\times m}, \ \dim \boldsymbol{G} = n \times m \\ \boldsymbol{H} &= [h_{ij}] \in \mathbb{R}^{m\times m}, \dim \boldsymbol{H} = n \times m \end{aligned}\right]$$
$$\implies \boldsymbol{K} := \boldsymbol{G} * \boldsymbol{H} = [k_{ij}] \in \mathbb{R}^{n\times m},$$
$$\dim \boldsymbol{K} = n \times m, \quad k_{ij} := g_{ij} h_{ij} \qquad (no \;\; summation).$$

"*two rectangular matrices of the same dimension define the elementwise product.*"
The product was introduced by Hadamard (1899). See also Moutard (1894), as well as Hadamard (1949) and Schur (1911) and Horn and Johnson (1991) and Styan (1973).

$\boxed{\text{Example:}}$

(a)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2\times 3}, \quad \boldsymbol{B} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix} \in \mathbb{Z}^{3\times 2} \Longrightarrow$$

(b)

$$\Longrightarrow \mathbf{A} \cdot \boldsymbol{B} = \begin{bmatrix} 28 & 34 \\ 64 & 79 \end{bmatrix} \in \mathbb{Z}^{2\times 2} \qquad (\textit{"integer numbers"}).$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2\times 3}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \in \mathbb{Z}^{3\times 2} \Longrightarrow \boldsymbol{B} \otimes \mathbf{A} = [b_{ij}\mathbf{A}]$$

$$= \left[\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \otimes \mathbf{A}\right] = \begin{bmatrix} 1 & 2 & 3 & 2 & 4 & 6 \\ 4 & 5 & 6 & 8 & 10 & 12 \\ 3 & 6 & 8 & 4 & 8 & 12 \\ 12 & 15 & 18 & 16 & 20 & 24 \\ 5 & 10 & 15 & 6 & 12 & 18 \\ 20 & 25 & 30 & 24 & 30 & 36 \end{bmatrix} \in \mathbb{Z}^{6\times 6}$$

(c)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2\times 3}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \in \mathbb{Z}^{3\times 3}$$

$$\Longrightarrow \boldsymbol{B} \odot \mathbf{A} = \left[\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} \otimes \begin{bmatrix} 2 \\ \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} \otimes \begin{bmatrix} 3 \\ \\ 6 \end{bmatrix}\right]$$

$$= \begin{bmatrix} 1 & 4 & 9 \\ 4 & 10 & 18 \\ 4 & 10 & 18 \\ 16 & 25 & 36 \\ 7 & 16 & 27 \\ 23 & 40 & 54 \end{bmatrix} \in \mathbb{Z}^{6\times 3}$$

(d)

$$\boldsymbol{G} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{Z}^{2\times 3}, \quad \boldsymbol{H} = \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix} \in \mathbb{Z}^{2\times 3}$$

$$\Longrightarrow \boldsymbol{G} * \boldsymbol{H} = [g_{ij}h_{ij}] = \begin{bmatrix} 2 & 6 & 12 \\ 20 & 30 & 42 \end{bmatrix} \in \mathbb{Z}^{2\times 3}.$$

A-61 Special Matrices: Helmert, Hankel, and Vandemonte

All the quoted textbooks review effectively special matrices of various types: symmetric, antisymmetric, diagonal, unity, zero, triangular, idempotent, normal, orthogonal, orthonormal, positive-definite, positive-semidefinite, permutation, and commutation matrices.

Here, we review only *orthonormal matrices* of various representations. The *Helmert representation* of orthonomal matrix is given in various forms. In addition, we give examples for an *orthogonal matrix*, for instance the *Hankel matrix* of power sums and the *Vandermonde matrix*.

Definition (orthogonal matrix) :

The matrix $\mathbf{A}$ is called orthogonal if $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are diagonal matrices. (The rows and columns of $\mathbf{A}$ are *orthogonal*.)

Definition (orthonormal matrix): The matrix $\mathbf{A}$ is called orthonormal if $\mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$. (The rows and columns of $\mathbf{A}$ are *orthonormal*.)

Facts (representation of a 22 orthonormal matrix) $\mathbf{X} \in \mathrm{SO}(2)$: A 2×2 orthonormal matrix $\mathbf{X} \in \mathrm{SO}(2)$ is an element of the special orthogonal group SO(2) defined by

$$\mathrm{SO}(2) := \{\mathbf{X} \in \mathbb{R}^{2\times 2} \,|\, \mathbf{X}'\mathbf{X} = \mathbf{I}_2 \text{ and } \det \mathbf{X} = +1\}$$

$$\left\{ \mathbf{X} = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \in \mathbb{R}^{2\times 2} \,\middle|\, \begin{matrix} x_1^2 + x_2^2 = 1 \\ x_1x_3 + x_2x_4 \\ x_3^2 + x_4^2 = 1 \end{matrix} = 0, x_1x_4 - x_2x_3 = +1 \right\}$$

(a)

$$\mathbf{X} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad \phi \in [0, 2\pi]$$

is a trigonometric representation of $\mathbf{X} \in \mathrm{SO}(2)$.

(b)

$$\mathbf{X} = \begin{bmatrix} x & \sqrt{1-x^2} \\ -\sqrt{1-x^2} & x \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad x \in [+1, -1]$$

is an algebraic representation of $\mathbf{X} \in SO(2)$

$$\left(x_{11}^2 + x_{12}^2 = 1,\ x_{11}x_{21} + x_{12}x_{22} = \ -x\sqrt{1-x^2} + x\sqrt{1-x^2} = 0,\ x_{21}^2 + x_{22}^2 = 1\right).$$

(c)

$$\mathbf{X} = \begin{bmatrix} \frac{1-x^2}{1+x^2} & +\frac{2x}{1+x^2} \\ -\frac{2x}{1+x^2} & \frac{1-x^2}{1+x^2} \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad x \in \mathbb{R}$$

is called a stereographic projection of $\mathbf{X}$ (stereographic projection of SO(2) $\mathbb{S}^1$ onto $\mathbb{L}^1$).
(d)

$$\mathbf{X} = (\mathbf{I}_2 + \mathbf{S})(\mathbf{I}_2 - \mathbf{S})^{-1}, \quad \mathbf{S} = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix},$$

where $\mathbf{S} = -\mathbf{S}'$ is a *skew matrix* (antisymmetric matrix), is called a *Cayley-Lipschitz representation* of $\mathbf{X} \in SO(2)$. (e) $\mathbf{X} \in SO(2)$ is a commutative group ("*Abel*") (*Example*: $\mathbf{X}_1 \in SO(2)$, $\mathbf{X}_2 \in SO(2)$, then $\mathbf{X}_1\mathbf{X}_2 = \mathbf{X}_2\mathbf{X}_1$) (SO($n$) for $n = 2$ is the only commutative group, SO($n | n \neq 2$) is not "*Abel*").

Facts (representation of an $n \times n$ orthonormal matrix) $\mathbf{X} \in SO(n)$:

An $n \times n$ orthonormal matrix $\mathbf{X} \in SO(n)$ is an element of the special orthogonal group SO(n) defined by

$$SO(n) := \{\mathbf{X} \in \mathbb{R}^{n\times n} | \mathbf{X}'\mathbf{X} = \mathbf{I}_n \text{ and} \det \mathbf{X} = +1\}.$$

As a differentiable manifold SO(n) inherits a *Riemann structure* from the *ambient space* $^{n^2}$ with a *Euclidean metric* (*vec* $\mathbf{X}' \in \mathbb{R}^{n^2}$, dim *vec* $\mathbf{X}' = n^2$). Any *atlas* of the special orthogonal group SO(n) has at least *four distinct charts* and there is one with exactly four charts. ("minimal atlas": *Lusternik-Schnirelmann category*) (a)

$$\mathbf{X} = (\mathbf{I}_n + \mathbf{S})(\mathbf{I}_n - \mathbf{S})^{-1},$$

where

$$\mathbf{S} = -\mathbf{S}'$$

is a *skew matrix* (antisymmetric matrix), is called a *Cayley-Lipschitz representation* of $\mathbf{X} \in \mathrm{SO}(n)$. ($n!/2(n-2)!$ is the number of independent parameters/coordinates of $\mathbf{X}$) (b) If each of the matrices $\mathbf{R}_1, \ldots, \mathbf{R}_k$ is an $n \times n$ *orthonormal matrix*, then their product

$$\mathbf{R}_1\mathbf{R}_2 \cdots \mathbf{R}_{k-1}\mathbf{R}_k \in \mathrm{SO}(n)$$

is an $n \times n$ *orthonormal matrix*.

Facts (orthonormal matrix: Helmert representation):

Let $\mathbf{a}' = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ represent any row vector such that $\mathbf{a}_i \neq \mathbf{0}$ $(i \in \{1, \ldots, n\})$ is any row vector whose elements are all nonzero. Suppose that we require an $n \times n$ *orthonormal matrix*, one row which is proportional to $\mathbf{a}'$. In what follows *one such matrix* $\mathbf{R}$ *is derived*. Let $[\mathbf{r}'_1, \ldots, \mathbf{r}'_n]$ represent the *rows* of $\mathbf{R}$ and take the *first row* $\mathbf{r}'_1$ to be the row of $\mathbf{R}$ that is proportional to $\mathbf{a}'$. Take the *second row* $\mathbf{r}'_2$ to be proportional to the n-dimensional row vector

$$[\mathbf{a}_1, -\mathbf{a}_1^2/\mathbf{a}_2, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}], \qquad (H2)$$

the *third row* $\mathbf{r}'_3$ proportional to

$$\left[\mathbf{a}_1, \mathbf{a}_2, -(\mathbf{a}_1^2 + \mathbf{a}_2^2)/\mathbf{a}_3, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}\right] \qquad (H3)$$

and more generally the first through nth rows $\mathbf{r}'_1, \ldots, \mathbf{r}'_n$ proportional to

$$\left[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{k-1}, -\sum_{i=1}^{k-1} \mathbf{a}_i^2/\mathbf{a}_k, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}\right] \qquad (Hn-1)$$

$$\text{for} \quad k \in \{2, \ldots, n\},$$

respectively confirm to yourself that the n-1 vectors $(\mathbf{H}_{n-1})$ are orthogonal to each other and to the vector $\mathbf{a}'$. In order to obtain explicit expressions for $\mathbf{r}'_1, \ldots, \mathbf{r}'_n$ it remains to *normalize* $\mathbf{a}'$ and the vectors $(\mathbf{H}_{n-1})$. The Euclidean norm of the kth of the vectors $(\mathbf{H}_{n-1})$ is

$$\left\{\sum_{i=1}^{k-1} \mathbf{a}_i^2 + \left(\sum_{i=1}^{k-1} \mathbf{a}_i^2\right)^2 \Big/ \mathbf{a}_k^2\right\}^{1/2} = \left\{\left(\sum_{i=1}^{k-1} \mathbf{a}_i^2\right)\left(\sum_{i=1}^{k} \mathbf{a}_i^2\right) \Big/ \mathbf{a}_k^2\right\}^{1/2}.$$

Accordingly for the *orthonormal vectors* $\mathbf{r}'_1, \ldots, \mathbf{r}'_n$ we finally find

$$\texttt{(1st row)} \qquad \mathbf{r}'_1 = \left[\sum_{i=1}^{n} \mathbf{a}_i^2\right]^{-1/2} (\mathbf{a}_1, \ldots, \mathbf{a}_n)$$

$$\text{(kth row)} \qquad \mathbf{r}'_k = \left[\frac{\mathbf{a}_k^2}{\left(\sum_{i=1}^{k-1}\mathbf{a}_i^2\right)\left(\sum_{i=1}^{k}\mathbf{a}_i^2\right)}\right]^{-1/2} \times \left(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{k-1}, -\sum_{i=1}^{k-1}\frac{\mathbf{a}_i^2}{\mathbf{a}_k}, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}\right)$$

$$(nth\ \ row) \qquad \mathbf{r}'_n = \left[\frac{\mathbf{a}_n^2}{\left(\sum_{i=1}^{n-1}\mathbf{a}_i^2\right)\left(\sum_{i=1}^{n}\mathbf{a}_i^2\right).}\right]^{-1/2} \left[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{n-1}, -\sum_{i=1}^{n-1}\frac{\mathbf{a}_i^2}{\mathbf{a}_n}\right].$$

The recipe is complicated: When $\mathbf{a}' = [1,\ 1,\ \ldots,\ 1,\ 1]$, the *Helmert factors* in the 1st row, ..., kth row,..., nth row simplify to

$$\mathbf{r}'_1 = n^{-1/2}[1, 1, \ldots, 1, 1] \in^n$$
$$\mathbf{r}'_k = [k(k-1)]^{-1/2}[1, 1, \ldots, 1, 1-k, 0, 0, \ldots, 0, 0] \in^n$$
$$\mathbf{r}'_n = [n(n-1)]^{-1/2}[1,\ 1,\ \ldots,\ 1,\ 1-n] \in \mathbb{R}^n.$$

The orthonormal matrix

$$\begin{bmatrix} \mathbf{r}'_1 \\ \mathbf{r}'_2 \\ \cdots \\ \mathbf{r}'_{k-1} \\ \mathbf{r}'_k \\ \cdots \\ \mathbf{r}'_{n-1} \\ \mathbf{r}'_n \end{bmatrix} \in SO(n)$$

is known as the *Helmert matrix* of order n. (Alternatively the transposes of such a matrix are called the *Helmert matrix*.)

$\boxed{\textit{Example}\text{ (Helmert matrix of order 3):}}$

$$\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{bmatrix} \in SO(3).$$

Check that the rows are orthogonal and normalized.

Example (Helmert matrix of order 4):

$$\begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix} \in SO(4).$$

Check that the rows are orthogonal and normalized.

Example (Helmert matrix of order n):

$$\begin{bmatrix} 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} & 1/\sqrt{n} \cdots & 1/\sqrt{n} & 1/\sqrt{n} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \;\; \cdots & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \;\; \cdots & 0 & 0 \\ \cdots & & & \cdots & & \\ \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \cdots \;\; \cdots & \frac{1-(n-1)}{\sqrt{(n-1)(n-2)}} & 0 \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots \;\; \cdots & \frac{1}{\sqrt{n(n-1)}} & \frac{1-n}{\sqrt{n(n-1)}} \end{bmatrix} \in \mathrm{SO}(n).$$

Check that the rows are orthogonal and normalized. An example is the nth row

$$\frac{1}{n(n-1)} + \cdots + \frac{1}{n(n-1)} + \frac{(1-n)^2}{n(n-1)} = \frac{n-1}{n(n-1)} + \frac{1-2n+n^2}{n(n-1)}$$
$$= \frac{n^2-n}{n(n-1)} = \frac{n(n-1)}{n(n-1)} = 1,$$

where $(n-1)$ terms $1/[n(n-1)]$ have to be summed.

Definition (orthogonal matrix):

A rectangular matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times m}$ is called "a *Hankel matrix*" if the $n+m-1$ distinct elements of $\mathbf{A}$,

$$\begin{bmatrix} a_{11} & & & \\ a_{21} & & & \\ \cdots & & & \\ a_{n-11} & & & \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

only appear in the first column and last row.

Example: Hankel matrix of power sums

Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ be a nm rectangular matrix ($n \leqslant m$) whose entries are power sums.

$$\mathbf{A} := \begin{bmatrix} \sum_{i=1}^{n} \alpha_i x_i & \sum_{i=1}^{n} \alpha_i x_i^2 & \cdots & \sum_{i=1}^{n} \alpha_i x_i^m \\ \sum_{i=1}^{n} \alpha_i x_i^2 & \sum_{i=1}^{n} \alpha_i x_i^3 & \cdots & \sum_{i=1}^{n} \alpha_i x_i^{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} \alpha_i x_i^n & \sum_{i=1}^{n} \alpha_i x_i^{n+1} & \cdots & \sum_{i=1}^{n} \alpha_i x_i^{n+m-1} \end{bmatrix}$$

$\mathbf{A}$ is a *Hankel matrix.*

Definition (Vandermonde matrix):

Vandermonde matrix: $\mathbf{V} \in \mathbb{R}^{n\times n}$

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix},$$

$$\det \mathbf{V} = \prod_{\substack{i,j\ n \\ i>j}} (x_i - x_j).$$

Example: Vandermonde matrix $\mathbf{V} \in \mathbb{R}^{3\times 3}$

$$\mathbf{V} := \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ x_1^2 & x_2^2 & x_3^2 \end{bmatrix}, \ \det \mathbf{V} = (x_2 - x_1)(x_3 - x_2)(x_3 - x_1).$$

Example: Submatrix of a Hankel matrix of power sums Consider the submatrix $\mathbf{P} = [a_1, a_2, \ldots, a_n]$ of the Hankel matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ ($n \leqslant m$) whose entries are power sums. The determinant of the power sums matrix $\mathbf{P}$ is

$$\det \mathbf{P} = \left(\prod_{i=1}^{n} \alpha_i\right)\left(\prod_{i=1}^{n} x_i\right)(\det \mathbf{V})^2,$$

where det $\mathbf{V}$ is the *Vandermonde determinant*. *Example*: Submatrix $\mathbf{P} \in \mathbb{R}^{3\times 3}$ of a 3×4 Hankel matrix of power sums ($n = 3, m = 4$)

$$\mathbf{A} = \begin{bmatrix} \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 & \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 \\ \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 \\ \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 & \alpha_1 x_1^6 + \alpha_2 x_2^6 + \alpha_3 x_3^6 \end{bmatrix}$$

$$\mathbf{P} = [a_1, a_2, a_3] \begin{bmatrix} \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 & \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 \\ \alpha_1 x_1^2 + \alpha_2 x_2^2 + \alpha_3 x_3^2 & \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 \\ \alpha_1 x_1^3 + \alpha_2 x_2^3 + \alpha_3 x_3^3 & \alpha_1 x_1^4 + \alpha_2 x_2^4 + \alpha_3 x_3^4 & \alpha_1 x_1^5 + \alpha_2 x_2^5 + \alpha_3 x_3^5 \end{bmatrix}.$$

A-62 Scalar Measures of Matrices

All reference textbooks review various scalar measures of matrices like

- Linear independence
- Column and row rank
- Rank identities

to which we refer. Permanently, we referred to a special technique called IPM ("**I**nverse **P**artitioned **Ma**trix") of a symmetric matrix we will review now.

Facts: (Inverse Partitional Matrix/IPM/of a symmetric matrix):

Let the symmetric matrix **A** be partitioned as

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A'}_{12} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{A'}_{11} = \mathbf{A}_{11}, \quad \mathbf{A'}_{22} = \mathbf{A}_{22}.$$

Then its Cayley inverse $\mathbf{A}^{-1}$ is symmetric and can be partitioned as well as

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A'}_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} [\mathbf{I} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A'}_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A'}_{12}]\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A'}_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \\ -(\mathbf{A}_{22} - \mathbf{A'}_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A'}_{12}\mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A'}_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{bmatrix}, \end{aligned}$$

$$\textit{if } \mathbf{A}_{11}^{-1} \textit{ exists},$$

$$\begin{aligned} \mathbf{A}^{-1} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A'}_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A'}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1} & -(\mathbf{A}_{11} - \mathbf{A'}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}(\mathbf{A}_{11} - \mathbf{A'}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12})^{-1} & [\mathbf{I} + \mathbf{A}_{22}^{-1}\mathbf{A'}_{12}(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1} \end{bmatrix}, \end{aligned}$$

$$\textit{if } \mathbf{A}_{22}^{-1} \textit{ exists}.$$

$$\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \quad \textit{and} \quad \mathbf{S}_{22} := \mathbf{A}_{11} - \mathbf{A}'_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}$$

are the minors determined by properly chosen rows and columns of the matrix $\mathbf{A}$ called "*Schur complements*" such that

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{I} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}'_{12})\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1} \\ -\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} & \mathbf{S}_{11}^{-1} \end{bmatrix}$$

$$\textit{if } \mathbf{A}_{11}^{-1} \textit{ exists},$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{S}_{22}^{-1} & -\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{S}_{22}^{-1} & [\mathbf{I} + \mathbf{A}_{22}^{-1}\mathbf{A}'_{12}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\textit{if } \mathbf{A}_{22}^{-1} \textit{ exists},$$

are representations of the *Cayley inverse* partitioned matrix $\mathbf{A}^{-1}$ in terms of "*Schur complements*".
The formulae $\mathbf{S}_{11}$ and $\mathbf{S}_{22}$ were first used by *J. Schur* (1917). The term "*Schur complements*" was introduced by Haynsworth (1968). Albert (1969) replaced the *Cayley inverse* $\mathbf{A}^{-1}$ by the *Moore-Penrose inverse* $\mathbf{A}^{+}$. For a survey we recommend Cottle (1974), Ouellette (1981), Carlson (1986) and Styan (1985).

Proof.

For the proof of the "inverse partitioned matrix" $\mathbf{A}^{-1}$ (Cayley inverse) of the partitioned matrix A of full rank we apply *Gauss elimination* (without pivoting).

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{A}'_{11} = \mathbf{A}_{11}, \quad \mathbf{A}'_{22} = \mathbf{A}_{22} \left[\begin{array}{l} \mathbf{A}_{11} \in \mathbb{R}^{m\times m}, \ \mathbf{A}_{12} \in \mathbb{R}^{m\times l} \\ \mathbf{A}'_{12} \in \mathbb{R}^{l\times m}, \ \mathbf{A}_{22} \in \mathbb{R}^{l\times l} \end{array} \right.$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}'_{12} & \mathbf{B}_{22} \end{bmatrix}, \quad \mathbf{B}'_{11} = \mathbf{B}_{11}, \quad \mathbf{B}'_{22} = \mathbf{B}_{22} \left[\begin{array}{l} \mathbf{B}_{11} \in \mathbb{R}^{m\times m}, \ \mathbf{B}_{12} \in \mathbb{R}^{m\times l} \\ \mathbf{B}'_{12} \in \mathbb{R}^{l\times m}, \ \mathbf{B}_{22} \in \mathbb{R}^{l\times l} \end{array} \right.$$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \Leftrightarrow \left[\begin{array}{ll} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{B}_{11}\mathbf{A}_{11} + \mathbf{B}_{12}\mathbf{A}'_{12} = \mathbf{I}_m & (1) \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{B}_{11}\mathbf{A}_{12} + \mathbf{B}_{12}\mathbf{A}_{22} = \mathbf{0} & (2) \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{B}'_{12}\mathbf{A}_{11} + \mathbf{B}_{22}\mathbf{A}'_{12} = \mathbf{0} & (3) \\ \mathbf{A}'_{12}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} = \mathbf{B}'_{12}\mathbf{A}_{12} + \mathbf{B}_{22}\mathbf{A}_{22} = \mathbf{I}_l. & (4) \end{array} \right.$$

$$\boxed{\textit{Case (a)}: \ \mathbf{A}_{11}^{-1} \ \textit{exists}}$$

"*forward step*"

$$\left.\begin{array}{l}\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \text{ (first left equation:multiply by } -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}) \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{0} \text{ (secondrightequation)}\end{array}\right]$$

$$\Leftrightarrow \left.\begin{array}{l}-\mathbf{A}'_{12}\mathbf{B}_{11} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{A}'_{12}\mathbf{A}_{11}^{-1} \\ \mathbf{A}'_{12}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}'_{12} = \mathbf{0}\end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{l}\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \\ (\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{B}'_{12} = -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\end{array}\right.$$

$$\Rightarrow \mathbf{B}'_{12} = -(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$

$$\mathbf{B}'_{12} = -\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$

or

$$\begin{bmatrix}\mathbf{I}_m & \mathbf{0} \\ -\mathbf{A}'_{12}\mathbf{A}_{11}^{-1} & \mathbf{I}_l\end{bmatrix}\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\end{bmatrix}.$$

Note the "*Schur complement*" $\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$.

"*backward step*"

$$\left.\begin{array}{l}\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}'_{12} = \mathbf{I}_m \\ \mathbf{B}'_{12} = -(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\end{array}\right]$$

$$\Rightarrow \mathbf{B}_{11} = \mathbf{A}_{11}^{-1}(\mathbf{I}_m - \mathbf{A}_{12}\mathbf{B}'_{12}) = (\mathbf{I}_m - \mathbf{B}_{12}\mathbf{A}'_{12})\mathbf{A}_{11}^{-1}$$

$$\mathbf{B}_{11} = [\mathbf{I}_m + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}'_{12}]\mathbf{A}_{11}^{-1}$$

$$\mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}'_{12}\mathbf{A}_{11}^{-1}$$

$$\mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0} \text{ (second left equation)}$$

$$\Rightarrow \mathbf{B}_{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}(\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$

$$\Leftrightarrow \mathbf{B}_{22} = (\mathbf{A}_{22} - \mathbf{A}'_{12}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$$

$$\mathbf{B}_{22} = \mathbf{S}_{11}^{-1}.$$

$$\boxed{\textit{Case (b)}: \ \mathbf{A}_{22}^{-1} \ \textit{exists}}$$

"*forward step*"

$$\left.\begin{array}{l}\mathbf{A}_{11}\mathbf{B}_{12}+\mathbf{A}_{12}\mathbf{B}_{22}=\mathbf{0}\ \text{(third right equation)}\\ \mathbf{A'}_{12}\mathbf{B}_{12}+\mathbf{A}_{22}\mathbf{B}_{22}=\mathbf{I}_l\ \text{(fourth left equation: multiply by } -\mathbf{A}_{12}\mathbf{A}_{22}^{-1})\end{array}\right]$$

$$\Leftrightarrow \left.\begin{array}{l}\mathbf{A}_{11}\mathbf{B}_{12}+\mathbf{A}_{12}\mathbf{B}_{22}=\mathbf{0}\\ -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}\mathbf{B}_{12}-\mathbf{A}_{12}\mathbf{B}_{22}=-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\end{array}\right]$$

$$\Leftrightarrow \left[\begin{array}{l}\mathbf{A'}_{12}\mathbf{B}_{12}+\mathbf{A}_{22}\mathbf{B}_{22}=\mathbf{I}_l\\ (\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})\mathbf{B}_{12}=-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\end{array}\right.$$

$$\Rightarrow \mathbf{B}_{12}=-(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}\mathbf{A'}_{12}\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{12}=-\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\quad \text{or}$$

$$\begin{bmatrix}\mathbf{I}_m & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\\ \mathbf{0} & \mathbf{I}_l\end{bmatrix}\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A'}_{12} & \mathbf{A}_{22}\end{bmatrix}=\begin{bmatrix}\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12} & \mathbf{0}\\ \mathbf{A'}_{12} & \mathbf{A}_{22}\end{bmatrix}.$$

Note the "*Schur complement*"

$$\mathbf{S}_{22}:=\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}.$$

"*backward step*"

$$\left.\begin{array}{l}\mathbf{A'}_{12}\mathbf{B}_{12}+\mathbf{A}_{22}\mathbf{B}_{22}=\mathbf{I}_l\\ \mathbf{B}_{12}=-(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\end{array}\right]$$

$$\Rightarrow \mathbf{B}_{22}=\mathbf{A}_{22}^{-1}(\mathbf{I}_l-\mathbf{A'}_{12}\mathbf{B'}_{12})=(\mathbf{I}_l-\mathbf{B'}_{12}\mathbf{A}_{12})\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{22}=[\mathbf{I}_l+\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}\mathbf{A}_{12}]\mathbf{A}_{22}^{-1}$$

$$\mathbf{B}_{22}=\mathbf{A}_{22}^{-1}+\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$

$$\mathbf{A'}_{12}\mathbf{B}_{11}+\mathbf{A}_{22}\mathbf{B'}_{12}=\mathbf{0}\ \text{(third left equation)}$$

$$\Rightarrow \mathbf{B'}_{12}=-\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}\mathbf{B}_{11}=-\mathbf{A}_{22}^{-1}\mathbf{A'}_{12}(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}$$

$$\Leftrightarrow \boxed{\begin{array}{c}\mathbf{B}_{11}=(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A'}_{12})^{-1}\\ \mathbf{B}_{11}=\mathbf{S}_{22}^{-1}.\end{array}} \qquad ♣$$

The representations $\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{21}=\mathbf{B'}_{12}, \mathbf{B}_{22}$ in terms of $\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}=\mathbf{A'}_{12}, \mathbf{A}_{22}$ have been derived by Banachiewicz (1937). Generalizations are referred to Ando (1979), Brunaldi and Schneider (1963), Burns et al. (1974), Carlson (1980), Meyer (1973) and Mitra (1982), Li and Mathias (2000). We leave the proof of the following fact as an exercise.

Fact (Inverse Partitioned Matrix /IPM/ of a quadratic matrix):

Let the quadratic matrix $\mathbf{A}$ be partitioned as

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Then its Cayley inverse $\mathbf{A}^{-1}$ can be partitioned as well as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{S}_{11}^{-1} \\ -\mathbf{S}_{11}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{S}_{11}^{-1} \end{bmatrix},$$

$$\textit{if } \mathbf{A}_{11}^{-1} \textit{ exists}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{S}_{22}^{-1} & -\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{S}_{22}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{S}_{22}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix},$$

$$\textit{if } \mathbf{A}_{22}^{-1} \textit{ exists}$$

and the "*Schur complements*" are defined by

$$\mathbf{S}_{11} := \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \quad \textit{and} \quad \mathbf{S}_{22} := \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}.$$

Facts: Cayley Inverse, DG matrix identity, SMW matrix identity, sum of two matrices

$$(\textit{DG id}) \qquad \mathbf{BD}(\mathbf{A} + \mathbf{CBD})^{-1} = (\mathbf{B}^{-1} + \mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}$$

(Duncan-Guttman matrix identity).

$$(\textit{SMW id}) \qquad (\mathbf{A} + \mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1}$$

(Sherman - Morrison - Woodbury matrix identity).

Duncan (1944) calls (DG) the *Sherman-Morrison-Woodbury matrix identity*. If the matrix $\mathbf{A}$ is singular consult Henderson and Searle (1981a), Henderson and Searle (1981b), Ouellette (1981), Hager (1989), Stewart (1977) and Riedel (1992). (SMW) has been noted by Duncan (1944) and Guttman (1946): The result is directly derived from the identity

$$(\mathbf{A} + \mathbf{CBD})(\mathbf{A} + \mathbf{CBD})^{-1} = \mathbf{I}$$

$$\Rightarrow \mathbf{A}(\mathbf{A} + \mathbf{CBD})^{-1} + \mathbf{CBD}(\mathbf{A} + \mathbf{CBD})^{-1} = \mathbf{I}$$

$$(\mathbf{A} + \mathbf{CBD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{CBD}(\mathbf{A} + \mathbf{CBD})^{-1}$$

$$\mathbf{A}^{-1} = (\mathbf{A} + \mathbf{CBD})^{-1} + \mathbf{A}^{-1}\mathbf{CBD}(\mathbf{A} + \mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1} = \mathbf{D}(\mathbf{A} + \mathbf{CBD})^{-1} + \mathbf{DA}^{-1}\mathbf{CBD}(\mathbf{A} + \mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1} = (\mathbf{I} + \mathbf{DA}^{-1}\mathbf{CB})\mathbf{D}(\mathbf{A} + \mathbf{CBD})^{-1}$$
$$\mathbf{DA}^{-1} = (\mathbf{B}^{-1} + \mathbf{DA}^{-1}\mathbf{C})\mathbf{BD}(\mathbf{A} + \mathbf{CBD})^{-1}$$
$$(\mathbf{B}^{-1} + \mathbf{DA}^{-1}\mathbf{C})^{-1}\mathbf{DA}^{-1} = \mathbf{BD}(\mathbf{A} + \mathbf{CBD})^{-1}.$$

Up to now we presented *IPM* of a symmetric matrix. Alternatively, we review rank factorization of a symmetric and positive semidefinite/definite matrix.

Facts (rank factorization):

(a) If the $n \times n$ matrix is *symmetric* and *positive semidefinite*, then its *rank factorization* is

$$\mathbf{A} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{G}'_1 & \mathbf{G}'_2 \end{bmatrix},$$

where $\mathbf{G}_1$ is a *lower triangular matrix* of the order $O(\mathbf{G}_1) = \operatorname{rk}\mathbf{A} \times \operatorname{rk}\mathbf{A}$ with $\operatorname{rk}\mathbf{G}_2 = \operatorname{rk}\mathbf{A}$, whereas $\mathbf{G}_2$ has the format $O(\mathbf{G}_2) = (n - \operatorname{rk}\mathbf{A}) \times \operatorname{rk}\mathbf{A}$. In this case we speak of a *Choleski decomposition*.

(b) In case that the matrix $\mathbf{A}$ is *positive definite*, *the matrix block* $\mathbf{G}_2$ is not needed anymore: $\mathbf{G}_1$ is *uniquely determined*. There holds

$$\mathbf{A}^{-1} = (\mathbf{G}_1^{-1})'\mathbf{G}_1^{-1}.$$

Various notions, for instance submatrices adjoint, traces and determinants, as scalar measures of a matrix are given by *D. A. Harville* (2001, examples, Sects. 2,5 and 13) to which we refer. We already introduced vector-valued matrix forms by means of the operations vec, vech and veck. Please pay attention to the vec-forms

(a) $\operatorname{vec}\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}' = (\mathbf{A} \otimes \mathbf{C})vec\mathbf{B}$
(b) $(\mathbf{A})'vec\mathbf{B} = \operatorname{tr}(\mathbf{A} \cdot \mathbf{B}')$
(c) $\operatorname{vec}\mathbf{xy}' = \mathbf{x} \otimes \mathbf{y}$

(d)
$$\begin{aligned} \operatorname{vec}(\mathbf{A} \cdot \mathbf{B}) &= (\mathbf{B}' \otimes \mathbf{I}_n)vec\mathbf{A} = (\mathbf{B}' \otimes \mathbf{A})vec\mathbf{I}_m \\ &= (\mathbf{I}_q \otimes \mathbf{A})vec\mathbf{B}, \quad for\ all\ \mathbf{A} \in \mathbb{R}^{n \times m}, \quad \mathbf{B} \in \mathbb{R}^{m \times q} \end{aligned}$$

given by Pukelsheim (1977), pp. 324, 326 or Harville (2001), examples, Sect. 16, 22 pp.

A-63 Three Basic Types of Generalized Inverses

In Chaps. 1, 3 and 5 we obtained the complete class of solution of equation "$\mathbf{Ax} = \mathbf{y}''$ and showed that every solution can be expressed in the form $\mathbf{x} = \mathbf{Gy}$ where $\mathbf{G}$ is the properly chosen *generalized inverse*. Here we inquire to find solutions for such generalized inverses which are of the following type

(i) $\quad rk\,\mathbf{A} = rk\,(\mathbf{AA}') = n$: *"minimum norm"*

Consistent system of observation equations: *Underdetermine*.

Theorem ("minimum norm": Rao and Mitra (1971), pp. 44–47)

Let $\mathbf{G}$ be a generalized inverse of the matrix $\mathbf{A}$ such that $\mathbf{G}_y$ is a *minimum norm solution* of the equation $\mathbf{Ax} = \mathbf{y}, rk\,\mathbf{A} = rk\,(\mathbf{AA}') = n$. Then it is necessary and sufficient that

(a) $\|\mathbf{x}\|^2 = \|\mathbf{G}_y\|^2 = \underset{\mathbf{Ax}=\mathbf{y}}{Min}$ *or*
(b) $\mathbf{AGA} = \mathbf{A}$ *and* $(\mathbf{GA})' = \mathbf{GA}$. or
(c) $\mathbf{A}_{mr}^{-} = \mathbf{A}'(\mathbf{AA}')^{-}$ (reflexive generalized inverse which is minimum norm).

If we use the weighted generalized inverse of the type $\left\{\underset{\mathbf{Ax}=\mathbf{y}}{Min}\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{Gx}\right\}$ we get the $\mathbf{G}_x$-MINOS solution of Chap. 1, case (a), case (b) and case (c).

(ii) $\quad rk\,\mathbf{A} = rk\,(\mathbf{A}'\mathbf{A}) = m$: *"generalized least-squares solution for* $\mathbf{Ax} = \mathbf{y}$*"*

Inconsistent system of linear observation equations: *overdetermine*.

Theorem ("Least-squares": Rao and Mitra (1971), pp. 48–50)

Let $\mathbf{G}$ be a matrix (not necessarily a generalized inverse) such that $\mathbf{G}_y$ is a *least-squares solution* of $\mathbf{Ax} = \mathbf{y}$ such that $rk\,\mathbf{A} = rk\,(\mathbf{A}'\mathbf{A}) = m$. Then it is necessary and sufficient that

(a) $\|\mathbf{y} - \mathbf{Ax}\|^2 = min$ (*inconsistent*)
or
(b) $\mathbf{AGA} = \mathbf{A}$ *and* $(\mathbf{GA})' = \mathbf{GA}$.
or
(c) $\mathbf{A}_{lr} = (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}$ (reflexive generalized inverse which is least-squares).

If we use the *weighted generalized inverse* of the type $\left\{\underset{\mathbf{x}}{Min}\|\mathbf{y} - \mathbf{Ax}\|^2_{\mathbf{G}_y} = (\mathbf{y} - \mathbf{Ax})'\mathbf{G}_y(\mathbf{y} - \mathbf{Ax})\right\}$ we enjoy receiving the $\mathbf{G}_y$-LESS solution of Chap. 3, case (a), case (b) and case (c).

(*iii*) $rk\,\mathbf{A} < min\{m, n\}$: *"generalized inverse for a minimum norm least-squares solution"*

Inconsistent system of linear observation equations with a datum defect: *overdetermine-underdetermined*. □

Here we start with a definition

Definition ("MINOLESS": Rao and Mitra (1971), p. 51)

$\mathbf{G}$ is said to be a minimum norm, least-squares generalized inverse of the matrix $\mathbf{A}$, if $\mathbf{G}$ is $\mathbf{A}_l^-$ and for any $\mathbf{y} \in \mathbb{Y}$

$$\|\mathbf{G}_y\|_m \leq \|\mathbf{x}\|_m \quad for\ all\,\mathbf{x} \in \{\mathbf{x} \mid \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_m \leq \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_m \quad for\ all\,\mathbf{z} \in \mathbb{X}$$

where $\|.\|_n$ and $\|.\|_m$ are norms in $\mathbb{X}^m$ and $\mathbb{Y}^n$ respectively. Finally we will refer to the basic results generating "MINOLESS".

Theorem ("MINOLESS": Rao and Mitra (1971), pp. 50–54,64)

Let $\mathbf{G}$ be a matrix such that $\mathbf{G}_y$ is a *minimum norm, least-squares solution* of the inconsistent equation $\mathbf{A}\mathbf{x} = \mathbf{y} + \mathbf{i}$. Then it is necessary and sufficient that $\mathbf{x} = \mathbf{A}^+\mathbf{y}$ namely, (a) $\mathbf{AGA} = \mathbf{A}$, (b) $(\mathbf{AG})' = \mathbf{AG}$, (c) $\mathbf{GAG}) = \mathbf{G}$, (d) $(\mathbf{GA})' = \mathbf{GA}$ (Penrose 1955). *or*

$$\mathbf{A}^+ = \mathbf{A}'\mathbf{A}[(\mathbf{A}'\mathbf{A})^2]^-\mathbf{A}' = (\mathbf{I} + \mathbf{A}'\mathbf{A})^-\mathbf{A}'[\mathbf{A}'\mathbf{A}(\mathbf{I} + \mathbf{A}'\mathbf{A})^-\mathbf{A}]^-\mathbf{A}^* \quad or$$

$$\mathbf{A}^+ = k \xrightarrow{lim} \infty \sum_{k=1}^{k} \mathbf{A}'(\mathbf{I} + \mathbf{A}\mathbf{A}')^{-k} = k \xrightarrow{lim} \infty \sum_{k=1}^{k} (\mathbf{I} + \mathbf{A}'\mathbf{A})^{-k}\mathbf{A}' \quad or$$

$$\mathbf{A}^+ = \lambda \xrightarrow{lim} 0^+\mathbf{A}'(\lambda\mathbf{I} + \mathbf{A}\mathbf{A}')^{-1} = \lambda \xrightarrow{lim} 0^+(\mathbf{I}\lambda + \mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$$

If we use the *weighted norms* $\|\mathbf{x}\|^2_{\mathbf{P}_x}$ and $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2_{\mathbf{P}_y}$, we receive the $\mathbf{G}_x$-minimum norm, $\mathbf{G}_y$- least squares solution of Chap. 5, case (a), case (b) and case (c).

A-7 Complex Algebra, Quaternion Algebra, Octonian Algebra, Clifford Algebra, Hurwitz Theorem

The greatest advantage of *Clifford algebra* is the following fact: We are able to write down the field equations namely "rot" and "div" equation of gravitational field and electromagnetic field *in one equation*. The related literature is very rich and we could recommend the book by Meetz and Engl (1980). A more detailed introduction into *Clifford algebra and its fascinating chess board* as well as *Clifford analysis* is listed in Grafarend (2004) with more than 300 *references*. From some point of view, the variona algebra presented here are rather out of fashion. Indeed

- Complex algebra, Clifford algebra $Cl(0,1)$
- Quaternion algebra, Clifford algebra $Cl(0,2)$

A-71 *Complex Algebra as a Division Algebra as well as a Composition Algebra,* Clifford algebra $Cl(0,1)$

Here we present to you "*complex algebra*" and refer it to *Clifford algebra* $Cl(0,1)$. The "*complex algebra*" $\mathbb{C}$ due to *C. F. Gauß* is a division algebra *over* $\mathbb{R}$ as well as a composition algebra.

$$\boldsymbol{x} \in \mathbb{C}$$

$$\boldsymbol{x} = \boldsymbol{e}_0 \boldsymbol{x}^0 + \boldsymbol{e}_1 \boldsymbol{x}^1 \quad \textit{subject to } \{x^0, x^1\} \in \mathbb{R}^2$$

$$\operatorname{span} \mathbb{C} = \{1, \boldsymbol{e}_1\}$$

$\boxed{1}$ $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^2$

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} + \boldsymbol{y}.$$

The axioms $(G1+)$, $(G2+)$, $(G3+)$, $(G4+)$ of an *Abelian group* apply.

$\boxed{2}$ $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$, $r, s \in \mathbb{R}$

$$\beta(r, \boldsymbol{x}) =: r \times \boldsymbol{x}.$$

The axioms $(D1+)$, $(D2+)$, $(D3)$ of *additive distributivity* apply.

$\boxed{3}$ One way to explicitly describe *a multiplicative group* with finitely many elements is to give a *table* listing the *multiplications* just representing the map $\gamma : \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}$.

$$\boxed{\textit{multiplication diagram, Cayley diagram}}$$

$$\begin{array}{c|cc} & 1 & \boldsymbol{e}_1 \\ \hline 1 & 1 & \boldsymbol{e}_1 \\ \boldsymbol{e}_1 & \boldsymbol{e}_1 & -1 \end{array}$$

Note that in the multiplication table each entry of a group appears exactly once in each row and column. The multiplication has to be read *from left to right* that is, the entry at the intersection of the row headed by $\boldsymbol{e}_1$ and the column headed by $\boldsymbol{e}_1$ is the product $\boldsymbol{e}_1 * \boldsymbol{e}_1$. Such a table is called a *Cayley diagram* of the multiplicative group.

Here note in addition the *associativity* of the internal multiplication given in the table. Such a "*complex algebra*" $\mathbb{C}$ is *not* a *Lie algebra* since neither $\boldsymbol{x} * \boldsymbol{x} = 0$ $(\boldsymbol{L}1)$ *nor* $(\boldsymbol{x} * \boldsymbol{y}) * z + +(\boldsymbol{y} * z) * \boldsymbol{x} + (z * \boldsymbol{x}) * \boldsymbol{y} = 0$ (Jacobi identity) $(\boldsymbol{L}2)$ hold. Just by means of the *multiplication table compute*

$$\boldsymbol{x} * \boldsymbol{x} = \mathbf{1}\{(x^0)^2 - (x^1)^2\} + 2\boldsymbol{e}_1 x^0 x^1 \neq 0.$$

$\boxed{4}$ *Begin* with the choice

$$\boldsymbol{x}^{-1} = \frac{1}{(x^0)^2 + (x^1)^2}(\mathbf{1}x^0 - \boldsymbol{e}_1 x^1)$$

in order to *end up* with

$$\boldsymbol{x} * \boldsymbol{x}^{-1} = (1x^0 + \boldsymbol{e}_1 x^1) * \frac{\mathbf{1}x^0 - \boldsymbol{e}_1 x^1}{(x^0)^2 + (x^1)^2} = 1$$

Accordingly $(\boldsymbol{G}1*)$, $(\boldsymbol{G}2*)$, $(\boldsymbol{G}3*)$ of a *division algebra* apply.

$\boxed{5}$ *Begin* with the choice

$$Q(\boldsymbol{x}) = Q(\mathbf{1}x^0 + \boldsymbol{e}_1 x^1) := (x^0)^2 + (x^1)^2$$

in order to prove (K1), (K2) and (K3). We only focus on (K2i):

$$\begin{aligned}
Q(\boldsymbol{x} * \boldsymbol{y}) &= Q(\boldsymbol{x}) \times Q(\boldsymbol{y}) \qquad (\textit{multiplicativity})\\
Q(\boldsymbol{x} * \boldsymbol{y}) &= Q\{\mathbf{1}(x^0 y^0 - x^1 y^1) + \boldsymbol{e}_1(x^1 y^0 + x^0 y^1)\}\\
&= (x^0)(y^0)^2 + (x^1)^2(y^1)^2 + (x^1)^2(y^0)^2 + (x^0)^2(y^1)^2\\
Q(\boldsymbol{x}) \times Q(\boldsymbol{y}) &= \{(x^0)^2 + (x^1)^2\} \times \{(y^0)^2 + (y^1)^2\}\\
&= (x^0)^2(y^0)^2 + (x^1)^2(y^1)^2 + (x^1)^2(y^0)^2 + (x^0)^2(\boldsymbol{y}^1)^2
\end{aligned}$$

$$\boxed{Q(\boldsymbol{x} * \boldsymbol{y}) = Q(\boldsymbol{x}) \times Q(\boldsymbol{y}) \qquad q.e.d.}$$

How can be one dream about such a complex algebra $\mathbb{C}$? Gauss (1887) had been motivated in his number theory to introduce *complex numbers* with $i := \sqrt{-1}$ as the "*imaginary unit*". Identify $\mathbf{1}x^0$ with the "*real part*" and $\boldsymbol{e}_1 x^1 = i x^1$ with the "*imaginary part*" of $\boldsymbol{x}$ and we are left with the standard theory of *complex numbers*. $\boldsymbol{x}^{-1}$ is based upon the *complex conjugate* $\mathbf{1}x^0 - \boldsymbol{e}_1 x^1$ of $\boldsymbol{x}$ being divided by the norm of $\boldsymbol{x}$. There is a remarkable *isomorphism* between *complex numbers* and *complex algebra*.

The proper *algebraic interpretation* of *complex numbers* is in terms of *Clifford algebra* $Cl(0,1)$. Observe $g(\boldsymbol{e}_1, \boldsymbol{e}_1) = -1$ which interprets the *binary operation* of the base vector which spans the *vector part* of a complex number. Now translate the multiplication table into the language of the *Clifford product*, namely

$$\boxed{\begin{array}{ll} 1 \overset{*}{\wedge} 1 = 1, & 1 \overset{*}{\wedge} \boldsymbol{e}_1 = \boldsymbol{e}_1 \\ \boldsymbol{e}_1 \overset{*}{\wedge} 1 = \boldsymbol{e}_1, & \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_1 = -1 \end{array}}$$

in order to convince yourself that the *Clifford algebra* $Cl(0,1)$ is algebraically isomorphic to the space of *complex numbers*. ♣

How can we relate complex numbers to *Clifford algebra* Cl_1? Observe $g(\boldsymbol{e}_1, \boldsymbol{e}_1) = -1$ which interprets the *binary operation* of the base vector which spans the *vector part* of a complex number. While the *scalar part* of a complex number is an element of $\boldsymbol{A}^0$, its *vector part* can be considered to be an element of $\boldsymbol{A}^1$. The *direct sum*

$$\boldsymbol{A}^0 \oplus \boldsymbol{A}^1$$

of spaces $\boldsymbol{A}^0$, $\boldsymbol{A}^1$ is algebraically isomorphic to the space of *complex numbers*, being an element of the *Clifford algebra* Cl_1. ♣

A-72 *Quaternion Algebra as a Division Algebra as well as a Composition Algebra,* Clifford algebra $Cl(0,2)$

Next we review "quaternion algebra" quoting *W. R. Hamilton* and refer it to *Clifford algebra* $Cl(0,2)$. The *"quaternion algebra"* $\mathbb{H}$ due to Hamilton (1843) is a *division algebra over* $\mathbb{R}$ as well as a *composition algebra*:

$$\boldsymbol{x} \in \mathbb{H}$$

$$\boldsymbol{x} = \mathbf{1}x^0 + \boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_3 x^3 \quad \textit{subject to} \quad \{x^0, x^1, x^2, x^3\} \in \mathbb{R}^4$$

$$\operatorname{span} \mathbb{H} = \{1, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$$

$\boxed{1}$

$$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^4$$

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} + \boldsymbol{y}.$$

The axioms $(G1+)$, $(G2+)$, $(G3+)$, $(G4+)$ of an *Abelian additive group* apply.

$\boxed{2}$

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^4, \quad r, s \in \mathbb{R}$$

$$\beta(r, \boldsymbol{y}) =: r \times \boldsymbol{x}.$$

The axioms $(D1+)$, $(D2+)$, $(D3)$ of *additive distributivity* apply.

$\boxed{3}$ One way to explicitly describe *a multiplicative group* with finitely many elements is to give a *table* listing the *multiplications* just representing the map $\gamma : \mathbb{H} \times \mathbb{H} \longrightarrow \mathbb{H}$.

multiplication table, Cayley diagram

$$\begin{array}{c|cccc} & 1 & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ \hline 1 & \mathbf{1} & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ \boldsymbol{e}_1 & \boldsymbol{e}_1 & -\mathbf{1} & \boldsymbol{e}_3 & -\boldsymbol{e}_2 \\ \boldsymbol{e}_2 & \boldsymbol{e}_2 & -\boldsymbol{e}_3 & -\mathbf{1} & \boldsymbol{e}_1 \\ \boldsymbol{e}_3 & \boldsymbol{e}_3 & \boldsymbol{e}_2 & -\boldsymbol{e}_1 & -\mathbf{1} \end{array}.$$

Note that in the multiplication table each antry of a group appears exactly once in each row and column. The multiplication has to be read *from left to right* that is, the entry at the intersection of the row headed by $\boldsymbol{e}_1$ and the column headed by $\boldsymbol{e}_2$ is the product $\boldsymbol{e}_1 * \boldsymbol{e}_2$. Such a table is called a *Cayley diagram* of the multiplicative group.
Here note in addition the *associativity* of the internal multiplication given by the table, e.g. $\boldsymbol{e}_1 * (\boldsymbol{e}_2 * \boldsymbol{e}_3) = \boldsymbol{e}_1 * \boldsymbol{e}_1 = -1 = \boldsymbol{e}_3 * \boldsymbol{e}_3 = (\boldsymbol{e}_1 * \boldsymbol{e}_2) * \boldsymbol{e}_3$ or

$$\boldsymbol{x} * \boldsymbol{y} = 1\left(x^\circ y^\circ - \sum_{k=1}^{3} x^k y^k\right) + \sum_{i,j,k} \boldsymbol{e}_k (x^\circ y^k + x^k y^\circ + \epsilon_{ij}^k x^i y^j)$$

such that $(\boldsymbol{x} * \boldsymbol{y}) * \boldsymbol{z} = \boldsymbol{x} * (\boldsymbol{y} * \boldsymbol{z})$.
Such a "*Hamilton algebra*" $\mathbb{H}$ is *not* a *Lie algebra* since *neither* $\boldsymbol{x} * \boldsymbol{x} = \mathbf{0}$ (L1) *nor* $(\boldsymbol{x} * \boldsymbol{y}) * \boldsymbol{z} + (\boldsymbol{y} * \boldsymbol{z}) * \boldsymbol{x} + (\boldsymbol{z} * \boldsymbol{x}) * \boldsymbol{y} = \mathbf{0}$ (L2) (*Jacobi identity*) hold. Just by means of the *multiplication table* compute

$$\begin{aligned} \boldsymbol{x} * \boldsymbol{x} &= 1\{(x^\circ)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2\} \\ &\quad + 2\boldsymbol{e}_1 x^\circ x^1 + 2\boldsymbol{e}_2 x^\circ x^2 + 2\boldsymbol{e}_3 x^\circ x^3 \neq \mathbf{0} \end{aligned}$$

$\boxed{4}$ *Begin* with the choice

$$\boldsymbol{x}^{-1} = \frac{1}{(x^\circ)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2}(1x^\circ - \boldsymbol{e}_1 x^1 - \boldsymbol{x}_2 x^2 - \boldsymbol{e}_2 x^3)$$

in order to *to end up* with

$$\boldsymbol{x} * \boldsymbol{x}^{-1} = [1x^\circ + \boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_3 x^3] * \frac{1x^\circ - \boldsymbol{e}_1 x^1 - \boldsymbol{e}_2 x^2 - \boldsymbol{e}_3 x^3}{(x^\circ)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2} = 1.$$

Accordingly $(G1*)$, $(G2*)$, $(G3*)$ of a division algebra apply.

$\boxed{5}$ *Begin* with the choice

$$Q(\boldsymbol{x}) = Q(1 + \boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_3 x^3) := (x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$$

in order to prove $(\boldsymbol{K}1)$. $(\boldsymbol{K}2)$ and $(\boldsymbol{K}3)$.

The laborious proofs are left as an *exercise*.
How can one dream about such a "quaternion algebra" $\mathbb{H}$? Hamilton (16 October 1843) invented *quaternion numbers* as outlined in a letter (1865) to his son *A. H. Hamilton* for the following reason:

"If I may be allowed to speak of *myself* in connexion with the subject, I might do so in a way which would bring *you* in, by referring to an *antequaternionic* time, when you were a mere *child*, but had caught from me the conception of a Vector, as represented by a *Triplet*; and indeed I happen to be able to put the finger of memory upon the year and month – October, 1843 – when having recently returned from visits to Cork and Parsonstown, connected with a Meeting of the British Association, the desire to discover the laws of the multiplication referred to regained with me a certain strength and carnestness, which had for years been dormant, but was then on the point of being gratified, and was occasionally talked of with you. Every morning in the early part of the above cited month, on my coming down to breakfast, your (then) little brother William Edwin, and yourself, used to ask me, "Well, Papa, can you *multiply* triplets"? Whereto I was always obliged to reply, with a sad shake of the head: "No, I can only *add* and subtract them."

But on the 16th day of the same month – which happened to be a Monday, and a Council day of the Royal Irish Academy – I was walking in to attend and preside, and your mother was walking with me, along the Royal Canal, to which she had perhaps driven; and although she talked with me now and then, yet an *under-current* of thought was going on in my mind, which gave at last a *result*, whereof it is not too much to say that I felt *at once* the importance. An *electric* circuit seemed to *close*; and a spark flashed forth. The herald (as I *foresaw, immediately*) of many long years to come of definitely directed thought and work, by *myself* if spared, and at all events on the part of *others*, if should even be allowed to live long enough distinctly to communicate the discovery. Nor could I resist the impulse – unphilosophical as it may have been – to cut with a knife on a stone of Brougham Bridge, as we passed it, the fundamental formula with the symbols, i, j, k; namely

$$\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = \boldsymbol{ijk} = -1$$

which contains the *Solution of the Problem*, but of course, as an inscription, has long since moldered away. A more durable notice remains, however, on the Council Books of the Academy for that day (October 16th, 1843), which records the fact, that I then asked for and obtained base to read a Paper on *Quaternion*, at the *First General Meeting* of the Session: which reading took place accordingly, on Monday the 13th of the November following."

Obviously *the vector part* $\boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_3 x^3 = \boldsymbol{i}\, x^1 + \boldsymbol{j}\, x^2 + \boldsymbol{k}\, x^3$ of a quaternion number replaces *the imaginary part* of a complex number, *the scalar part* $1x^0$ *the real part*. The *quaternian conjugate*

$$1x^\circ - \sum_{k=1}^{3} e_k x^k =: \boldsymbol{x}^*$$

substitutes the *complex conjugate* of ta complex number, leading to the *quaternion inverse*

$$\boldsymbol{x}^{-1} = \frac{\boldsymbol{x}^*}{Q(\boldsymbol{x})}.$$

The proper algebraic interpretation of *quaternion numbers* is in terms of *Clifford algebra* $Cl(0,2)$. If $n = \dim \mathbb{X} = 2$ is the dimension of the *linear space* $\mathbb{X}$ which we base the *Clifford algebra* on, its basis elements are

$$\begin{gathered}
\{1, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2\} \\
\textit{subject to} \\
\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 + \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_1 = 0, \\
\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_1 = g(\boldsymbol{e}_1, \boldsymbol{e}_1)1 = -1\,,\ \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2 = g(\boldsymbol{e}_2, \boldsymbol{e}_2)1 = -1, \\
(\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2)^2 = (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2) \overset{*}{\wedge} (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2) = -\boldsymbol{e}_2 \overset{*}{\wedge} (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_1) \overset{*}{\wedge} \boldsymbol{e}_1 = +\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2 = -1 \\
(\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2) \overset{*}{\wedge} \boldsymbol{e}_1 = -\boldsymbol{e}_2 \overset{*}{\wedge} (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_1) = \boldsymbol{e}_2 \\
(\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2) \overset{*}{\wedge} \boldsymbol{e}_2 = \boldsymbol{e}_1 \overset{*}{\wedge} (\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2) = -\boldsymbol{e}_1 \\
\textit{and} \\
\boldsymbol{e}_3 := \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2
\end{gathered}$$

by *classical notation*. Obviously

$$\boldsymbol{x} = 1x^\circ + \boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 x^3 \in Cl(0,2)$$

is an element of *Clifford algebra* $Cl(0,2)$.
There is an algebraic isomorphism between the *quaternion algebra* $\mathbb{H}$ of vectors and the *quaternion algebra of matrices*, namely either $\mathbb{M}(\mathbb{R}^{4\times 4})$ of 4×4 *real matrices or* $\mathbb{M}(\mathbb{C}^{2\times 2})$ of 2×2 *complex matrices*.

Firstly we define the 4×4 real matrix basis $\boldsymbol{E}$ and decompose it into the four constituents $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, $\boldsymbol{\Sigma}_3$ of 4×4 *real Pauli matrices* which form a multiplicative group of the multiplication table of Hamilton type

$$\boldsymbol{E} := \begin{bmatrix} 1 & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ -\boldsymbol{e}_1 & 1 & -\boldsymbol{e}_3 & \boldsymbol{e}_2 \\ -\boldsymbol{e}_2 & \boldsymbol{e}_3 & 1 & -\boldsymbol{e}_1 \\ -\boldsymbol{e}_3 & -\boldsymbol{e}_2 & \boldsymbol{e}_1 & 1 \end{bmatrix} = 1\boldsymbol{\Sigma}_0 + \boldsymbol{e}_1\boldsymbol{\Sigma}_1 + \boldsymbol{e}_2\boldsymbol{\Sigma}_2 + \boldsymbol{e}_3\boldsymbol{\Sigma}_3 \in \mathbb{R}^{4\times 4}$$

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_1 := \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_3 := \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix},$$

$\boxed{\text{multiplication table, Cayley diagram}}$

	$\boldsymbol{\Sigma}_0$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_3$
$\boldsymbol{\Sigma}_0$	$\boldsymbol{\Sigma}_0$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_3$
$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_1$	$-\boldsymbol{\Sigma}_0$	$\boldsymbol{\Sigma}_3$	$-\boldsymbol{\Sigma}_2$
$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_2$	$-\boldsymbol{\Sigma}_3$	$-\boldsymbol{\Sigma}_0$	$\boldsymbol{\Sigma}_1$
$\boldsymbol{\Sigma}_3$	$\boldsymbol{\Sigma}_3$	$\boldsymbol{\Sigma}_2$	$-\boldsymbol{\Sigma}_1$	$-\boldsymbol{\Sigma}_0$

or

$$\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_i \quad \textit{for all } i \in \{1, 2, 3\}$$
$$\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j = -\delta_{ij}\boldsymbol{\Sigma}_0 + \epsilon_{ijk}\boldsymbol{\Sigma}_k \quad \textit{for all } i, j, k \in \{1, 2, 3\}$$

Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{M}\,(\mathbb{R}^{4\times 4}, \textit{Hamilton})$ constituted by means of

$$\begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ -a_2 & a_1 & -a_4 & a_3 \\ -a_3 & a_4 & a_1 & -a_2 \\ -a_4 & -a_3 & a_2 & a_1 \end{bmatrix} =: \boldsymbol{A}$$

$$\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ -b_2 & b_1 & -b_4 & b_3 \\ -b_3 & b_4 & b_1 & -b_2 \\ -b_4 & -b_3 & b_2 & b_1 \end{bmatrix} =: \boldsymbol{B}$$

such that the *Cayley-product*

$$\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ -c_2 & c_1 & -c_4 & c_3 \\ -c_3 & c_4 & c_1 & -c_2 \\ -c_4 & -c_3 & c_2 & c_1 \end{bmatrix} =: \boldsymbol{C} \in \mathbb{M} \quad (\mathbb{R}^{4\times 4},\ \textit{Hamilton})$$

$$c_1 = a_1 b_1 - a_2 b_2 - a_3 b_3 - a_4 b_4$$

$$c_2 = a_1 b_2 + a_2 b_1 + a_3 b_4 - a_4 b_3$$

$$c_3 = a_1 b_3 - a_2 b_4 + a_3 b_1 + a_4 b_2$$

$$c_4 = a_1 b_4 + a_2 b_3 - a_3 b_2 + a_4 b_1$$

fulfilling the axioms $(G1\circ)$, $(G2\circ)$, $(G3\circ)$ of a *non-Abelian multiplicative group*, namely

$$\begin{array}{lll} (G1\circ) & (\boldsymbol{A}\boldsymbol{B})\boldsymbol{C} = A(BC) & (\textit{associativity}) \\ (G2\circ) & \boldsymbol{A}\boldsymbol{I} = \boldsymbol{A} & (\textit{identity}) \\ (G3\circ) & \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} & (\textit{inverse}), \end{array}$$

but $(G4\circ)$ does *not* apply, in particular

$$\det \boldsymbol{A} = a_1^2 + a_2^2 + a_3^2 + a_4^2, \quad \det \boldsymbol{B} = b_1^2 + b_2^2 + b_3^2 + b_4^2$$
$$\det(\boldsymbol{A}\boldsymbol{B}) = (\det \boldsymbol{A})(\det \boldsymbol{B}).$$

Secondly we define the 2×2 complex matrix basis E and decompose it into the four constituents Σ^0, Σ^1, Σ^2, Σ^3 of 2×2 *complex Pauli matrices* which form a multiplicative group of the multiplication table of Hamilton type

$$\boldsymbol{E} := \begin{bmatrix} 1 + i\boldsymbol{e}_1 & \boldsymbol{e}_2 + i\boldsymbol{e}_3 \\ -\boldsymbol{e}_2 + i\boldsymbol{e}_3 & 1 - i\boldsymbol{e}_1 \end{bmatrix} = 1\boldsymbol{\Sigma}^0 + \boldsymbol{e}1\boldsymbol{\Sigma}^1 + \boldsymbol{e}_2\boldsymbol{\Sigma}^2 + \boldsymbol{e}_3\boldsymbol{\Sigma}^{\boldsymbol{3}} \in \mathbb{C}^{2\times 2}$$

$$\boldsymbol{\Sigma}^0 := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}^1 := \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$

$$\boldsymbol{\Sigma}^2 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \boldsymbol{\Sigma}^3 := \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$$

$\boxed{\text{multiplication table, Cayley diagram}}$

$$\begin{array}{c|cccc} & \Sigma^0 & \Sigma^1 & \Sigma^2 & \Sigma^3 \\ \hline \Sigma^0 & \Sigma^0 & \Sigma^1 & \Sigma^2 & \Sigma^3 \\ \Sigma^1 & \Sigma^1 & -\Sigma^0 & \Sigma^3 & -\Sigma^2 \\ \Sigma^2 & \Sigma^2 & -\Sigma^3 & -\Sigma^0 & \Sigma^1 \\ \Sigma^3 & \Sigma^3 & \Sigma^2 & -\Sigma^1 & -\Sigma^0 \end{array}.$$

Note that $E \in \mathbb{C}^{2\times 2}$ is "*Hermitean*". Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{M}(\mathbb{C}^{2\times 2}, \mathit{Hamilton})$ constituted by means of

$$\begin{bmatrix} a_1 + i a_2 & a_3 + i a_4 \\ -a_3 + i a_4 & a_1 - i a_2 \end{bmatrix} =: \boldsymbol{A}$$

$$\begin{bmatrix} b_1 + i b_2 & b_3 + i b_4 \\ -b_3 + i b_4 & b_1 - i b_2 \end{bmatrix} =: \boldsymbol{B}$$

such that the *Cayley-product*

$$\begin{aligned} \boldsymbol{A}\boldsymbol{B} &= \begin{bmatrix} c_1 + i c_2 & c_3 + i c_4 \\ -c_3 + i c_4 & c_1 - i c_2 \end{bmatrix} =: \boldsymbol{C} \in \mathbb{M}(\mathbb{C}^2,\ \mathit{Hamilton}) \\ \det(\boldsymbol{A}B) &= (\det \boldsymbol{A})(\det \boldsymbol{B}) = \left(a_1^2 + a_2^2 + a_3^2 + a_4^2\right)\left(b_1^2 + b_2^2 + b_3^2 + b_4^2\right) \\ &= c_1^2 + c_2^2 + c_3^2 + c_4^2 \\ c_1 + i c_2 &= (a_1 + i a_2)(b_1 + i b_2) + (a_3 + i a_4)(-b_3 + i b_4) \\ &= a_1 b_1 - a_2 b_2 - a_3 b_3 - a_4 b_4 + i(a_1 b_2 + a_2 b_1 + a_3 b_4 - a_4 b_3) \\ c_3 + i c_4 &= (a_1 + i a_2)(b_3 + i b_4) + (a_3 + i a_4)(b_1 - i b_2) \\ &= a_1 b_3 - a_2 b_4 + a_3 b_1 + a_4 b_2 + i(a_1 b_4 + a_2 b_3 - a_3 b_2 + a_4 b_1) \end{aligned}$$

fulfilling the axioms ($G1\circ$), $G2\circ$), ($G3\circ$) of a *non-Abelian multiplicative group*. The *spinor*

$$\boldsymbol{s} := \begin{bmatrix} 1 + i\boldsymbol{e}_1 \\ \boldsymbol{e}_2 + i\boldsymbol{e}_3 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

as *a vector of length zero* relates to the 2×2 *complex matrix basis* by

$$E = \begin{bmatrix} \boldsymbol{s}_1 & \boldsymbol{s}_2 \\ -\boldsymbol{s}_2^* & \boldsymbol{s}_1^* \end{bmatrix}. \qquad ♣$$

A-73 *Octanian Algebra as a Non-Associative Algebra as well as a Composition Algebra,* Clifford algebra *with Respect to* $\mathbb{H} \times \mathbb{H}$

The *octanian algebra* $\mathbb{O}$ also called "*the algebra of octaves*" due to Graves (1843) and Cayley (1845) is a *composition algebra over* $\mathbb{R}$ as well as a *non-associative algebra*:

$$\boldsymbol{x} \in \mathbb{O}$$

$$\boldsymbol{x} = 1x^{\circ} + \boldsymbol{e}_1 x^1 + \boldsymbol{e}_2 x^2 + \boldsymbol{e}_3 x^3 + \boldsymbol{e}_4 x^4 + \boldsymbol{e}_5 x^5 + \boldsymbol{e}_6 x^6 + \boldsymbol{e}_7 x^7 \quad \textit{subject to}$$
$$\{x^{\circ}, x^1, \ldots, x^6, x^7\} \in \mathbb{R}^8$$

$$\operatorname{span} \mathbb{O} = \{1, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4, \boldsymbol{e}_5, \boldsymbol{e}_6, \boldsymbol{e}_7\}$$

$\boxed{1}$

$$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^8$$

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) =: \boldsymbol{x} + \boldsymbol{y}$$

The axioms $(G1+)$, $(G2+)$, $(G3+)$, $(G4+)$ of an *Abelian additive group* apply.

$\boxed{2}$

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^8, \quad r, s \in \mathbb{R}$$

$$\beta(r, \boldsymbol{x}) =: r \times \boldsymbol{x}$$

The axioms $(D1+)$, $(D2+)$, $(D3)$ of *additive distributivity* apply.

$\boxed{3}$ One way to explicitly describe *a multiplicative group* with finitely many elements is to give a *table* listing the *multiplications* just representing the map $\gamma : \mathbb{O} \times \mathbb{O} \longrightarrow \mathbb{O}$.

multiplication table, Cayley diagram

	1	$\boldsymbol{e}_1$	$\boldsymbol{e}_2$	$\boldsymbol{e}_3$	$\boldsymbol{e}_4$	$\boldsymbol{e}_5$	$\boldsymbol{e}_6$	$\boldsymbol{e}_7$
1	1	$\boldsymbol{e}_1$	$\boldsymbol{e}_2$	$\boldsymbol{e}_3$	$\boldsymbol{e}_4$	$\boldsymbol{e}_5$	$\boldsymbol{e}_6$	$\boldsymbol{e}_7$
$\boldsymbol{e}_1$	$\boldsymbol{e}_1$	-1	$\boldsymbol{e}_3$	$-\boldsymbol{e}_2$	$\boldsymbol{e}_5$	$-\boldsymbol{e}_4$	$-\boldsymbol{e}_7$	$\boldsymbol{e}_6$
$\boldsymbol{e}_2$	$\boldsymbol{e}_2$	$-\boldsymbol{e}_3$	-1	$\boldsymbol{e}_1$	$-\boldsymbol{e}_6$	$\boldsymbol{e}_7$	$-\boldsymbol{e}_4$	$-\boldsymbol{e}_5$
$\boldsymbol{e}_3$	$\boldsymbol{e}_3$	$\boldsymbol{e}_2$	$-\boldsymbol{e}_1$	-1	$\boldsymbol{e}_7$	$-\boldsymbol{e}_6$	$\boldsymbol{e}_5$	$-\boldsymbol{e}_4$
$\boldsymbol{e}_4$	$\boldsymbol{e}_4$	$-\boldsymbol{e}_5$	$-\boldsymbol{e}_6$	$-\boldsymbol{e}_7$	-1	$-\boldsymbol{e}_1$	$-\boldsymbol{e}_2$	$-\boldsymbol{e}_3$
$\boldsymbol{e}_5$	$\boldsymbol{e}_5$	$\boldsymbol{e}_4$	$-\boldsymbol{e}_7$	$\boldsymbol{e}_6$	$-\boldsymbol{e}_1$	-1	$-\boldsymbol{e}_3$	$\boldsymbol{e}_2$
$\boldsymbol{e}_6$	$\boldsymbol{e}_6$	$\boldsymbol{e}_7$	$\boldsymbol{e}_4$	$-\boldsymbol{e}_5$	$-\boldsymbol{e}_2$	$\boldsymbol{e}_3$	-1	$-\boldsymbol{e}_1$
$\boldsymbol{e}_7$	$\boldsymbol{e}_7$	$-\boldsymbol{e}_6$	$\boldsymbol{e}_5$	$\boldsymbol{e}_4$	$-\boldsymbol{e}_3$	$-\boldsymbol{e}_2$	$\boldsymbol{e}_1$	-1

Note that in the multiplication table each entry of a group appears exactly once in each row and column. The multiplication has to be read *from left to right* that is, the entry at the intersection of the row headed by $\boldsymbol{e}_5$ is the product $\boldsymbol{e}_3 * \boldsymbol{e}_5$. Such a table is called a *Cayley diagram* of the multiplicative group.
Note the *non-associativity* of the internal multiplication given by the table, e.g. $\boldsymbol{e}_2*(\boldsymbol{e}_3*\boldsymbol{e}_4) \neq (\boldsymbol{e}_2*\boldsymbol{e}_3)*\boldsymbol{e}_4$, namely by means of $\boldsymbol{e}_3*\boldsymbol{e}_4 = \boldsymbol{e}_7$, $\boldsymbol{e}_2*(\boldsymbol{e}_3*\boldsymbol{e}_4) = \boldsymbol{e}_2 * \boldsymbol{e}_7 = -\boldsymbol{e}_5$ *versus* $\boldsymbol{e}_2 * \boldsymbol{e}_3 = \boldsymbol{e}_1$, $(\boldsymbol{e}_2 * \boldsymbol{e}_3) * \boldsymbol{e}_4 = \boldsymbol{e}_1 * \boldsymbol{e}_4 = +\boldsymbol{e}_5$. Such an "octonian algebra" $\mathbb{O}$ is *not* a *Lie algebra* since neither $\boldsymbol{x} * \boldsymbol{x} = \boldsymbol{0}$ (L1) *nor* $(\boldsymbol{x} * \boldsymbol{y})\boldsymbol{z} + (\boldsymbol{y} * \boldsymbol{z}) * \boldsymbol{x} + (\boldsymbol{z} * \boldsymbol{x}) * \boldsymbol{y} = \boldsymbol{0}$ (L2) (*Jacobi identity*) hold. Just by means of the multiplication table compute

$$(\boldsymbol{e}_2 * \boldsymbol{e}_3) * \boldsymbol{e}_4 + (\boldsymbol{e}_3 * \boldsymbol{e}_4) * \boldsymbol{e}_2 + (\boldsymbol{e}_4 * \boldsymbol{e}_2) * \boldsymbol{e}_3 = \boldsymbol{e}_5 \neq \boldsymbol{0}.$$

$\boxed{4}$ does not apply.

$\boxed{5}$ Begin with the choice

$$\begin{aligned} Q(\boldsymbol{x}) &= Q(1x^\circ + \boldsymbol{e}_1 x^1 + \cdots + \boldsymbol{e}_6 x^6 + \boldsymbol{e}_7 x^7) \\ &= (x^\circ)^2 + (x^1)^2 + \cdots + (x^6)^2 + (x^7)^2 \end{aligned}$$

in order to prove (K1), (K2), (K3). The laborious proofs are left as an *exercise*.

The proper algebraic interpretation of octonian numbers is in terms of *Clifford algebra*, namely with respect to the *eight dimensional* set $\mathbb{H} \times \mathbb{H} =: \mathbb{H}^2$ where $\mathbb{H}$ is the usual skew field of *Hamilton's* quaternions, algebraically isomorphic to $Cl(0,2)$. Indeed it would have been temptating to base "*octonian algebra*" $\mathbb{O}$ on $Cl(0,3)$, $\dim Cl(0,3) = 2^3 = 8$, but its *generic* elements

$$\{1, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2, \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3, \boldsymbol{e}_3 \overset{*}{\wedge} \boldsymbol{e}_1, \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3\}$$

are *not* representing the *octonian multiplication table* e.g.

$$\begin{aligned} &\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3)^2 = g(\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3, \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3) \\ &\quad = (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3) \overset{*}{\wedge} (\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3) \\ &\quad = -\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_3 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_3 = \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} (\boldsymbol{e}_3 \overset{*}{\wedge} \boldsymbol{e}_3) \\ &\quad = -\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 = -\boldsymbol{e}_1 \overset{*}{\wedge} (\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_2) \overset{*}{\wedge} \boldsymbol{e}_1 = -(\boldsymbol{e}_2 \overset{*}{\wedge} \boldsymbol{e}_1) = +1 \end{aligned}$$

In contrast, let us introduce the *pair*

$$\boldsymbol{x} := (\boldsymbol{a}, \boldsymbol{b}) \in \{\mathbb{X} \,|\, \boldsymbol{a} \in \mathbb{H}, \boldsymbol{b} \in \mathbb{H}\}$$

$\boxed{1}$

$$\boldsymbol{x} \in \mathbb{H}^2, \quad \boldsymbol{x}^{'} \in \mathbb{H}^2, \quad \boldsymbol{x}^{''} \in \mathbb{H}^2$$

$$\alpha(\boldsymbol{x}, \boldsymbol{x}^{'} =: \boldsymbol{x} + \boldsymbol{x}^{'})$$
$$\boldsymbol{x} + \boldsymbol{x}^{'} = (\boldsymbol{a} + \boldsymbol{a}^{'}, \boldsymbol{b} + \boldsymbol{b}^{'}).$$

The axioms $(G1+)$, $(G2+)$, $(G3+)$, $(G4+)$ of an *Abelian additive group* apply.

$\boxed{2}$

$$\boldsymbol{x} . \boldsymbol{x}^{'} \in \mathbb{H}^2, \quad r, r^{'} \in \mathbb{R}$$

$$\beta(r, \boldsymbol{x}) =: r \times \boldsymbol{x}$$
$$r \times \boldsymbol{x} = (r \times \boldsymbol{a}, r \times \boldsymbol{b}).$$

The axioms $(D1+)$, $(D2+)$, $(D3)$ of *additive distributivity* apply.

$\boxed{3}$

$$\boldsymbol{x} \in \mathbb{H} \times \mathbb{H} = \mathbb{H}^2, \quad \boldsymbol{x}^{'} \in \mathbb{H} \times \mathbb{H} = \mathbb{H}^2$$

$$\gamma(\boldsymbol{x}, \boldsymbol{x}^{'}) =: \boldsymbol{x} * \boldsymbol{x}^{'}$$
$$\boldsymbol{x} * \boldsymbol{x}^{'} =: (\boldsymbol{a}\boldsymbol{a}^{'} - \overline{\boldsymbol{b}}^{'}\boldsymbol{b}, \boldsymbol{b}^{'}\boldsymbol{a} + \boldsymbol{b}\overline{\boldsymbol{a}}^{'})$$

$\overline{\boldsymbol{a}}$, $\overline{\boldsymbol{b}}$ denote the conjugate of $\boldsymbol{a} \in \mathbb{H}$, $\boldsymbol{b} \in \mathbb{H}$, respectively. If $(\boldsymbol{a}, \boldsymbol{b})$, $(\boldsymbol{a}^{'}, \boldsymbol{b}^{'})$ are represented by

$$\left(1\alpha^{\circ} + \sum_{i=1}^{3} \boldsymbol{e}_i \alpha^i, 1\beta^{\circ} + \sum_{j=1}^{3} \boldsymbol{e}_j \beta^j\right), \left(1\alpha^{'\circ} + \sum_{i'=1}^{3} \boldsymbol{e}_{i'} \alpha^{'i}, 1\beta^{'\circ} + \sum_{j=1}^{3} \boldsymbol{e}_{j'} \beta^{'j}\right)$$

respectively, where

$$\operatorname{span} \mathbb{H} = \{1, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_2 = \boldsymbol{e}_3\} \quad or$$
$$\operatorname{span} \mathbb{H} = \{1^{'}, \boldsymbol{e}_{1'}, \boldsymbol{e}_{2'}, \boldsymbol{e}_{1'} \overset{*}{\wedge} \boldsymbol{e}_{2'} = \boldsymbol{e}_{3'}\} = \{1^{'}, \boldsymbol{e}_5, \boldsymbol{e}_6, \boldsymbol{e}_7\}$$

the "*octonian product*" $\boldsymbol{x} * \boldsymbol{x}^{'}$ results in

$$\boxed{\begin{aligned} \boldsymbol{x} * \boldsymbol{x}^{'} = \Bigg(& 1\left[\alpha^{\circ}\alpha^{'\circ} - \beta^{\circ}\beta^{'\circ} - \sum_{k=1}^{3}(\alpha^k \alpha^{`k} - \beta^k \beta^{'k})\right] \\ & + \sum_{i,j,k=1}^{3} \boldsymbol{e}_k [\alpha^{\circ}\beta^k + \alpha^k \beta^{\circ} + \epsilon_{ij}^k (\alpha^i \alpha^{'j} - \beta^j \beta^{'i})], \\ & \times 1\left[\alpha^{\circ}\beta^{'\circ} - \alpha^{'\circ}\beta^{\circ} - \sum_{k=1}^{3}(\alpha^k \beta^{`k} - \alpha^{'k}\beta^k)\right] \\ & + \sum_{i,j,k=1}^{3} \boldsymbol{e}_k [\alpha^{\circ}\beta^{'k} + \alpha^k \beta^{'\circ} + \epsilon_{ij}^k (\alpha^j \beta^{'i} - \alpha^{'j}\beta^i)] \Bigg). \end{aligned}}$$

the axioms $(D1 * +)$, $(D2 * \times)$ of *distributivity* apply. The pair $(1, 0)$ is the neutral element.

$\boxed{4}$ *Begin* with the choice of

$$\text{(1st) the transpose } \overline{\boldsymbol{x}} := \overline{\boldsymbol{a}}, -\boldsymbol{b}) \text{ of } \boldsymbol{x} \in \mathbb{H}^2$$

$$\text{(2nd) } \boldsymbol{x} * \overline{\boldsymbol{x}} = Q(\boldsymbol{x}) \text{ or } \boldsymbol{x} * \overline{\boldsymbol{x}} = (\boldsymbol{a}\overline{\boldsymbol{a}} + \boldsymbol{b}\overline{\boldsymbol{b}}, \overline{0})$$

$$\text{(3rd) } \boldsymbol{x}^{-1} = \frac{\overline{\boldsymbol{x}}}{Q(\boldsymbol{x})} \text{ if } \boldsymbol{x} \neq \mathbf{0}$$

in order to *end up* with

$$\boldsymbol{x} * \boldsymbol{x}^{-1} = 1.$$

A historical perspective of octonian numbers is given by Van Der Waerden (1950). Reference is made to Graves (1848) and A. Caley: Collected Mathematical Papers, vol. 1, p. 127 and vol. 11, pp. 368–371. ♣

The exceptional role of the examples on *complex, quaternion and octonian algebra* illustrating Clifford algebra $Cl(0, 1)$, $Cl(0, 2)$ as well as Clifford algebra with respect to $\mathbb{H}^2$ is *established* by the following theorems:

Theorem ("*Hurwitz' theorem of composition algebras*"):

A complete list of composition algebras over $\mathbb{R}$ *consists of*

(a) the real numbers $\mathbb{R}$,

(b) the complex numbers $\mathbb{C}$,

(c) the quaternions $\mathbb{H}$,

(d) the octonians $\mathbb{O}$.

Theorem ("*Frobenius' theorem of division algebras*"):

The only finite-dimensional division algebra over $\mathbb{R}$ *are*

(α) the real numbers $\mathbb{R}$,

(β) the complex numbers $\mathbb{C}$,

(γ) the quaternions $\mathbb{H}$.

Historical Aside

For details consult the historical texts Hurwitz (1898), Frobenius (1878) as well as Hazlett (1917). A more recent reference is Jacobson (1974), pp. 425 and 430.

A-74 Clifford Algebra

We already took advantage of the notion of *Clifford algebra* when we referred to complex algebra, *quaternion algebra*, and *octonian algebra*. Here we finally confront you with the definition of "*orthogonal clifford algebra* $Cl(p,q)$". But on our way to *Clifford algebra* we have to generalize at first the notion of a basis, in particular its bilinear form.

Theorem (bilinear form):

Suppose that the bracket $\langle\cdot|\cdot\rangle$ *or* $g(\cdot,\cdot) : \mathbb{X}\times\mathbb{X}^* \longrightarrow \mathbb{R}$ *is a bilinear form on a finite dimensional linear space* $\mathbb{X}$*, e. g. a vector space, over the field* $\mathbb{R}$ *of real numbers, in addition* $\mathbb{X}^*$ *its dual space such that* $n = \dim \mathbb{X}^* = \dim\mathbb{X}$*. There exists a basis* $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ *such that*

(a) $\langle \boldsymbol{e}_i \,|\, \boldsymbol{e}_j\rangle = 0 \;\; or \;\; g(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0 \;\; for \;\; i \neq j$

(b) $$\left[\begin{array}{l} \langle \boldsymbol{e}_{i_1} \,|\, \boldsymbol{e}_{i_1}\rangle = +1 \;\; or \;\; g(\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_1}) = +1 \;\; for \;\; 1 \leq i_1 < p\,, \\ \langle \boldsymbol{e}_{i_2} \,|\, \boldsymbol{e}_{i_2}\rangle = -1 \;\; or \;\; g(\boldsymbol{e}_{i_2}, \boldsymbol{e}_{i_2}) = -1 \;\; for \;\; p+1 \leq i_2 \leq p+q = r\,, \\ \langle \boldsymbol{e}_{i_3} \,|\, \boldsymbol{e}_{i_3}\rangle = \;\; 0 \;\; or \;\; g(\boldsymbol{e}_{i_3}, \boldsymbol{e}_{i_3}) = 0 \;\; for \;\; r+1 \leq i_3 \leq n \end{array}\right.$$

holds.

The numbers r and p are determined exclusively by the bilinear form. r is called the *rank*, $r - p = q$ is called the *index* and the *ordered pair* (p,q) the *signature*. The theorem assures that any two spaces of the same dimension with *bilinear forms of the same signature* are isometrically isomorphic. A *scalar product* ("inner product") in this context is a *nondegenerate bilinear* form, i.e., a form with rank equal to the dimension of $\mathbb{X}$. When dealing with low dimensional spaces as we do, we will often indicate *the signature* with a series of plus and minus signs and zeroes where appropriate. For example, the signature of $\mathbb{R}^4_1$ may be written $(+++-)$ instead of $(3,1)$. If the bilinear form is *nondegenerate*, a basis with the properties listed in corresponding *Theorem* is called an *orthonormal basis* ("unimodular") for $\mathbb{X}$ with respect to the bilinear form.

Definition (orthogonal Clifford algebra $Cl(p,q)$):

The orthogonal Clifford algebra $Cl(p,q)$ *is the algebra of polynomials generated by the direct sum of the space of multilinear functions*

$$\oplus^n_{m=0} \overset{*\,m}{\wedge} (\mathbb{X}^*)$$

on a linear space $\mathbb{X}$*, respectively its dual* $\mathbb{X}^*$ *over the field of real numbers* $\mathbb{R}^n_p$ *of dimension*

$$\dim \oplus_{m=0}^{n} \overset{*}{\wedge}{}^{m} (\mathbb{X}^*) = 2^n$$

and signature (p, q)*, namely*

$$\boxed{\begin{aligned} &1 f_0 + \sum_{i_1=1}^{n=\dim \mathbb{X}^*} e^{i_1} f_{i_1} + \sum_{i_1,i_2=1}^{n=\dim \mathbb{X}^*} e^{i_1} \overset{*}{\wedge} e^{i_2} f_{i_1 i_2} \\ &+ \sum_{i_1,i_2,i_3=1}^{n=\dim \mathbb{X}^*} e^{i_1} \overset{*}{\wedge} e^{i_2} \overset{*}{\wedge} e^{i_3} f_{i_1 i_2 i_3} + \cdots + \sum_{i_1,\ldots,i_n=1}^{n=\dim \mathbb{X}^*} e^{i_1} \overset{*}{\wedge} \cdots \overset{*}{\wedge} e^{i_n} f_{i_1 \cdots i_n} \end{aligned}}$$

subject to the Clifford product*, also called the* Clifford dualizer*,*

$$\boxed{\begin{aligned} &\text{(i) } e_i \overset{*}{\wedge} e_j = -e_j \overset{*}{\wedge} e_i \quad for \;\; i \neq j \\ &\text{(ii)} \left[\begin{aligned} & e_{i_1} \overset{*}{\wedge} e_{i_1} = g(e_{i_1}, e_{i_1}) = +1 \;\; for \;\; 1 \leq i_1 \leq p\,, \\ & e_{i_2} \overset{*}{\wedge} e_{i_2} = g(e_{i_2}, e_{i_2}) = -1 \;\; for \;\; p+1 \leq i_2 \leq p+q = r\,, \\ & e_{i_3} \overset{*}{\wedge} e_{i_3} = g(e_{i_3}, e_{i_3}) = \;\; 0 \;\; for \;\; r+1 \leq i_3 \leq n\,, \end{aligned}\right. \end{aligned}}$$

or

$$\boldsymbol{e}_1 \overset{*}{\wedge} \boldsymbol{e}_j + \boldsymbol{e}_j \overset{*}{\wedge} \boldsymbol{e}_i$$

$$= 2 g(\boldsymbol{e}_i, \boldsymbol{e}_j) \delta_{ij} \, 1 \;\; subject \;\; to \left[\begin{aligned} & g(\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_1}) = +1 \; for \; 1 \leq i_1 \leq p \\ & g(\boldsymbol{e}_{i_2}, \boldsymbol{e}_{i_2}) = -1 \; for \; p+1 \leq i_2 \leq p+q = r \\ & g(\boldsymbol{e}_{i_3}, \boldsymbol{e}_{i_3}) = 0 \; for \; r+1 \leq i_3 \leq n \end{aligned}\right.$$

1 *being the neutral element. If* $\boldsymbol{e}_k \overset{*}{\wedge} \boldsymbol{e}_k = 0$ *or* $g(e_k, e_k) = 0$ *holds uniformly the orthogonal* Clifford algebra $Cl(p, q)$ *reduces to the polynomial algebra of* antisym-metric multilinear functions

$$\oplus_{m=0}^{n} A^m = \oplus_{m=0}^{n} \Lambda^m (\mathbb{X}^*) = \boldsymbol{\Lambda} (\mathbb{X}^*)$$

$$\dim \, \oplus_{m=0}^{n} A^m = \dim \, \oplus_{m=0}^{n} \Lambda^m (\mathbb{X}^*) = \dim (\mathbb{X}^*) = 2^n$$

represented by

$$\boxed{\begin{aligned} &1 f_0 + \frac{1}{1!} \sum_{i_1=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} f_{i_1} + \frac{1}{2!} \sum_{i_1,i_2=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} f_{i_1 i_2} \\ &+ \frac{1}{3!} \sum_{i_1,i_2,i_3=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} \wedge \boldsymbol{e}^{i_2} \wedge \boldsymbol{e}^{i_3} f_{i_1 i_2 i_3} + \cdots + \frac{1}{n!} \sum_{i_1,\ldots,i_n=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} \wedge \cdots \wedge \boldsymbol{e}^{i_n} f_{i_1 \cdots i_n}\,. \end{aligned}}$$

Example: *Clifford product:* $\mathbf{x} * \mathbf{y}$, $\mathbf{x} * \mathbf{y} \in \mathbb{X}$, $sign\,\mathbb{X} = (3, 0)$

Let $\mathbf{x} * \mathbf{y} \in \mathbb{X}$ be a real three-dimensional vector space $\mathbb{X}$ of signature (3,0). then $\mathbf{x} * \mathbf{y}$ (read: "**x** *Clifford* **y**") with respect to a set of the bases $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ accounts for

$$\mathbf{x} \overset{*}{\wedge} \mathbf{y} = \sum_{i=1}^{3}\sum_{j=1}^{3} \mathbf{e}_i \overset{*}{\wedge} \mathbf{e}_j x^i y^j = \sum_{i=1}^{3}\sum_{j=1}^{3} x^i y^j \mathbf{e}_i \overset{*}{\wedge} \mathbf{e}_j$$

$$\mathbf{x} \overset{*}{\wedge} \mathbf{y} = (\mathbf{e}_1 x^1 + \mathbf{e}_2 x^2 + \mathbf{e}_3 x^3) \overset{*}{\wedge} (\mathbf{e}_1 y^1 + \mathbf{e}_2 y^2 + \mathbf{e}_3 y^3)$$

$$\begin{aligned} \mathbf{x} \overset{*}{\wedge} \mathbf{y} = {} & 1(x^1 y^1 + x^2 y^2 + x^3 y^3) + \mathbf{e}_1 \wedge \mathbf{e}_2 (x^1 y^2 - x^2 y^1) \\ & + \mathbf{e}_2 \wedge \mathbf{e}_3 (x^2 y^3 - x^3 y^2) + \mathbf{e}_3 \wedge \mathbf{e}_1 (x^3 y^1 - x^1 y^3) \end{aligned}$$

Indeed $\mathbf{x} \overset{*}{\wedge} \mathbf{y}$ as a *Clifford number* is decomposed into a scalar part and an antisymmetric tensor part with respect to the bilinear basis $\{\mathbf{e}_1 \wedge \mathbf{e}_2, \mathbf{e}_2 \wedge \mathbf{e}_3, \mathbf{e}_3 \wedge \mathbf{e}_1\}$.

Tensor algebra or the algebra of multilinear functions, namely

$$\begin{aligned} & 1 f_0 + \sum_{i_1=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} f_{i_1} + \sum_{i_1, i_2=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i} \otimes \boldsymbol{e}^{i_2} f_{i_1 i_2} \\ & + \sum_{i_1, i_2, i_3=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} \otimes \boldsymbol{e}^{i_2} \otimes \boldsymbol{e}^{i_3} f_{i_1 i_2 i_3} + \cdots + \sum_{i_1, \ldots, i_n=1}^{n=\dim \mathbb{X}^*} \boldsymbol{e}^{i_1} \otimes \cdots \otimes \boldsymbol{e}^{i_n} f_{i_1 \cdots i_n} \\ & \in \oplus_{m=0}^{n} \otimes (\mathbb{X}^*) = \otimes(\mathbb{X}) = \boldsymbol{T}^+ \oplus \boldsymbol{T}^- \\ & \textit{subject to} \begin{bmatrix} \boldsymbol{T}^+ = \oplus_{h=0} \otimes^{2h} (\mathbb{X}) \quad (\textit{“even”}) \\ \boldsymbol{T}^- = \oplus_{k=0} \otimes^{2k+1} (\mathbb{X}) \quad (\textit{“odd”}) \end{bmatrix} \end{aligned}$$

is the sum of two spaces, $\boldsymbol{T}^+$ and $\boldsymbol{T}^-$, respectively, in particular

$$Cl^+ \ni 1 f_0 + \frac{1}{2!} \sum_{i_1, i_2=1}^{n} \boldsymbol{e}^{i_1} \overset{*}{\wedge} \boldsymbol{e}^{i_2} f_{i_1 i_2} + \frac{1}{4!} \sum_{i_1, i_2, i_3, i_4}^{n} \boldsymbol{e}^{i_1} \overset{*}{\wedge} \boldsymbol{e}^{i_2} \overset{*}{\wedge} \boldsymbol{e}^{i_3} \overset{*}{\wedge} \boldsymbol{e}^{i_4} f_{i_1 i_2 i_3 i_4} + \cdots,$$

$$Cl^- \ni \frac{1}{1!} \sum_{i_1=1}^{n} \boldsymbol{e}^{i_1} f_{i_1} + \frac{1}{3!} \sum_{i_1, i_2, i_3=1}^{n} \boldsymbol{e}^{i_1} \overset{*}{\wedge} \boldsymbol{e}^{i_2} \overset{*}{\wedge} \boldsymbol{e}^{i_3} f_{i_1 i_2 i_3} + \cdots.$$

Obviously Cl^+ as well as Cl^- are *subalgebras* of Cl. Let the *Clifford numbers* z be divided into $z^+ \in Cl^+$ and $z^- \in Cl^-$, then the properties

$$\boxed{z^+ \overset{*}{\wedge} z^+ \in Cl^+, \quad z^- \overset{*}{\wedge} z^- \in Cl^+, \quad z^+ \overset{*}{\wedge} z^- \in Cl^-}$$

prove that $Cl(p,q)$ is graded over the *cydric group* $\mathbb{Z}_2 = \{0, 1\}$. ♣

Appendix B
Sampling Distributions and Their Use: Confidence Intervals and Confidence Regions

We begin with sampling distributions by two sections on the *transformation of random variables*. A *first confidence* interval for *Gauss-Laplace normal distribution* observations $\mu_1 \sigma^2$ *known* is constructed, including its famous *Three Sigma Rule*. The *second confidence* interval of *Gauss-Laplace i.i.d. observations* is derived for the *sample mean* $\widehat{\mu}$ BLUE of μ_1 when the *variance is known*. We take advantage of the *forward-backward Helmert transformation* leading to *Helmert chi-square* χ^2 probability distribution (Helmert, 1875). In contrast, the sampling distributions from *Gauss-Laplace i.i.d. normal distributions* within the third confidence interval is analyzed for *the mean, but an unknown variance*. Our basis is student's *t-distribution* on the random variable $\sqrt{n}(\widehat{\mu} - \mu)/\widehat{\sigma}$ where the sample mean $\widehat{\mu}$ – is BLUE of μ whereas the *sample confidence* for the "true" mean μ, variance σ^2 unknown which is based on *student's probability distribution*. In detail, we derive the related *Uncertainty Principle* generated by the *Magic Triangle* of (a) length of the confidence interval, (b) coefficient of negative confidence called the uncertainty number and (c) the number of observations. For more details we refer to Lemma B.12 of *W.S. Gosset*. The sampling from the *Gauss-Laplace normal distribution*, namely the *fourth confidence interval* for the *variance* is reviewed: We introduce the missing confidence interval for the *"true" variance* σ^2. Table B.18 as a flow chart summarizes the various steps in computing a confidence interval for the variance. This time the related *Uncertainty Principle* is built on (a) the coefficient of complementary confidence α, also called uncertainty number (b) the number of observations, n, and (c) the length $\delta\sigma^2(n, \alpha)$ of the confidence interval the *"true" variance* σ^2. The most detailed sampling theory from multidimensional *Gauss-Laplace normal distribution*, namely for the *confidence region for the fixed parameters* in the *linear Gauss–Markov*, is stated by an *extensive Example B.19*. We need the *Principle Component Analysis* PCA, also called *Singular Value Decomposition* (SVD) or *Eigenvalue Analysis* (EIGEN). Theorem B.17 summarizes the *marginal distributions special linear Gauss–Markov model with datum defect*.
In our appendix we concentrate on *multidimensional variance analysis*, namely sampling form the *multivariate Gauss-Laplace normal distribution*. In short

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2,
© Springer-Verlag Berlin Heidelberg 2012

introduction we review the distribution of sample mean and variance-covariance. We define the *k-dimensional noncentral WHISHART* distribution by (B377) and the *characteristic function* of a *vector-valued random variate of type Gauss-Laplace distribution*. Our highlight is the distribution of $x_N^\wedge$ and S_N of type *WHISHART*. Another highlight is the *distribution related to the correlation coefficients* originated by R.A. Fisher (1915), programmed by F.N. David (1954).

B-1 A First Vehicle: Transformation of Random Variables

If the probability density function (*pdf*) of a random vector $\mathbf{y} = [y_1, \ldots, y_n]'$ is *known*, but we want to derive the probability density function (*pdf*) of a random vector $\mathbf{x} = [x_1, \ldots, x_n]'$ (*pdf*) which is generated by an *injective mapping* x =g(y) or $x_i = g_i[y_1, \ldots, y_n]$ for all $i \in \{1, \ldots, n\}$ we need the results of Lemma B.1.

Lemma B.1. (transformation of pdf):

Let the random vector $\mathbf{y} := [y_1, \ldots, y_n]'$ be transformed into the random vector $\mathbf{x} = [x_1, \ldots, x_n]'$ by an *injective mapping* $\mathbf{x} = g(\mathbf{y})$ or $x_i = g_i[y_1, \ldots, y_n]$ for all $i \in \{1, \ldots, n\}$ which is of continuity class $\mathbb{C}^1$ (first derivatives are continuous). Let the Jacobi matrix $\mathbf{J}_x : = (\partial g_i / \partial y_i)$ be *regular* ($\det \mathbf{J}_x \neq 0$), then the inverse transformation $\mathbf{y} = g - 1(\mathbf{x})$ or $y_i = g_i^{-1}[x_1, \ldots, x_n]$ is *unique*. Let $f_x(x_1, \ldots, x_n)$ be the unknown *pdf*, but $f_y(y_1, \ldots, y_n)$ the given *pdf*, then

$$f_x(x_1, \ldots, x_n) = f(g_1^{-1}(x_1, \ldots, x_n), \ldots, g_1^{-1}(x_1, \ldots, x_n)) \left|\det \mathbf{J}_y\right|$$

with respect to the Jacobi matrix

$$\mathbf{J}_y := \left[\frac{\partial y_i}{\partial x_j}\right] = \left[\frac{\partial g_i^{-1}}{\partial x_j}\right]$$

for all $i, j \in \{1, \ldots, n\}$ holds. Before we sketch the proof we shall present two examples in order to make you more familiar with the notation.

Example B.1. ("counter example"):

The vector-valued random variable (y_1, y_2) is transformed into the vector-valued random variable (x_1, x_2) by means of

$$x_1 = y_1 + y_2, \quad x_2 = y_1^2 + y_2^2$$

$$\mathbf{J}_x : = \left[\frac{\partial x}{\partial y'}\right] = \begin{bmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2y_1 & 2y_2 \end{bmatrix}$$

$$x_1^2 = y_1^2 + 2y_1y_2 + y_2^2,\ x_2 + 2y_1y_2 = y_1^2 + 2y_1y_2 + y_2^2$$
$$x_1^2 = x_2 + 2y_1y_2,\ y_2 = (x_1^2 - x_2)/(2y_1)$$
$$x_1 = y_1 + y_2 = y_1 + \frac{1}{2}\frac{x_1^2 - x_2}{y_1}, \quad x_1y_1 = y_1^2 + \frac{1}{2}(x_1^2 - x_2)$$
$$y_1^2 - x_1y_1 + \frac{1}{2}(x_1^2 - x_2) = 0$$
$$y_1^{\pm} = -\frac{1}{2}x_1 \pm \sqrt{\frac{x_1^2}{4} - \frac{1}{2}(x_1^2 - x_2)}$$
$$y_2^{\pm} = \frac{1}{2}x_1 \mp \sqrt{\frac{x_1^2}{4} - \frac{1}{2}(x_1^2 - x_2)}.$$

At first we have computed the Jacobi matrix $\mathbf{J}_x$, secondly we aimed at an inversion of the direct transformation $(y_1, y_2) \mapsto (x_1, x_2)$. As the detailed inversion step proves, namely the solution of a quadratic equation, the mapping $\mathbf{x} = g(\mathbf{y})$ is not injective.

Example B.2.

Suppose (x_1, x_2) is a random variable having *pdf*

$$f_x(x_1, x_2) = \begin{bmatrix} \exp(-x_1 - x_2),\ x_1 \geq 0, \quad x_2 \geq 0 \\ 0 \qquad\qquad\qquad\qquad ,\text{otherwise.} \end{bmatrix}$$

We require to find the *pdf* of the random variable

$$(x_1 + x_2, x_2/x_1).$$

The transformation

$$\boxed{y_1 = x_1 + x_2,\ y_2 = \frac{x_2}{x_1}}$$

has the inverse

$$\boxed{x_1 = \frac{y_1}{1 + y_2},\ x_2 = \frac{y_1y_2}{1 + y_2}.}$$

The transformation provides a *one-to-one mapping* between points in the *first quadrant* of the (x_1, x_2) – plane $\mathbb{P}_x^2$ *and* in the *first quadrant* of the (y_1, y_2) – plane $\mathbb{P}_y^2$. The absolute value of the *Jacobian* of the transformation for all points in the *first quadrant* is

$$\left|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} (1+y_2)^{-1} & -y_1(1+y_2)^{-2} \\ y_2(1+y_2)^{-1} & y_1(1+y_2)^{-2}[(1+y_2)-y_2] \end{vmatrix}$$

$$= y_1(1+y_2)^{-3} + y_1 y_2 (1+y_2)^{-3} = \frac{y_1}{(1+y_2)^2}.$$

Hence we have found for the *pdf* of (y_1, y_2)

$$f_y(y_1,y_2) = \begin{bmatrix} \exp(-y_1)\frac{y_1}{(1+y_2)^2}, & y_1 > 0,\ y_2 > 0 \\ 0 & , \text{otherwise.} \end{bmatrix}$$

Incidentally it should be noted that y_1 and y_2 are *independent random variables*, namely

$$\boxed{f_y(y_1,y_2) = f_1(y_1)f(y_2) = y_1 \exp(-y_1)(1+y_2)^{-2}.} \qquad \clubsuit$$

Proof. The probability that the random variables $y_1, \ldots, y_n$ take on values in the region Ω_y is given by

$$\int_{\Omega_y} \cdots \int f_y(y_1,\ldots,y_n)dy_1 \cdots dy_n.$$

If the random variables of this integral are transformed by the function $x_i = g_i(y_1,\ldots,y_n)$ for all $i \in \{1,\ldots,n\}$ which map the region Ω_y onto the regions Ω_x, we receive

$$\int_{\Omega_y} \cdots \int f_y(y_1,\ldots,y_n)dy_1 \cdots dy_n$$
$$= \int_{\Omega_x} \cdots \int f_y(g_1^{-1}(x_1,\ldots,x_n),\ldots,g_n^{-1}(x_1,\ldots,x_n)) \left|\det J_y\right| dx_1 \cdots dx_n$$

from the standard theory of transformation of hypervolume elements, namely

$$dy_1 \cdots dy_n = |\det \mathbf{J}_y| dx_1 \cdots dx_n$$

or

$$*(dy_1 \wedge \cdots \wedge dy_n) = |\det \mathbf{J}_y| * (dx_1 \wedge \cdots \wedge dx_n).$$

Here we have taken advantage of the *oriented hypervolume element* $dy_1 \wedge \cdots \wedge dy_n$ (*Grassmann product*, skew product, wedge product) and the *Hodge star operator* $*$ applied to the n – differential form $dy_1 \wedge \cdots \wedge dy_n \in \Lambda^n$ (the exterior algebra Λ^n).

The star $* : \Lambda^p \rightarrow \Lambda^{n-p}$ in $\mathbb{R}^n$ maps a p – differential form onto a $(n-p)$ – differential form, in general. Here $p = n, n - p = 0$ applies. Finally we define

$$f_x(x_1, \ldots, x_n) := f(g_1^{-1}(x_1, \ldots, x_n), \ldots, g_n^{-1}(x_1, \ldots, x_n))|\det \mathbf{J}_y|$$

as a function which is certainly *non-negative* and *integrated* over Ω_x to one. ♣

In applying the transformation theorems of *pdf* we meet quite often the problem that the function $x_i = g_i(y_1, \ldots, y_n)$ for all $i \in \{1, \ldots, n\}$ is given but not the inverse function $y_i = g_i^{-1}(x_1, \ldots, x_n)$ for all $i \in \{1, \ldots, n\}$. Then the following results are helpful.

Corollary B.1. *(*Jacobian*):*

If the inverse Jacobian $|\det \mathbf{J}_x| = |\det(\partial g_i/\partial y_j)|$ is given, we are able to compute

$$|\det \mathbf{J}_y| = |\det \frac{\partial g_i^{-1}(x_1, \ldots, x_n)}{\partial x_j}| = |\det \mathbf{J}|^{-1} = |\det \frac{\partial g_i(y_1, \ldots, y_n)}{\partial y_j}|^{-1}.$$

Example B.3. (Jacobian):

Let us continue Example B.2. The *inverse map*

$$y = \begin{bmatrix} g_1^{-1}(y_1, y_2) \\ g_2^{-1}(y_1, y_2) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_2/x_1 \end{bmatrix} \Rightarrow \frac{\partial y}{\partial x'} = \begin{bmatrix} 1 & 1 \\ -x_2/x_1^2 & 1/x_1 \end{bmatrix}$$

$$j_y = |\mathbf{J}_y| = \left|\frac{\partial y}{\partial x'}\right| = \frac{1}{x_1} + \frac{x_2}{x_1^2} = \frac{x_1 + x_2}{x_1^2}$$

$$j_x = |\mathbf{J}_x| = j_y^{-1} = |\mathbf{J}_y|^{-1} = \left|\frac{\partial x}{\partial y'}\right| = \frac{x_1^2}{x_1 + x_2} = \frac{x_1^2}{y_1}$$

allows us to compute the *Jacobian* $\mathbf{J}_x$ from $\mathbf{J}_y$. The *direct map*

$$x = \begin{bmatrix} g_1(y_1, y_2) \\ g_2(y_1, y_2) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1/(1 + y_2) \\ y_1 y_2/(1 + y_2) \end{bmatrix}$$

leads us to the final version of the *Jacobian*.

$$\boxed{j_x = |\mathbf{J}_x| = \frac{y_1}{(1 + y_2)^2}.}$$

For the special case that the *Jacobi matrix* is given in a *partitioned form*, the results of Corollary B.3 are useful.

Corollary B.2. *(Jacobian):*

If the *Jacobi matrix* $\mathbf{J}_x$ is given in the partitioned form

$$|\mathbf{J}_x| = \left(\frac{\partial g_i}{\partial y_j}\right) = \begin{bmatrix}\mathbf{U}\\ \mathbf{V}\end{bmatrix},$$

then

$$\begin{aligned}
\det \mathbf{J}_x &= \sqrt{|\det \mathbf{J}_x \mathbf{J}'_x|} = \sqrt{\det(\mathbf{U}\mathbf{U}')\det[\mathbf{V}\mathbf{V}' - (\mathbf{V}\mathbf{U}')(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V}']}\\
&\qquad \textit{if } \det(\mathbf{U}\mathbf{U}') \neq 0\\
\det \mathbf{J}_x &= \sqrt{|\det \mathbf{J}_x \mathbf{J}'_x|} = \sqrt{\det(\mathbf{V}\mathbf{V}')\det[\mathbf{U}\mathbf{U}' - \mathbf{U}\mathbf{V}'(\mathbf{V}\mathbf{V}')^{-1}\mathbf{V}\mathbf{U}']},\\
&\qquad \textit{if } \det(\mathbf{V}\mathbf{V}') \neq 0\\
|\det \mathbf{J}_y| &= |\det \mathbf{J}_x|^{-1}.
\end{aligned}$$

Proof. The Proof is based upon the *determinantal relations* of a partitioned matrix of type

$$\det\begin{bmatrix}\mathbf{A} & \mathbf{U}\\ \mathbf{V} & \mathbf{D}\end{bmatrix} = \det \mathbf{A} \det(\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{U})\ if\ \det \mathbf{A} \neq 0$$

$$\det\begin{bmatrix}\mathbf{A} & \mathbf{U}\\ \mathbf{V} & \mathbf{D}\end{bmatrix} = \det \mathbf{D} \det(\mathbf{A} - \mathbf{U}\mathbf{D}^{-1}\mathbf{U})\ if\ \det \mathbf{D} \neq 0$$

$$\det\begin{bmatrix}\mathbf{A} & \mathbf{U}\\ \mathbf{V} & \mathbf{D}\end{bmatrix} = \mathbf{D} \det \mathbf{A} - \mathbf{V}(\mathrm{adj}\mathbf{A})\mathbf{U},$$

which have been introduced by Frobenius (1908) and Schur (1917).

B-2 A Second Vehicle: Transformation of Random Variables

Previously we analyzed the transformation of the *pdf* under an *injective map* of random variables $y \mapsto g(y) = x$. Here we study the transformation of *polar coordinates* $[\phi_1, \phi_2, \ldots, \phi_{n-1}, r] \in \mathbb{Y}$ as parameters of an *Euclidian observation space* to *Cartesian coordinates* $[y_1, \ldots, y_n] \in \mathbb{Y}$. In addition we introduce the hypervolume element of a sphere $\mathbb{S}^{n-1} \subset \mathbb{Y}$, $\dim \mathbb{Y} = n$. First, we give three examples. Second, we summarize the general results in Lemma B.4.

Example B.4. (polar coordinates: "2d"):

Table B.1 collects characteristic elements of the transformation of polar coordinates (ϕ_1, r) of type "*longitude, radius*" to Cartesian coordinates (y_1, y_2), their domain and range, the planar elements dy_1, dy_2 as well as the circle $\mathbb{S}^1$ embedded into $\mathbb{E}^2 := \{\mathbb{R}^2, \delta_{kl}\}$, equipped with the canonical metric $I_2 = [\delta_{kl}]$ and its total measure of arc ω_1.

Table B.1

Cartesian and polar coordinates of a two-dimensional observation space, total measure of the arc of the circle

$$(\phi_1, r) \in [0, 2\pi] \times]0, \infty[, \ \ (y_1, y_2) \in \mathbb{R}^2$$

$$dy_1 dy_2 = r dr d\phi_1$$

$$\mathbb{S}^1 := \{\mathbf{y} \in \mathbb{R}^2 | y_1^2 + y_2^2 = 1\}$$

$$\omega_1 = \int_0^{2\pi} d\phi_1 = 2\pi.$$

Example B.5. (polar coordinates: "3d"):

Table B.2 is a collectors' item for characteristic elements of the transformation of polar coordinates (ϕ_1, ϕ_2, r) of type "*longitude, latitude, radius*" to Cartesian coordinates (y_1, y_2, y_3), their domain and range, the volume element dy_1, dy_2, dy_3 *as well as* of the sphere $\mathbb{S}^2$ embedded into $\mathbb{E}^3 := \{\mathbb{R}^3, \delta_{kl}\}$ equipped with the canonical metric $I_3 = [\delta_{kl}]$ and its total measure of surface ω_2.

Table B.2

Cartesian and polar coordinates of a three-dimensional observation space, total measure of the surface of the circle

$$y_1 = r \cos\phi_2 \cos\phi_1, \ \ y_2 = r \cos\phi_2 \sin\phi_1, \ \ y_3 = r \sin\phi_2$$

$$(\phi_1, \phi_2, r) \in [0, 2\pi] \times]-\frac{\pi}{2}, \frac{\pi}{2}[\times]0, r[, \ \ (y_1, y_2) \in \mathbb{R}^2$$

$$(y_1, y_2, y_3), \in \mathbb{R}^3$$

$$dy_1 dy_2 dy_3 = r^2 dr \cos\phi_2 d\phi_1 d\phi_2$$

$$\mathbb{S}^2 := \{\mathbf{y} \in \mathbb{R}^{\not\models} | y_1^2 + y_2^2 + y_3^2 = 1\}$$

$$\omega_2 = \int_0^{2\pi} d\phi_1 \int_{-\pi/2}^{+\pi/2} d\phi_2 \cos\phi_2 = 4\pi.$$

Example B.6. (polar coordinates: "4d"):

Table B.3 is a collection of characteristic elements of the transformation of polar coordinates $(\phi_1, \phi_2, \phi_3, r)$ to Cartesian coordinates (y_1, y_2, y_3, y_4), their domain and range, the hypervolume element dy_1, dy_2, dy_3, dy_4 as well as of the 3 - sphere $\mathbb{S}^3$ embedded into $\mathbb{E}^4 := \{\mathbb{R}^4, \delta_{kl}\}$ equipped with the canonical metric $I_4 = [\delta_{kl}]$ and its total measure of hypersurface.

Table B.3

Cartesian and polar coordinates of a four-dimensional observation space total measure of the hypersurface of the 3-sphere

$$y_1 = r\cos\phi_3\cos\phi_2\cos\phi_1,\ \ y_2 = r\cos\phi_3\cos\phi_2\sin\phi_1,$$

$$y_3 = r\cos\phi_3\sin\phi_2,\ \ y_4 = r\sin\phi_3$$

$$(\phi_1, \phi_2, \phi_3, r) \in [0, 2\pi]\times]-\frac{\pi}{2}, \frac{\pi}{2}[\times]-\frac{\pi}{2}, \frac{\pi}{2}[\times]0, 2\pi[$$

$$dy_1 dy_2 dy_3 dy_4 = r^3\cos^2\phi_3\cos\phi_2 dr d\phi_3 d\phi_2 d\phi_1$$

$$\mathbf{J}_y := \frac{\partial(y_1, y_2, y_3, y_4)}{\partial(\phi_1, \phi_2, \phi_3, r)}$$

$$= \begin{bmatrix} -r\cos\phi_3\cos\phi_2\sin\phi_1 & -r\cos\phi_3\sin\phi_2\cos\phi_1 & -r\sin\phi_3\cos\phi_2\cos\phi_1 & \cos\phi_3\cos\phi_2\cos\phi_1 \\ +r\cos\phi_3\cos\phi_2\cos\phi_1 & -r\cos\phi_3\sin\phi_2\sin\phi_1 & -r\sin\phi_3\cos\phi_2\sin\phi_1 & \cos\phi_3\cos\phi_2\sin\phi_1 \\ 0 & +r\cos\phi_3\cos\phi_2 & -r\sin\phi_3\sin\phi_2 & \cos\phi_3\cos\phi_2 \\ 0 & 0 & r\cos\phi_3 & \sin\phi_3 \end{bmatrix}$$

$$|\det\mathbf{J}_y| = r^3\cos^2\phi_3\cos\phi_2$$

$$\mathbb{S}^3 := \{y \in \mathbb{R}^4 | y_1^2 + y_2^2 + y_3^2 + y_4^2 = 1\}$$

$$\omega_3 = 2\pi^2.$$

Lemma B.2. *(polar coordinates, hypervolume element, hypersurface element):*

$$Let \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \cdots \\ y_{n-3} \\ y_{n-2} \\ y_{n-1} \\ y_n \end{bmatrix} = r \begin{bmatrix} \cos\phi_{n-1}\cdot\cos\phi_{n-2}\cdot\cos\phi_{n-3}\cdot\cdots\cdot\cos\phi_2\cdot\cos\phi_1 \\ \cos\phi_{n-1}\cdot\cos\phi_{n-2}\cdot\cos\phi_{n-3}\cdot\cdots\cdot\cos\phi_2\cdot\sin\phi_1 \\ \cos\phi_{n-1}\cdot\cos\phi_{n-2}\cdot\cos\phi_{n-3}\cdot\cdots\cdot\sin\phi_2 \\ \cos\phi_{n-1}\cdot\cos\phi_{n-2}\cdot\cdots\cdot\cos\phi_3 \\ \cdots \\ \cos\phi_{n-1}\cdot\cos\phi_{n-2}\cdot\sin\phi_{n-3} \\ \cos\phi_{n-1}\cdot\cos\phi_{n-2} \\ \cos\phi_{n-1}\cdot\sin\phi_{n-2} \\ \sin\phi_{n-1} \end{bmatrix}$$

be a transformation of polar coordinates $(\phi_1, \phi_2, \ldots, \phi_{n-2}, \phi_{n-1}, r)$ to Cartesian coordinates $(y_1, y_2, \ldots, y_{n-1}, y_n)$, their domain and range given by

$$(\phi_1, \phi_2, \ldots, \phi_{n-2}, \phi_{n-1}, r) \in [0, 2\pi] \times]-\frac{\pi}{2}, +\frac{\pi}{2}[\times \cdots \times]-\frac{\pi}{2},$$
$$+\frac{\pi}{2}[\times]-\frac{\pi}{2}, +\frac{\pi}{2}[\times]0, \infty[,$$

then the *local hypervolume element*

$$dy_1 \cdots dy_n$$
$$= r^{n-1} dr \cos^{n-2}\phi_{n-1} \cos^{n-3}\phi_{n-2} \ldots \cos^2\phi_3 \cos\phi_2 d\phi_{n-1} d\phi_{n-2} \ldots d\phi_3 d\phi_2 d\phi_1,$$

as well as the *global hypersurface element*

$$\omega_{n-1} = \frac{2 \cdot \pi^{(n-1)/2}}{\Gamma(\frac{n-1}{2})} := \int_{-\pi/2}^{+\pi/2} \cos\phi_{n-1}^{n-2} d\phi_{n-1} \cdots \int_{-\pi/2}^{+\pi/2} \cos\phi_2 d\phi_2 \int_0^{2\pi} d\phi_1,$$

where $\gamma(X)$ is the *gamma function*. Before we care for the proof, let us define *Euler's gamma* function.

Definition B.1. (Euler's gamma function):

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad (x > 0)$$

is Euler's gamma function which enjoys the recurrence relation

$$\Gamma(x+1) = x\Gamma(x)$$

subject to

$$\Gamma(1) = 1 \qquad or \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(2) = 1 \qquad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

$$\ldots \qquad \ldots$$

$$\Gamma(n+1) = n! \qquad \Gamma\left(\frac{p}{q}\right) = \frac{p-q}{q}\Gamma\left(\frac{p-q}{q}\right)$$

if n is integer, $n \in \mathbb{Z}^+$ *if* $\frac{p}{q}$ *is a rational number,* $p/q \in \mathbb{Q}^+$.

Example B.7. (Euler's gamma function):

(a) $\Gamma(1) = 1$

(b) $\Gamma(2) = 1$

(c) $\Gamma(3) = 1 \cdot 2 = 2$

(d) $\Gamma(4) = 1 \cdot 2 \cdot 3 = 6$

(i) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

(ii) $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$

(iii) $\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi}$

(iv) $\Gamma\left(\frac{7}{2}\right) = \frac{5}{2}\Gamma\left(\frac{5}{2}\right) = \frac{15}{8}\sqrt{\pi}.$

Proof. Our proof of Lemma B.4 will be based upon computing the image of the tangent space $\mathbb{T}_y\mathbb{S}^{n-1} \subset \mathbb{E}^n$ of the hypersphere $\mathbb{S}^{n-1} \subset \mathbb{E}^n$. Let us embed the hypersphere $\mathbb{S}^{n-1}$ parameterized by $(\phi_1, \phi_2, \ldots, \phi_{n-2}, \phi_{n-1})$ in $\mathbb{E}^n$ parameterized by $(y_1, \ldots, y_n)$, namely $y \in \mathbb{E}^n$,

$$\begin{aligned}
y = {} & \mathbf{e}_1 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2\cos\phi_1 \\
& + \mathbf{e}_2 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2\sin\phi_1 + \cdots \\
& + \mathbf{e}_{n-1} r \cos\phi_{n-1}\sin\phi_{n-2} \\
& + \mathbf{e}_n r \sin\phi_{n-1}.
\end{aligned}$$

Note that ϕ_1 is a parameter of *type* longitude, $0 \leq \phi_1 \leq 2\pi$, while $\phi_2, \ldots, \phi_{n-1}$ are parameters of *type* latitude, $-\pi/2 < \phi_2 < +\pi/2, \ldots, -\pi/2 < \phi_{n-1} < +\pi/2$ (*open intervals*). The images of the tangent vectors which span the local tangent space are given in the orthonormal n- leg $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{n-1}, \mathbf{e}_n | 0\}$ by

$$\begin{aligned}
g_1 := D_{\phi_1} y = & -\mathbf{e}_1 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2\sin\phi_1 \\
& + \mathbf{e}_2 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2\cos\phi_1 \\
g_2 := D_{\phi_2} y = & -\mathbf{e}_1 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\sin\phi_2\cos\phi_1 \\
& - \mathbf{e}_2 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\sin\phi_2\sin\phi_1 \\
& + \mathbf{e}_3 r \cos\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2 \\
& \cdots \\
g_{n-1} := D_{\phi_{n-1}} y = & -\mathbf{e}_1 r \sin\phi_{n-1}\cos\phi_{n-2}\cdots\cos\phi_2\cos\phi_1 - \cdots \\
& - \mathbf{e}_{n-1} r \sin\phi_{n-1}\sin\phi_{n-2} \\
& + \mathbf{e}_n r \cos\phi_{n-1}
\end{aligned}$$

$$
\begin{aligned}
g_n := D_{\phi_n} y = {} & \mathbf{e}_1 r \cos\phi_{n-1} \cos\phi_{n-2} \cdots \cos\phi_2 \cos\phi_1 \\
& + \mathbf{e}_2 r \cos\phi_{n-1} \cos\phi_{n-2} \cdots \sin\phi_2 \sin\phi_1 + \cdots \\
& + \mathbf{e}_{n-1} r \cos\phi_{n-1} \cos\phi_{n-2} \\
& + \mathbf{e}_n r \sin\phi_{n-1} = y/r.
\end{aligned}
$$

$\{g_1, \ldots, g_{n-1}\}$ span the image of the tangent space in $\mathbb{E}^n$. g_n is the *hypersphere normal vector*, $\|g_n\| = 1$. From the inner products $\langle g_i | g_j \rangle = g_{ij}$, $i, j \in \{1, \ldots, n\}$, we derive the *Gauss matrix of the metric* $\mathbf{G} := [g_{ij}]$.

$$
\begin{aligned}
\langle g_1 | g_1 \rangle &= r^2 \cos^2\phi_{n-1} \cos^2\phi_{n-2} \cdots \cos^2\phi_3 \cos^2\phi_2 \\
\langle g_2 | g_2 \rangle &= r^2 \cos^2\phi_{n-1} \cos^2\phi_{n-2} \cdots \cos^2\phi_3 \\
& \cdots \\
\langle g_{n-1} | g_{n-1} \rangle &= r^2, \\
\langle g_n | g_n \rangle &= 1.
\end{aligned}
$$

The off-diagonal elements of the *Gauss matrix of the metric* are zero. Accordingly

$$
\sqrt{\det \mathbf{G}_n} = \sqrt{\det \mathbf{G}_{n-1}} = r^{n-1} (\cos\phi_{n-1})^{n-2} (\cos\phi_{n-2})^{n-3} \cdots (\cos\phi_3)^2 \cos\phi_2.
$$

The square root $\sqrt{\det \mathbf{G}_n}$, $\sqrt{\det \mathbf{G}_{n-1}}$ elegantly represents the *Jacobian determinant*

$$
\mathbf{J}_y := \frac{\partial(y_1, y_2, \ldots, y_n)}{\partial(\phi_1, \phi_2, \ldots, \phi_{n-1}, r)} = \sqrt{\det \mathbf{G}_n}.
$$

Accordingly we have found the *local hypervolume element* $\sqrt{\det \mathbf{G}_n}\; dr\, d\phi_{n-1}\, d\phi_{n-2} \cdots d\phi_3\, d\phi_2\, d\phi_1$. For the *global hypersurface element* ω_{n-1}, we integrate

$$
\int_0^{2\pi} d\phi_1 = 2\pi
$$

$$
\int_{-\pi/2}^{+\pi/2} \cos\phi_2 d\phi_2 = [\sin\phi_2]_{-\pi/2}^{+\pi/2} = 2
$$

$$
\int_{-\pi/2}^{+\pi/2} \cos^2\phi_3 d\phi_3 = \frac{1}{2} [\cos\phi_3 \sin\phi_3 - \phi_3]_{-\pi/2}^{+\pi/2} = -\pi/2
$$

$$\int_{-\pi/2}^{+\pi/2} \cos^3 \phi_4 d\phi_4 = \frac{1}{3}[\cos^2 \phi_4 \sin \phi_4 - 2 \sin \phi_4]_{-\pi/2}^{+\pi/2} = -\frac{4}{3}$$

$$\dots$$

$$\int_{-\pi/2}^{+\pi/2} (\cos \phi_{n-1})^{n-2} d\phi_{n-1} = \frac{1}{n-2}[(\cos \phi_{n-1})^{n-3}]_{-\pi/2}^{+\pi/2}$$

$$+\frac{1}{n-3} \int_{-\pi/2}^{+\pi/2} (\cos \phi_{n-1})^{n-4} d\phi_{n-1}$$

recursively. As soon as we substitute the gamma function, we arrive at ω_{n-1}.

B-3 A First Confidence Interval of Gauss–Laplace Normally Distributed Observations μ, σ^2 Known, the Three Sigma Rule

The first confidence interval of *Gauss-Laplace normally distributed observations constrained* to (μ, σ^2) known, will be computed as an introductory example. An application is the *Three Sigma Rule*.
In the empirical sciences, estimates of certain quantities derived from observations are often given in the form of the estimate *plus or minus a certain amount*. For instance, the distance between a benchmark on the Earth's surface and a satellite orbiting the Earth may be estimated to be

$$\boxed{(20, 101, 104.132 \pm 0.023)m}$$

with the idea that the *first factor* is very unlikely to be outside the range

$$\boxed{20, 101, 104.155\, m \text{ to } 20, 101, 104.109m}.$$

A cost accountant for a publishing company in trying to allow for all factors which enter into the cost of producing a certain book,

actual production costs, proportion of plant overhead, proportion of executive salaries,

may estimate the cost to be $21 \pm 1, 1$ *Euro per volume* with the implication that the correct cost very probably lies between 19.9 and 22.1 Euro per volume. The *Bureau of Labor Statistics* may estimate the number of unemployed in a certain area to be 2.4 ± 0.3 million at a given time though intuitively it should be between 2.1 and 2.7 million. What we are saying is that in practice *we are quite accustomed* to seeing estimates in the form of *intervals*. In order to give precision to these ideas we shall consider a particular example. Suppose that a random sample $x \in \{\mathbb{R},\ pdf\}$ is taken from a *Gauss-Laplace normal distribution* with known mean μ and known variance σ^2. We ask the key question.

What is the probability γ of the random interval $(\mu - c\sigma, \mu + c\sigma)$ to cover the mean μ as a *quantile* c of the standard deviation σ?

To put this question into a mathematical form we write the *probabilistic two-sided interval identity*.

$$\boxed{\begin{gathered} P(x_1 < \mathrm{X} < x_2) = P(\mu - c\sigma < \mathrm{X} < \mu + c\sigma) = \gamma, \\ \int\limits_{x_1=\mu-c\sigma}^{x_2=\mu+c\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \gamma \end{gathered}}$$

with a *left boundary* $l = x_1$ and a *right boundary* $r = x_2$. The length of the interval is $x_2 - x_1 = r - l$. The *center* of the interval is $(x_1 + x_2)/2$ or μ. Here we have taken advantage of the *Gauss-Laplace pdf* in generating the *cumulative* probability

$$\boxed{\begin{aligned} &\mathrm{P}(x_1 < \mathrm{X} < x_2) = \mathrm{F}(x_2) - \mathrm{F}(x_1) \\ &\quad \mathrm{F}(x_2) - \mathrm{F}(x_1) = \mathrm{F}(\mu + c\sigma) - \mathrm{F}(\mu + c\sigma). \end{aligned}}$$

Typical values for the confidence coefficient γ are $\gamma = 0.95$ ($\gamma = 95\%$ or $1-\gamma = 5\%$ negative confidence), $\gamma = 0.99$ ($\gamma = 1\%$ or $1 - \gamma = 1\%$ negative confidence) or $\gamma = 0.999$ ($\gamma = 999‰$ or $1 - \gamma = 1‰$ negative confidence).
Consult Fig. B.1 for a geometric interpretation. The confidence coefficient γ is a measure of the probability mass between $x_1 = \mu - c\sigma$ and $x_2 = \mu + c\sigma$. For a given confidence coefficient γ

$$\boxed{\int\limits_{x_1}^{x_2} f(x|\mu, \sigma^2) dx = \gamma}$$

establishes an *integral equation*. To make this point of view to be better under-stood let us transform the integral equations to its *standard form*.

$$\boxed{\begin{gathered} x \mapsto z = \tfrac{1}{\sigma}(x - \mu) \Leftrightarrow x = \mu + \sigma z \\ \int\limits_{x_1=\mu-c\sigma}^{x_2=\mu+c\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \int\limits_{-c}^{+c} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz = \gamma \\ \int\limits_{x_1}^{x_2} f(x|\mu, \sigma^2) dx = \int\limits_{-c}^{+c} f(z|0,\ 1) dz = \gamma. \end{gathered}}$$

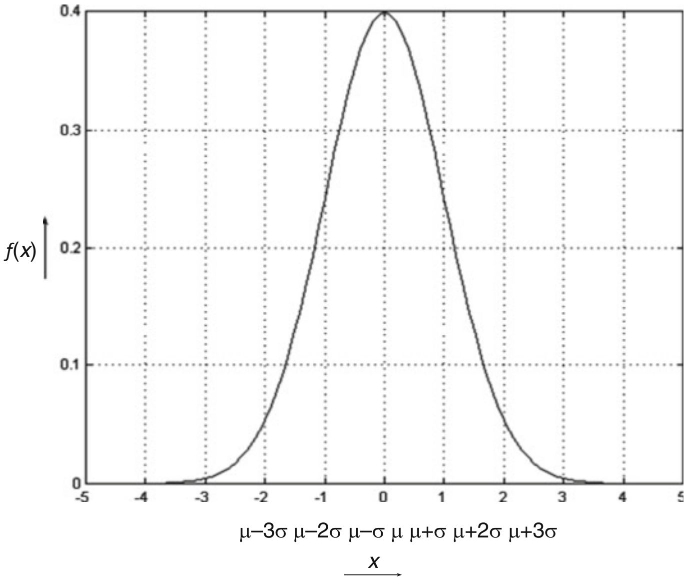


Fig. B.1 Probability mass in a two-sided confidence interval $x_1 < \mathrm{X} < x_2$ or $\mu - c\sigma < \mathrm{X} < \mu + c\sigma$, three cases: (**a**) $c = 1$, (**b**) $c = 2$ and (**c**) $c = 3$

The *special Helmert transformation* maps x to z, now being *standard Gauss-Laplace normal*:σ^{-1} is the *dilatation factor*, also called scale variation, but μ the translation parameter.
The *Gauss-Laplace pdf* is symmetric, namely $f(-x) = f(+x)$ or $f(-z) = f(+z)$. Accordingly we can write the integral identity

$$\boxed{\begin{aligned}&\int_{x_1}^{x_2} f(x|\mu,\sigma^2)dx = 2\int_0^{x_2} f(x|\mu,\sigma^2)dx = \gamma\\ \Leftrightarrow &\int_{-c}^{+c} f(z|0,1)dz = 2\int_0^{c} f(z|0,1)dz = \gamma.\end{aligned}}$$

The classification of *integral equations* tells us that

$$\boxed{\gamma(z) = 2\int_0^z f(z^*)dz^*}$$

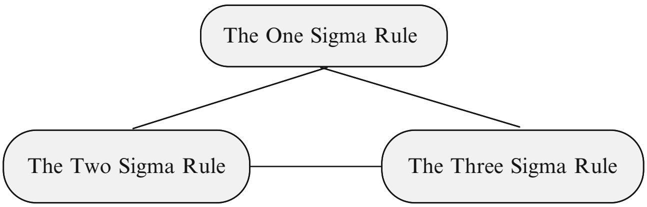
is a *linear Volterra integral equation the first kind*.

In case of a *Gauss-Laplace standard normal pdf*, such an integral equation is solved by a table. In a *forward computation*

$$\boxed{\mathrm{F}(z) := \int_{-\infty}^{z} f(z^*|0,1)dz^* \quad or \quad \Phi(z) := \int_{-\infty}^{z} \frac{1}{2\pi}\exp\left(-\frac{1}{2}z^{*^2}\right)dz^*}$$

are tabulated in a regular grid. For a given value $F(z_1)$ or $F(z_2)$, z_1 or z_2 are determined by interpolation. *C.F. Gauss* did *not* use such a procedure. He took advantage of the Gauss inequality which has been reviewed in this context by Pukelsheim (1994). There also the *Vysochanskii-Petunin inequality* has been discussed. We follow here a two-step procedure. *First*, we divide the domain $z \in [0, \infty]$ into two intervals $z \in [0, 1]$ and $z \in [1, \infty]$. In the first interval $f(z)$ is isotonic, differentiable and *convex*, $f''(z) = f(z)(z^2 - 1) < 0$, while in the second interval isotonic, differentiable and concave, $f''(z) = f(z)(z^2-1) > 0$. $z = 1$ is the point of inflection. *Second*, we setup Taylor series of $f(z)$ in the interval $z \in [0, 1]$ at the point $z = 0$, while in the interval $z \in [1, \infty]$ at the point $z = 1$ and $z \in [2, \infty]$ at the point $z = 2$.

Three examples of such a forward solution of the characteristic linear *Volterra integral equation of the first kind* will follow. They establish:



Box B.1. (Operational calculus applied to the Gauss-Laplace probability distribution):

"*generating differential equations*"

$$f''(z) + 2f'(z) + f(z) = 0$$

subject to

$$\int_{-\infty}^{+\infty} f(z)dz = 1$$

"*recursive differentiation*"

$$
\begin{aligned}
f(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \\
f'(z) &= -z f(z) =: g(z) \\
f''(z) &= g'(z) = -f(z) - zg(z) = (z^2 - 1) f(z) \\
f'''(z) &= 2z f(z) + (z^2 - 1) g(z) = (-z^3 + 3z) f(z) \\
f^{(4)}(z) &= (-3z^2 + 3) f(z) + (-z^3 + 3z) g(z) = (z^4 - 6z^2 + 3) f(z) \\
f^{(5)}(z) &= (4z^3 - 12z) f(z) + (z^4 - 6z^2 + 3) g(z) = (-z^5 + 10z^3 - 15z) f(z) \\
f^{(6)}(z) &= (-5z^4 + 30z^2 - 15) f(z) + (-z^5 + 10z^3 - 15z) g(z) \\
&= (z^6 - 15z^4 + 45z^2 - 15) f(z) \\
f^{(7)}(z) &= (6z^5 - 60z^3 + 90z) f(z) + (z^6 - 15z^4 + 45z^2 - 15) g(z) \\
&= (-z^7 + 21z^5 - 105z^3 + 105z) f(z) \\
f^{(8)}(z) &= (-7z^6 + 105z^4 - 315z^2 + 105) f(z) \\
&\quad + (-z^7 + 21z^5 - 105z^3 + 105z) f(z) \\
&= (z^8 - 28z^6 + 210z^4 - 420z^2 + 105) f(z) \\
f^{(9)}(z) &= (8z^7 - 168z^5 + 840z^3 - 840z) f(z) \\
&\quad + (z^8 - 28z^6 + 210z^4 - 420z^2 + 105) g(z) \\
&= (-z^9 + 36z^7 - 378z^5 + 1260z^3 - 945z) f(z) \\
f^{(10)}(z) &= (-9z^8 + 252z^6 - 1890z^4 + 3780z^2 - 945) f(z) + \\
&\quad + (-z^9 + 36z^7 - 378z^5 + 1260z^3 - 945) g(z) \\
&= (z^{10} - 45z^8 + 630z^6 - 3150z^4 + 4725z^2 - 945) f(z)
\end{aligned}
$$

"*upper triangle representation of the matrix transforming* $f(z) \to f^n(z)$"

$$
\begin{bmatrix} f(z) \\ f'(z) \\ f''(z) \\ f'''(z) \\ f^{(4)}(z) \\ f^{(5)}(z) \\ f^{(6)}(z) \\ f^{(7)}(z) \\ f^{(8)}(z) \\ f^{(9)}(z) \\ f^{(10)}(z) \end{bmatrix}
= f(z)
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & -6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -15 & 0 & 10 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
-15 & 0 & 45 & 0 & -15 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 105 & 0 & -105 & 0 & 21 & 0 & -1 & 0 & 0 & 0 \\
105 & 0 & -420 & 0 & 210 & 0 & -28 & 0 & 1 & 0 & 0 \\
0 & -945 & 0 & 1260 & 0 & -378 & 0 & 36 & 0 & -1 & 0 \\
-945 & 0 & -4725 & 0 & -3150 & 0 & 630 & 0 & -45 & 0 & 1
\end{bmatrix}
\begin{bmatrix} 1 \\ z \\ z^2 \\ z^3 \\ z^4 \\ z^5 \\ z^6 \\ z^7 \\ z^8 \\ z^9 \\ z^{10} \end{bmatrix} .
$$

B-31 *The Forward Computation of a First Confidence Interval of Gauss–Laplace Normally Distributed Observations: μ, σ^2 Known*

We can avoid solving the linear *Volterra integral equation* of the first kind if we push forward the integration for a fixed value z.

Example B.8. (Series expansion of the Gauss-Laplace integral, first interval):

Let us solve the integral

$$\gamma(z=1) := 2\int_0^1 f(z^*)dz^*$$

in the *first interval* $0 \le z \le 1$ by *Taylor expansion* with respect to the successive differentiation of $f(z)$ outlined in Box B.1 and the specific derivatives $f^n(0)$ given in Table B.4. Based on those auxiliary results, Box B.2 presents us the detailed interpretation. First, we expand $\exp(-z^2/2)$ up to order $O(14)$. The specific Taylor series are uniformly convergent. Accordingly, in order to compute the integral, second we integrate termwise up to order $O(15)$. For the specific value $z = 1$, we have computed the *coefficient of confidence* $\gamma(1) = 0.683$. The result

$$\boxed{\mathrm{P}(\mu - \sigma < \mathrm{X} < \mu + \sigma) = 0.683}$$

is known as the *One Sigma Rule*.

68.3% of the sample are in the interval $]\mu - 1\sigma, \mu + 1\sigma[$, 0.317% outside. If we make three experiments, one experiment is outside the one interval.

Box B.2. A specific integral

"*expansion of the exponential function*"

$$\exp x = 71 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \mathcal{O}(\mathrm{n}) \quad \forall\, |\mathrm{x}| < \infty$$

$$\mathrm{x} := -\frac{1}{2}z^2$$

$$\exp\left(-\frac{1}{2}z^2\right) = 1 - \frac{1}{2}z^2 + \frac{1}{8}z^4 - \frac{1}{48}z^6 + \frac{1}{384}z^8 - \frac{1}{3840}z^{10} + \frac{1}{46080}z^{12} + \mathcal{O}(14)$$

"*the specific integral*"

$$\int_0^z f(z^*)dz^* = \frac{1}{\sqrt{2\pi}}\int_0^z \exp\left(-z^{*^2}\right)dz^* = \frac{1}{\sqrt{2\pi}}\left(z - \frac{1}{6}z^3 + \frac{1}{40}z^5 - \frac{1}{336}z^7 + \frac{1}{3456}z^9 - \frac{1}{42240}z^{11} + \frac{1}{599040}z^{13} + \mathcal{O}(15)\right)$$

"*the specific values $z = 1$*"

$$\begin{aligned}\gamma(1) &= 2\int_0^1 f(z)dz \\ &= \frac{2}{\sqrt{2\pi}}\left(1 - \frac{1}{6} + \frac{1}{40} - \frac{1}{336} + \frac{1}{3456} - \frac{1}{42240} + \frac{1}{599040} + \mathcal{O}(15)\right) \\ &= \frac{2}{\sqrt{2\pi}}(1 - 0.166{,}667 + 0.025{,}000) \\ &\quad -0.002{,}976 + 0.000{,}289 \\ &\quad -0,000{,}024 + 0.000{,}002) \\ &= \frac{2}{\sqrt{2\pi}}0.855{,}624 = 0.682{,}689 \\ &= 0.683\end{aligned}$$

"*coefficient of confidence*"

$$0.683 = 1 - 317.311 * 10^{-3} = 1 - \frac{1}{3}.$$

Table B.4 (Special values of derivatives- Gauss-Laplace probability distribution):

$$z = 0$$

$$\frac{1}{\sqrt{2\pi}} = 0.398,942$$

$$f(0) = \frac{1}{\sqrt{2\pi}},\ f'(0) = 0,$$

$$f''(0) = -\frac{1}{\sqrt{2\pi}},\ \frac{1}{2!}f''(0) = -0.199,471$$

$$f'''(0) = 0, \;\; f^{(4)}(0) = +\frac{3}{\sqrt{2\pi}}, \;\; \frac{1}{4!} f^{(4)}(0) = +0.049, 868$$

$$f^{(5)}(0) = 0, \;\; f^{(6)}(0) = -\frac{15}{\sqrt{2\pi}}, \;\; \frac{1}{6!} f^{(6)}(0) = -0.008, 311.$$

Example B.9. (Series expansion of the Gauss-Laplace integral, 2nd interval):

Let us solve the integrals

$$\gamma(z=2) := 2\int_0^2 f(z^*)dz^* = 2\int_0^1 f(z^*)dz^* + 2\int_1^2 f(z^*)dz^*$$

$$\gamma(z=2) \;=\; \gamma(z=1) + 2\int_1^2 f(z^*)dz^*,$$

namely in the second interval $1 \leq z \leq 2$. *First*, we setup *Taylor series* of $f(z)$ "around the point $z = 1$". The derivatives of $f(z)$ "at the point z=*1*" are collected up to order 10 in Table B.5. *Second*, we integrate the *Taylor series* termwise and receive the specific integral of Box B.3. Note that termwise integrated is permitted since the Taylor series are uniformly convergent. The detailed computation up to order O(12) has led us to the *coefficient of confidence* $\gamma(2) = 0.954$. The result

$$\mathrm{P}(\mu - 2\sigma < \mathrm{X} < \mu + 2\sigma) = 0.954$$

is known as the *Two Sigma Rule*.

95.4% of the sample interval $]\mu - 2\sigma, \mu + 2\sigma[$, 0.046% outside. If we make 22 experiments, *one experiment* is outside the 2σ interval.

Box B.3. (A specific integral):

"*expansion of the exponential function*"

$$f(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

$$f(z) \;=\; f(1) + \frac{1}{1!}f'(1)(z-1) + \frac{1}{2!}f''(1)(z-1)^2 + \cdots + \frac{1}{10!}f^{10}(1)(z-1)^{10} + \mathrm{O}(11)$$

"*the specific integral*"

$$\int_1^2 f(z^*)dz^* = f(1)(z-1) + \frac{1}{2}\frac{1}{1!}f'(1)(z-1)^2 + \frac{1}{3}\frac{1}{2!}f''(1)(z-1)^3$$
$$+ \frac{1}{4}\frac{1}{3!}f'''(1)(z-1)^4 + \frac{1}{5}\frac{1}{4!}f^{(4)}(1)(z-1)^5$$
$$+ \frac{1}{6}\frac{1}{5!}f^{(5)}(1)(z-1)^6 + \cdots + \frac{1}{11}\frac{1}{10!}f^{(10)}(1)(z-1)^{11} + O(12)$$

case $z = 2$

$$\begin{aligned}\gamma(2) = \gamma(1) + 2\int_1^2 f(z)dz =&\gamma(1) + 2(0.241,971 \\ &- 0.120,985 + 0.020,122 - 0.004,024 \\ &- 0.002,016 + 0.000,768 + 0.000,120 \\ &- 0.000,088 - 0.000,002 - 0.000,050 + O(12)) \\ =& 0.682,690 + 0.271,632 \\ =& 0.954\end{aligned}$$

"*coefficient of confidence*"

$$0.954 = 1 - 45.678 * 10^{-3} = 1 - \frac{1}{22}.$$

Table B.5 (Special values of derivatives – Gauss-Laplace probability distribution):

$$z = 1$$
$$\frac{1}{\sqrt{2\pi}} = 0.398,942, \ \exp\left(-\frac{1}{2}\right) = 0.606,531$$
$$f(1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right) = 0.241,971$$
$$f'(1) = -f(1) = -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right) = -0.241,971, \ \ f''(1) = 0$$
$$f'''(1) = 2f(1) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\right) = 0.482,942$$

$$\frac{1}{3!}f'''(1) = +0.080,490$$
$$f^{(4)}(1) = -2f(1) = -0.482,942$$
$$\frac{1}{4!}f^{(4)}(1) = -0.020,122$$
$$f^{(5)}(1) = -6f(1) = -\frac{6}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\right) = -1.451,826$$
$$\frac{1}{5!}f^{(5)}(1) = -0.012,098$$
$$f^{(6)}(1) = 16f(1) = \frac{16}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\right) = 3.871,536$$
$$\frac{1}{6!}f^{(6)}(1) = 0.005,377$$
$$f^{(7)}(1) = 20f(1) = \frac{20}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\right) = 4.829,420$$
$$\frac{1}{7!}f^{(7)}(1) = 0.000,958$$
$$f^{(8)}(1) = -132f(1) = -31.940,172$$
$$\frac{1}{8!}f^{(8)}(1) = -0.000,792$$
$$f^{(9)}(1) = -28f(1) = -6.775,188$$
$$\frac{1}{9!}f^{(9)}(1) = -0.000,019$$
$$f^{(10)}(1) = -8234f(1) = -1992.389$$
$$\frac{1}{10!}f^{(10)}(1) = -0.000,549.$$

Example B.10. (Series expansion of the Gauss-Laplace integral, third interval):

Let us solve the integrals

$$\gamma(z=3) := 2\int_0^3 f(z)dz = 2\int_0^1 f(z)dz + 2\int_1^2 f(z)dz + 2\int_2^3 f(z)dz$$
$$\gamma(z=3) \;=\; \gamma(z=1) + \gamma(z=2) + 2\int_2^3 f(z)dz,$$

namely in the third interval $2 \leq z \leq 3$. *First*, we setup *Taylor series* of *f(z)* "around the point $z = 2$". The derivatives of $f(z)$ "at the point $z = 2$" are collected up to order 10 in Table B.6. Second, we integrate the *Taylor series* termwise and receive the specific integral of Box B.4. Note that termwise integration is permitted since the *Taylor series* are uniformly convergent. The detailed computation up to order O(12) leads us to the *coefficient of confidence* $\gamma(3) = 0.997$. The result

$$P(\mu - 3\sigma < \mathrm{X} < \mu + 3\sigma) = 0.997$$

is known as the *Three Sigma Rule*.

99.7% of the sample are in the interval $]\mu - 3\sigma, \ \mu + 3\sigma[$, 0.003% outside. If we make 377 experiments, *one experiment* is outside the 3σ interval.

Table B.6 (Special values of derivatives Gauss-Laplace probability distribution):

$$z = 2$$

$$\frac{1}{\sqrt{2\pi}} = 0.398,942, \quad \exp(-2) = 0.135,335$$

$$f(2) = \frac{1}{\sqrt{2\pi}} \exp(-2) = 0.053,991$$

$$f'(2) = -2f(2) = -0.107,982$$

$$f''(2) = 3f(2), \quad \frac{1}{2!} f''(2) = +0.080,986$$

$$f'''(2) = -2f(2), \quad \frac{1}{3!} f'''(2) = -0.017,997$$

$$f^{(4)}(2) = -5f(2), \quad \frac{1}{4!} f^{(4)}(2) = -0.011,248$$

$$f^{(5)}(2) = 18f(2), \quad \frac{1}{5!} f^{(5)}(2) = +0.008,099$$

$$f^{(6)}(2) = -11f(2), \quad \frac{1}{6!} f^{(6)}(2) = -0.000,825$$

$$f^{(7)}(2) = -86f(2), \quad \frac{1}{7!} f^{(7)}(2) = -0.000,921$$

$$f^{(8)}(2) = +249f(2), \quad \frac{1}{8!} f^{(8)}(2) = +0.000,333$$

$$f^{(9)}(2) = 190f(2), \quad \frac{1}{9!} f^{(9)}(2) = +0.000,028$$

$$f^{(10)}(2) = -2621f(2), \quad \frac{1}{10!} f^{(10)}(2) = -0.000,039.$$

Box B.4 (A specific integral)

"*expansion of the exponential function*"

$$f(z) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$

$$f(z) = f(2) + \frac{1}{1!}f'(2)(z-2) + \frac{1}{3!}f''(2)(z-2)^2 + \cdots + \frac{1}{10!}f^{(10)}(2)(z-2)^{10} + \mathcal{O}(11)$$

"*the specific integral*"

$$\int_2^z f(z^*)dz^* = f(2)(z-2) + \frac{1}{2}\frac{1}{1!}f'(2)(z-2)^2 + \frac{1}{3}\frac{1}{2!}f''(2)(z-2)^3 + \frac{1}{4}\frac{1}{3!}f'''(2)(z-2)^4 + \frac{1}{5}\frac{1}{4!}f^{(4)}(2)(z-2)^5 + \frac{1}{6}\frac{1}{5!}f^{(5)}(2)(z-2)^6 + \cdots + \frac{1}{11}\frac{1}{10!}f^{(10)}(2)(z-2)^{11} + \mathcal{O}(12)$$

case $z = 3$

$$\gamma(3) = \gamma(1) + \gamma(2) + 2\int_2^3 f(z)dz$$

$$= 0.682,690 + 0.271,672 + 2(0.053,991 - 0.053,991 + 0.026,995 - 0.004,499 - 0.002,250 + 0.001,350 - 0.000,118 + 0.000,037 + 0.000,003 - 0.000,004 + O(12)) = 0.682,690 + 0.271,632 + 0.043,028 = 0.997$$

"coefficient of confidence"

$$0.997 = 1 - 2.65 * 10^{-3} = 1 - \frac{1}{377}.$$

B-32 The Backward Computation of a First Confidence Interval of Gauss–Laplace Normally Distributed Observations: μ, σ^2 Known

Finally we solve the *Volterra integral equation* of the first kind by the technique of *series inversion*, also called *series reversion*. Let us recognize that the interval integration of a Taylor series expanded Gauss-Laplace normal density distribution led us to a univariate homogeneous polynomial of arbitrary order. Such a *univariate homogeneous polynomial* $\mathbf{y} = a_1 x + a_2 x^2 + \cdots + a_n x^n$ ("*input*") can be reversed as a univariate homogeneous polynomial $\mathbf{x} = b_1 y + b_2 y^2 + \cdots + b_n y^n$ ("*output*") as outlined in Table B.7. Consult Abramowitz and Stegun (1965), p. 16 for a review, but Grafarend (1996) for a derivation based upon Computer Algebra.

Table B.7 (Series inversion (Grafarend, 1996)):

"*input*: univariate homogeneous polynomial"

$$\mathbf{y} = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6 + a_7 x^7 + O(x^8)$$

"*output*: reverse univariate homogeneous polynomial"

$$\mathbf{x} = b_1 y + b_2 y^2 + b_3 y^3 + b_4 y^4 + b_5 y^5 + b_6 y^6 + b_7 y^7 + O(y^8)$$

"*coefficient relation*"

(a) $a_1 b_1 = 1$

(b) $a_1^3 b_2 = -a_2$

(c) $a_1^5 b_3 = 2a_2^2 - a_1 a_3$

(d) $a_1^7 b_4 = 5a_1 a_2 a_3 - a_1^2 a_4 - 5a_2^3$

(e) $a_1^9 b_5 = 6a_1^2 a_2 a_4 + 3a_1^2 a_3^2 + 14a_2^4 - a_1^3 a_5 - 21 a_1 a_2^2 a_3$

(f) $a_1^{11} b_6 = 7a_1^3 a_2 a_5 + 7a_1^3 a_2 a_4 + 84 a_1 a_2^3 a_3 - a_1^4 a_6 - 28 a_1 a_2 a_3^2 - 42 a_2^5 - 28 a_1^2 a_2^2 a_4$

(g) $a_1^{13} b_7 = 6a_1^4 a_2 a_6 + 8a_1^4 a_2 a_5 + 4a_1^4 a_4^2 + 120 a_1^2 a_2^3 a_4 + 180 a_1^2 a_2^2 a_3^2 + 132 a_2^6$
$- a_1^5 a_7 - 36 a_1^3 a_2^2 a_5 - 72 a_1^3 a_2 a_3 a_4 - 12 a_1^3 a_3^3 - 330 a_1 a_2^4 a_3.$

Example B.11. (Solving the Volterra integral equation of the first kind):

Let us define the *coefficient of confidence* $\gamma = 0.90$ or 90%. We want to know the quantile $c_{0.90}$ which determines the probability identity

$$P(\mu - c\sigma < \mathrm{X} < \mu + c\sigma) = 0.90.$$

If you follow the detailed computation of Table B.5, namely the input as well as the output data up to order O(5), you find the *quantile*

$$c_{0.90} = 1.64,$$

as well as the confidence interval

$$P(\mu - 1.64\sigma < \mathrm{X} < \mu + 1.64\sigma) = 0.90.$$

90% of the sample are in the interval $]\mu - 1.64\sigma, \mu + 1.64\sigma[$, 10% outside. If we make 10 experiments one experiment is outside the 1.62σ interval.

Table B.8 (Series inversion quantile $c_{0.90}$):

(i) *input*

$$\gamma(z) = 2\int_0^1 f(z^*)dz^* + 2\int_0^z f(z^*)dz^* = 0.682,689 + 2[f(1)(z-1)$$

$$+ \frac{1}{2}\frac{1}{1!}f'(1)(z-1)^2 + \cdots + \frac{1}{n}\frac{1}{(n-1)!}f^{(n-1)}(1)(z-1)^n + O(n+1)]$$

$$\frac{1}{2}[\gamma(z) - 0.682,689] = f(1)(z-1) + \frac{1}{2}\frac{1}{1!}f'(1)(z-1)^2 + \cdots$$

$$+ \frac{1}{n}\frac{1}{(n-1)!}f^{(n-1)}(1)(z-1)^n + O(n+1)]$$

$$\mathbf{y} = a_1 x + a_2 x^2 + \cdots + a_n x^n + O(n+1)$$

$$x := z - 1$$

$$y := (0.900,000 - 0.682,689)/2 = 0.108,656$$

$$a_1 := f(1) = 0.241,971$$

$$a_2 := \frac{1}{2}\frac{1}{1!}f'(1) = -0.241,971$$

$$a_3 := \frac{1}{3}\frac{1}{2!}f''(1) = 0$$

$$a_4 := \frac{1}{4}\frac{1}{3!}f'''(1) = 0.020,125$$

$$a_5 := \frac{1}{5}\frac{1}{4!}f^{(4)}(1) = -0.004,024$$

$$\ldots$$

$$a_n := \frac{1}{n}\frac{1}{(n-1)!}f^{(n-1)}(1).$$

(ii) *output*

$$b_1 = \frac{1}{a_1} = 4.132,726$$

$$b_2 = -\frac{1}{a_1^3}a_2 = 8.539,715$$

$$b_3 = \frac{1}{a_1^5}(2a_2^2 - a_1 a_3) = 35.292,308$$

$$b_4 = \frac{1}{a_1^7}(5a_1 a_2 a_3 - a_1^2 a_4 - 5a_2^3) = 158.070$$

$$b_5 = \frac{1}{a_1^9}(6a_1^2 a_2 a_4 + 3a_1^2 a_3^2 + 14a_2^4 - a_1^3 a_5 - 21 a_1 a_2^2 a_3) = 475.452,152$$

$$b_1 y = 0.449,045, \;\; b_2 y^2 = 0.100,821,$$

$$b_3 y^3 = 0.045,273, \;\; b_4 y^4 = 0.022,032,$$

$$b_5 y^5 = 0.007,201$$

$$\mathbf{x} = b_1 y + b_2 y^2 + b_3 y^3 + b_4 y^4 + b_5 y^5 + O(6) = 0.624,372$$

$$\boxed{\mathbf{z} = \mathbf{x} + 1 = 1.624,372 = c_{0.90}.}$$

At this end we would like to give some sample references on computing the "inverse error function"

$$y = F(x) := \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz =: erfx \quad versus$$

$$x = F^{-1}(y) = inverfy,$$

namely Carlitz (1963) and Strecok (1968).

B-4 Sampling from the Gauss–Laplace Normal Distribution: A Second Confidence Interval for the Mean, Variance Known

The second confidence interval of *Gauss-Laplace i.i.d. observations* will be constructed for the mean $\widehat{\mu}$ BLUUE of μ, when the variance σ^2 is known. "n" is the size of the sample, namely to agree to the number of observations. Before we present the general sampling distribution we shall work through two examples. *Example B.12* has been chosen for a sample size $n = 2$, while *Example B.12* for $n = 3$ observations. Afterwards the general result is obvious and sufficiently motivated.

Example B.12. (Gauss-Laplace i.i.d. observations, observation space $\mathbb{Y}$, dim $\mathbb{Y}$)

In order to derive the marginal distributions of $\widehat{\mu}$ BLUUE of μ and $\widehat{\sigma}^2$ BIQUUE of σ^2 for a two dimensional Euclidean observation space we have to introduce

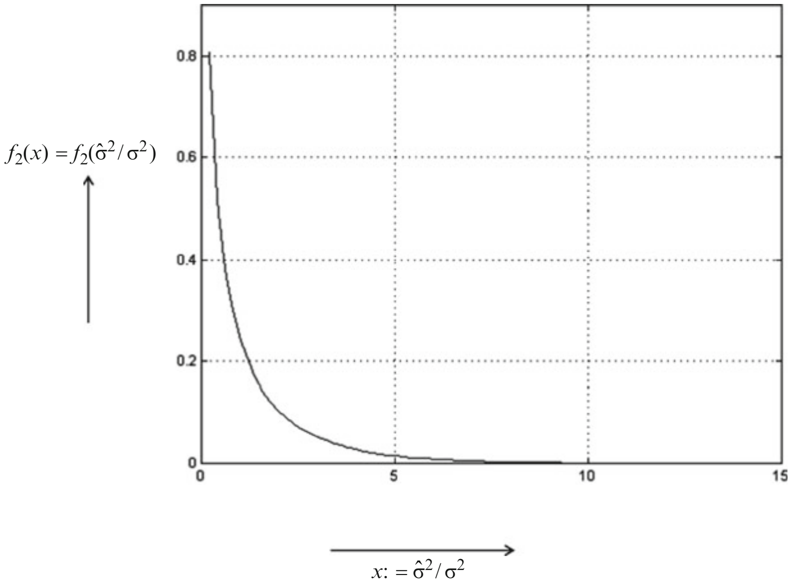


Fig. B.2 Special Helmert pdf of χ_p^2 for one degree of freedom, $p = 1$

various images. *First*, we define the probability function of two Gauss-Laplace i.i.d. observations and consequently implement the basic $(\widehat{\mu}, \widehat{\sigma}^2)$ decomposition into the *pdf*. *Second*, we separate the quadratic form $(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu)$ into the quadratic form $(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})$, the vehicle to introduce $\widehat{\sigma}^2$ BIQUUE of σ^2, and into the quadratic form $(\widehat{\mu} - \mu)^2$ the vehicle to bring in $\widehat{\mu}$ BLUUE of μ. *Third*, we aim at transforming the quadratic form $(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) = \mathbf{y}'\ \mathbf{My}$, $\operatorname{rk} \mathbf{M} = 1$, into the special form $1/2(y_1 - y_2)^2 =: x$. *Fourth*, we generate the marginal distributions $f_1(\widehat{\mu}\,|\mu, \sigma^2/2)$ of the mean $\widehat{\mu}$ BLUUE of μ and $f_2(\sigma^2)$ of the sample variance $\widehat{\sigma}^2$ BIQUUE of σ^2. The basic results of the example are collected in Corollary B.6. The special Helmert pdf χ^2 with one degree of freedom is plotted in Fig. B.2, but the *special Gauss-Laplace normal pdf* of variance $\sigma^2/2$ in Fig. B.3.

The first action item

Let us assume an experiment of *two Gauss-Laplace i.i.d.* observations. Their pdf is given by

$$f(y_1, y_2) = f(y_1)f(y_2),$$

$$f(y_1, y_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}[(y_1 - \mu)^2 + (y_2 - \mu)^2]\right),$$

$$f(y_1, y_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu)\right).$$

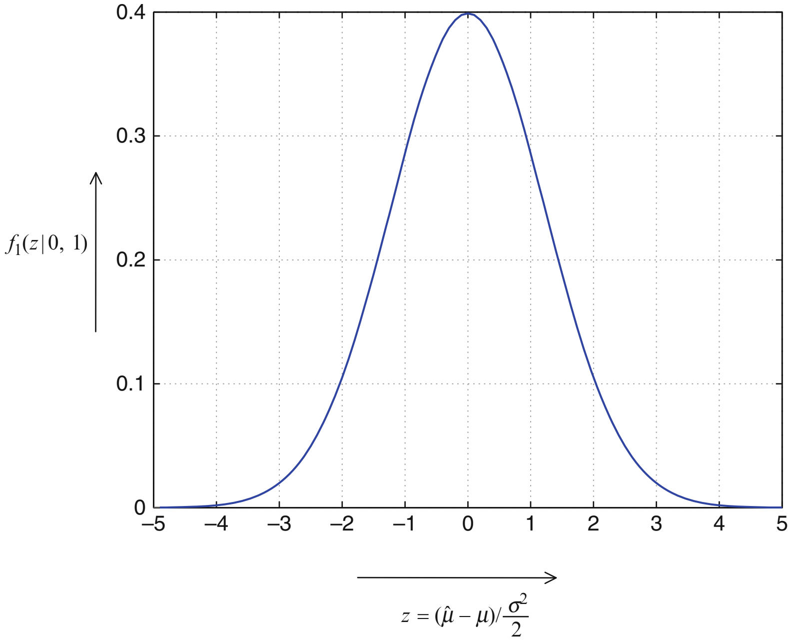


Fig. B.3 Special *Gauss-Laplace normal* pdf of $(\widehat{\mu} - \mu)/\sigma^2/2$

The second action item

The coordinates of the observation vector have been denoted by $[y_1, y_2]' = \mathbf{y} \in \mathbb{Y}$, $\dim \mathbb{Y} = 2$. The *quadratic form* $(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu)$ allows the *fundamental decomposition*

$$\begin{aligned}
\mathbf{y} - \mathbf{1}\mu &= (\mathbf{y} - \mathbf{1}\widehat{\mu}) + \mathbf{1}(\widehat{\mu} - \mu) \\
(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) &= (\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) + \mathbf{1}'\mathbf{1}(\widehat{\mu} - \mu)^2 \\
(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) &= \widehat{\sigma}^2 + 2(\widehat{\mu} - \mu)^2.
\end{aligned}$$

Here, $\widehat{\mu}$ is BLUUE of μ and $\widehat{\sigma}^2$ BIQUUE of σ^2. The detailed computation proves our statement.

$$\begin{aligned}
&[(y_1 - \widehat{\mu}) + (\widehat{\mu} - \mu)]^2 + [(y_2 - \widehat{\mu}) + (\widehat{\mu} - \mu)]^2 \\
&\quad = (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 + 2(\widehat{\mu} - \mu)^2 + 2(\widehat{\mu} - \mu)[(y_1 - \widehat{\mu}) + (y_2 - \widehat{\mu})] \\
&\quad = (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 + 2(\widehat{\mu} - \mu)^2 + 2\widehat{\mu}(y_1 - \widehat{\mu}) + 2\widehat{\mu}(y_2 - \widehat{\mu}) \\
&\qquad - 2(y_1 - \widehat{\mu})\mu - 2(y_2 - \widehat{\mu})\mu
\end{aligned}$$

$$\boxed{\begin{aligned}\widehat{\mu} &= \frac{1}{2}(y_1 + y_2) \\ \widehat{\sigma}^2 &= (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2.\end{aligned}}$$

As soon as we substitute $\widehat{\mu}$ and $\widehat{\sigma}^2$ we arrive at

$$\begin{aligned}(y_1 - \mu)^2 + (y_2 - \mu)^2 &= [(y_1 - \widehat{\mu}) + (\widehat{\mu} - \mu)]^2 + [(y_2 - \widehat{\mu}) + (\widehat{\mu} - \mu)]^2 \\ &= \widehat{\sigma}^2 + 2(\widehat{\mu} - \mu)^2,\end{aligned}$$

since the residual terms vanish:

$$\begin{aligned}&2\widehat{\mu}(y_1 - \widehat{\mu}) + 2\widehat{\mu}(y_2 - \widehat{\mu}) - 2(y_1 - \widehat{\mu})\mu - 2(y_2 - \widehat{\mu})\mu \\ &\quad = 2\widehat{\mu}y_1 - 2\widehat{\mu}^2 + 2\widehat{\mu}y_2 - 2\widehat{\mu}^2 - 2\mu y_1 + 2\mu\widehat{\mu} - 2\mu y_2 + 2\mu\widehat{\mu} \\ &\quad = 2\widehat{\mu}(y_1 + y_2) - 4\widehat{\mu}^2 - 2\mu(y_1 + y_2) + 4\mu\widehat{\mu} \\ &\quad = 4\widehat{\mu}^2 - 4\widehat{\mu}^2 - 4\mu\widehat{\mu} + 4\mu\widehat{\mu} = 0.\end{aligned}$$

The third action item

The cumulative *pdf*

$$d\mathrm{F} = f(y_1, y_2)dy_1 dy_2 = f_1(\widehat{\mu}) f_2(x) d\widehat{\mu} dx = f_1(\widehat{\mu}) f_2\left(\frac{\widehat{\sigma}^2}{\sigma^2}\right) d\widehat{\mu}\, d\frac{\widehat{\sigma}^2}{\sigma^2}$$

has to be decomposed into the *first pdf* $f_1(\widehat{\mu}\,|\mu\,,\sigma^2/n)$ representing the *pdf* of the *sample mean* $\widehat{\mu}$ and the *second pdf* $f_2(x)$ of the new variable $x := (y_1 - y_2)^2/(2\sigma^2) = \widehat{\sigma}^2/\sigma^2$ representing the sample variance $\widehat{\sigma}^2$, normalized by σ^2.

How can the second decomposition $f_1 f_2$ be understood?

Let us replace the quadratic form $(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) = \widehat{\sigma}^2 + 2(\widehat{\mu} - \mu)^2$ in the cumulative *pdf*

$$\begin{aligned}d\mathrm{F} &= f(y_1, y_2)dy_1 dy_2 = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu)\right) dy_1 dy_2 \\ &= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}[\widehat{\sigma}^2 + 2(\widehat{\mu} - \mu)^2]\right) dy_1 dy_2\end{aligned}$$

$$\begin{aligned}d\mathrm{F} &= f(y_1, y_2)dy_1 dy_2 \\ &= \frac{1}{\sqrt{2\pi}} * \frac{1}{\frac{\sigma}{\sqrt{2}}}\exp\left(-\frac{1}{2}\frac{(\widehat{\mu} - \mu)^2}{\sigma^2/2}\right) * \frac{1}{\sqrt{2\pi}} * \frac{1}{\sqrt{2}\sigma}\exp\left(-\frac{1}{2}\frac{\widehat{\sigma}^2}{\sigma^2}\right) dy_1 dy_2.\end{aligned}$$

The *quadratic form* $\widehat{\sigma}^2$, conventionally given in terms of the residual vector $\mathbf{y} - \mathbf{1}\widehat{\mu}$ will be rewritten in terms of the coordinates $[y_1, y_2]' = \mathbf{y}$ of the observation vector.

$$\widehat{\sigma}^2 = (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 = \left[y_1 - \frac{1}{2}(y_1 + y_2)\right]^2 + \left[y_2 - \frac{1}{2}(y_1 + y_2)\right]^2$$

$$\widehat{\sigma}^2 = \frac{1}{4}(y_1 - y_2)^2 + \frac{1}{4}(y_2 - y_1)^2 = \frac{1}{2}(y_1 - y_2)^2.$$

The fourth action item

The new variable $x := \frac{1}{2\sigma^2}(y_1 - y_2)^2$ will be introduced in the cumulative pdf $dF = f(y_1, y_2)dy_1dy_2$. The new surface element $d\widehat{\mu}dx$ will be related to the old surface element dy_1dy_2.

$$d\widehat{\mu}\,dx = \left|\det\begin{bmatrix} D_{y_1}\widehat{\mu} & D_{y_2}\widehat{\mu} \\ D_{y_1}x & D_{y_2}x \end{bmatrix}\right| dy_1dy_2 = |\mathbf{J}|\,dy_1dy_2$$

$$D_{y_1}\widehat{\mu} := \frac{\partial\widehat{\mu}}{\partial y_1} = \frac{1}{2}, \quad D_{y_2}\widehat{\mu} := \frac{\partial\widehat{\mu}}{\partial y_2} = -\frac{1}{2}$$

$$D_{y_1}x := \frac{\partial x}{\partial y_1} = \frac{y_1 - y_2}{\sigma^2}, \quad D_{y_2}x := \frac{\partial x}{\partial y_2} = -\frac{y_1 - y_2}{\sigma^2}.$$

The absolute value of the *Jacobi determinant* amounts to

$$|\mathbf{J}| = \left|\det\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{y_1-y_2}{\sigma^2} & -\frac{y_1-y_2}{\sigma^2} \end{bmatrix}\right| = \frac{y_1 - y_2}{\sigma^2}, \quad |\mathbf{J}|^{-1} = \frac{\sigma^2}{y_1 - y_2}.$$

In consequence, we have derived

$$\boxed{dy_1dy_2 = \frac{\sigma^2}{y_1 - y_2}d\widehat{\mu}\,dx = \frac{\sigma}{\sqrt{2}\sqrt{x}}d\widehat{\mu}dx}$$

based upon

$$x = \frac{1}{2\sigma^2}(y_1 - y_2)^2 \Rightarrow \sqrt{2x} = \frac{1}{\sigma}(y_1 - y_2).$$

In collecting all detailed partial results we can formulate a corollary.

Corollary B.3. *(marginal probability distributions of $\widehat{\mu}$, σ^2 given, and $\widehat{\sigma}^2$):*

The *cumulative pdf* of a set of two observations is represented by

$$dF = f(y_1, y_2)dy_1dy_2 = f_1(\widehat{\mu}\,|\mu, \sigma^2/2)\,f_2(x)d\widehat{\mu}dx \quad subject \;\; to$$

$$f_1(\widehat{\mu}\,|\mu, \;,\sigma^2/2) := \frac{1}{\sqrt{2\pi}} * \frac{1}{\frac{\sigma}{\sqrt{2}}} \exp\left(-\frac{1}{2}\frac{(\widehat{\mu}-\mu)^2}{\sigma^2/2}\right)$$

$$f_2(x) = f_2\left(\frac{\widehat{\sigma}^2}{\sigma^2}\right) := \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{x}}\exp\left(-\frac{1}{2}x\right) \quad subject \;\; to$$

$$\int_{-\infty}^{+\infty} f_1(\widehat{\mu})\,d\widehat{\mu} = 1 \quad and \quad \int_{0}^{+\infty} f_2(x)\,dx = 1.$$

$f_1(\widehat{\mu})$ is the *pdf* of the *sample mean* $\widehat{\mu} = (y_1 + y_2)/2$ and $f_2(x)$ the pdf of the *sample variance* $\widehat{\sigma}^2 = (y_1 - y_2)^2/2 = \sigma^2 x$. $f_1(\widehat{\mu})$ is a *Gauss-Laplace pdf* with mean μ and variance σ^2/n, while $f_2(x)$ is a Helmert χ^2 with one degree of freedom.

Example B.13. (Gauss-Laplace i.i.d. observations, observation space $\mathbb{Y}$, dim $\mathbb{Y} = 3$):

In order to derive the marginal distribution of $\widehat{\mu}$ BLUUE of μ and $\widehat{\sigma}^2$ BIQUUE of σ^2 for a three-dimensional Euclidean observation space, we have to act in various scenes. First, we introduce the probability function of three *Gauss-Laplace i.i.d.* observations and consequently implement the $(\widehat{\mu}, \widehat{\sigma}^2)$ decomposition in the *pdf. Second*, we force the quadratic form $(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})$ to be decomposed into $\widehat{\sigma}^2$ and $(\widehat{\mu}-\mu)^2$, actually a way to introduce the sample variance $\widehat{\sigma}^2$ BIQUUE of $\widehat{\sigma}^2$ and the sample mean $\widehat{\mu}$ BLUUE of μ. *Third*, we succeed to transform the quadratic form $(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) = \mathbf{yMy}$, rk$\mathbf{M} = 2$ into the *canonical form* $z_1^2 + z_2^2$ by means of $[z_1, z_2]' = \mathbf{H}[y_1, y_2, y_3]'$. *Fourth*, we produce the *right inverse* $\mathbf{H}_k^- = \mathbf{H}'$ in order to invert $\mathbf{H}$, namely $[y_1, y_2, y_3]' = \mathbf{H}'[z_1, z_2]'$. *Fifth*, in order to transform the original quadratic form $\sigma^{-2}(\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu})$ into the *canonical form* $z_1^2 + z_2^2 + z_3^2$ we review the *general Helmert transformation* $\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mu)$ and *its inverse* $\mathbf{y} - \mu = \sigma\mathbf{H}'\mathbf{z}$ subject to $\mathbf{H} \in SO(3)$ and identify its parameters *translation, rotation and dilatation* (scale). *Sixth*, we summarize the marginal probability distributions $f_1(\widehat{\mu}\,|\mu\,, \sigma^2/3)$ of the sample mean $\widehat{\mu}$ BLUUE of μ and $f_2(2\widehat{\sigma}^2/\sigma^2)$ of the sample variance $\widehat{\sigma}^2$ BIQUUE of σ^2. The special Helmert pdf χ^2 with two degrees of freedom is plotted in Fig. B.4 while the special Gauss-Laplace normal pdf of variance $\sigma^2/3$ in Fig. B.5. The basic results of the example are collected in Corollary B.7.

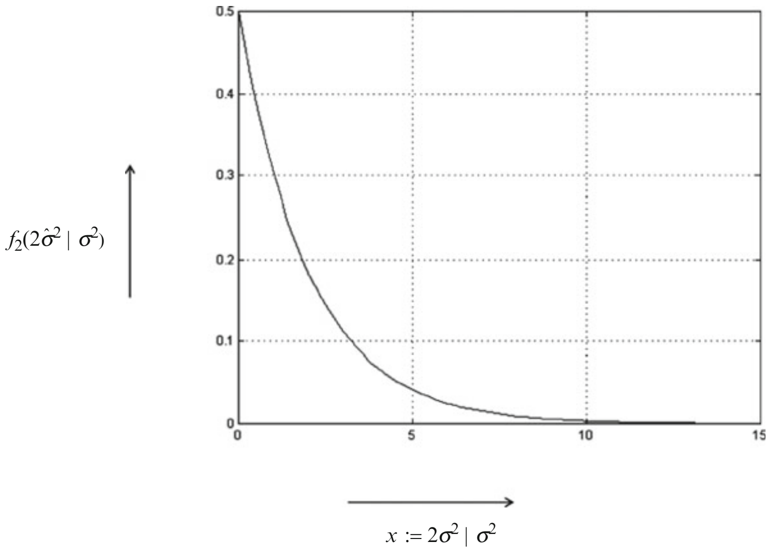


Fig. B.4 Special Helmert pdf of χ_p^2 for two degrees of freedom, $p = 2$

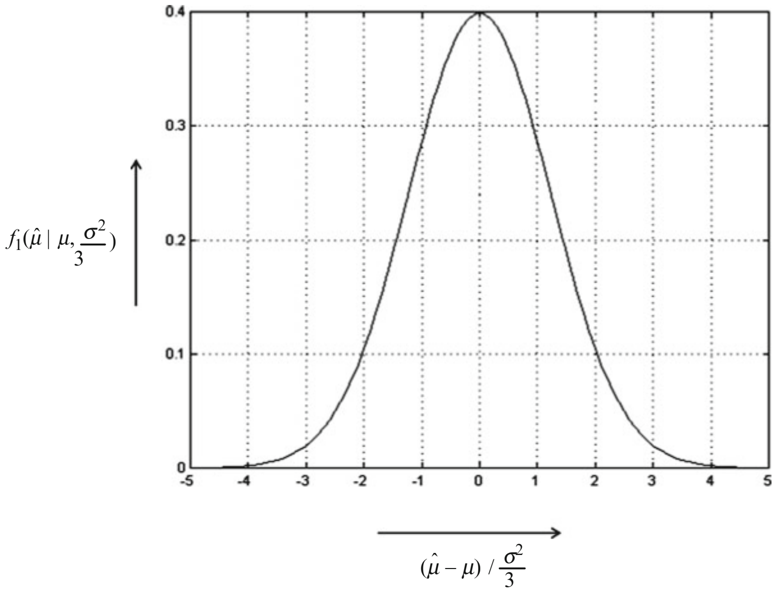


Fig. B.5 Special Gauss-Laplace normal pdf of $(\widehat{\mu} - \mu)/\frac{\sigma^2}{3}$

The first action item

Let us assume an experiment of *three Gauss-Laplace* i.i.d. observations. Their *pdf* is given by

$$\boxed{\begin{aligned} f(y_1, y_2, y_3) &= f(y_1)f(y_2)f(y_3),\\ f(y_1, y_2, y_3) &= (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}[(y_1-\mu)^2+(y_2-\mu)^2+(y_3-\mu)^2]\right),\\ f(y_1, y_2, y_3) &= (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right). \end{aligned}}$$

The coordinates of the observation vector have been denoted by $[y_1, y_2, y_3]' = \mathbf{y} \in \mathbb{Y}$, $\dim \mathbb{Y} = 2$.

The second action item

The coordinates of the observation vector have been denoted by $= \mathbf{y} \in \mathbb{Y}$, $\dim \mathbb{Y} = 2$. The *quadratic form* $(y_1 - 1\mu)'(y_2 - 1\mu)$ allows the *fundamental decomposition*

$$\mathbf{y}-\mathbf{1}\mu = (\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}(\widehat{\mu}-\mu),$$

$$(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = (\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}'\mathbf{1}(\widehat{\mu}-\mu)^2$$

$$\boxed{(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = 2\widehat{\sigma}^2 + 3(\widehat{\mu}-\mu)^2.}$$

Here, $\widehat{\mu}$ is BLUUE of μ and $\widehat{\sigma}^2$ BIQUUE of σ^2. The detailed computation proves our statement.

$$\begin{aligned} &[(y_1-\widehat{\mu})+(\widehat{\mu}-\mu)]^2 + [(y_2-\widehat{\mu})+(\widehat{\mu}-\mu)]^2 + [(y_3-\widehat{\mu})+(\widehat{\mu}-\mu)]^2\\ &\quad= (y_1-\widehat{\mu})^2 + (y_2-\widehat{\mu})^2 + (y_3-\widehat{\mu})^2 + 3(\widehat{\mu}-\mu)^2 + 2\widehat{\mu}(y_1-\widehat{\mu})\\ &\qquad + 2\widehat{\mu}(y_2-\widehat{\mu}) + 2\widehat{\mu}(y_3-\widehat{\mu}) - 2(y_1-\widehat{\mu})\mu - 2(y_2-\widehat{\mu})\mu - 2(y_3-\widehat{\mu})\mu \end{aligned}$$

$$\boxed{\begin{aligned} \widehat{\mu}\ \text{BLUUE of}\ \mu:\ \widehat{\mu} &= \frac{1}{3}(y_1+y_2+y_3)\\ \widehat{\sigma}^2\ \text{BIQUUE of}\ \sigma^2:\widehat{\sigma}^2 &= \frac{1}{2}[(y_1-\widehat{\mu})^2+(y_2-\widehat{\mu})^2+(y_3-\widehat{\mu})^2]. \end{aligned}}$$

As soon as we substitute $\widehat{\mu}$ and $\widehat{\sigma}^2$ we arrive at

$$\begin{aligned} &(y_1-\mu)^2 + (y_2-\mu)^2 + (y_3-\mu)^2\\ &\quad= [(y_1-\widehat{\mu})^2 + (\widehat{\mu}-\mu)]^2 + [(y_2-\widehat{\mu})^2 + (\widehat{\mu}-\mu)^2]^2 + [(y_3-\widehat{\mu})^2 + (\widehat{\mu}-\mu)^2]^2\\ &\quad= 2\widehat{\sigma}^2 + 3(\widehat{\mu}-\mu)^2. \end{aligned}$$

The third action item

Let us begin with transforming the cumulative probability function

$$
\begin{aligned}
d\mathrm{F} &= f(y_1,\ y_2,\ y_3)dy_1dy_2dy_3 \\
&= (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right)dy_1dy_2dy_3 \\
&= (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}[2\widehat{\sigma}^2+3(\widehat{\mu}-\mu)^2]\right)dy_1dy_2dy_3
\end{aligned}
$$

into its canonical form. In detail, we transform the quadratic form $\widehat{\sigma}^2$ BIQUUE of σ^2 canonically.

$$
\begin{aligned}
y_1-\widehat{\mu} &= y_1-\frac{1}{3}y_1-\frac{1}{3}(y_2+y_3)=\frac{2}{3}y_1-\frac{1}{3}(y_2+y_3)=\frac{2}{3}y_1-\frac{1}{3}y_2-\frac{1}{3}y_3 \\
y_2-\widehat{\mu} &= y_2-\frac{1}{3}y_2-\frac{1}{3}(y_1+y_3)=-\frac{1}{3}y_1+\frac{2}{3}y_2-\frac{1}{3}y_3 \\
y_3-\widehat{\mu} &= y_3-\frac{1}{3}y_3-\frac{1}{3}(y_1+y_2)=-\frac{1}{3}y_1-\frac{1}{3}y_2+\frac{2}{3}y_3
\end{aligned}
$$

as well as

$$
\begin{aligned}
(y_1-\widehat{\mu})^2 &= \frac{4}{9}y_1^2+\frac{1}{9}y_2^2+\frac{1}{9}y_3^2-\frac{4}{9}y_1y_2+\frac{2}{9}y_2y_3-\frac{4}{9}y_3y_1 \\
(y_2-\widehat{\mu})^2 &= \frac{1}{9}y_1^2+\frac{4}{9}y_2^2+\frac{1}{9}y_3^2-\frac{4}{9}y_1y_2-\frac{4}{9}y_2y_3+\frac{2}{9}y_3y_1 \\
(y_3-\widehat{\mu})^2 &= \frac{1}{9}y_1^2+\frac{1}{9}y_2^2+\frac{4}{9}y_3^2+\frac{2}{9}y_1y_2-\frac{4}{9}y_2y_3-\frac{4}{9}y_3y_1
\end{aligned}
$$

and

$$
\boxed{(y_1-\widehat{\mu})^2+(y_2-\widehat{\mu})^2+(y_3-\widehat{\mu})^2=\frac{2}{3}(y_1^2+y_2^2+y_3^2-y_1y_2-y_2y_3-y_3y_1).}
$$

We shall prove

$$
(\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu})=\mathbf{y}'\mathbf{M}\mathbf{y}=z_1^2+z_2^2,\quad \mathrm{rk}\mathbf{M}=2,\quad \mathbf{M}\in\mathbb{R}^{3\times 3}
$$

that the symmetric matrix $\mathbf{M}$,

$$
\mathbf{M}=\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}
$$

has rank 2 or rank deficiency 1. Just compute $|\mathbf{M}| = 0$ as a determinant identity as well as the subdeterminant

$$\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3 \neq 0$$

$$(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) = \mathbf{y}'\mathbf{M}\mathbf{y} = \frac{1}{3}[y_1, y_2, y_3] \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

How to transform a degenerate quadratic form to a canonical form?

Helmert (1875) had the bright idea to implement what we call nowadays *the forward Helmert transformation*

$$z_1 = \frac{1}{\sqrt{1 \cdot 2}}(y_1 - y_2)$$

$$z_2 = \frac{1}{\sqrt{2 \cdot 3}}(y_1 + y_2 - 2y_3) \quad \textit{or}$$

$$z_1^2 = \frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2)$$

$$z_2^2 = \frac{1}{6}(y_1^2 + y_2^2 + 4y_3^2 + 2y_1y_2 - 4y_2y_3 - 4y_3y_1)$$

$$\boxed{z_1^2 + z_2^2 = \frac{2}{3}(y_1^2 + y_2^2 + y_3^2 - y_1y_2 - y_2y_3 - y_3y_1)}.$$

Indeed we found

$$(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) = (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 + (y_3 - \widehat{\mu})^2 = z_1^2 + z_2^2.$$

In algebraic terms, a representation of the *rectangular Helmert matrix* is

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & -\frac{1}{\sqrt{1\cdot 2}} & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$\mathbf{z} = \mathbf{H}_{23}\mathbf{y}, \quad \mathbf{H}_{23} := \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & -\frac{1}{\sqrt{1\cdot 2}} & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} \end{bmatrix}.$$

The *rectangular Helmert matrix is right orthogonal*,

$$\mathbf{H}_{23}\mathbf{H}'_{23} = \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & -\frac{1}{\sqrt{1\cdot 2}} & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} \\ -\frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} \\ 0 & -\frac{2}{\sqrt{2\cdot 3}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathrm{I}_2.$$

The fourth action item

By means of the *forward Helmert transformation* we could prove that $z_1^2 + z_2^2$ represents the quadratic form $(\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu})$. Unfortunately, the forward Helmert transformation only allows indirectly by a canonical transformation. What would be needed is the *inverse Helmert transformation* $\mathbf{z} \mapsto \mathbf{y}$, also called *backward Helmert transformation*. The *rectangular Helmert matrix* has the disadvantage to have *no Cayley inverse*. Fortunately, its *right inverse*

$$\mathbf{H}_R^- := \mathbf{H}'_{23}(\mathbf{H}_{23}\mathbf{H}'_{23})^{-1} = \mathbf{H}'_{23} \in \mathbb{R}^{3\times 2}$$

solves our problem. The *inverse Helmert transformation*

$$\mathbf{y} = \mathbf{H}_R^-\mathbf{z} = \mathbf{H}'_{23}\mathbf{z}$$

brings

$$(\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu}) = \frac{2}{3}(y_1^2 + y_2^2 + y_3^2 - y_1y_2 - y_2y_3 - y_3y_1)$$

via

$$\mathbf{y} = \mathbf{H}'_{23}\mathbf{z} \quad or \quad \left[\begin{array}{l} y_1 = \frac{1}{\sqrt{1\cdot 2}}z_1 + \frac{1}{\sqrt{2\cdot 3}}z_2 \\ y_2 = -\frac{1}{\sqrt{1\cdot 2}}z_1 + \frac{1}{\sqrt{2\cdot 3}}z_2 \\ y_3 = -\frac{2}{\sqrt{2\cdot 3}}z_2, \end{array}\right.$$

$$y_1^2 + y_2^2 + y_3^2 = \frac{1}{2}z_1^2 + \frac{1}{6}z_2^2 + \frac{1}{\sqrt{3}}z_1z_2 + \frac{1}{2}z_2^2 + \frac{1}{6}z_2^2 - \frac{1}{\sqrt{3}}z_1z_2 + \frac{2}{3}z_2^2,$$

$$y_1y_2 + y_2y_3 + y_3y_1 = -\frac{1}{2}z_1^2 + \frac{1}{6}z_2^2 + \frac{1}{\sqrt{3}}z_1z_2 - \frac{1}{3}z_2^2 - \frac{1}{\sqrt{3}}z_1z_2 - \frac{1}{2}z_2^2,$$

$$\frac{2}{3}(y_1^2 + y_2^2 + y_3^2 - y_1y_2 - y_2y_3 - y_3y_1) = \frac{2}{3}\left(\frac{3}{2}z_1^2 + \frac{3}{2}z_2^2\right) = z_1^2 + z_2^2,$$

into the canonical form.

The fifth action item

Let us go back to the partitioned *pdf* in order to inject the canonical representation of the deficient quadratic form $\mathbf{y}'\mathbf{M}\mathbf{y}$, $\mathbf{M} \in \mathbb{R}^{3\times 3}$, $rk M = 2$. Here we meet *first* the problem to transform

$$dF = (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}[2\widehat{\sigma}^2 + 3(\widehat{\mu}-\mu)^2]\right)dy_1dy_2dy_3$$

by an extended vector $[z_1, z_2, z_3]' =: \mathbf{z}$ into the *canonical form* $dF = (2\pi)^{-3/2}$ $\exp\left(-\frac{1}{2}(z_1^2+z_2^2+z_3^2)\right)dz_1dz_2dz_3$, which is generated by the general *forward Helmert transformation*

$$\boxed{\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mathbf{1}\mu)}$$

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \frac{1}{\sigma}\begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{1\cdot 2}} & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} y_1-\mu \\ y_2-\mu \\ y_3-\mu \end{bmatrix}$$

or its *backward Helmert transformation*, also called the *general inverse Helmert transformation*

$$\begin{bmatrix} y_1-\mu \\ y_2-\mu \\ y_3-\mu \end{bmatrix} = \sigma\begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

$$\boxed{\mathbf{y} - \mathbf{1}\mu = \sigma\mathbf{H}'\mathbf{z}}$$

thanks to the *orthonormality* of the *quadratic Helmert matrix* $\mathbf{H}_3$, namely $\mathbf{H}_3\mathbf{H}'_3 = \mathbf{H}'_3\mathbf{H}_3 = \mathbf{I}_3$ or $\mathbf{H}_3^{-1} = \mathbf{H}'_3$.
Secondly, notice the transformation of the *volume element*

$$dy_1dy_2dy_3 = d(y_1-\mu)d(y_2-\mu)d(y_3-\mu) = \sigma^3dz_1dz_2dz_3,$$

which is due to the *Jacobi determinant*

$$\mathbf{J} = \sigma^3\left|\mathbf{H}'\right| = \sigma^3\left|\mathbf{H}\right| = \sigma^3.$$

Let us prove that

$$\boxed{z_3 := \sqrt{3}\sigma^{-1}(\widehat{\mu}-\mu) = \frac{\widehat{\mu}-\mu}{\frac{\sigma}{\sqrt{3}}}}$$

brings the *first marginal density*

$$\boxed{f_1 = \left(\widehat{\mu}|\mu, \frac{\sigma^2}{3}\right) := \frac{1}{\sqrt{2\pi}} * \frac{1}{\frac{\sigma}{\sqrt{3}}} * \exp\left(-\frac{1}{2\sigma^2}3(\widehat{\mu}-\mu)^2\right)}$$

into the canonical form

$$\boxed{f_1(z_3\,|0,1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_3^2\right).}$$

Let us compute $\widehat{\mu}-\mu$ as well as $3(\widehat{\mu}-\mu)^2$ which concludes the proof.

$$z_3 = \sigma^{-1}\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]\begin{bmatrix} y_1-\mu \\ y_2-\mu \\ y_3-\mu \end{bmatrix}$$

$$\left.\begin{array}{l} z_3 = \sigma^{-1}\dfrac{y_1+y_2+y_3-3\mu}{\sqrt{3}} \\ \dfrac{y_1+y_2+y_3}{3} = \widehat{\mu} \Rightarrow \; y_1+y_2+y_3 = 3\widehat{\mu} \end{array}\right]$$

$$\Rightarrow z_3 = 3\sigma^{-1}\frac{\widehat{\mu}-\mu}{\sqrt{3}}, \quad z_3^2 = \frac{1}{\sigma^2}3(\widehat{\mu}-\mu)^2.$$

Indeed the *extended Helmert matrix* $\mathbf{H}_3$ is ingenious to decompose

$$\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu}) = z_1^2+z_2^2+z_3^2$$

into a *canonical quadratic form relating* $z_1^2+z_2^2$ to $\widehat{\sigma}^2$ and z_3^2 to $(\widehat{\mu}-\mu)^2$. At this point, we have to interpret the *general Helmert transformation* $\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y}-\mu)$:

Structure elements of the Helmert transformation

scale or dilatation	σ^{-1}
rotation	$\mathbf{H}$
translation	μ

$\sigma^{-1} \in \mathbb{R}^+$ produces a *dilatation* or a *scale change*, $\mathbf{H} \in SO(3) := \{\mathbf{H} \in \mathbb{R}^{3\times 3}\,|\mathbf{H}'\mathbf{H} = \mathbf{I}_3$ and $|\mathbf{H}| = +1\}$ a *rotation* (3 parameters) and $\mathbf{1}\mu \in \mathbb{R}^3$ a *translation*. Please, prove for yourself that the *quadratic Helmert matrix* is orthonormal, that is $\mathbf{H}\mathbf{H}' = \mathbf{H}'\mathbf{H} = \mathbf{I}_3$ and $|\mathbf{H}| = +1$.

The sixth action item

Finally we are left with the problem to split the *cumulative pdf* into *one part* $f_1(\widehat{\mu})$ which is a marginal distribution of the arithmetic mean $\widehat{\mu}$ BLUUE of μ and *another part* $f_2(\widehat{\mu})$ which is a marginal distribution of the standard deviation $\widehat{\sigma}$, $\widehat{\sigma}^2$ BIQUUE of σ^2, *Helmert's* χ_2^2 with two degrees of freedom.
First let us introduce *polar coordinates* (ϕ_1, r) which represent the *Cartesian coordinates* $z_1 = r\cos\phi_1$, $z_2 = r\sin\phi_1$. The index 1 is needed for later generalization to *higher dimension*. As a longitude, the domain of ϕ_1 is $\phi_1 \in [0, 2\pi]$ or $0 \le \phi_1 \le 2\pi$. The new random variable $z_1^2 + z_2^2 = \|z\|^2 =: x$ or radius r relates *to Helmert's*

$$\boxed{\chi^2 = z_1^2 + z_2^2 = \frac{2\widehat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})}.$$

Secondly, the marginal distribution of the arithmetic mean $\widehat{\mu}$, $\widehat{\mu}$ BLUUE of μ,

$$f_1\left(\widehat{\mu}\,|\mu\,,\,\frac{\sigma}{\sqrt{3}}\right)d\widehat{\mu} = f_1(z_3\,|0,\,1)dz_3 \;\sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{3}}\right)$$

is a *Gauss-Laplace normal distribution with mean* μ *and variance* $\sigma^2/3$ generated by

$$dF_1 = f_1\left(\widehat{\mu}\,|\mu\,,\,\frac{\sigma}{\sqrt{3}}\right)d\widehat{\mu} = f_1(z_3\,|0,\,1)dz_3$$

$$= (2\pi)^{-1/2}\exp\left(-\frac{1}{2}z_3^2\right)dz_3\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(2\pi)^{-1}\exp\left(-\frac{1}{2}(z_1^2+z_2^2)\right)dz_1dz_2 \quad or$$

$$f_1\left(\widehat{\mu}\,|\mu\,,\,\frac{\sigma}{\sqrt{3}}\right)d\widehat{\mu} = (2\pi)^{-1/2}\frac{\sqrt{3}}{\sigma}\exp\left(-\frac{3}{2\sigma^2}(\widehat{\mu}-\mu)^2\right)d\widehat{\mu}.$$

Third, the marginal distribution of the sample variance $2\widehat{\sigma}^2/\sigma^2 = z_1^2 + z_2^2 =: x$, *Helmert's* χ^2 distribution for $p = n - 1 = 2$ degrees of freedom,

$$\boxed{f_2(2\widehat{\sigma}^2/\sigma^2) = f_2(x) = \frac{1}{2^{p/2}\Gamma(\frac{p}{2})}x^{\frac{p}{2}-1}\exp\left(-\frac{x}{2}\right)}$$

is generated by

$$dF_2 = \int_{-\infty}^{+\infty}(2\pi)^{-1/2}\exp\left(-\frac{1}{2}z_3^2\right)dz_3\int_0^{2\pi}d\phi_1(2\pi)^{-1}\frac{1}{2}\exp\left(-\frac{1}{2}x\right)dx$$

$$dF_2 = (2\pi)^{-1}\omega_2 \frac{1}{2}\exp\left(-\frac{1}{2}x\right)dx \quad \textit{subject to } \omega_2 = \int_0^{2\pi} d\phi_1 = 2\pi$$

$$\boxed{dF_2 = \frac{1}{2}\exp\left(-\frac{1}{2}x\right)dx}$$

and

$$x := z_1^2 + z_2^2 = r^2 \Rightarrow dx = 2rdr, \quad dr = \frac{dx}{2r}$$

$$\boxed{dz_1 dz_2 = rdrd\phi_1 = \frac{1}{2}dxd\phi_1}$$

is the transformation of the *surface element* dz_1dz_2. In collecting all detailed results let us formulate a *corollary*.

Corollary B.7. *(marginal probability distributions of $\widehat{\mu}$, σ^2 given, and of $\widehat{\sigma}^2$):*

The *cumulative pdf* of a set of three observations is represented by

$$dF = f(y_1, y_2, y_3)dy_1dy_2dy_3 = f_1\left(\widehat{\mu}|\mu, \frac{\sigma^2}{3}\right)f_2(x)d\widehat{\mu}dx$$

subject to

$$f_1\left(\widehat{\mu}|\mu, \frac{\sigma^2}{3}\right) := \frac{1}{\sqrt{2\pi}} * \frac{1}{\sigma/\sqrt{3}} * \exp\left(-\frac{1}{2}\frac{(\widehat{\mu}-\mu)^2}{\sigma^2/3}\right)$$

$$f_2(x) = \frac{1}{2}\exp\left(-\frac{1}{2}x\right)$$

subject to

$$x := z_1^2 + z_2^2 = 2\widehat{\sigma}^2/\sigma^2, \quad dx = \frac{2}{\sigma^2}d\widehat{\sigma}^2$$

and

$$\int_{-\infty}^{\infty} f_1(\widehat{\mu})\,d\widehat{\mu} = 1 \quad \textit{versus} \quad \int_0^{\infty} f_2(x)\,dx = 1.$$

$f_1(\widehat{\mu})$ is the *pdf* of the *sample mean* $\widehat{\mu} = (y_1 + y_2 + y_3)/3$ and $f_2(x)$ the pdf of the *sample variance* $\widehat{\sigma}^2 = (\mathbf{y}' - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})/2$ normalized by σ^2. $f_1(\widehat{\mu})$ is a

Gauss-Laplace pdf with mean $\widehat{\mu}$ and variance $\sigma^2/3$, while $f_2(x)$ a *Helmert* χ^2 with two degrees of freedom.

In summary, an experiment with three Gauss-Laplace i.i.d. observations *is characterized by two marginal probability densities, one for the mean* $\widehat{\mu}$ BLUUE of μ and another one for $\widehat{\sigma}^2$ BIQUUE of σ^2:

Marginal probability densities $n = 3$, Gauss-Laplace i.i.d. observations

$$\widehat{\mu} : f_1\left(\widehat{\mu}\,|\mu\,,\ \frac{\sigma}{\sqrt{3}}\right) d\widehat{\mu} = (2\pi)^{-1/2}\frac{1}{\sigma/\sqrt{3}}\exp\left(-\frac{1}{2}\frac{(\widehat{\mu}-\mu)^2}{\sigma^2/3}\right) d\widehat{\mu}$$

or

$$f_1(z_3\,|0,\ 1)dz_3 = (2\pi)^{-1/2}\exp\left(-\frac{1}{2}z_3^2\right) dz_3$$

$$z_3 := \frac{\widehat{\mu}-\mu}{3\sigma}$$

$$\widehat{\sigma}^2 : f_2(2\widehat{\sigma}^2/\sigma^2)d\widehat{\sigma}^2 = \frac{\widehat{\sigma}^2}{\sigma^2}\exp\left(-\frac{\widehat{\sigma}^2}{\sigma^2}\right) d\widehat{\sigma}^2$$

or

$$f_2(x)dx = \frac{1}{2}\exp\left(-\frac{1}{2}x\right) dx$$

$$x := 2\frac{\widehat{\sigma}^2}{\sigma^2}$$

B-41 Sampling Distributions of the Sample Mean $\widehat{\mu}$, σ^2 Known, and of the Sample Variance $\widehat{\sigma}^2$

The two examples have prepared us for the general sampling distribution of the sample mean $\widehat{\mu}$, σ^2 known, and of the sample variance $\widehat{\sigma}^2$ for *Gauss-Laplace i.i.d. observations*, namely samples of size n. By means of Lemma B.8 on the *rectangular Helmert transformation* and Lemma B.9 on the *quadratic Helmert transformation* we prepare for Theorem B.10 which summaries both the *pdfs* for $\widehat{\mu}$ BLUUE of μ, σ^2 known, for the standard deviation $\widehat{\sigma}$ and for $\widehat{\sigma}^2$ BIQUUE of σ^2. Corollary B.11 focusses on the *pdf* of $\widetilde{\sigma} = q\widehat{\sigma}$ where $\widetilde{\sigma}$ is an *unbiased estimation* of the standard deviation σ, namely $\mathbf{E}\{\widetilde{\sigma}\} = \sigma$.

Lemma B.3. (rectangular Helmert transformation):

The *rectangular Helmert matrix* $\mathbf{H}_{n-1,n} \in \mathbb{R}^{n\times(n-1)}$ transforms the degenerate quadratic form

$$(n-1)\widehat{\sigma}^2 := (\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu}) = \mathbf{y}'\mathbf{M}\mathbf{y},\ \mathrm{rk}\mathbf{M}{=}\,n{-}1$$

subject to

$$\widehat{\mu} = \frac{1}{n}\mathbf{1}'\mathbf{y}$$

into the canonical form

$$(n-1)\widehat{\sigma}^2 = \mathbf{z}'_{n-1}\mathbf{z}_{n-1} = z_1^2 + \cdots + z_{n-1}^2.$$

The *special Helmert transformation* $\mathbf{y}_n \mapsto \mathbf{H}_{n-1,n}\mathbf{y}_n = \mathbf{z}_{n-1}$ is represented by

$$\mathbf{H}_{n-1,n} := \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{1\cdot 2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \cdots & -\frac{n-1}{\sqrt{(n-1)(n-2)}} & 0 \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & -\frac{n}{\sqrt{n(n-1)}} \end{bmatrix}.$$

The *inverse Helmert transformation* $\mathbf{z} \mapsto \mathbf{y} = \mathbf{H}_R^- \mathbf{z} = \mathbf{H}'\mathbf{z}$ or $y_n = \mathbf{H}'_{n-1\times n}\mathbf{z}_n$ is based on its *right inverse* which thanks to the right orthogonality $\mathbf{H}_{n-1,n}\mathbf{H}'_{n,n-1} = \mathbf{I}_{n-1}$ coincides with its transpose,

$$\mathbf{H}_R^- = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1} = \mathbf{H}' \in \mathbb{R}^{n\times(n-1)}.$$

Lemma B.4. (quadratic Helmert transformation):

The *quadratic Helmert matrix* $\mathbf{H} \in \mathbb{R}^{n\times n}$, also called *extended Helmert matrix* or *augmented Helmert matrix*, transforms the quadratic form

$$\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = \frac{1}{\sigma^2}[(\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}(\widehat{\mu}-\mu)]'[(\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}(\widehat{\mu}-\mu)]$$

subject to

$$\widehat{\mu} = \frac{1}{n}\mathbf{1}'\mathbf{y}$$

by means of

$$\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y}-\mathbf{1}\mu) \ \ or \ \ \mathbf{y} = \sigma\mathbf{H}'(\mathbf{z}-\mathbf{1}\mu)$$

into the canonical form

$$\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = \mathbf{z}_1^2 + \cdots + \mathbf{z}_{n-1}^2 + \mathbf{z}_n^2 = \sum_{j=1}^{n-1}\mathbf{z}_j^2 + \mathbf{z}_n^2$$

$$\mathbf{z}_n^2 = n(\widehat{\mu}-\mu)^2.$$

Such a *Helmert transformation* $\mathbf{y} \mapsto \mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mathbf{1}\mu)$ is represented by displaylines

$$\mathbf{H} :=$$

$$\begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{1\cdot 2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{3\cdot 4}} & \frac{1}{\sqrt{3\cdot 4}} & \frac{1}{\sqrt{3\cdot 4}} & -\frac{3}{\sqrt{34}} & \cdots & 0 & 0 \\ \cdots & & & & \cdots & & \\ \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \frac{1}{\sqrt{(n-1)(n-2)}} & \cdots & -\frac{n-1}{\sqrt{(n-1)(n-2)}} & 0 \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} & -\frac{n}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \end{bmatrix}$$

Since the *quadratic Helmert matrix* is *orthonormal*, the inverse Helmert transformation is generated by

$$\mathbf{z} \mapsto \mathbf{y} - \mathbf{1}\mu = \sigma \mathbf{H}' z.$$

The proofs for Lemmas B.8 and B.9 are based on generalizations of the special cases for $n = 2$, Example B.11, and for $n = 3$, Example B.12. Any proof will be omitted here.
The highlight of this paragraph is the following theorem.

Theorem B.10. (marginal probability distribution of $(\widehat{\mu}, \sigma^2)$ and $\widehat{\sigma}^2$):

The *cumulative pdf* of a set of n observations is represented by

$$\begin{aligned} dF &= f(y_1, \ldots, y_n)dy_1 \cdots dy_n \\ &= f_1(\widehat{\mu}) f_4(\widehat{\sigma}) d\widehat{\mu} d\widehat{\sigma} = f_1(\widehat{\mu}) f_2(\widehat{\sigma}^2) d\widehat{\mu} d\widehat{\sigma}^2 \end{aligned}$$

as the product of the *marginal pdf* $f_1(\widehat{\mu})$ of the sample mean $\widehat{\mu} = n^{-1}\mathbf{1y}$ and the *marginal pdf* $f_2(\widehat{\sigma})$ of the sample standard deviation $\widehat{\sigma} = \sqrt{(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})/(n-1)}$, also called *r.m.s.* (*root mean square error*), or the *marginal pdf* $f_2(\widehat{\sigma}^2)$ of the sample variance $\widehat{\sigma}^2$. Those *marginal pdfs* are represented by

$$dF_1 = f_1(\widehat{\mu}) d\widehat{\mu}$$

$$dF_4 = f_4(\widehat{\sigma}) d\widehat{\sigma} \text{ and } dF_2 = f_2(\widehat{\sigma}^2) d\widehat{\sigma}^2,$$

(a) *sample mean* $\widehat{\mu}$

$$f_1(\widehat{\mu}) = f_1\left(\widehat{\mu}|\mu, \frac{\sigma^2}{n}\right) := \frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(\widehat{\mu}-\mu)^2}{\sigma^2/n}\right]$$

$$\mathbf{z} := \frac{\sqrt{n}}{\sigma}(\widehat{\mu}-\mu) : f_1(z)dz = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mathbf{z}^2\right) dz$$

(b) *sample r.m.s.* $\widehat{\sigma}$

$$p := n-1$$

$$dF_4 = f_4(\widehat{\sigma})d\widehat{\sigma}$$

$$f_4(\widehat{\sigma}) = \frac{2p^p}{\sigma^p 2^{p/2}\Gamma(\frac{p}{2})}\widehat{\sigma}^{p-1} \exp\left(-\frac{p}{2}\frac{\widehat{\sigma}^2}{\sigma^2}\right)$$

$$x := \sqrt{n-1}\frac{\widehat{\sigma}}{\sigma} = \frac{\sqrt{p}}{\sigma}\widehat{\sigma}$$

$$f_4(x)dx = \frac{2}{2^{p/2}\Gamma(\frac{p}{2})}x^{p-1} \exp\left(-\frac{1}{2}x^2\right) dx$$

$$dF_4 = f_4(x)dx$$

(c) *sample variance* $\widehat{\sigma}^2$

$$p := n-1$$

$$f_2(\widehat{\sigma}^2) = \frac{1}{\sigma^p 2^{p/2}\Gamma(\frac{p}{2})} p^{p/2}\widehat{\sigma}^{p-2} \exp\left(-\frac{1}{2}p\frac{\widehat{\sigma}^2}{\sigma^2}\right),$$

$$x := (n-1)\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{p}{\sigma^2}\widehat{\sigma}^2 :$$

$$f_2(x)dx = \frac{1}{2^{p/2}\Gamma(\frac{p}{2})}x^{\frac{p}{2}-1} \exp\left(-\frac{1}{2}x\right) dx.$$

$f_1(\widehat{\mu}|\mu, \frac{\sigma^2}{n})$ as the marginal *pdf* of the *sample mean* BLUUE of μ is a *Gauss-Laplace pdf* with mean μ and variance σ^2/n. $f_2(x),\ x := \sqrt{p}\widehat{\sigma}/\sigma$, is the *standard pdf* of the normalized root-mean-square error with p degrees of freedom. In contrast, $f_2(x),\ x := p\widehat{\sigma}^2/\sigma^2$ is a *Helmert Chi Square* χ_p^2 *pdf* with $p = n-1$ degrees of freedom.

Before we present a sketch of a proof of Theorem B.10 which will be run with *five action items* and a special reference to the first and second vehicle, we give some historical comments. Kullback (1934) refers the marginal pdf $f_1(\widehat{\mu})$ of the "*arithmetic mean*" $\widehat{\mu}$ to Poisson (1827), Haussdorff (1901) and Irwins (1937). He has also solved the problem to find the marginal pdf of the "*geometric mean*". The marginal pdf $f_2(\widehat{\sigma}^2)$ of the sample variance $\widehat{\sigma}^2$ has been originally derived by Helmert (1875), Helmert (1876a), Helmert (1876b). A historical discussion of *Helmert's distribution* is offered by David (1957), Kruskal (1946), Lancaster (1965), Lancaster (1966), Pearson (1931) and Sheynin (1995).

The marginal pdf $f_4(\widehat{\sigma})$ has not found any interest in practice so far. The reason may be found in the effect that $\widehat{\sigma}$ is not an unbiased estimate of the *standard deviation* σ, namely $E\{\widehat{\sigma}\} = \sigma$. According to Czüber (1891), p. 162, Rosèen (1948), p. 37, Schmetterer (1956), p. 203, Stoom (1967), p. 199, 218, Fisz (1971), p. 240 and Richter and Mammitzsch (1973), p. 42 have documented that

$$\widetilde{\sigma} = \sqrt{\frac{p}{2}} \frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)} \widehat{\sigma} = q\widehat{\sigma}$$

is an *unbiased estimation* $\widetilde{\sigma}$ of the *standard deviation* σ, namely $E\{\widehat{\sigma}\} = \sigma$. $p = n - 1$ again denotes the number of *degrees of freedom*. Schaffrin (1979), p. 240 has proven that

$$\widehat{\sigma}_p := \frac{\sqrt{(\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\mu)}}{\sqrt{p - \frac{1}{2}}} = \widehat{\sigma} \frac{\sqrt{2p}}{\sqrt{2p-1}} \sqrt{p - \frac{1}{2}}$$

is an *asymptotic* ("*from above*") *unbiased* estimation $\widehat{\sigma}_p$ of the standard deviation σ. Let us implement $\widetilde{\sigma}$ BLUUE of σ into the marginal *pdf* $f(\widehat{\sigma})$:

Corollary B.8. *(marginal probability distributions of $\widetilde{\sigma}$ for Gauss-Laplace i.i.d. observations, $E\{\widetilde{\sigma}\} = \sigma$):*

The *marginal pdf* of $\widetilde{\sigma}$, an unbiased estimation of the standard deviation, is represented by

$$dF_4 = f_4(\widetilde{\sigma})d\widetilde{\sigma}$$

$$f_4(\widetilde{\sigma}) = \frac{2p^p}{\sigma^p 2^{p/2}} \frac{\Gamma^p\left(\frac{p+1}{2}\right)}{\left(\frac{p}{2}\right)^{\frac{p}{2}} \Gamma^{p+1}\left(\frac{p}{2}\right)} \widetilde{\sigma}^{p-1} \exp\left(-\left(\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}\right)^2 \widetilde{\sigma}^2\right)$$

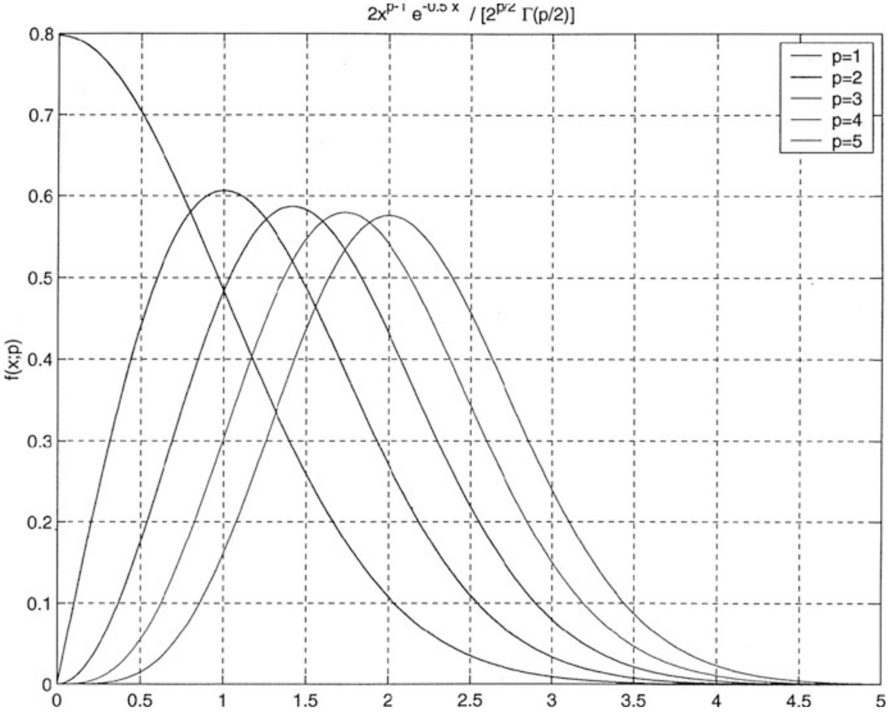


Fig. B.6 Marginal pdf for the sample standard deviation $\widetilde{\sigma}$ (r.m.s.)

and

$$
\begin{aligned}
dF_4 &= f_4(x)dx \\
x &:= \frac{\sqrt{2}}{\sigma}\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}\widehat{\sigma} \\
f_4(x) &= \frac{2}{2^{p/2}\Gamma\left(\frac{p}{2}\right)}x^{p-1}\exp\left(-\frac{1}{2}x^2\right)
\end{aligned}
$$

subject to

$$
E\{x\} = \sqrt{2}\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}.
$$

Figure B.6 illustrates the marginal pdf for "*degrees of freedom*" $p \in \{1, 2, 3, 4, 5\}$.

Proof.

The first action item

The *pdf* of n *Gauss-Laplace* i.i.d. observations is given by

$$f(y_1, \ldots, y_n) = f(y_1)\cdots f(y_n) = \prod_{i=1}^{n} f(y_i)$$

$$f(y_1, \ldots, y_n) = (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right).$$

The coordinates of the observation space $\mathbb{Y}$ have been denoted by $[y_1, \ldots, y_n]' = \mathbf{y}$. Note $\mathit{dim}\ \mathbb{Y} = n$.

The second action item

The quadratic form $(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) > 0$ allows the fundamental decomposition

$$\begin{aligned}
\mathbf{y}-\mathbf{1}\mu &= (\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}(\widehat{\mu}-\mu),\\
(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) &= (\mathbf{y}-\mathbf{1}\widehat{\mu})'(\mathbf{y}-\mathbf{1}\widehat{\mu}) + \mathbf{1}'\mathbf{1}(\widehat{\mu}-\mu)^2\\
\mathbf{1}'\mathbf{1} &= n
\end{aligned}$$

$$\boxed{(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = (n-1)\widehat{\sigma}^2 + n(\widehat{\mu}-\mu)^2.}$$

Here, $\widehat{\mu}$ is BLUUE of μ and $\widehat{\sigma}^2$ BIQUUE of σ^2. The decomposition of the quadratic $(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)$ into the *sample variance* $\widehat{\sigma}^2$ and the square $(\widehat{\mu}-\mu)^2$ of the shifted *sample mean* $\widehat{\mu}-\mu$ has already been proved for $n = 2$ and n=*3*. The general result is obvious.

The third action item

Let us transform the cumulative probability into its canonical forms.

$$\begin{aligned}
dF &= f(y_1, \ldots, y_n)dy_1\cdots dy_n\\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right) dy_1\cdots dy_n\\
&= (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)\widehat{\sigma}^2 + n(\widehat{\mu}-\mu)^2\right]\right) dy_1\cdots dy_n\\
&\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y}-\mathbf{1}\mu) \ \ \mathit{or}\ \ \mathbf{y}-\mathbf{1}\mu = \sigma\mathbf{H}'\mathbf{z}\\
&\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu) = z_1^2 + z_2^2 + \cdots + z_{n-1}^2 + z_n^2.
\end{aligned}$$

Here, we have substituted the divert Helmert transformation (quadratic Helmert matrix $\mathbf{H}$) and its inverse. Again σ^{-1} is the scale factor, also called *dilatation*,

$\mathbf{H}$ an orthonormal matrix, also called *rotation matrix*, and $\mathbf{1}\mu \in \mathbb{R}^n$ the translation, also called *shift*.

$$\begin{aligned}
dF &= f(y_1, \ldots, y_n) dy_1 \cdots dy_n \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_n^2 \frac{1}{(2\pi)^{(n-1)/2}}\right) \exp\left(-\frac{1}{2}(z_1^2 + \cdots + z_n^2)\right) dz_1 dz_2 \cdots dz_{n-1} dz_n
\end{aligned}$$

based upon

$$dy_1 dy_2 \cdots dy_{n-1} dy_n = \sigma^n dz_1 dz_2 \cdots dz_{n-1} dz_n$$
$$J = \sigma^n |\mathbf{H}'| = \sigma^n |\mathbf{H}| = \sigma^n.$$

$\mathbf{J}$ again denotes the absolute value of the *Jacobian determinant* introduced by the *first vehicle*.

The fourth action item

First, we identify the marginal distribution of the sample mean $\widehat{\mu}$.

$$dF_1 = f_1\left(\widehat{\mu}|\mu, \frac{\sigma^2}{n}\right) d\widehat{\mu}.$$

According to the specific structure of the *quadratic Helmert matrix* z_n is generated by

$$z_n = \sigma^{-1}\left[\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right]\begin{bmatrix} y_1 - \mu \\ \vdots \\ y_n - \mu \end{bmatrix} = \sigma^{-1}\frac{y_1 + \cdots + y_n - n\mu}{\sqrt{n}},$$

upon substituting

$$\widehat{\mu} = \frac{1}{n}\mathbf{1}'\mathbf{y} = \frac{y_1 + \cdots + y_n}{n} \Rightarrow y_1 + \cdots + y_n = n\widehat{\mu} \Rightarrow$$

$$\boxed{z_n = \sigma^{-1}\frac{n(\widehat{\mu} - \mu)}{\sqrt{n}} \Rightarrow z_n^2 = n\frac{(\widehat{\mu} - \mu)^2}{\sigma^2},\ dz_n = \frac{1}{\sigma}\sqrt{n}\, d\widehat{\mu}.}$$

Let us implement dz_n in the marginal distribution.

$$dF_1 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_n^2\right) dz_n *$$

$$*\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(2\pi)^{-(n-1)/2}\exp\left[-\frac{1}{2}(z_1^2+\cdots+z_{n-1}^2)\right]dz_1\cdots dz_{n-1},$$

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(2\pi)^{-(n-1)/2}\exp\left[-\frac{1}{2}(z_1^2+\cdots+z_{n-1}^2)\right]dz_1\cdots dz_{n-1}=1$$

$$dF_1=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_n^2\right)dz_n$$

$$\boxed{dF_1=\frac{1}{\sqrt{2\pi}}\frac{1}{\frac{\sigma}{\sqrt{n}}}\exp\left(-\frac{1}{2}\frac{(\widehat{\mu}-\mu)^2}{\frac{\sigma^2}{n}}\right)d\widehat{\mu}=f_1\left(\widehat{\mu}|\mu,\frac{\sigma^2}{n}\right)d\widehat{\mu}.}$$

The fifth action item

Second, we identify the marginal distribution of the sample variance $\widehat{\sigma}^2$. We depart the *ansatz*

$$dF_2=f_2(\widehat{\sigma})d\widehat{\sigma}=f_2(\widehat{\sigma}^2)d\widehat{\sigma}^2$$

in order to determine the *marginal distribution* $f_2(\widehat{\sigma})$ of the sample root-mean-square errors $\widehat{\sigma}$ and the *marginal distribution* $f_2(\widehat{\sigma}^2)$ of the sample variance $\widehat{\sigma}^2$. A first version of the marginal probability distribution dF_2 is

$$dF_2=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}z_n^2\right)dz_n*$$

$$*\frac{1}{(2\pi)^{(n-1)/2}}\exp\left[-\frac{1}{2}(z_1^2+\cdots+z_{n-1}^2)\right]dz_1\cdots dz_{n-1}.$$

Transform the Cartesian coordinates $(z_1+\cdots+z_{n-1})\in\mathbb{R}^{n-1}$ to spherical coordinates $(\Phi_1,\Phi_2,\ldots,\Phi_{n-2},r_{n-1})$. From the operational point of view, $p=n-1$, the number of "*degrees of freedom*", is an optional choice. Let us substitute the *global hypersurface* element ω_{n-1} or ω_p into dF_2, namely

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}z_n^2\right)dz_n=1$$

$$dF_2=\frac{1}{2^{(n-1)/2}}r^{n-2}\exp\left(-\frac{1}{2}r^2\right)dr*$$

$$* \int_{-\pi/2}^{+\pi/2} \cos^{n-3}\phi_{n-2} d\phi_{n-2} \int_{-\pi/2}^{+\pi/2} \cos^{n-4}\phi_{n-3} d\phi_{n-3} \cdots \int_{-\pi/2}^{+\pi/2} \cos\phi_2 d\phi_2 \int_0^{2\pi} d\phi_1$$

$$dF_2 = \frac{1}{(2\pi)^{p/2}} r^{p-1} \exp\left(-\frac{1}{2}r^2\right) dr*$$

$$* \int_{-\pi/2}^{+\pi/2} \cos^{p-2}\phi_{p-1} \int_{-\pi/2}^{+\pi/2} \cos^{p-3}\phi_{p-2} d\phi_{p-2} \cdots \int_{-\pi/2}^{+\pi/2} \cos^2\phi_3 d\phi_3 \int_{-\pi/2}^{+\pi/2} \cos\phi_2 d\phi_2 \int_0^{2\pi} d\phi_1$$

$$\omega_{n-1} = \omega_p = \frac{2}{\Gamma\left(\frac{n-1}{2}\right)} \pi^{(n-1)/2} = \frac{2}{\Gamma\left(\frac{p}{2}\right)} \pi^{p/2}$$

$$= \int_{-\pi/2}^{\pi/2} \cos^{p-2}\phi_{p-1} d\phi_{p-1} \int_{-\pi/2}^{\pi/2} \cos^{p-3}\phi_{p-2} d\phi_{p-2} \cdots$$

$$\int_{-\pi/2}^{\pi/2} \cos^2\phi_3 d\phi_3 \int_{-\pi/2}^{\pi/2} \cos\phi_2 d\phi_2 \int_0^{2\pi} d\varphi_1$$

$$dF_2 = \frac{\omega_p}{(2\pi)^{p/2}} r^{p-1} \exp\left(-\frac{1}{2}r^2\right) dr$$

$$\boxed{dF_2 = \frac{2}{2^{p/2}\Gamma\left(\frac{p}{2}\right)} r^{p-1} \exp\left(-\frac{1}{2}r^2\right) dr}.$$

The *marginal distribution* of the r.m.s. $f_2(\widehat{\sigma})$ is generated as soon as we substitute the radius $r = \sqrt{z_1^2 + \cdots + z_{n-1}^2} = \sqrt{n-1}\,\widehat{\sigma}_1/\sigma$. *Alternatively the marginal distribution* of the *sample variance* $f_2(\widehat{\sigma}^2)$ is produced when we substitute the radius square $r^2 = z_1^2 + \cdots + z_{n-1}^2 = (n-1)\widehat{\sigma}^2/\sigma^2$.

Project A

$$r = \sqrt{n-1}\,\widehat{\sigma}/\sigma = \sqrt{p}\frac{\widehat{\sigma}}{\sigma} \Rightarrow dr = \frac{\sqrt{p}}{\sigma} d\widehat{\sigma}$$

$$dF_2 = f_2(\widehat{\sigma}) d\widehat{\sigma}$$

$$\boxed{f_2(\widehat{\sigma}) = \frac{2p^p}{\sigma^p 2^{p/2}} \widehat{\sigma}^{p-1} \exp\left(-\frac{1}{2}\frac{p}{\sigma^2}\widehat{\sigma}^2\right)}.$$

Indeed, $f_2(\widehat{\sigma})$ establishes the *marginal distribution* of the root-mean-square error $\widehat{\sigma}$ with $p = n - 1$ degrees of freedom.

Project B

$$x := r^2 = (n-1)\frac{\widehat{\sigma}^2}{\sigma^2} = p\frac{\widehat{\sigma}^2}{\sigma^2} =: \chi_p^2 \Rightarrow \left[\begin{array}{l} dx = 2rdr \\ dr = \dfrac{dx}{2r} = \dfrac{1}{2}\dfrac{dx}{\sqrt{x}} \end{array}\right.$$

$$\begin{aligned} r^{p-1}dr &= \frac{1}{2}x^{\frac{p}{2}-1}dx \\ dF_2 &= f_2(x)dx \end{aligned}$$

$$\boxed{f_2(x) := \frac{1}{2^{p/2}\Gamma(\frac{p}{2})}x^{\frac{p}{2}-1}\exp\left(-\frac{1}{2}x\right)}.$$

Finally, we have derived *Helmert's Chi Square* χ_p^2 distribution $f_2(x)dx = dF_2$ by substituting r^2, r^{p-1} and dr in factor of $x := r^2$ and $dx = 2rdr$.

Project C
Replace the radical coordinate squared $r^2 = (n-1)\widehat{\sigma}^2/\sigma^2 = p\widehat{\sigma}^2/\sigma^2$ by rescaling on the basis p/σ^2

$$\begin{aligned} x &= r^2 = z_1^2 + \cdots + z_{n-1}^2 = z_1^2 + \cdots + z_p^2 = (n-1)\frac{\widehat{\sigma}^2}{\sigma^2} = \frac{p}{\sigma^2}\widehat{\sigma}^2 \\ dx &= \frac{p}{\sigma^2}d\widehat{\sigma} \end{aligned}$$

within *Helmert's* χ_p^2 with $p = n - 1$ degrees of freedom

$$dF_2 = f_2(\widehat{\sigma}^2)d\widehat{\sigma}^2$$

$$\boxed{f_2(\widehat{\sigma}^2) = \frac{1}{\sigma^p 2^{p/2}\Gamma(\frac{p}{2})}p^{p/2}\widehat{\sigma}^{p-2}\exp\left(-\frac{1}{2}\frac{p}{\sigma^2}\widehat{\sigma}^2\right)}.$$

Recall that $f_2(\widehat{\sigma}^2)$ establishes the *marginal distribution* of the *sample variance* $\widehat{\sigma}^2$ with $p = n - 1$ degrees of freedom. *Both*, the marginal *pdf* $f_2(\widehat{\sigma})$ of the sample standard deviation $\widehat{\sigma}$, also called root-mean-square error, and the marginal pdf $f_2(\widehat{\sigma}^2)$ of the sample variance $\widehat{\sigma}^2$ document the dependence on the variance σ^2 and its power σ^p.

"Here is our journey's end."(*W. Shakespeare*: Hamlet) ♣

B-42 The Confidence Interval for the Sample Mean, Variance Known

An application of Theorem B.10 is Lemma B.12 where we construct the confidence interval for the sample mean $\widehat{\mu}$, BLUUE of μ, variance σ^2 known, on the basis of its sampling distribution. Example B.14 is an example of a random sample of size $n = 4$.

Lemma B.12. (confidence interval for the sample mean, variance known):

The sampling distribution of the sample mean $\widehat{\mu} = n^{-1}\,\mathbf{1}'\mathbf{y}$, BLUUE of μ, is *Gauss-Laplace* normal, $\widetilde{\mu} \sim \mathcal{N}(\widehat{\mu}|\mu, \sigma^2/n)$, if the observations y_i, $i \in \{1, \ldots, n\}$, are *Gauss-Laplace* i.i.d. The "true" mean μ is an element of the two-sided confidence interval.

$$\mu \in]\widehat{\mu} - \frac{\sigma}{\sqrt{n}} c_{1-\alpha/2}, \widehat{\mu} + \frac{\sigma}{\sqrt{n}} c_{1-\alpha/2}[$$

with confidence

$$P\left\{\widehat{\mu} - \frac{\sigma}{\sqrt{n}} c_{1-\alpha/2} < \mu < \widehat{\mu} + \frac{\sigma}{\sqrt{n}} c_{1-\alpha/2}\right\} = 1 - \alpha$$

of level $1 - \alpha$. For three values of the coefficient of confidence $\gamma = 1 - \alpha$, Table B.9 is a list of associated quantiles $c_{1-\alpha/2}$.

Example B.14. (confidence interval for the sample mean $\widehat{\mu}$, σ^2 known):

Suppose that a *random sample*

$$\boxed{y := \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 3.4 \\ 0.6 \\ 5.6 \end{bmatrix}, \ \widehat{\mu} = 2.7, \ \sigma^2 = 9, \ \text{r.m.s.} : 3}$$

of *four observation* is known from a *Gauss-Lapace normal distribution* with unknown mean μ and a known standard deviation $\sigma = 3$. $\widehat{\mu}$ BLUUE of μ is the *arithmetic mean*. We intend to determine upper and lower limits which are rather certain to contain the unknowns parameter μ between them. Previously, for samples of size 4 we have known that the random variable

$$\boxed{z = \frac{\widehat{\mu} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\widehat{\mu} - \mu}{\frac{3}{2}}}$$

is normally distributed with mean zero and unit variance. $\widehat{\mu}$ is the sample mean 2.7 and 3/2 is $\sigma/\sqrt{n}$. The probability $\gamma = 1 - \alpha$ that z will be between any two arbitrarily chosen numbers $c_1 = -c$ and $c_2 = c$ is

$$P\{-c_1 < z < +c_2\} = \int_{c_1}^{c_2} f(z)dz = \gamma = 1 - \alpha,$$

$$P\{-c < 2 < +c\} = \int_{-c}^{+c} f(z)dz = \gamma = 1 - \alpha,$$

$$\int_{-\infty}^{c} f(z)dz = 1 - \frac{\alpha}{2},$$

$$\int_{-\infty}^{-c} f(z)dz = \frac{\alpha}{2}.$$

γ is the *coefficient of confidence*, α the *coefficient of negative confidence*, also called *complementary coefficient of confidence*. The four representations of the probability $\gamma = 1 - \alpha$ to include z in the *confidence interval* $-c < z < +c$ have led to the *linear Voltera integral equation of the first kind*

$$\int_{-\infty}^{c} f(z)dz = 1 - \frac{\alpha}{2} = \frac{1+\gamma}{2}.$$

Three values of the coefficient of confidence γ or its compliment α are popular and listed in Table B.9.

Table B.9 (Values of the coefficient of confidence):

γ	0.950	0.990	0.999
α	0.050	0.010	0.001
$\frac{\alpha}{2}$	0.025	0.005	0.000, 5
$1 - \frac{\alpha}{2} = \frac{1+\gamma}{2}$	0.975	0.995	0.999, 5

In solving the *linear Voltera integral equation of the first kind*

$$\int_{-\infty}^{z} f(z^*)dz^* = 1 - \frac{1}{2}\alpha(z) = \frac{1}{2}[1 + \gamma(z)].$$

Table B.10 collects the quantiles $c_{1-\alpha/2}$ given the coefficients of confidence on their complements which we listed in Table B.10.

Table B.10 (Quantiles for the confidence interval of the sample mean, variance unknown):

$1-\frac{\alpha}{2}=\frac{1+\gamma}{2}$	γ	α	$c_{1-\alpha/2}$
0.975	0.95	0.05	1.960
0.995	0.99	0.01	2.576
0.999, 5	0.999	0.001	3.291

Given the quantiles $c_{1-\alpha/2}$, we are going to construct the confidence interval for the sample mean $\widehat{\mu}$, the variance σ^2 to be known. For this purpose, we convert the forward transformations $\widehat{\mu} \to z = \sqrt{n}(\widehat{\mu}-\mu)/\sigma$ to μ.

$$\boxed{\begin{gathered}\widehat{\mu}-\frac{\sigma}{\sqrt{n}}z=\mu\\ \widehat{\mu}_1 := \widehat{\mu}-\frac{\sigma}{\sqrt{n}}c_{1-\alpha/2} < \mu < \widehat{\mu}+\frac{\sigma}{\sqrt{n}}c_{1-\alpha/2} =: \widehat{\mu}_2.\end{gathered}}$$

The interval $\widehat{\mu}_1 < \mu < \widehat{\mu}_2$ for the fixed value $z = c_{1-\alpha/2}$ contains the "true" mean μ with probability $\gamma = 1-\alpha$.

$$\boxed{\begin{aligned}&P\left\{\widehat{\mu}-\frac{\sigma}{\sqrt{n}}c_{1-\frac{\alpha}{2}} < \mu < \widehat{\mu}+\frac{\sigma}{\sqrt{n}}c_{1-\frac{\alpha}{2}}\right\}\\ &\quad=\int_{-c}^{c} f(z)dz=\int_{\widehat{\mu}_1}^{\widehat{\mu}_2} f\left(\widehat{\mu}|\mu,\frac{\sigma^2}{n}\right)d\widehat{\mu}=\gamma=1-\alpha\end{aligned}}$$

since

$$\int_{-\infty}^{\widehat{\mu}_2} f(\widehat{\mu})d\widehat{\mu} = \int_{-\infty}^{\widehat{\mu}+\frac{\sigma}{\sqrt{n}}c_{1-\alpha/2}} f(\widehat{\mu})d\widehat{\mu} = \int_{-\infty}^{c_{1-\alpha/2}} f(\widehat{\mu})d\widehat{\mu} = 1-\frac{\alpha}{2}.$$

An animation of the coefficient of confidence and the probability functions of a confidence interval is offered by Figs. B.7 and B.8.

$$\widehat{\mu}_1 = \widehat{\mu} - c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \widehat{\mu}_2 = \widehat{\mu} + c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Let us specify all the integrals to our example

$$\boxed{\int_{-\infty}^{c_{1-\alpha/2}} f(z)dz = 1-\alpha/2}$$

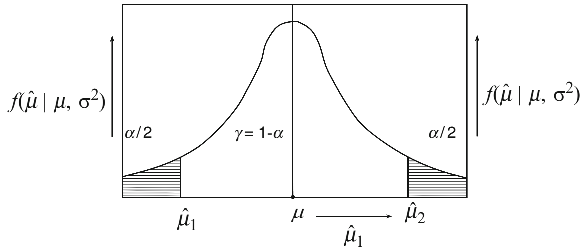


Fig. B.7 Two-sided confidence interval $\mu \in]\widehat{\mu}_1, \widehat{\mu}_2[$, $f(\widehat{\mu}|\mu, \sigma^2/n)$ *pdf*:

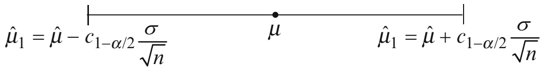


Fig. B.8 Two-sided confidence interval, quantile $c_{1-\alpha/2}$

$$\boxed{\int_{-\infty}^{1.960} f(z)dz = 0.975, \quad \int_{-\infty}^{2.576} f(z)dz = 0.995, \quad \int_{-\infty}^{3.291} f(z)dz = 0.999,5.}$$

Those data lead to a triplet of confidence intervals.

$$case\,(a) = 0.95, = 0.05, \quad c_{1-\alpha/2} = 1.960$$

$$P\left\{2.7 - \frac{3}{2}1.96 < \mu < 2.7 + \frac{3}{2}1.96\right\} = 0.95$$

$$P\{-0.24 < \mu < +5.64\} = 0.95$$

$$case\,(b) = 0.99, = 0.01, \quad c_{1-\alpha/2} = 2.576$$

$$P\left\{2.7 - \frac{3}{2}2.576 < \mu < 2.7 + \frac{3}{2}2.576\right\} = 0.99$$

$$P\{-1.164 < \mu < 6.564\} = 0.99$$

$$case\,(c) = 0.999, = 0.001, \quad c_{1-\alpha/2} = 3.291$$

$$P\left\{2.7 - \frac{3}{2}3.291 < \mu < 2.7 + \frac{3}{2}3.291\right\} = 0.999$$

$$\boxed{P\{-2.236 < \mu < 7.636\} = 0.999}.$$

With probability 95% the "true" mean μ is an element of the interval $]-0.24, +5.64[$. In contrast, with probability 99% the "true" mean μ is an element of the larger interval $]-1.164, +6.564[$. Finally, with probability 99.9% the "true" mean μ is an element of the largest interval $[-2.236, +7.636]$.

B-5 Sampling from the Gauss–Laplace Normal Distribution: A Third Confidence Interval for the Mean, Variance Unknown

In order to derive the sampling distributions for the sample mean, variance unknown, of Gauss-Laplace i.i.d. observation B51 introduces two examples (two and three observations, respectively, for generating *Student's t distribution*. Lemma B.13 reviews Student's t-distribution of the random variable $\sqrt{n'}(\widehat{\mu}-\mu)/\widehat{\sigma}$ where the sample mean $\widehat{\mu}$ is BLUUE of μ, whereas the sample variance $\widehat{\sigma}^2$ is BIQUUE of σ^2. B52 by means of Lemma B.13 introduces the confidence interval for the "true" mean μ variance σ^2 unknown, which is based on *Student's probability distribution*. For easy computation, Table B.12 is its *flow chart*. B53 discusses The *Uncertainly Principle* generated by The Magic Triangle of (a) length of confidence interval, (b) coefficient of negative confidence, also called the uncertainty number, and (c) the number of observations. Various figures and examples pave the way for the *routine analyst's use* of the confidence interval for the mean, variance unknown.

B-51 Student's Sampling Distribution of the Random Variable $(\widehat{\mu}-\mu)/\widehat{\sigma}$

Two examples for $n=2$ or $n=3$ *Gauss-Laplace* i.i.d. observations keep us to derive Student's t-distribution for the random variable $\sqrt{n}(\widehat{\mu}-\mu)/\widehat{\sigma}$ where $\widehat{\mu}$ is BLUUE of μ, whereas $p=n-1$ is BIQUUE of 2. Lemma B.12 and its proof is the highlight of this paragraph in generating the *sampling probability distribution* of *Student's t*.

Example B.15 (Student's t-distribution for two Gauss-Laplace i.i.d. observations):

First, assume an experiment of *two Gauss-Laplace i.i.d. observations* called y_1 and y_2: We want to prove that $(y_1+y_2)/2$ and $(y_1-y_2)2/2$ or the sample mean $\widehat{\mu}$ and the sample variance $\widehat{\sigma}^2$ are stochastically *independent*. y_1 and y_2 are elements of the *joint pdf*.

$$f(y_1,y_2)=f(y_1)f(y_2)$$

$$f(y_1,y_2)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}[(y_1-\mu)^2+(y_2-\mu)^2]\right)$$

$$f(y_1,y_2)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right).$$

The *quadratic form* $(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)$ is decomposed into the sample variance $\widehat{\sigma}^2$, BIQUUE of σ^2, and the deviate of the sample mean $\widehat{\mu}$, BLUUE of μ, from μ by means of the fundamental separation

$$\begin{aligned}
\mathbf{y} - \mathbf{1}\mu &= \mathbf{y} - \mathbf{1}\widehat{\mu} + \mathbf{1}(\widehat{\mu} - \mu) \\
(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) &= (\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) + \mathbf{1}'\mathbf{1}(\widehat{\mu} - \mu)^2 \\
(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) &= \widehat{\sigma}^2 + 2(\widehat{\mu} - \mu)^2
\end{aligned}$$

$$\boxed{f(y_1, y_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\widehat{\sigma}^2\right) \exp\left(-\frac{2}{2\sigma^2}(\widehat{\mu} - \mu)^2\right).}$$

The joint *pdf* $f(y_1, y_2)$ is transformed into a special form if we replace

$$\begin{aligned}
\widehat{\mu} &= (y_1 + y_2)/2, \\
\widehat{\sigma}^2 &= (y_1 - \widehat{\mu})^2 + (y_2 - \widehat{\mu})^2 = \frac{1}{2}(y_1 - y_2)^2,
\end{aligned}$$

namely

$$\boxed{f(y_1, y_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\frac{1}{2}(y_1 - y_2)^2\right) \exp\left(-\frac{1}{\sigma^2}(\frac{y_1 + y_2}{2} - \mu)^2\right).}$$

Obviously the product decomposition of the joint pdf documented that (y1+y2)/2 and (y1-y2)2/2 or $\widehat{\mu}$ and $\widehat{\sigma}^2$ are *independent random variables.*
Second, we intend to derive the pdf of *Student's random variable* $t := \sqrt{2}(\widehat{\mu} - \mu)/\widehat{\sigma}$, the deviate of the sample mean $\widehat{\mu}$ form the "true" mean μ, normalized by the sample standard deviations $\widehat{\sigma}$. Let us introduce the *direct Helmert transformation*

$$z_1 = \frac{1}{\sigma}\frac{y_1 - y_2}{\sqrt{2}} = \frac{\widehat{\sigma}}{\sigma}, \quad z_2 = \frac{\sqrt{2}}{\sigma}\left(\frac{y_1 + y_2}{2} - \mu\right) = \frac{2}{\sigma}\frac{\widehat{\mu} - \mu}{\sqrt{2}},$$

or

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \frac{1}{\sigma}\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} y_1 - \mu \\ y_2 - \mu \end{bmatrix},$$

as well as the inverse

$$\begin{bmatrix} y_1 - \mu \\ y_2 - \mu \end{bmatrix} = \sigma\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

which brings the $jointpdf \;\; dF = f(y_1, y_2)dy_1dy_2 \; = \; f(z_1, z_2)dz_1dz_2$ into the *canonical form.*

$$\boxed{dF = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_1^2\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_2^2\right) dz_1 dz_2}$$

$$\frac{1}{\sigma^2} dy_1 dy_2 = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} dz_1 dz_2 = dz_1 dz_2.$$

The *Helmert random variable* $x := z_1^2$ or $z_1 = \sqrt{x}$ replaces the random variable z_1 such that

$$dz_1 dz_2 = \frac{1}{2}\frac{1}{\sqrt{x}} dx dz_2$$

$$\boxed{dF = \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{x}} \exp\left(-\frac{1}{2}x\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_2^2\right) dx dz_2}$$

is the joint pdf of x and z_2. *Finally*, we introduce *Student's random variable*

$$t := \sqrt{2}\frac{\widehat{\mu} - \mu}{\widehat{\sigma}}$$

$$z_2 = \frac{\sqrt{2}}{\sigma}(\widehat{\mu} - \mu) \Rightarrow \widehat{\mu} - \mu = \frac{\sigma}{\sqrt{2}} z_2$$

$$z_1 = \sqrt{x} = \frac{\widehat{\sigma}}{\sigma} \Rightarrow \widehat{\sigma} = \sigma z_1 = \sigma\sqrt{x}$$

$$\boxed{t = \frac{z_2}{\sqrt{x}} \Leftrightarrow z_2 = \sqrt{x}\, t \Rightarrow z_2^2 = x\, t^2}.$$

Let us transform $dF = f(x, z2) dx dz_2$ to $dF = f(t, x) dt dx$, namely from the joint *pdf* of the *Helmert random variable* x and the *Gauss-Laplace normal variate* z_2 *to* the *joint pdf* of the *Student random variable* t and the *Helmert random variable* x.

$$dz_2 dx = \begin{vmatrix} D_t z_2 & D_x z_2 \\ D_t x & D_x x \end{vmatrix} dt\, dx$$

$$dz_2 dx = \begin{vmatrix} \sqrt{x} & \frac{\sqrt{2}}{2} x^{-3/2} t \\ 0 & 1 \end{vmatrix} dt\, dx$$

$$dz_2 dx = \sqrt{x}\, dt\, dx$$

$$dF = f(t, x) dt dx$$

$$f(t, x) = \frac{1}{2\pi}\frac{1}{2} \exp\left[-\frac{1}{2}(1 + t^2)x\right].$$

The *marginal distribution of Student's random variable* t, namely $dF_3 = f_3(t)dt$, is generated by

$$f_3(t) := \frac{1}{2\pi}\frac{1}{2}\int_0^\infty \exp\left[-\frac{1}{2}(1+t^2)x\right]dx$$

subject to the standard integral

$$\int_0^\infty \exp(-\beta x)dx = \left[-\frac{1}{\beta}\exp(-\beta x)\right]_0^\infty = \frac{1}{\beta}$$

$$\beta := \frac{1}{2}(1+t^2), \quad \frac{1}{\beta} = \frac{2}{1+t^2}$$

such that

$$\boxed{\begin{aligned} f_3(t) &= \frac{1}{2\pi}\frac{1}{1+t^2}, \\ dF_3 &= \frac{1}{2\pi}\frac{1}{1+t^2}dt \end{aligned}}$$

and characterized by a *pdf* $f_3(t)$ which is reciprocal to $(1+t^2)$.

Example B.16. (Student's t-distribution for three Gauss-Laplace i.i.d. observations):

First, assume an experiment of three *Gauss-Laplace i.i.d. observations* called y_1, y_2 and y_3: We want to derive the *joint pdf* $f(y_1, y_2, y_3)$ in terms of the sample mean $\widehat{\mu}$, BLUUE of μ, and the sample variance $\widehat{\sigma}^2$, BIQUUE of σ^2.

$$f(y_1, y_2, y_3) = f(y_1)f(y_2)f(y_3)$$

$$f(y_1, y_2, y_3) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{1}{2\sigma^2}[(y_1-\mu)^2 + (y_2-\mu)^2 + (y_3-\mu)^2]\right)$$

$$f(y_1, y_2, y_3) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)\right).$$

The *quadratic form* $(\mathbf{y}-\mathbf{1}\mu)'(\mathbf{y}-\mathbf{1}\mu)$ is decomposed into the sample variance $\widehat{\sigma}^2$ and the deviate sample mean $\widehat{\mu}$ from the "true" mean μ by means of the fundamental separation

$$\mathbf{y} - \mathbf{1}\mu = \mathbf{y} - \mathbf{1}\widehat{\mu} + \mathbf{1}(\widehat{\mu} - \mu)$$
$$(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) = (\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu}) + \mathbf{1}'\mathbf{1}(\widehat{\mu} - \mu)^2$$
$$(\mathbf{y} - \mathbf{1}\mu)'(\mathbf{y} - \mathbf{1}\mu) = 2\widehat{\sigma}^2 + 3(\widehat{\mu} - \mu)^2$$

$$\boxed{\begin{aligned} dF &= f(y_1, y_2, y_3) dy_1 dy_2 dy_3 \\ &= \frac{1}{(2\pi)^{3/2}\sigma^3} \exp\left(-\frac{1}{2\sigma^2} 2\widehat{\sigma}^2\right) \exp\left(-\frac{3}{2\sigma^2}(\widehat{\mu} - \mu)^2\right) dy_1 dy_2 dy_3. \end{aligned}}$$

Second, we intend to derive the *pdf of Student's random variable* $t := \sqrt{3}(\widehat{\mu} - \mu)/\widehat{\sigma}$, the deviate of the sample mean $\widehat{\mu}$ from the "true" mean μ, normalized by the sample standard deviation $\widehat{\sigma}$. Let us introduce the direct *Helmert transformation*

$$z_1 = \frac{1}{\sigma\sqrt{2}}(y_1 - \mu + y_2 - \mu) = \frac{1}{\sigma\sqrt{2}}(y_1 + y_2 - 2\mu)$$
$$z_2 = \frac{1}{\sigma\sqrt{2 \cdot 3}}(y_1 - \mu + y_2 - \mu - 2y_3 + 2\mu) = \frac{1}{\sigma\sqrt{2 \cdot 3}}(y_1 + y_2 - 2y_3)$$
$$z_3 = \frac{1}{\sigma\sqrt{3}}(y_1 - \mu + y_2 - \mu + y_3 - \mu) = \frac{1}{\sigma\sqrt{3}}(y_1 + y_2 + y_3 - 3\mu)$$

or

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \frac{1}{\sigma} \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{1\cdot 2}} & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & -\frac{2}{\sqrt{2\cdot 3}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} y_1 - \mu \\ y_2 - \mu \\ y_3 - \mu \end{bmatrix}$$

as well as its inverse

$$\begin{bmatrix} y_1 - \mu \\ y_2 - \mu \\ y_3 - \mu \end{bmatrix} = \sigma \begin{bmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

in general

$$\mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mu) \quad \textit{versus } (\mathbf{y} - \mathbf{1}\mu) = \sigma\mathbf{H}'\mathbf{z},$$

which help us to bring the *joint pdf* $dF(y_1, y_2, y_3)dy_1dy_2dy_3 = f(z_1, z_2, z_3)dz_1dz_2dz_3$ into the *canonical form.*

$$\boxed{dF = \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{1}{2}(z_1^2 + z_2^2)\right) \exp\left(-\frac{1}{2}z_3^2\right) dz_1 dz_2 dz_3}$$

$$\frac{1}{\sigma^3} dy_1 dy_2 dy_3 = \begin{vmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3}} \end{vmatrix} dz_1 dz_2 dz_3 = dz_1 dz_2 dz_3.$$

The *Helmert random variable* $x := z_1^2 + z_2^2$ or $\sqrt{x} = \sqrt{z_1^2 + z_2^2}$ replaces the random variable $\sqrt{z_1^2 + z_2^2}$ as soon as we introduce polar coordinates $z_1 = r\cos\phi_1$, $z_2 = r\sin\phi_1$, $z_1^2 + z_2^2 = r^2 =: x$ and compute the *marginal pdf*

$$dF(z_3, x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_3^2\right) dz_3 \frac{1}{2\pi}\frac{1}{2} \exp\left(-\frac{1}{2}x\right) dx \int_0^{2\pi} d\phi_1,$$

by means of

$$\begin{aligned} x &:= z_1^2 + z_2^2 = r^2 \Rightarrow dx = 2r dr, \quad dr = \frac{dx}{2r}, \\ dz_1 dz_2 &= r dr d\phi_1 = \frac{1}{2} dx d\phi_1, \\ dF(z_3, x) &= \frac{1}{2}\exp\left(-\frac{1}{2}x\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_3^2\right) dx dz_3, \end{aligned}$$

the *joint pdf* of x and z. *Finally*, we inject *Student's random variable*

$$\boxed{t := \sqrt{3}\frac{\widehat{\mu} - \mu}{\widehat{\sigma}},}$$

decomposed into

$$\begin{aligned} z_3 &= \frac{\sqrt{3}}{\sigma}(\widehat{\mu} - \mu) \Rightarrow \widehat{\mu} - \mu = \frac{\sigma}{\sqrt{3}} z_3 \\ \sqrt{z_1^2 + z_2^2} &= \sqrt{x} = \sqrt{2}\frac{\widehat{\sigma}}{\sigma} \Rightarrow \widehat{\sigma} = \frac{\sigma}{\sqrt{2}}\sqrt{z_1^2 + z_2^2} = \frac{\sigma}{\sqrt{2}}\sqrt{x} \end{aligned}$$

$$\boxed{t = \frac{z_3}{\sqrt{x}}\sqrt{2} \Leftrightarrow z_3 = \frac{1}{\sqrt{2}}\sqrt{x}\, t \Rightarrow z_3^2 = \frac{1}{2}x\, t^2.}$$

Let us transform $dF = f(x, z_3)dxdz_3\, to\, dF = f(t, x)dtdx$. Alternatively we may say that we transform the *joint pdf* of the *Helmert random variable* x and the *Gauss-Laplace normal variate* z_3 to the *joint pdf* of the Student random variable t and the *Helmert random variable* x.

$$dz_3 dx = \begin{vmatrix} D_t z_3 & D_x z_3 \\ D_t x & D_x x \end{vmatrix} dtdx$$

$$dz_3 dx = \begin{vmatrix} \frac{\sqrt{x}}{\sqrt{2}} & \frac{1}{2\sqrt{2}} x^{-3/2} t \\ 0 & 1 \end{vmatrix} dtdx$$

$$dz_3 dx = \frac{\sqrt{x}}{\sqrt{2}} dtdx$$

$$dF = f(t, x)dtdx$$

$$\boxed{f(t, x) = \frac{1}{2\sqrt{2}} \frac{1}{\sqrt{2\pi}} \sqrt{x} \exp\left[-\frac{1}{2}(1 + \frac{t^2}{2})x\right].}$$

The *marginal distribution of Student's random variable* t, namely $dF_3 = f_3(t)dt$, is generated by

$$f_3(t) := \frac{1}{2\sqrt{2}} \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{x} \exp\left[-\frac{1}{2}\left(1 + \frac{t^2}{2}\right)x\right] dx$$

subject to the standard integral

$$\int_0^\infty x^\alpha \exp(-\beta x)dx = \frac{\Gamma(\alpha + 1)}{\beta^{\alpha+1}}, \quad \alpha = \frac{1}{2}, \quad \beta = \frac{1}{2}\left(1 + \frac{t^2}{2}\right)$$

such that

$$\int_0^\infty \sqrt{x} \exp\left[-\frac{1}{2}\left(1 + \frac{t^2}{2}\right)x\right] dx = 2^{3/2} \frac{\Gamma(\frac{3}{2})}{(1 + \frac{t^2}{2})^{3/2}}$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

$$\boxed{\begin{aligned} f_3(t) &= \Gamma\left(\frac{3}{2}\right) \frac{1}{\sqrt{2\pi}} \frac{1}{(1+\frac{t^2}{2})^{3/2}}, \\ dF_3 &= \frac{\sqrt{2}}{4(1 + \frac{t^2}{2})^{3/2}} dt. \end{aligned}}$$

Again *Student's t-distribution* is reciprocal $t(1 + t2/2)3/2$.

Lemma B.13. Student's t-distribution for the derivate of the mean $(\widehat{\mu}-\mu)/\widehat{\sigma}$ (Gosset, 1908)):

Let the random vector of observations $\mathbf{y} = [y_1, \ldots, y_n]'$ be *Gauss-Laplace i.i.d. Student's random variable*

$$t := \frac{\widehat{\mu}-\mu}{\widehat{\sigma}}\sqrt{n},$$

where the sample mean $\widehat{\mu}$ is BLUUE of μ and the sample variance $\widehat{\sigma}^2$ is BIQUUE of σ^2, associated to the *pdf*

$$f(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})}\frac{1}{\sqrt{p\pi}}\frac{1}{(1+\frac{t^2}{p})^{(p+1)/2}}.$$

$p = n-1$ is the "*degree of freedom*" of Student's distribution $f_p(t)$. Gosset (1908) published the t-distribution of the ratio $\sqrt{n}(\widehat{\mu}-\mu)/\widehat{\sigma}$ under the pseudonym "Student": The probable error of a mean.

Proof.

The *joint probability distribution* of the random variable $z_n := \sqrt{n}(\widehat{\mu}-\mu)/\sigma$ and the *Helmert random variable* $x := z_1^2 + \cdots + z_{n-1}^2 = (n-1)\widehat{\sigma}^2/\sigma^2$ represented by

$$dF = f_1(z_n) f_2(x) dz_n dx$$

due to the effect that z_n and $z_1^2 + \cdots + z_{n-1}^2$ or $\widehat{\mu}-\mu$ and $(n-1)\widehat{\sigma}^2$ are stochastically independent. Let us take reference to the specific *pdfs* with n and $n-1 = p$ degrees of freedom

$$dF_1 = f_1(z_n)dz_n \quad \textit{and} \quad dF_2 = f_2(x)dx,$$

$$f_1(z_n) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_n^2\right) \quad \textit{and} \quad f_2(x) = \frac{1}{\Gamma(\frac{p}{2})}\left(\frac{1}{2}\right)^{p/2} x^{(p-2)/2}\exp\left(-\frac{1}{2}x\right),$$

or

$$dF_1 = f_1(\widehat{\mu})d\widehat{\mu} \quad \textit{and} \quad dF_2 = f_2(\widehat{\sigma}^2)d\widehat{\sigma}^2,$$

$$f_1(\widehat{\mu}) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\exp\left(-\frac{n}{2}\frac{(\widehat{\mu}-\mu)^2}{\sigma^2}\right) \quad \textit{and} \quad \begin{aligned} f_2(\widehat{\sigma}^2) = {} & \frac{1}{\sigma^p 2^{p/2}\Gamma(p/2)} p^{p/2}\widehat{\sigma}^{p-2} \\ & \exp\left(-\frac{1}{2}\frac{p}{\sigma^2}\widehat{\sigma}^2\right) \end{aligned}$$

we derived earlier. Here, let us introduce *Student's random variable*

$$\boxed{t := \frac{z_n}{\sqrt{x}}\sqrt{n-1} = \frac{\widehat{\mu}-\mu}{\widehat{\sigma}}\sqrt{n} \quad or \quad z_n = \frac{\sqrt{x}}{\sqrt{n-1}}t = \frac{\sqrt{x}}{\sqrt{p}}t.}$$

By means of the *Jacobi matrix* **J** and its *absolute Jacobi determinant* $|\mathbf{J}|$

$$\begin{bmatrix} dt \\ dx \end{bmatrix} = \begin{bmatrix} D_{z_n}t & D_x t \\ D_{z_n}x & D_x x \end{bmatrix} \begin{bmatrix} dz_n \\ dx \end{bmatrix} = \mathbf{J} \begin{bmatrix} dz_n \\ dx \end{bmatrix}$$

$$\mathbf{J} = \begin{bmatrix} \frac{\sqrt{p}}{\sqrt{x}} & -\frac{1}{2}x^{-3/2}z_n\sqrt{p} \\ 0 & 1 \end{bmatrix}, \quad |\mathbf{J}| = \frac{\sqrt{p}}{\sqrt{x}}, \quad |\mathbf{J}|^{-1} = \frac{\sqrt{x}}{\sqrt{p}}$$

we transform the surface element $dz_n dx$ to the surface element $|\mathbf{J}|^{-1}dtdx$, namely

$$\boxed{dz_n dx = \frac{\sqrt{x}}{\sqrt{p}}dtdx.}$$

Lemma B.14. (Gamma Function):

Our final action item is to calculate the *marginal probability distributions* $dF_3 = f_3(t)dt$ of *Student's random variable* t, namely

$$dF_3 = f_3(t)dt$$

$$f_3(t) := \frac{1}{\sqrt{2p\pi}}\frac{1}{\Gamma(p/2)}\left(\frac{1}{2}\right)^{p/2}\int_0^\infty x^{(p-1)/2}\exp\left(-\frac{1}{2}\left(1+\frac{t^2}{p}\right)x\right)dx.$$

Consult (Gröbner and Hofreiter, 1973, p. 55), for the standard integral

$$\int_0^\infty x^\alpha \exp(-\beta x)dx = \frac{\alpha!}{\beta^{\alpha+1}} = \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \quad and$$

$$\int_0^\infty x^{(p-1)/2}\exp\left[-\frac{1}{2}\left(1+\frac{t^2}{p}\right)x\right]dx = 2^{(p+1)/2}\frac{\Gamma(\frac{p+1}{2})}{(1+\frac{t^2}{p})^{(p+1)/2}},$$

where $p = n-1$ is the rank of the *quadratic Helmert matrix* $\mathbf{H}_n$. Notice a result of the *gamma function* $\Gamma(\frac{p+1}{2}) = (\frac{p-1}{2})!$. In summary, substituting the standard integral

$$dF_3 = f_3(t)dt$$

$$f_3(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{\sqrt{p\pi}} \frac{1}{(1 + \frac{t^2}{p})^{(p+1)/2}}$$

resulted in *Student's t distribution*, namely the pdf of *Student's random variable* $(\widehat{\mu} - \mu)/\widehat{\sigma}$.

B-52 The Confidence Interval for the Mean, Variance Unknown

Lemma B.12 is the basis for the construction of the *confidence interval of the "true" mean*, variance unknown, which we summarize in Lemma B.13, Example B.17 contains all details for computing such a confidence interval, namely Table B.11, a collection of the most popular values of the coefficient of confidence, as well as Tables B.12, B.13 and B.14, listing the *quantiles* for the confidence interval of the Student random variable with $p = n - 1$ degrees of freedom. Figures B.9 and B.10 illustrate the probability of two-sided confidence interval for the mean, variance unknown, and the limits of the confidence interval. Table B.15 as a *flow chart* paves the way for the "*fast computation*" of the confidence interval for the "true" mean, variance unknown.

Lemma B.15. (confidence interval for the sample mean, variance unknown):

The random variable $t = \sqrt{n}(\widehat{\mu} - \mu)/\widehat{\sigma}$ characterized by the ratio of the deviate of the sample mean $\widehat{\mu} = n^{-1}\mathbf{1}'\mathbf{y}$, BLUUE of μ, from the "true" mean μ and the standard deviation $\widehat{\sigma}$, $\widehat{\sigma}^2 = (\mathbf{y} - \mathbf{1}\widehat{\mu})'(\mathbf{y} - \mathbf{1}\widehat{\mu})/(n - 1)$ BIQUUE of the "true" variance σ^2, has the *Student t-distribution* with $p = n - 1$ "degrees of freedom". The "true" mean μ is an element of the two-sided confidence interval

$$\mu \in]\widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{n}} c_{1-\alpha/2}, \quad \widehat{\mu} + \frac{\widehat{\sigma}}{\sqrt{n}} c_{1-\alpha/2}[$$

with confidence

$$\mathrm{P}\left\{\widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{n}} c_{1-\alpha/2} < \mu < \widehat{\mu} + \frac{\widehat{\sigma}}{\sqrt{n}} c_{1-\alpha/2}\right\} = 1 - \alpha$$

of level $1-\alpha$. For three values of the coefficient of confidence $\gamma = 1-\alpha$, Table B.12 is a list of associated quantiles $c_{1-\alpha/2}$.

Example B.17. (confidence interval for the example mean $\widehat{\mu}$, σ^2 unknown):

Suppose that a *random sample*

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 3.4 \\ 0.6 \\ 5.6 \end{bmatrix}, \quad \widehat{\mu} = 2.7, \quad \widehat{\sigma}^2 = 5.2, \quad \widehat{\sigma} = 2.3$$

of *four observations* is characterized by the sample mean $\widehat{\mu} = 2.7$ and the sample variance $\widehat{\sigma}^2 = 5.2$. $2(2.7 - \mu)/2.3 = \sqrt{n}(\widehat{\mu} - \mu)/\widehat{\sigma} = t$ has *Student's pdf* with $p = n-1 = 3$ degrees of freedom. The probability $\gamma = 1-\alpha$ that t will be between any two arbitrarily chosen numbers $c_1 = -c$ and $c_2 = +c$ is

$$\mathrm{P}\{-c_1 < t < c_2\} = \int\limits_{c_1}^{c_2} f(t)dt = \gamma = 1 - \alpha$$

$$\mathrm{P}\{-c < t < c\} = \int\limits_{-c}^{+c} f(t)dt = \gamma = 1 - \alpha$$

$$\mathrm{P}\{-c < t < c\} = \int\limits_{-\infty}^{c} f(t)dt - \int\limits_{-\infty}^{-c} f(t)dt = \gamma = 1 - \alpha$$

$$\int\limits_{-\infty}^{c} f(t)dt = 1 - \frac{\alpha}{2}, \quad \int\limits_{-\infty}^{-c} f(t)dt = \frac{\alpha}{2}$$

γ is the *coefficient of confidence*, α the coefficient of negative confidence, also called *complementary* coefficient of confidence. The four representations of the probability $\gamma = 1-\alpha$ to include t in the confidence interval $-c < t < +c$ have led to the linear Volterra integral equation of the first kind

$$\int\limits_{-\infty}^{c} f(t)dt = 1 - \frac{\alpha}{2} = \frac{1}{2}(1 + \gamma).$$

Three values of the coefficient of confidence γ or its complement α are popular and listed in Table B.10.

Table B.11 (Values of the coefficient of confidence):

γ	0.950	0.990	0.999
α	0.050	0.010	0.001
$\frac{\alpha}{2}$	0.025	0005	0.000, 5
$1 - \frac{\alpha}{2} = \frac{1+\gamma}{2}$	0.975	0.995	0.999, 5

In solving the *linear Volterra integral equation of the first kind*

$$\int_{-\infty}^{t} f(t^*)dt^* = 1 - \frac{1}{2}\alpha(t) = \frac{1}{2}[1 + \gamma(t)],$$

which depends on the degrees of freedom $p = n - 1$, Tables B.12–B.14 collect the *quantiles* $c_{1-\alpha/2}/\sqrt{n}$ given the coefficients of confidence or their complements which we listed in Table B.11.

Table B.12 (Quantiles $c_{1-\alpha/2}/\sqrt{n}$ for the confidence interval of the Student random variable with $p = n - 1$ degrees of freedom):

$$1 - \alpha/2 = (1 + \gamma)/2 = 0.975, \quad \gamma = 0.95, \quad \alpha = 0.05$$

p	n	$\frac{1}{\sqrt{n}}c_{1-\alpha/2}$	p	n	$\frac{1}{\sqrt{n}}c_{1-\alpha/2}$
1	2	8.99	14	15	0.554
2	3	2.48	19	20	0.468
3	4	1.59	24	25	0.413
4	5	1.24	29	30	0.373
5	6	1.05	39	40	0.320
6	7	0.925	49	50	0.284
7	8	0.836	99	100	0.198
8	9	0.769	199	200	0.139
9	10	0.715	499	500	0.088

Table B.13 (Quantiles $c_{1-\alpha/2}/\sqrt{n}$ for the confidence interval of the Student random variable with $p = n - 1$ degrees of freedom):

$$1 - \alpha/2 = (1 + \gamma)/2 = 0.995, \quad \gamma = 0.990, \quad \alpha = 0.010$$

p	n	$\frac{1}{\sqrt{n}}c_{1-\alpha/2}$	p	n	$\frac{1}{\sqrt{n}}c_{1-\alpha/2}$
1	2	45.01	14	15	0.769
2	3	5.73	19	20	0.640
3	4	2.92	24	25	0.559
4	5	2.06	29	30	0.503
5	6	1.65	39	40	0.428
6	7	1.40	49	50	0.379
7	8	1.24	99	100	0.263
8	9	1.12	199	200	0.184
9	10	1.03	499	500	0.116

Table B.14 (Quantiles $c_{1-\alpha/2}/\sqrt{n}$ for the confidence interval of the Student random variable with $p = n - 1$ degrees of freedom):

$$1 - \alpha/2 = (1 + \gamma)/2 = 0.999.5, \quad \gamma = 0.999, \quad \alpha = 0.001$$

p	n	$c_{1-\alpha/2}$	$c_{1-\alpha/2}/\sqrt{n}$	p	n	$c_{1-\alpha/2}$	$c_{1-\alpha/2}/\sqrt{n}$
1	2	636.619	450.158	14	15	4.140	1.069
2	3	31.598	18.243	19	20	3.883	0.868
3	4	12.941	6.470	24	25	3.725	0.745
4	5	8.610	3.851	29	30	3.659	0.668
5	6	6.859	2.800	30	41	3.646	0.655
6	7	5.959	2.252	40	41	3.551	0.555
7	8	5.405	1.911	60	61	3.460	0.443
8	9	5.041	1.680	120	121	3.373	0.307
9	10	4.781	1.512	∞	∞	3.291	0

Since *Student's pdf* depends on $p = n - 1$, we have tabulated $c_{1-\alpha/2}/\sqrt{n}$ for the confidence interval we are going to construct. The Student's random variable $t = \sqrt{n}(\widehat{\mu} - \mu)/\widehat{\sigma}$ is solved for μ, evidently our motivation to introduce the confidence interval $\widehat{\mu}_1 < \mu < \widehat{\mu}_2$

$$t = \sqrt{n}\frac{\widehat{\mu} - \mu}{\widehat{\sigma}} \Rightarrow \widehat{\mu} - \mu = \frac{\widehat{\sigma}}{\sqrt{n}}t \Rightarrow \mu = \widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{n}}t$$

$$\widehat{\mu}_1 := \widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{n}}c_{1-\alpha/2} < \mu < \widehat{\mu} + \frac{\widehat{\sigma}}{\sqrt{n}}c_{1-\alpha/2} =: \widehat{\mu}_2.$$

The interval $\widehat{\mu}_1 < \mu < \widehat{\mu}_2$ for the fixed value $t = c_{1-\alpha/2}$ contains the "true" mean μ with probability $\gamma = 1 - \alpha$.

$$\mathrm{P}\left\{\widehat{\mu} - \frac{\widehat{\sigma}}{\sqrt{\mathrm{n}}}c_{1-\alpha/2} < \mu < \widehat{\mu} + \frac{\widehat{\sigma}}{\sqrt{\mathrm{n}}}c_{1-\alpha/2}\right\}$$

$$= \int_{-c}^{+c} f(t)dt = \gamma = 1 - \alpha$$

because of

$$\int_{-\infty}^{c} f(t)dt = \int_{-\infty}^{c_{1-\alpha/2}} f(t)dt = 1 - \frac{\alpha}{2}.$$

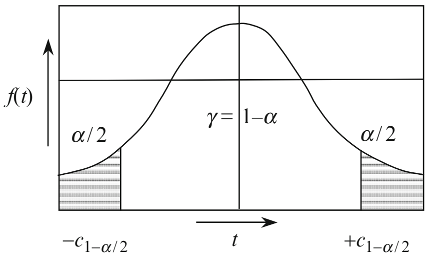


Fig. B.9 Two-sided confidence interval $\mu \in]\widehat{\mu}_1, \widehat{\mu}_2[$, Student's pdf for $p = 3$ degrees of freedom $(n = 4)$ $\widehat{\mu}_1 = \widehat{\mu} - \widehat{\sigma} c_{1-\alpha/2}/\sqrt{n}$, $\widehat{\mu}_2 = \widehat{\mu} + \widehat{\sigma} c_{1-\alpha/2}/\sqrt{n}$

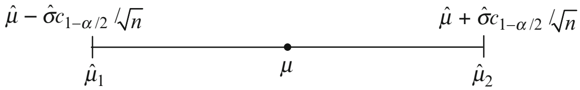


Fig. B.10 Two-sided confidence interval for the "true" mean μ, quantile $c_{1-\alpha/2}$

Figures B.9 and B.10 illustrate the coefficient of confidence and the probability function of a confidence interval.
Let us specify all the integrals to our example.

$$\int_{-\infty}^{c_{1-\alpha/2}} f(t)dt = 1 - \alpha/2$$

$$\int_{-\infty}^{c_{0.975}} f(t)dt = 0.975, \quad \int_{-\infty}^{c_{0.995}} f(t)dt = 0.995, \quad \int_{-\infty}^{c_{0.999,5}} f(t)dt = 0.999,5.$$

These data substituted into Tables B.12–B.14 lead to the triplet of confidence intervals for the "true" mean.

$$\boxed{\begin{gathered} case\,(a) : \gamma = 0.95, \quad \alpha = 0.05, \quad 1 - \alpha/2 = 0.975 \\ p = 3, \quad n = 4, \quad c_{1-\alpha/2}/\sqrt{n} = 1.59 \end{gathered}}$$

$$\mathrm{P}\{2.7 - 2.3 * 1.59 < \mu < 2.7 + 2.3 * 1.59\} = 0.95$$

$$\boxed{\mathrm{P}\{-0.957 < \mu < 6.357\} = 0.95.}$$

$$\boxed{\begin{gathered} case\,(b) : \gamma = 0.99, \quad \alpha = 0.01, \quad 1 - \alpha/2 = 0.995 \\ p = 3, \quad n = 4, \quad c_{1-\alpha/2}/\sqrt{n} = 2.92 \end{gathered}}$$

$$P\{2.7 - 2.3 * 2.92 < \mu < 2.7 + 2.3 * 2.92\} = 0.99$$

$$\boxed{P\{-4.016 < \mu < 9.416\} = 0.99.}$$

$$\boxed{case\,(c) : \gamma = 0.999, \quad \alpha = 0.001, \quad 1 - \alpha/2 = 0.999, 5}$$

$$p = 3, n = 4, c_{1-\alpha/2}/\sqrt{n} = 6.470$$
$$P\{2.7 - 2.3 * 6.470 < \mu < 2.7 + 2.3 * 6.470\} = 0.999$$

$$\boxed{P\{-12.181 < \mu < 17.581\} = 0.999.}$$

The results may be summarized as follows:

> With probability 95% the "true" mean μ is an element of the interval $]-0.957, +6.357[$. In contrast, with probability 99% the "true" mean μ is an element of the larger interval $]-4.016, +9.416[$. Finally, with probability 99.9is an element of the *largest interval* $]-12.181, +17.581[$. If we compare the confidence intervals for the mean μ, σ^2 *known*, versus σ^2 *unknown*, we realize much larger intervals for the second model. Such a result is not too much a surprise, since the model "σ^2 *unknown*" is much weaker than the model "σ^2 *known*".

Table B.15 (Flow chart: Confidence interval for the mean μ, σ^2 unknown):

γ

Choose a coefficient of confidence γ according to *Table B11*.
Solve the linear Volterra integral equation of the first kind $F(c_{1-\alpha/2}) = 1-\alpha/2 = (1+\gamma)/2$ by *Table B12*, *Table B13* or *Table B14* for a *Student pdf* with $p = n-1$ degrees of freedom: read $c_{1-\alpha/2}/\sqrt{n}$.
Compute the sample mean $\hat{\mu}$ and the sample standard deviation $\hat{\sigma}$.
Compute $\hat{\sigma}c_{1-\alpha/2}/\sqrt{n}$ and $\hat{\mu}-\hat{\sigma}c_{1-\alpha/2}/\sqrt{n}$, $\hat{\mu}+\hat{\sigma}c_{1-\alpha/2}/\sqrt{n}$.

B-53 The Uncertainty Principle

Figure B.11 is the graph of the function

$$\Delta\mu(n;\alpha) := 2\widehat{\sigma}c_{1-\alpha/2}/\sqrt{n},$$

where $\Delta\mu$ is the length of the confidence interval of the "true" mean μ, σ^2 unknown. The independent variable of the function is the number of observations n. The function $\Delta\mu(n;\alpha)$ is plotted for fixed values of the coefficient of complementary confidence α, namely $\alpha = 10\%,\ 5\%,\ 1\%$. For reasons given later the coefficient of complementary confidence α is called *uncertainty number*. The graph function $\Delta\mu(n;\alpha)$ illustrates two important facts.

Fact #1: For a constant uncertainty number α, the length of the confidence interval $\Delta\mu$ is smaller the larger the number of observations n is chosen.

Fact #2: For a constant number of observations n, the smaller the number of uncertainly α is chosen, the larger is the confidence interval $\Delta\mu$.

Evidently, the diversive influences of (a) the length of the confidence interval $\Delta\mu$, (b) the uncertainly number α and (c) the number of observations n, which we collected in *The Magic Triangle of Figure B.12*, constitute The *Uncertainly Principle*, in formula

$$\Delta\mu(n-1) \geq: k_{\mu\alpha},$$

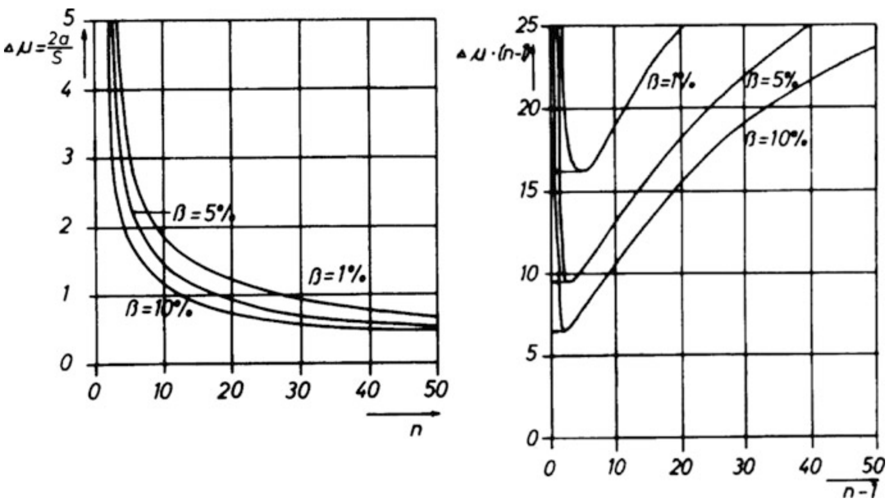


Fig. B.11 Length of the confidence interval for the mean against the number of observations

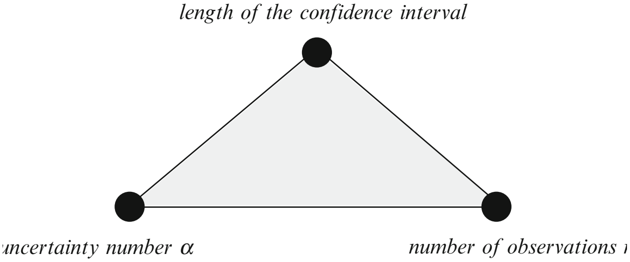


Fig. B.12 The Magic Triangle, constituents: (**a**) uncertainty number α (**b**), number of observations n, (**c**) length of the confidence interval $\Delta\mu$

where $k_{\mu\alpha}$ is called the *quantum number for the mean* μ which depends on the uncertainty number α. Table B.15.2 is a list of those quantum numbers. Let us interpret the uncertainty relation $\Delta\mu(n-1) \geq k_{\mu\alpha}$. The product $\Delta\mu(n-1)$ defines geometrically a *hyperbola* which we approximated out of the graph of Fig. B.12. Given the uncertainty number α, the product $\Delta\mu(n-1)$ has a smallest number, here denoted by $k_{\mu\alpha}$. *For instance*, choose $\alpha = 1\%$ such that $k_{\mu\alpha}/(n-1) \leq \Delta\mu$ or $16.4/(n-1) \leq \Delta\mu$. For n taken values 2, 11, 101, we get the *inequalities* $8.2 \leq \Delta\mu$, $1.64 \leq \Delta\mu$, $0.164 \leq \Delta\mu$.

Table B.15.2 (Coefficient of complementary confidence α, uncertainty number α, versus quantum number of the mean $k_{\mu\alpha}$ (Grafarend, 1970)):

α	$k_{\mu\alpha}$
10%	6.6
5%	9.6
1%	16.4
$\to 0$	$\to \infty$

B-6 Sampling from the Gauss–Laplace Normal Distribution: A Fourth Confidence Interval for the Variance

Theorem B.10 already supplied us with the sampling distribution of the sample variance, namely with the probability density function $f(x)$ of *Helmert's random variable* $x = (n-1)\widehat{\sigma}^2/\sigma^2$ of *Gauss-Laplace i.i.d.* observations n of sample variance $\widehat{\sigma}^2$, BIQUUE of σ^2. *B61* introduces accordingly the so far missing confidence interval for the "*true*" *variance* σ^2. Lemma B.15 contains the details, followed by Example B.16. Tables B.17–B.19 contain the properly chosen coefficients of complementary confidence and their quantiles

$$c_1(p;\alpha/2),\ \ c_2(p;1-\alpha/2),$$

dependent on the "*degrees of freedom*" $p = n - 1$. Table B.20 as a flow chart summarizes various steps in computing a confidence interval for the variance σ^2. *B62* reviews *The Uncertainty Principle* which is built on (a) the coefficient of complementary confidence α, also called uncertainty number, (b) the number of observations n and (c) the length $\Delta\sigma^2(n;\alpha)$ of the confidence interval for the "*true*" *variance* σ^2.

B-61 The Confidence Interval for the Variance

Lemma B.15 summaries the construction of a two-sided confidence interval for the "true" variance based upon *Helmert's Chi Square distribution* of the random variable $(n-1)\widehat{\sigma}^2/\sigma^2$ where $\widehat{\sigma}^2$ is BIQUUE of σ^2. Example B.18 introduces a random sample of size $n = 100$ with an empirical variance $\widehat{\sigma}^2 = 20.6$. Based upon coefficients of confidence and complementary confidence given in Table B.14, related confidence intervals are computed. The associated *quantiles* for *Helmert's Chi Square distribution* are tabulated in Table B.15 ($\gamma = 0.95$), Table B.18 ($\gamma = 0.99$) and Table B.19 ($\gamma = 0.998$). Finally, Table B.20 as a flow chart pares the way for the "*fast computation*" of the confidence interval for the "*true*" *variance* σ^2.

Lemma B.16. (confidence interval for the variance):

The random variable $x = (n-1)\widehat{\sigma}^2/\sigma^2$, also called *Helmert's* χ^2, characterized by the ration of the sample variance $\widehat{\sigma}^2$ BIQUUE of σ^2, and the *"true" variance* σ^2, has the χ^2_φ *pdf* of $p = n - 1$ degrees of freedom, if the random observations $y_i, i \in \{1, \ldots n\}$ are *Gauss-Laplace i.i.d.* The *"true" variance* σ^2 is an element of the two-sided confidence interval

$$\sigma^2 \in \left]\frac{(n-1)\widehat{\sigma}^2}{c_2(p;1-\alpha/2)}, \frac{(n-1)\widehat{\sigma}^2}{c_1(p;\alpha/2)}\right[$$

with confidence

$$\mathrm{P}\left\{\frac{(n-1)\widehat{\sigma}^2}{c_2(p;1-\alpha/2)} < \sigma^2 < \frac{(n-1)\widehat{\sigma}^2}{c_1(p;\alpha/2)}\right\} = 1-\alpha = \gamma$$

of level $1 - \alpha$. Tables D.8, D.18 and D.16 list the quantiles $c_1(p;1-\alpha/2)$ and $c_2(p;\alpha/2)$ associated to three values of Table B.19 of the coefficient of complementary confidence $1-\alpha$.
In order to make yourself more familiar with *Helmert's* Chi Square distribution we recommend to solve the problems of *Exercise B.1*.

Exercise B.1. (Helmert's Chi Square χ^2_φ distribution):

Helmert's random variable $x := (n-1)\widehat{\sigma}^2/\sigma^2 = p\widehat{\sigma}^2/\sigma^2$ has the non-symmetric χ^2_φ pdf. Prove that the first four central moments are,

(a) $\pi_1 = 0, \mathrm{E}\{x\} = \mu = p$
(b) $\pi_2 = \sigma_x^2 = 2p$
(c) $\pi_3 = (2p)^{3/2}(8/p)^{1/4}$ (coefficient of skewness $\pi_3^2/\pi_2^2 = \sqrt{8/p}$)
(d) $\pi_4 = 6\sigma^4(1+2p)$ (coefficient of curtosis $\pi_4/\pi_2^2 - 3 = 3 + 12/p$).

Guide

$$(a) \quad \left.\begin{array}{l} \mathrm{E}\{x\} = \dfrac{p}{\sigma^2}\mathrm{E}\{\widehat{\sigma}^2\} \\ \mathrm{E}\{\widehat{\sigma}^2\} = \sigma^2 (\text{unbiasedness}) \end{array}\right] \Rightarrow \mathrm{E}\{x\} = p$$

$$(b) \quad \left.\begin{array}{l} \mathrm{D}\{\widehat{\sigma}^2\} = \dfrac{2}{n-1}\sigma^4 = \dfrac{2}{p}\sigma^4 \\ \mathrm{D}\{x\} = \dfrac{p^2}{\sigma^4}\mathrm{D}\{\widehat{\sigma}^2\} \end{array}\right] \Rightarrow \mathrm{D}\{x\} = 2p.$$

Example B.18. (confidence interval for the sample variance $\widehat{\sigma}^2$):

Suppose that a random sample $(y_1, \dots y_n) \in \mathbb{Y}$ of size $n = 100$ has led to an empirical variance $\widehat{\sigma}^2 = 20.6$. $x = (n-1)\widehat{\sigma}^2/\sigma^2$ or $99 * 20.6/\sigma^2 = 2039.4/\sigma^2$ has Helmert's pdf with $p = n-1 = 99$ degrees of freedom. The probability $\gamma = 1-\alpha$ that x will be between $c_1(p;\alpha/2)$ and $c_2(p;1-\alpha/2)$ is

$$\mathrm{P}\{c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)\} = \int_{c_1}^{c_2} f(x)dx = \gamma = 1-\alpha$$

$$\mathrm{P}\{c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)\} = \int_0^{c_2} f(x)dx - \int_0^{c_1} f(x)dx = \gamma = 1-\alpha$$

$$\int_0^{c_2} f(x)dx = \mathrm{F}(c_2) = 1-\alpha/2 = (1+\gamma)/2,$$

$$\int_0^{c_1} f(x)dx = \mathrm{F}(c_1) = \alpha/2 = (1-\gamma)/2$$

$$\mathrm{P}\{c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)\} = \mathrm{F}(c_2) - \mathrm{F}(c_1) = (1+\gamma)/2 - (1-\gamma)/2 = \gamma$$

$$c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2) \Leftrightarrow c_1(p;\alpha/2) < (n-1)\frac{\widehat{\sigma}^2}{\sigma^2} < c_2(p;1-\alpha/2)$$

or

$$\boxed{\begin{aligned}&\frac{1}{c_2} < \frac{\sigma^2}{(n-1)\widehat{\sigma}^2} < \frac{1}{c_1} \Leftrightarrow \frac{(n-1)\widehat{\sigma}^2}{c_2} < \sigma^2 < \frac{(n-1)\widehat{\sigma}^2}{c_1}\\ &\mathrm{P}\left\{\frac{(n-1)\widehat{\sigma}^2}{c_2(p;1-\alpha/2)} < \sigma^2 < \frac{(n-1)\widehat{\sigma}^2}{c_1(p;\alpha/2)}\right\} = 1-\alpha = \gamma.\end{aligned}}$$

Since *Helmert's pdf* $f(x;p)$ is now-symmetric there arises the question how to distribute the confidence γ or the complementary confidence $\alpha = \gamma - 1$ on the confidence interval limits c_1 and c_2, respectively. If we setup $\mathrm{F}(c_1) = \alpha/2$, we define a cumulative probability half of the complementary confidence. If $\mathrm{F}(c_2) - \mathrm{F}(c_1) = \mathrm{P}\{c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)\} = 1-\alpha = \gamma$ is the *cumulative probability contained in the interval* $c_1 < x < c_2$, we derive $\mathrm{F}(c_2) = 1-\alpha/2$. Accordingly $c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)$ is the confidence interval based upon the quantile c_1 with cumulative probability $\alpha/2$ and the quantile c_2 with cumulative probability $1-\alpha/2$. The four representations of the cumulative probability of the confidence inter-val $c_1 < x < c_2$ establish *two linear Volterra integral equations of the first kind*

$$\int_0^{c_1} f(x)dx = \alpha/2 \quad \textit{and} \quad \int_0^{c_2} f(x)dx = 1-\alpha/2,$$

dependent on the degree of freedom $p = n-1$ of *Helmert's pdf* $f(x,p)$.
As soon as we have established the confidence interval $c_1(p;\alpha/2) < x < c_2(p;1-\alpha/2)$ for *Helmert's random variable* $x = (n-1)\widehat{\sigma}^2/\sigma^2 = p\widehat{\sigma}^2/\sigma^2$, we are left with the problem of how to generate a confidence interval for the *"true" variance* σ^2, the sample variance $\widehat{\sigma}^2$ given. If we take the reciprocal interval $c_2^{-1} < x^{-1} < c_1^{-1}$ for *Helmert's inverse random variable* $1/x = \sigma^2/[(n-1)\widehat{\sigma}^2]$, we are able to multiply both sides by $(n-1)\widehat{\sigma}^2$. In summary, a confidence interval which corresponds to $c_1 < x < c_2$ is $(n-1)\widehat{\sigma}^2/c_2 < \sigma^2 < (n-1)\widehat{\sigma}^2/c_1$.
Three values of the coefficient of confidence γ or of complementary confidence $\alpha = 1-\gamma$ which are most popular we list in Table B.16.

Table B.16 (Values of the coefficient of confidence γ,$c_1(p;\alpha/2)$ and $c_2(p;1-\alpha/2)$):

α	0.950	0.990	0.998
$\alpha/2$	0.050	0.010	0.002
$\alpha/2$	0.025	0.005	0.001
$1-\alpha/2$	0.975	0.995	0.999

In solving the linear Volterra integral equations of the first kind

$$\int_0^{c_1} f(x)dx = \mathrm{F}(c_1) = \alpha/2 \quad \textit{and} \quad \int_0^{c_2} f(x)dx = \mathrm{F}(c_2) = 1-\alpha/2,$$

which depend on the degrees of freedom $p = n - 1$, Tables B.17–B.19 collect the quantiles $c_1(p;\alpha/2)$ and $c_2(p;1-\alpha/2)$ for given values p and α.

Table B.17 (Quantiles $c_1(p;\alpha/2)$, $c_2(p;1-\alpha/2)$ for the confidence interval of the Helmert random variable with $p = n - 1$ degrees of freedom):

$$\alpha/2 = 0.025, \quad 1-\alpha/2 = 0.975, \quad \alpha = 0.05, \quad \gamma = 0.95$$

p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$	p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$
1	2	0.000	5.02	14	15	5.63	26.1
2	3	0.506	7.38	19	20	8.91	32.9
3	4	0.216	9.35	24	25	12.4	39.4
4	5	0.484	11.1	29	30	16.0	45.7
5	6	0.831	12.8	39	40	23.7	58.1
6	7	1.24	14.4	49	50	31.6	70.2
7	8	1.69	16.0	99	100	73.4	128
8	9	2.18	17.5				
9	10	2.70	19.0				

Table B.18 (Quantiles $c_1(p;\alpha/2)$, $c_2(p;1-\alpha/2)$ for the confidence interval of the Helmert random variable with $p = n - 1$ degrees of freedom):

$$\alpha/2 = 0.005, \quad 1-\alpha/2 = 0.995, \quad \alpha = 0.01, \quad \gamma = 0.99$$

p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$	p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$
1	2	0.000	7.88	8	9	1.34	22.0
2	3	0.010		9	10	1.73	23.6
3	4	0.072					
4	5	0.207		14	15	4.07	31.3
				19	20	6.84	38.6
5	6	0.412		24	25	9.89	45.6
6	7	0.676		29	30	13.1	52.3
7	8	0.989					
				39	40	20	65.5
				49	50	27.2	78.2
				99	100	66.5	139

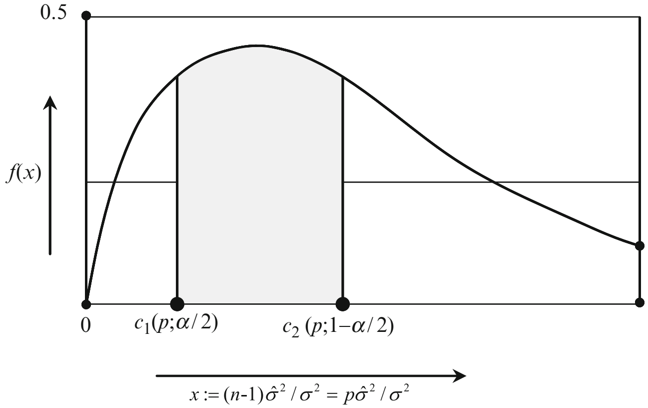


Fig. B.13 Two-sided confidence interval $\sigma^2 \in]p\sigma^2/c_2,\ p\sigma^2/c_1[$ *Helmert's pdf*, f(x)

Table B.19 (Quantiles $c_1(p;\alpha/2)$, $c_2(p;1-\alpha/2)$ for the confidence interval of the Helmert random variable with $p = n - 1$ degrees of freedom):

$$\alpha/2 = 0.001, \quad 1-\alpha/2 = 0.999, \quad \alpha = 0.002, \quad \gamma = 0.998$$

p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$	p	n	$c_1(p;\alpha/2)$	$c_2(p;1-\alpha/2)$
1	2	0.00	10.83	9	10	1.15	27.88
2	3	0.00	13.82	14	15	3.04	36.12
3	4	0.02	16.27	19	20	5.41	43.82
4	5	0.09	18.47	24	25	8.1	51.2
5	6	0.21	20.52	29	30	11.0	58.3
6	7	0.38	22.46	50	51	24.7	86.7
7	8	0.60	24.32	70	71	39.0	112.3
8	9	0.86	26.13	99	100	60.3	147.4
				100	101	61.9	149.4

Those data collected in Tables B.17–B.19 lead to the triplet of confidence intervals for the "true" variance

$$\boxed{\begin{array}{l} case\,(a): \gamma = 0.95, \quad \alpha = 0.05, \quad \alpha/2 = 0.025, \quad 1-\alpha/2 = 0.975 \\ p = 99, \quad n = 100, \quad c_1(p;\alpha/2) = 73.4, \quad c_2(p;1-\alpha/2) = 128 \end{array}}$$

$$\mathrm{P}\left\{\frac{99 * 20.6}{128} < \sigma^2 < \frac{99 * 20.6}{73.4}\right\} = 0.95$$

$$\mathrm{P}\{15.9 < \sigma^2 < 27.8\} = 0.95.$$

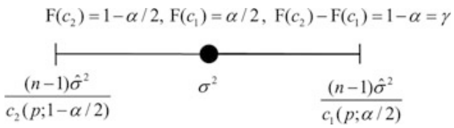


Fig. B.14 Two-sided confidence interval for the "true" variance σ^2, quantiles $c_1(p;\alpha/2)$ and $c_2(p;1-\alpha/2)$

$$\boxed{\begin{array}{l} case\,(b):\gamma = 0.99, \quad \alpha = 0.01, \quad \alpha/2 = 0.005, \quad 1-\alpha/2 = 0.995 \\ p = 99, \quad n = 100, \quad c_1(p;\alpha/2) = 66.5, \quad c_2(p;1-\alpha/2) = 139 \end{array}}$$

$$\mathrm{P}\left\{\frac{99 * 20.6}{139} < \sigma^2 < \frac{99 * 20.6}{66.5}\right\} = 0.99$$

$$\mathrm{P}\{14.7 < \sigma^2 < 30.7\} = 0.99.$$

$$\boxed{\begin{array}{l} case\,(c):\gamma = 0.998, \quad \alpha = 0.002, \quad \alpha/2 = 0.001, \quad 1-\alpha/2 = 0.999 \\ p = 99, \quad n = 100, \quad c_1(p;\alpha/2) = 60.3, \quad c_2(p;1-\alpha/2) = 147.3 \end{array}}$$

$$\mathrm{P}\left\{\frac{99 * 20.6}{147.3} < \sigma^2 < \frac{99 * 20.6}{60.3}\right\} = 0.996$$

$$\mathrm{P}\{13.8 < \sigma^2 < 33.8\} = 0.998.$$

The results can be summarized as follows. With probability 95%, the "true" variance σ^2 is an element of the interval [15.9, 27.8]. In contrast, with probability 99%, the "true" variance is an element of the larger interval [14.7, 30.7]. Finally, with probability 99.8% the "true" variance is an element of the largest interval [13.8, 33.8]. If we compare the confidence intervals for the variance σ^2, we realize much larger intervals for smaller complementary confidence namely 5%, 1% and 0.2%. Such a result is subject of The Uncertainty Principle.

Table B.20 (Flow chart – confidence interval for the variance σ^2):

Choose a coefficient of confidence γ according to *Table B16*.

Solve thelinear Volterra integral equation of the first kind
$F(c_1(p;\alpha/2)) - \alpha/2,\ F(c_2(p;1-\alpha/2)) - 1 - \alpha/2$
by *Table B17*, *Table B18* or *Table B19* for a *Helmert pdf* with
$p = n-1$ degrees of freedom.

Compute the sample variance $\hat{\sigma}^2$and the quantiles

$$\sigma^2 \in \left] \frac{(n-1)\hat{\sigma}^2}{c_2(p;\alpha/2)} \quad \frac{(n-1)\hat{\sigma}^2}{c_1(p;\alpha/2)} \right[.$$

B-62 The Uncertainty Principle

Figure B.15 is the graph of the function

$$\boxed{\Delta\sigma^2(n;\alpha) := (n-1)\widehat{\sigma}^2\left(\frac{1}{c_1(n-1;\alpha/2)} - \frac{1}{c_2(n-1;1-\alpha/2)}\right),}$$

where $\Delta\sigma^2$ is the length of the confidence interval of the "true" variance σ^2. The independent variable of the functions is the number of observations n. The function $\Delta\sigma^2(n;\alpha)$ is plotted for fixed values of the coefficient of complementary confidence α, namely $\alpha = 5\%, 1\%, 0.2\%$. For reasons given later on the coefficient of complementary confidence α is called *uncertainty number*. The graph of the function $\Delta\sigma^2(n;\alpha)$ illustrates two important facts.

- Fact #1: For a contrast uncertainty number α, the length of the confidence interval $\Delta\sigma^2$ is smaller, the larger number of observations n is chosen.
- Fact #2: For a contrast number of observations n, the smaller the number of uncertainty α is chosen, the larger is the confidence interval $\Delta\sigma^2$.

Evidently, the divisive influences of (a) the length of the confidence interval $\Delta\sigma^2$, (b) the uncertainty number α and (c) the number of observations n, which we collect in *The Magic Triangle of Fig. B.16*, constitute The *Uncertainty Principle*, formulated by the inequality

$$\Delta\sigma^2(n-1) \geq k_{\sigma^2\alpha},$$

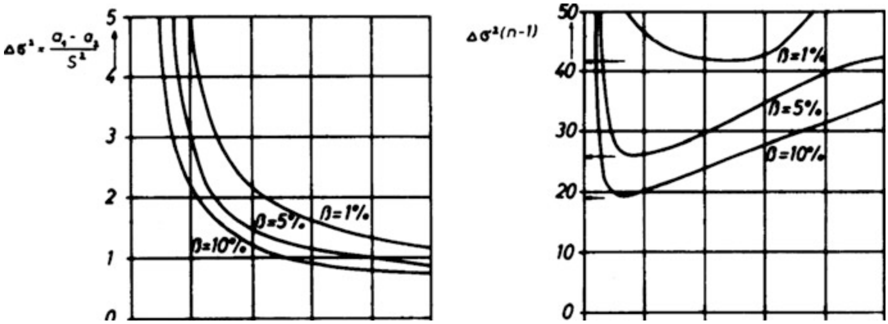


Fig. B.15 Length of the confidence interval for the variance against the number of observations

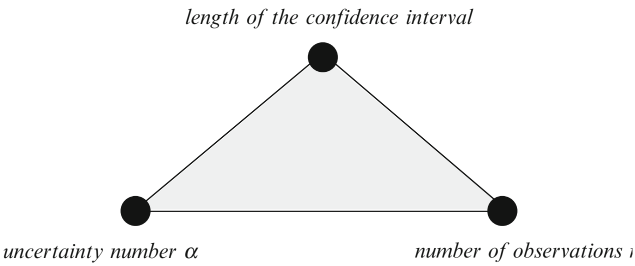


Fig. B.16 The magic triangle, constituents: (**a**) uncertainty number α (**b**), number of observations n, (**c**) length of the confidence interval $\Delta\sigma^2$

where $k_{\sigma^2\alpha}$ is called quantum number for the variance σ^2. The quantum number depends on the uncertainty number α. Let us interpret the uncertainty relation $\Delta\sigma^2(n-1) \geq k_{\sigma^2\alpha}$. The product $\Delta\sigma^2(n-1)$ defines geometrically a *hyperbola* which we approximated to the graph of Fig. B.15. Given the uncertainty number α, the product $\Delta\sigma^2(n-1)$ has a smallest number denoted by $k_{\sigma^2\alpha}$. For *instance*, choose $\alpha = 1\%$ such that $k_{\sigma^2\alpha}/(n-1) \leq \Delta\sigma^2$ or $42/(n-1) \leq \Delta\sigma^2$. For the number of observations n, for instance $n = 2, 11, 101$, we find the *inequalities* $42 \leq \Delta\sigma^2$, $4.2 \leq \Delta\sigma^2$, $0.42 \leq \Delta\sigma^2$.

Table B.21 (Coefficient of complementary confidence α, uncertainty number α, *versus* quantum number of the mean $k_{\sigma^2\alpha}$ (Grafarend, 1970):

α	$k_{\sigma^2\alpha}$
10%	19.5
5%	25.9
1%	42.0
$\to 0$	$\to \infty$

At the end, pay attention to the quantum numbers $k_{\sigma^2\alpha}$ listed in Table B.21.

B-7 Sampling from the Multidimensional Gauss–Laplace Normal Distribution: The Confidence Region for the Fixed Parameters in the Linear Gauss–Markov Model

Example B.19. (special linear Gauss–Markov model, marginal probability distributions):

For a simple linear *Gauss–Markov model* $E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi}$, $\mathbf{A} \in \mathbb{R}^{n\times m}$, rk$\mathbf{A} = m$, namely $n = 3$, $m = 2$, $D\{\mathbf{y}\} = \mathbf{I}\sigma^2$ of *Gauss-Laplace i.i.d. observations*, we are going to compute

- The sample pdf of $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$
- The sample pdf of $\widehat{\sigma}^2$ BIQUUE of σ^2.

We follow the action within seven items. *First*, we identify the pdf of *Gauss-Laplace i.i.d. observations* $\mathbf{y} \in \mathbb{R}^n$. *Second*, we review the estimations of $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$ as well as $\widehat{\sigma}^2$ BIQUUE of σ^2. *Third*, we decompose the Euclidean norm of the observation space $\|\mathbf{y} - E\{\mathbf{y}\}\|^2$ into the Euclidean norms $\|\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\xi}}\|^2$ and $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2$. *Fourth*, we present the eigenspace $\mathbf{A}'\mathbf{A}$ analysis and the eigenspace synthesis of the associated matrices $\mathbf{M} := \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{N} := \mathbf{A}'\mathbf{A}$ within the Eulidean norms $\|\mathbf{y} - \mathbf{A}\boldsymbol{\xi}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}$ and $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}} = (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$, respectively. The eigenspace representation leads us to canonical random variables $(z_1, \dots, z_{n-m})$ relating to the norm $\|\mathbf{y} - \mathbf{A}\boldsymbol{\xi}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}$ and $(z_{n-m+1}, \dots, z_n)$ relating to the norm $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{N}}$ which are *standard Gauss-Laplace normal*. *Fifth*, we derive the cumulative probability of *Helmert's random variable* $x := (n - \text{rk}\mathbf{A})\widehat{\sigma}^2/\sigma^2 = (n - m)\widehat{\sigma}^2/\sigma^2$ and of the unknown parameter vector $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$ or its canonical counterpart $\widehat{\boldsymbol{\eta}}$ BLUUE of $\boldsymbol{\eta}$, *multivariate Gauss-Laplace normal*. *Action six* generates the marginal pdf of $\widehat{\boldsymbol{\xi}}$ or $\widehat{\boldsymbol{\eta}}$, both BLUUE of $\boldsymbol{\xi}$ or $\boldsymbol{\eta}$, respectively. *Finally, action seven* leads us to *Helmert's Chi Square* distribution χ^2_p with $p = n - \text{rk}\mathbf{A} = n - m$ (here: $n - m = 1$) "*degrees of freedom*".

The first action item

Let us assume an experiment of *three Gauss-Laplace i.i.d. observations* $[y_1, y_2, y_3]' = \mathbf{y}$ which constitute the coordinates of the observation space Y, dim The observationsy_i, $i \in \{1, 2, 3\}$ are related to a parameter space Ξ with coordinates $[\xi_1, \xi_2] = \boldsymbol{\xi}$ in the sense to generate a straight line $y\{k\} = \xi_1 + \xi_2 k$. The *fixed effects* ξ_j, $j \in \{1, 2\}$, define geometrically a straight line, statistically a *special linear Gauss–Markov model of one variance component*, namely *the first moment*

$$E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi} \sim E\left\{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right\} = \begin{bmatrix} 1 & k_1 \\ 1 & k_2 \\ 1 & k_3 \end{bmatrix}\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \quad \mathrm{rk}\mathbf{A} = 2.$$

The central second moment

$$D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2 \sim D\{\mathbf{y}\} = D\left\{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}\right\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\sigma^2, \quad \sigma^2 < 0.$$

k represents the *abscissa* as a fixed random, $\mathbf{y}$ the *ordinate* as the observation, naturally a random effect. *Samples* of the straight line are taken at $k_1 = 0$, $k_2 = 1$, $k_3 = 2$, this calling for $y(k_1) = y(0) = y_1$, $y(k_2) = y(1) = y_2$, $y(k_3) = y(2) = y_3$, respectively. $E\{\mathbf{y}\}$ is a consistent equation. Alternatively, we may say $E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A})$. The matrix $\mathbf{A} \in \mathbb{R}^{3\times 2}$ is rank deficient by $p = n - \mathrm{rk}\mathbf{A} = 1$, also called "*degree of freedom*". The dispersion matrix $D\{\mathbf{y}\}$, the central moment of second order, is represented as a linear model, too, namely by one-variance component σ^2. The joint probability function of the *three Gauss-Laplace i.i.d. observations*

$$dF = f(y_1, y_2, y_3)dy_1dy_2dy_3 = f(y_1)f(y_2)f(y_3)dy_1dy_2dy_3$$

$$f(y_1, y_2, y_3) = (2\pi)^{-3/2}\sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\})\right)$$

will be transformed by means of the *special linear Gauss–Markov model with one-variance component*.

The second action item

For such a transformation, we need $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$ and $\widehat{\sigma}^2$ BIQUUE σ^2.

$$\widehat{\boldsymbol{\xi}}\ \mathit{BLUUE\ of}\ \boldsymbol{\xi}: \left[\begin{array}{l} \widehat{\boldsymbol{\xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}, \\ \mathbf{A}'\mathbf{A} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}, \quad (\mathbf{A}'\mathbf{A})^{-1} = \dfrac{1}{6}\begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix}, \\ \widehat{\boldsymbol{\xi}} = \dfrac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \end{array}\right.$$

$$\widehat{\sigma}^2\ \mathit{BIQUUE\ of}\ \sigma^2: \left[\begin{array}{l} \widehat{\sigma}^2 = \dfrac{1}{n - \mathrm{rk}\mathbf{A}}(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \\ \quad = \dfrac{1}{n - \mathrm{rk}\mathbf{A}}\mathbf{y}'(\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y}, \\ \widehat{\sigma}^2 = \dfrac{1}{6}(y_1^2 - 4y_1y_2 + 2y_1y_3 + 4y_2^2 - 4y_2y_3 + y_3^2). \end{array}\right.$$

The third action item

The quadratic form $(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\})$ allows the *fundamental decomposition*

$$\mathbf{y} - E\{\mathbf{y}\} = \mathbf{y} - \mathbf{A}\boldsymbol{\xi} = \mathbf{y} - \mathbf{A}\boldsymbol{\xi} + \mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$$
$$(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = (\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})$$
$$= (\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$$

$$\boxed{(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = (n - \mathrm{rk}\mathbf{A})\widehat{\boldsymbol{\sigma}}^2 + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})}$$

$$(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \widehat{\boldsymbol{\sigma}}^2 + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}).$$

The fourth action item

In order to bring the quadratic form $(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = (n - \mathrm{rk}\mathbf{A})\widehat{\sigma}^2 + +(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$ into a *canonical form*, we introduce the generalised *forward and backward Helmert transformation*

$$\boxed{\begin{gathered} \mathbf{H}\mathbf{H}' = \mathbf{I}_n \\ \mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - E\{\mathbf{y}\}) = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \quad \textit{and} \\ \mathbf{y} - E\{\mathbf{y}\} = \sigma\mathbf{H}'\mathbf{z} \end{gathered}}$$

$$(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \sigma^2\mathbf{z}'\mathbf{H}\mathbf{H}'\mathbf{z} = \sigma^2\mathbf{z}'\mathbf{z}$$

$$\boxed{\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \mathbf{z}'\mathbf{z} = z_1^2 + z_2^2 + z_3^2.}$$

How to relate the sample variance $\widehat{\sigma}^2$ and the sample quadratic form $(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$ to the canonical quadratic form $\mathbf{z}'\mathbf{z}$?

Previously, for the example of direct observations in the *special linear Gauss–Markov model* $E\{\mathbf{y}\} = \mathbf{1}\mu$, $D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2$ we succeeded to relate $z_1^2 + \cdots + z_{n-1}^2$ to $\widehat{\sigma}^2$ and z_n^2 to $(\widehat{\mu} - \mu)^2$. Here the sample variance $\widehat{\sigma}^2$, BIQUUE σ^2, *as well as* the quadratic form of the deviate of the sample parameter vector $\widehat{\boldsymbol{\xi}}$ from the "true" parameter vector $\boldsymbol{\xi}$ have been represented by

$$\widehat{\sigma}^2 = \frac{1}{n - \mathrm{rk}\mathbf{A}}(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) = \frac{1}{n - \mathrm{rk}\mathbf{A}}\mathbf{y}'[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{y}$$
$$\mathrm{rk}[\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'] = n - \mathrm{rk}\mathbf{A} = n - m = 1 \quad \textit{versus}$$
$$(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}), \quad \mathrm{rk}(\mathbf{A}'\mathbf{A}) = \mathrm{rk}\mathbf{A} = m.$$

The eigenspace of the matrices **M** and **N**, namely

$$\mathbf{M} := \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A} \quad \textit{and} \quad \mathbf{N} := \mathbf{A}'\mathbf{A}$$

$$= \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \qquad = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix},$$

will be analyzed.

The eigenspace analysis of the	The eigenspace analysis of the
matrix $\mathbf{M}$ $j \in \{1, \ldots, n\}$	*matrices* $\mathbf{N}, \mathbf{N}^{-}1$ $i \in \{1, \ldots, m\}$
$\mathbf{V}'\mathbf{M}\mathbf{V}$ $= \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \mathbf{M}\,[\mathbf{V}_1, \mathbf{V}_2]$	$\mathbf{U}'\mathbf{N}\mathbf{U} = Diag(\gamma_1, \ldots, \gamma_m) = \boldsymbol{\Lambda}_{\mathbf{N}}$ $\mathbf{U}'\mathbf{N}^{-1}\mathbf{U} = Diag(\lambda_1, \ldots, \lambda_m) = \boldsymbol{\Lambda}_{\mathbf{N}^{-1}}$
$= Diag(\mu_1, \ldots, \mu_{n-m}, 0, \ldots, 0) = \boldsymbol{\Lambda}_{\mathbf{M}}$	$\gamma_1 = \frac{1}{\lambda_1}, \ldots, \gamma_m = \frac{1}{\lambda_m}$ $\lambda_1 = \frac{1}{\gamma_1}, \ldots, \lambda_m = \frac{1}{\gamma_m}$
$\mathrm{rk}\mathbf{M} = n - m$	$\mathrm{rk}\mathbf{N} = \mathrm{rk}\mathbf{N}^{-1} = m$

Orthonormality of the eigencolumns	Orthonormality of the eigencolumns
$\mathbf{V}'_1\mathbf{V}_1 = \mathbf{I}_{n-m}, \quad \mathbf{V}'_2\mathbf{V}_2 = \mathbf{I}_m$ $\mathbf{V}'_1\mathbf{V}_2 = 0 \in \mathbb{R}^{(n-m)\times m}$ $\begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n-m} & 0 \\ 0 & \mathbf{I}_m \end{bmatrix}$	$\mathbf{U}'\mathbf{U} = \mathbf{I}_m$

$$\boxed{\begin{aligned} v_1'v_1 &= 1 \\ v_1'v_2 &= 0 \\ &\ldots \\ v_{n-1}'v_n &= 0 \\ v_n'v_n &= 1 \end{aligned}} \qquad \boxed{\begin{aligned} u_1'u_1 &= 1 \\ u_1'u_2 &= 0 \\ &\ldots \\ u_{m-1}'u_m &= 0 \\ u_m'u_m &= 1 \end{aligned}}$$

$$\mathbf{V} \in SO(n) : eigencolumns : \qquad \mathbf{U} \in SO(m) : eigencolumns :$$

$$(\mathbf{M} - \mu_j \mathbf{I}_n)v_j = 0 \qquad (\mathbf{N} - \gamma_j \mathbf{I}_n)u_j = 0$$

$$eigenvalues \qquad eigenvalues$$

$$|\mathbf{M} - \mu_j \mathbf{I}_n| = 0 \qquad |\mathbf{N} - \gamma_i \mathbf{I}_m| = 0$$

in particular

$$\textit{eigenspace analysis of the matrix } \mathbf{M},$$
$$rkM = n - m,\ \mathbf{M} \in \mathbb{R}^{3\times 3},\ \mathbf{A} \in \mathbb{R}^{3\times 2},\ rk\mathbf{M} = 1$$
$$\boldsymbol{\Lambda}_{\mathbf{M}} = Diag(1, 0, 0)$$
$$\mathbf{V} = [\mathbf{V}_1,\ \mathbf{V}_2] = \begin{bmatrix} 0.4082 & 0.7024 & 0.5830 \\ -0.8165 & 0.5667 & -0.1109 \\ 0.4082 & 0.4307 & -0.8049 \end{bmatrix}$$
$$\mathbf{V}_1 \in \mathbb{R}^{3\times 1}\ \mathbf{V}_2 \in \mathbb{R}^{3\times 2}$$

$$\textit{eigenspace analysis of the matrix } \mathbf{N},$$
$$rkN = m,\ \mathbf{N} \in \mathbb{R}^{2\times 2},\ \mathrm{rk}\mathbf{N} = 2$$
$$\boldsymbol{\Lambda}_{\mathbf{N}} = Diag(0.8377, 7.1623)$$
$$\mathbf{U} = \begin{bmatrix} 0.8112 & 0.5847 \\ -0.5847 & 0.8112 \end{bmatrix}$$
$$\mathbf{U} \in \mathbb{R}^{2\times 2}$$

to be completed by

$$\textit{eigenspace synthesis of the matrix } \mathbf{M}$$
$$\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{V}'$$
$$\mathbf{M} = [\mathbf{V}_1, \mathbf{V}_2]\boldsymbol{\Lambda}_{\mathbf{M}} \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix}$$
$$= [\mathbf{V}_1, \mathbf{V}_2]\ Diag(\mu_1, \ldots, \mu_{n-m}, 0, \ldots, 0) \begin{bmatrix} \mathbf{V}'_1 \\ \mathbf{V}'_2 \end{bmatrix}$$
$$\mathbf{M} = \mathbf{V}_1\ Diag(\mu_1, \ldots, \mu_{n-m})\mathbf{V}'_1$$

$$\textit{eigenspace synthesis of the matrix } \mathbf{N}, \mathbf{N} - 1$$
$$\mathbf{N} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{N}}\mathbf{U}', \quad \mathbf{N}^{-1} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{N}^{-1}}\mathbf{U}'$$
$$\mathbf{N} = \mathbf{U}\ Diag(\gamma_1, \ldots, \gamma_m)\mathbf{U}'$$
$$\mathbf{N}^{-}\mathbf{1} = \mathbf{U}\ Diag(\lambda_1, \ldots, \lambda_m)\mathbf{U}'$$

in particular

$$\mathbf{M} = \begin{bmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{bmatrix} \mu_1 \begin{bmatrix} 0.4082 & -0.8165 & 0.4082 \end{bmatrix} \quad \mathbf{N} = \begin{bmatrix} 0.8112 & 0.5847 \\ -0.5847 & 0.8112 \end{bmatrix} \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \begin{bmatrix} 0.8112 & -0.5847 \\ 0.5847 & 0.8112 \end{bmatrix}$$

$$\mu_1 = 1 \quad \textit{versus} \quad \gamma_1 = 0.8377, \quad \gamma_2 = 7.1623$$

$$\mathbf{N}^{-1} = \begin{bmatrix} 0.8112 & 0.5847 \\ -0.5847 & 0.8112 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 0.8112 & -0.5847 \\ 0.5847 & 0.8112 \end{bmatrix}$$
$$\mu_1 = 1 \quad \textit{versus} \quad \lambda_1 = \gamma_1^{-1} = 1.1937, \quad \lambda_2 = \gamma_2^{-1} = 0.1396.$$

The non-vanishing eigenvalues of the matrix **M** have been denoted by $(\mu_1, \ldots, \mu_{n-m})$, m eigenvalues are zero such that *eigen*, $0, \ldots, 0)$. The eigenvalues of the regular matrix **N** span eigen $(\mathbf{N}) = (\gamma_1, \ldots, \gamma_m)$. Since the dispersion matrix $D\{\widehat{\boldsymbol{\xi}}\} = (\mathbf{A}'\mathbf{A})^{-1}\sigma^2 = \mathbf{N}^{-1}\sigma^2$ is generated by the inverse of the matrix $\mathbf{A}'\mathbf{A} = \mathbf{N}$, we have computed, in addition, the eigenvalues of the matrix $\mathbf{N}^{-1}$ by means of $eigen(\mathbf{N}^{-1}) = (\lambda_1, \ldots, \lambda_m)$. The eigenvalues of N and $\mathbf{N}^{-1}$, respectively, are related by

$$\gamma_1 = \lambda_1^{-1}, \ldots, \gamma_m = \lambda_m^{-1} \quad or \quad \lambda_1 = \gamma_1^{-1}, \ldots, \lambda_m = \gamma_m^{-1}.$$

In the example, the matrix **M** had only one non-vanishing eigenvalue $\mu_1 = 1$. In contrast, the regular matrix **N** was characterized by two eigenvalues $\gamma_1 = 0.8377$ and $\gamma_2 = 7.1623$, its inverse matrix $\mathbf{N}^{-1}$ by $\lambda_1 = 1.1937$ and $\lambda_2 = 0.1396$.
The two quadratic forms, namely

$$\mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{V}\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{V}'\mathbf{y} \quad \textit{versus} \quad (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}),$$

build up the original quadratic form

$$\begin{aligned}\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) &= \frac{1}{\sigma^2}\mathbf{y}'\mathbf{M}\mathbf{y} + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ &= \frac{1}{\sigma^2}\mathbf{y}'\mathbf{V}\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{V}'\mathbf{y} + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\boldsymbol{\Lambda}_{\mathbf{N}}\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\end{aligned}$$

in terms of the *canonical random variables*

$$\mathbf{V}'\mathbf{y} = \mathbf{y}^* \Leftrightarrow \mathbf{y} = \mathbf{V}\mathbf{y}^* \quad \textit{and} \quad \mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}$$

such that

$$\begin{aligned}\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) &= \frac{1}{\sigma^2}(\mathbf{y}^*)'\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{y}^* + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\boldsymbol{\Lambda}_{\mathbf{N}}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \\ \frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) &= \frac{1}{\sigma^2}\sum_{j=1}^{n-m}(\mathbf{y}_j^*)^2\mu_j + \frac{1}{\sigma^2}\sum_{i=1}^{m}(\widehat{\boldsymbol{\eta}}_i - \boldsymbol{\eta}_i)^2\gamma_i \\ \frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) &= z_1^2 + \cdots + z_{n-m}^2 + z_{n-m+1}^2 + \cdots + z_n^2.\end{aligned}$$

The quadratic form $\mathbf{z}'\mathbf{z}$ splits up into two terms, namely

$$\boxed{\begin{array}{lcl} z_1^2 + \cdots + z_{n-m}^2 & & z_{n-m+1}^2 + \cdots + z_n^2 \\ \quad = \dfrac{1}{\sigma^2}\displaystyle\sum_{j=1}^{n-m}(y_j^*)^2\mu_j & \textit{and} & \quad = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{m}(\widehat{\eta}_i - \eta_i)^2\gamma_j \end{array}}$$

here

$$\begin{aligned} z_1^2 &= \frac{1}{\sigma^2}y_{1*}^2 = \frac{\widehat{\sigma}^2}{\sigma^2} \quad \textit{and} \\ z_2^2 + z_3^2 &= \frac{1}{\sigma^2}[(\widehat{\eta}_1 - \eta_1)^2\gamma_1 + (\widehat{\eta}_2 - \eta_2)^2\gamma_2], \quad \textit{or} \\ z_1^2 &= \frac{1}{6\sigma^2}(y_1^2 - 4y_1y_2 + 2y_1y_3 + 4y_2^2 - 4y_2y_3 + y_3^2) \quad \textit{and} \end{aligned}$$

$$z_2^2 + z_3^2 = \frac{1}{\sigma^2}[0.8377(\widehat{\eta}_1 - \eta_1)^2 + 7.1623(\widehat{\eta}_2 - \eta_2)^2] = \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}).$$

The fifth action item

We are left with the problem to transform the *cumulative probability* $dF = f(y_1, y_2, y_3)dy_1dy_2dy_3$ into the canonical form $dF = f(z_1, z_2, z_3)dz_1dz_2dz_3$. Here we take advantage of Corollary B.3. *First*, we introduce *Helmert's random variable* $x := z_1^2$ and the random variables $\widehat{\xi}_1$ and $\widehat{\xi}_2$ of the unknown parameter vector $\widehat{\boldsymbol{\xi}}$ of fixed effects $(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = z_2^2 + z_3^2 = \|\mathbf{z}\|^2_{\mathbf{A}'\mathbf{A}}$ if we denote $\mathbf{z} := [z_2, z_3]'$.

$$dz_1dz_2 = \sqrt{\det \mathbf{A}'\mathbf{A}}\, d\widehat{\xi}_1 d\widehat{\xi}_2 = \sqrt{6}\, d\widehat{\xi}_1 d\widehat{\xi}_2,$$

according to Corollary B.3 is special representation of the surface element by means of the matrix of the metric $\mathbf{A}'\mathbf{A}$. In summary, the volume element

$$dz_1dz_2dz_3 = \frac{dx}{\sqrt{x}}\sqrt{\det \mathbf{A}'\mathbf{A}}\, d\widehat{\xi}_1 d\widehat{\xi}_2$$

leads to the *first* canonical representation of the *cumulative probability*

$$\boxed{dF = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x\right)\frac{dx}{\sqrt{x}}\frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right] d\widehat{\xi}_1 d\widehat{\xi}_2.}$$

The left *pdf* establishes *Helmert's pdf* of $x = z_1^2 = \widehat{\sigma}^2/\sigma^2$, $dx = \sigma^{-2}d\widehat{\sigma}^2$. In contrast, the right *pdf* characterizes the *bivariate Gauss-Laplace pdf* of $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2$. Unfortunately, the *bivariate Gauss-Laplace* $\mathbf{A}'\mathbf{A}$ *normal pdf* is not given in the canonical form. Therefore, *second* we do correlate $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}}$ by means of *eigenspace synthesis*.

$$\mathbf{A}'\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{A}'\mathbf{A}}\mathbf{U}' = \mathbf{U}\, Diag(\gamma_1, \gamma_2)\mathbf{U}' = \mathbf{U}\, Diag\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right)\mathbf{U}'$$

$$\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}} := (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\, Diag\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right)\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$$

$$|\mathbf{A}'\mathbf{A}|^{1/2} = \sqrt{\gamma_1\gamma_2} = \frac{1}{\sqrt{\lambda_1\lambda_2}}$$

$$\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta} := \mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \Leftrightarrow \widehat{\boldsymbol{\xi}} - \boldsymbol{\xi} = \mathbf{U}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})$$

$$\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}} = (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})'\, Diag\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right)(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \|\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}\|^2_D$$

$$\boxed{\begin{aligned}\|\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}} &= (\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\begin{bmatrix}3 & 3\\ 3 & 5\end{bmatrix}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\\ &= (\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})'\begin{bmatrix}\frac{1}{1.1937} & 0\\ 0 & \frac{1}{0.1396}\end{bmatrix}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}) = \|\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}\|^2_{\mathbf{D}}.\end{aligned}}$$

By means of the canonical variables $\widehat{\boldsymbol{\eta}} = \mathbf{U}'\widehat{\boldsymbol{\xi}}$ we derive the cumulative probability

$$\begin{aligned}dF &= \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{\sqrt{x}}\exp\left(-\frac{1}{2}x\right)dx\frac{1}{2\pi\sigma^2\sqrt{\lambda_1\lambda_2}}\\ &\quad\times\exp\left[-\frac{1}{2\sigma^2}\left(\frac{(\widehat{\eta}_1-\eta_1)^2}{\lambda_1}+\frac{(\widehat{\eta}_2-\eta_2)^2}{\lambda_2}\right)\right]d\widehat{\eta}_1 d\widehat{\eta}_2 \quad \text{or}\\ dF &= f(x)f(\widehat{\eta}_1)f(\widehat{\eta}_2)dx\,d\widehat{\eta}_1 d\widehat{\eta}_2.\end{aligned}$$

Third, we prepare ourselves for the *cumulative probability* $dF = f(z_1, z_2, z_3)dz_1 dz_2dz_3$. We depart from the representation of the volume element

$$dz_1dz_2dz_3 = \frac{dx}{2\sqrt{x}}\sqrt{\det\mathbf{A}'\mathbf{A}}\,d\widehat{\xi}_1 d\widehat{\xi}_2$$

subject to

$$\boxed{\begin{aligned}&x = z_1^2 = \frac{1}{\sigma^2}\widehat{\sigma}^2 = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{M}\mathbf{y}\\ &\widehat{\boldsymbol{\xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}.\end{aligned}}$$

The *Helmert random variable is a quadratic form* of the coordinates (y_1, y_2, y_3) of the observation vector y. In contrast, $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$ is a linear form of the coordinates (y_1, y_2, y_3) of observation vector $\mathbf{y}$. The transformation of the volume element

$$dx\,d\widehat{\xi}_1 d\widehat{\xi}_2 = |\mathbf{J}_x|dy_1dy_2dy_3$$

is based upon the Jacobi matrix $\mathbf{J}_x(y_1, y_2, y_3)$

$$\mathbf{J}_x = \begin{bmatrix}D_1x & D_2x & D_3x\\ D_1\widehat{\xi}_1 & D_2\widehat{\xi}_1 & D_3\widehat{\xi}_1\\ D_1\widehat{\xi}_2 & D_2\widehat{\xi}_2 & D_3\widehat{\xi}_2\end{bmatrix} = \begin{bmatrix}a'\\ (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\end{bmatrix}\begin{array}{l}1\times 3\\ 2\times 3\end{array}$$

$$\mathbf{a} = \begin{bmatrix} D_1 x \\ D_2 x \\ D_3 x \end{bmatrix} = \frac{2}{\sigma^2}\mathbf{My} = \frac{1}{3\sigma^2}\begin{bmatrix} y_1 - 2y_2 + y_3 \\ -2y_1 + 4y_2 - 2y_3 \\ y_1 - 2y_2 + y_3 \end{bmatrix}$$

$$\mathbf{x} = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{My} = \frac{1}{6\sigma^2}(y_1^2 - 4y_1y_2 + 2y_1y_3 + 4y_2^2 - 4y_2y_3 + y_3^2)$$

$$(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}$$

$$\mathbf{J}_x = \frac{1}{6\sigma^2}\begin{bmatrix} 2y_1 - 4y_2 + 2y_3 & -4y_1 + 8y_2 - 4y_3 & 2y_1 - 4y_2 + 2y_3 \\ 5 & 2 & -1 \\ -3 & 0 & 3 \end{bmatrix}$$

$$\det \mathbf{J}_x = \sqrt{\det(\mathbf{A}'\mathbf{A})^{-1} \det[\mathbf{a}'\mathbf{a} - \mathbf{a}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{a}]}$$

$$\det \mathbf{J}_x = \frac{\sqrt{\det[\mathbf{a}'\mathbf{a} - \mathbf{a}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{a}]}}{\sqrt{\det(\mathbf{A}'\mathbf{A})}}$$

$$\mathbf{a}'\mathbf{a} = \frac{4}{\sigma^4}\mathbf{y}'\mathbf{M}'\mathbf{My} = \frac{4}{\sigma^4}\mathbf{y}'\mathbf{My}$$

$$\mathbf{a}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{a} = \frac{4}{\sigma^4}\mathbf{y}'\mathbf{M}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{My}$$

$$= \frac{4}{\sigma^4}\mathbf{y}'[\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'[\mathbf{I}_3 - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}']\mathbf{y} = 0$$

$$\det \mathbf{J}_x = \frac{2\sqrt{\mathbf{y}'\mathbf{My}}}{\sigma^2\sqrt{\det(\mathbf{A}'\mathbf{A})}}$$

$$|\det \mathbf{J}_y| = |\det \mathbf{J}_x|^{-1} = \frac{\sigma^2}{2}\frac{\sqrt{\det(\mathbf{A}'\mathbf{A})}}{\sqrt{\mathbf{y}'\mathbf{My}}}.$$

The various representations of the Jacobian will finally lead us to the special form of the volume element

$$\boxed{\frac{1}{2}dx\, d\widehat{\xi}_1\, d\widehat{\xi}_2 = \frac{1}{\sigma^2}\frac{\sqrt{\mathbf{y}'\mathbf{My}}}{|\mathbf{A}'\mathbf{A}|^{1/2}}dy_1dy_2dy_3}$$

and the *cumulative probability*

$$\boxed{\begin{aligned} dF &= \frac{1}{\sigma^3(2\pi)^{3/2}} \exp\left[-\frac{1}{2}(z_1^2+z_2^2+z_3^2)\right] dz_1dz_2dz_3 \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}x\right) \frac{dx}{\sqrt{x}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right] d\widehat{\xi}_1 d\widehat{\xi}_2 \\ &= \frac{1}{\sigma^3(2\pi)^{3/2}} \exp\left[-\frac{1}{2}(\mathbf{y}-E\{\mathbf{y}\})'(\mathbf{y}-E\{\mathbf{y}\})\right] dy_1dy_2dy_3. \end{aligned}}$$

The sixth action item

The first target is to generate the marginal pdf of the unknown parameter vector $\widehat{\boldsymbol{\xi}}$, BLUUE of $\boldsymbol{\xi}$.

$$dF_1 = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x\right) \frac{dx}{\sqrt{x}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) d\widehat{\xi}_1 d\widehat{\xi}_2$$

$$dF_1 = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x\right) \frac{dx}{\sqrt{x}} \frac{1}{2\pi\sigma^2\sqrt{\lambda_1\lambda_2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{2}\frac{(\widehat{\eta}_i-\eta_i)^2}{\lambda_i}\right) d\widehat{\eta}_1 d\widehat{\eta}_2$$

Let us substitute the *standard integral*

$$\int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x\right) \frac{dx}{\sqrt{x}} = 1,$$

in order to have derived the marginal probability

$$\begin{aligned} dF_1 &= f_1(\widehat{\boldsymbol{\xi}}|\boldsymbol{\xi}, (\mathbf{A}'\mathbf{A})^{-1}\sigma^2) d\widehat{\xi}_1 d\widehat{\xi}_2 \\ f_1(\widehat{\boldsymbol{\xi}}) &:= \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) \\ dF_1 &= f_1(\widehat{\boldsymbol{\eta}}|\boldsymbol{\eta}, \boldsymbol{\Lambda}_{\mathbf{N}^{-1}}\sigma^2) d\widehat{\eta}_1 d\widehat{\eta}_2 \\ f_1(\widehat{\boldsymbol{\eta}}) &= \frac{1}{2\pi\sigma^2} \frac{1}{\sqrt{\lambda_1\lambda_2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{2}\frac{(\widehat{\eta}_i-\eta_i)^2}{\lambda_i}\right). \end{aligned}$$

The seventh action item

The *second target* is to generate the *marginal pdf of Helmert's random variable* $x := \widehat{\sigma}^2/\sigma^2$, $\widehat{\sigma}^2$ BIQUUE σ^2, namely *Helmert's Chi Square pdf* χ_p^2 with $p = n - rkA$ (here $p = 1$) "*degree of freedom*".

$$dF_2 = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x\right) \frac{dx}{\sqrt{x}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}}\right) d\widehat{\xi}_1 d\widehat{\xi}_2.$$

Let us substitute the integral

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}}\right) d\widehat{\xi}_1 d\widehat{\xi}_2 = 1$$

in order to have derived the marginal distribution

$$dF_2 = f_2(x)dx, \quad 0 \le x \le \infty$$

$$f_2(x) = \frac{1}{2^{p/2}\Gamma(p/2)} x^{\frac{p-2}{2}} \exp\left(-\frac{1}{2}x\right), \quad \textit{subject to}$$

$$p = n - \mathrm{rk}\mathbf{A} = n - m, \quad \textit{here}: \ p = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$f_2(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} \exp\left(-\frac{1}{2}x\right).$$

The results of the example will be generalized in Lemma B.

Theorem B.11. (marginal probability distributions, special linear Gauss-Markov model):

$$\begin{matrix} E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi} \\ D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2 \end{matrix} \quad \textit{subject to} \quad \left[\begin{matrix} \mathbf{A} \in \mathbb{R}^{n\times m}, \quad \mathrm{rk}\mathbf{A} = m, \quad E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A}) \\ \sigma^2 \in \mathbb{R}^+ \end{matrix}\right.$$

defines a *special linear Gauss–Markov model* of fixed effects $\boldsymbol{\xi} \in \mathbb{R}^m$ and $\sigma^2 \in \mathbb{R}^+$ based upon *Gauss-Laplace i.i.d. observations* $\mathbf{y} := [y_1, \ldots, y_n]'$.

$$\widehat{\boldsymbol{\xi}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \quad \textit{subject to} \quad \left[\begin{matrix} E\{\widehat{\boldsymbol{\xi}}\} = \boldsymbol{\xi} \\ D\{\widehat{\boldsymbol{\xi}}\} = (\mathbf{A}'\mathbf{A})^{-1}\sigma^2 \end{matrix}\right.$$

and

$$\widehat{\sigma}^2 = \frac{1}{n - \mathrm{rk}\mathbf{A}}(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \quad \textit{subject to} \quad \left[\begin{matrix} E\{\widehat{\sigma}^2\} = \widehat{\sigma}^2 \\ D\{\widehat{\sigma}^2\} = \dfrac{2\sigma^4}{n - \mathrm{rk}\mathbf{A}} \end{matrix}\right.$$

identify $\widehat{\boldsymbol{\xi}}$ BLUUE $\boldsymbol{\xi}$ and $\widehat{\sigma}^2$ BIQUUE of σ^2. The *cumulative pdf* of the multidimensional *Gauss-Laplace probability distribution* of the observation vector $\mathbf{y} = [y_1, \ldots, y_n]' \in \mathbb{Y}$

$$
\begin{aligned}
& f(\mathbf{y}|E\{\mathbf{y}\}, D\{\mathbf{y}\} = \mathbf{I}_n\sigma^2)dy_1 \cdots dy_n \\
&= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\})\right) dy_1 \cdots dy_n \\
&= f_1(\widehat{\boldsymbol{\xi}}) f_2(\widehat{\sigma}^2) d\widehat{\xi}_1 \cdots d\widehat{\xi}_m d\widehat{\sigma}^2
\end{aligned}
$$

can be split into two *marginal pdfs* $f_1(\widehat{\boldsymbol{\xi}})$ of $\widehat{\boldsymbol{\xi}}$, BLUUE of $\boldsymbol{\xi}$, and $f_2(\widehat{\sigma}^2)$ of $\widehat{\sigma}^2$, BIQUUE of σ^2.

(i) $\widehat{\boldsymbol{\xi}}$ *BLUUE of* $\boldsymbol{\xi}$
The *marginal pdf* of $\widehat{\boldsymbol{\xi}}$, *BLUUE of* $\boldsymbol{\xi}$, is represented by
(1st version)

$$
\begin{aligned}
dF_1 &= f_1(\widehat{\boldsymbol{\xi}}) d\xi_1 \cdots d\xi_m \\
f_1(\widehat{\boldsymbol{\xi}}) &= \frac{1}{(2\pi)^{m/2}\sigma^m} \left|\mathbf{A}'\mathbf{A}\right|^{1/2} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right) d\xi_1 \cdots d\xi_m \quad \textit{or}
\end{aligned}
$$

(2nd version)

$$
\begin{aligned}
dF_1 &= f_1(\widehat{\boldsymbol{\eta}}) d\eta_1 \cdots d\eta_m \\
f_1(\widehat{\eta}) &= \frac{1}{(2\pi)^{m/2}(\sigma^2)^{m/2}} (\lambda_1\lambda_2 \cdots \lambda_{m-1}\lambda_m)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m} \frac{(\widehat{\eta}_i - \eta)^2}{\lambda_i}\right),
\end{aligned}
$$

by means of *Principal Component Analysis* PCA also called *Singular Value Decomposition* (SVD) or *Eigenvalue Analysis* (EIGEN) of $(\mathbf{A}'\mathbf{A})^{-1}$,

$$
\begin{array}{l}
\boldsymbol{\eta} = \mathbf{U}'_\xi \boldsymbol{\xi} \\
\widehat{\boldsymbol{\eta}} = \mathbf{U}'_\xi \widehat{\boldsymbol{\xi}}
\end{array}
\quad \textit{subject to} \quad
\left[
\begin{array}{l}
\mathbf{U}'_\xi (\mathbf{A}'\mathbf{A})^{-1} \mathbf{U}_\xi = \boldsymbol{\Lambda} = Diag(\lambda_1, \lambda_2, \ldots, \lambda_{m-1}, \lambda_m) \\
\mathbf{U}'_\xi \mathbf{U}_\xi = \mathbf{I}_m, \quad \det \mathbf{U}_\xi = +1
\end{array}
\right.
$$

$$
\begin{aligned}
& f_1(\widehat{\boldsymbol{\eta}}|\boldsymbol{\eta}, \boldsymbol{\Lambda}\sigma^2) = f(\widehat{\eta}_1) f(\widehat{\eta}_2) \cdots f(\widehat{\eta}_{m-1}) f(\widehat{\eta}_m) \\
& \qquad f(\widehat{\eta}_i) = \frac{1}{\sigma\sqrt{\lambda_i}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\frac{(\widehat{\eta}_i - \eta_i)^2}{\lambda_i}\right) \quad \forall i \in \{1, \ldots, m\}.
\end{aligned}
$$

The *transformed fixed effects* $(\widehat{\eta}_1, \ldots, \widehat{\eta}_m)$, BLUUE of $(\eta_1, \ldots, \eta_m)$, are mutually independent and *Gauss-Laplace normal*

$$\widehat{\eta}_i \sim \mathcal{N}(\eta_i|\sigma^2\lambda_i) \quad \forall i \in \{1, \ldots, m\}.$$

(third version)

$$z_i := \frac{\widehat{\eta}_i - \eta}{\sqrt{\sigma^2 \lambda_i}} : f_1(z_i) dz_i = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z_i^2\right) dz_i \quad \forall i \in \{1, \ldots, m\}.$$

(ii) $\widehat{\sigma}^2$ *BIQUUE* σ^2

The *marginal pdf* of $\widehat{\sigma}^2$, BIQUUE σ^2, is represented by
(first version)

$$p = n - \text{rk}\mathbf{A}$$
$$dF_2 = f_2(\widehat{\sigma}^2) d\widehat{\sigma}^2$$
$$f_2(\widehat{\sigma}^2) = \frac{1}{\sigma^p 2^{p/2} \Gamma(p/2)} p^{p/2} \widehat{\sigma}^{p-2} \exp\left(-\frac{1}{2} p \frac{\widehat{\sigma}^2}{\sigma^2}\right).$$

(second version)

$$dF_2 = f_2(x) dx$$
$$x := (n - \text{rk}\mathbf{A}) \frac{\widehat{\sigma}^2}{\sigma^2} = \frac{p}{\sigma^2} \widehat{\sigma}^2 = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})$$
$$f_2(x) = \frac{1}{2^{p/2} \Gamma(p/2)} x^{\frac{p}{2}-1} \exp\left(-\frac{1}{2} x\right).$$

f2(x) as the *standard pdf* of the normalized sample variance is a *Helmert Chi Square* χ_p^2 pdf with $p = n - \text{rk}\mathbf{A}$ "*degree of freedom*".

Proof.

The first action item

First, let us decompose the quadratic form $\|\mathbf{y} - E\{\mathbf{y}\}\|^2$ into estimates $\widehat{E\{\mathbf{y}\}}$ of $E\{\mathbf{y}\}$.

$$\mathbf{y} - E\{\mathbf{y}\} = \mathbf{y} - \widehat{E\{\mathbf{y}\}} + (\widehat{E\{\mathbf{y}\}} - E\{\mathbf{y}\})$$
$$\mathbf{y} - E\{\mathbf{y}\} = \mathbf{y} - \mathbf{A}\widehat{\xi} + \mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$$

and

$$(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = (\mathbf{y} - \widehat{E\{\mathbf{y}\}})'(\mathbf{y} - \widehat{E\{\mathbf{y}\}}) + (\widehat{E\{\mathbf{y}\}} - E\{\mathbf{y}\})'(\widehat{E\{\mathbf{y}\}} - E\{\mathbf{y}\})$$
$$\|\mathbf{y} - E\{\mathbf{y}\}\|^2 = \|\mathbf{y} - \widehat{E\{\mathbf{y}\}}\|^2 + \|\widehat{E\{\mathbf{y}\}} - E\{\mathbf{y}\}\|^2$$
$$\boxed{(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = (\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})}$$
$$\boxed{\|\mathbf{y} - E\{\mathbf{y}\}\|^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\xi}\|^2 + \|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}}}.$$

Here, we took advantage of the *orthogonality relation*.

$$\begin{aligned}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi})' &= (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'(\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} \\ &= (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'(\mathbf{A}' - \mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} = 0.\end{aligned}$$

The second action item

Second, we implement $\widehat{\sigma}^2$ BIQUUE of σ^2 into the decomposed quadratic form.

$$\begin{aligned}\|\mathbf{y} - \mathbf{A}\boldsymbol{\xi}\|^2 &= (\mathbf{y} - \mathbf{A}\boldsymbol{\xi})'(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) = \mathbf{y}'(\mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}\mathbf{y} = (n - \mathrm{rk}\mathbf{A})\widehat{\sigma}^2 \\ \|\mathbf{y} - E\{\mathbf{y}\}\|^2 &= (n - \mathrm{rk}\mathbf{A})\widehat{\sigma}^2 + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\end{aligned}$$

$$\boxed{\|\mathbf{y} - E\{\mathbf{y}\}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y} + (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}).}$$

The matrix of the *normal equations* $\mathbf{N} := \mathbf{A}'\mathbf{A}$, $\mathrm{rk}\mathbf{N} = \mathrm{rk}\mathbf{A}'\mathbf{A} = \mathrm{rk}\mathbf{A} = m$, and the matrix of the *variance component estimation* $\mathbf{M} := \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, $\mathrm{rk}\mathbf{M} = n - \mathrm{rk}\mathbf{A} = n - m$ have been introduced since their rank forms the basis of the *generalized forward and backward Helmert transformation*.

$$\begin{gathered}\mathbf{H}\mathbf{H}' = \mathbf{I}_n \\ \mathbf{z} = \sigma^{-1}\mathbf{H}(\mathbf{y} - E\{\mathbf{y}\}) = \sigma^{-1}\mathbf{H}(\mathbf{y} - \mathbf{A}\boldsymbol{\xi}) \quad \textit{and} \\ \mathbf{y} - E\{\mathbf{y}\} = \sigma\mathbf{H}'\mathbf{z} \\ \frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \mathbf{z}'\mathbf{H}'\mathbf{H}\mathbf{z} = \mathbf{z}'\mathbf{z} \\ \frac{1}{\sigma^2}\|\mathbf{y} - E\{\mathbf{y}\}\|^2 = \|\mathbf{z}\|^2.\end{gathered}$$

The standard canonical variable $\mathbf{z} \in \mathbb{R}^n$ has to be associated with norms $\|\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\xi}}\|$ and $\|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_{\mathbf{A}'\mathbf{A}}$.

The third action item

Third, we take advantage of the eigenspace representation of the matrices $(\mathbf{M}, \mathbf{N})$ and their associated norms.

$$\boxed{\begin{array}{ccc}\mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{V}\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{V}'\mathbf{y} & \textit{versus} & (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\boldsymbol{\Lambda}_{\mathbf{N}}\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ \boldsymbol{\Lambda}_{\mathbf{M}} = \mathit{Diag}(\mu_1, \ldots, \mu_{n-m}, 0, \ldots, 0) & \textit{versus} & \boldsymbol{\Lambda}_{\mathbf{N}} = \mathit{Diag}(\gamma_1, \ldots, \gamma_m). \\ \in \mathbb{R}^n = \mathbb{R}^{n-m} \times \mathbb{R}^m & & \in \mathbb{R}^m\end{array}}$$

m eigenvalues of the matrix $\mathbf{M}$ are zero, but $n - \mathrm{rk}\mathbf{A} = n - m$ is the number of its non-vanishing eigenvalues which we denote by $(\mu_1, \ldots, \mu_{n-m})$. In contrast, $m = \mathrm{rk}\mathbf{A}$ is the n-m number of eigenvalues of the matrix **N**, all non-zero. *The canonical random variable*

$$\mathbf{V}'\mathbf{y} = \mathbf{y}^* \Leftrightarrow \mathbf{y} = \mathbf{V}\mathbf{y}^* \quad \textit{and} \quad \mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) = \widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}$$

lead to

$$\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \frac{1}{\sigma^2}(\mathbf{y}^*)\boldsymbol{\Lambda}_{\mathbf{M}}\mathbf{y}^* + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})'\boldsymbol{\Lambda}_{\mathbf{N}}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})$$

$$\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = \frac{1}{\sigma^2}\sum_{j=1}^{n-m}(y_j^*)^2\mu_j + \frac{1}{\sigma^2}\sum_{i=1}^{m}(\widehat{\eta}_i - \eta_i)^2\gamma_i$$

$$\frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}) = z_1^2 + \cdots + z_{n-m}^2 + z_{n-m+1}^2 + \cdots + z_n^2$$

subject to

$$z_1^2 + \cdots + z_{n-m}^2 = \tfrac{1}{\sigma^2}\sum_{j=1}^{n-m}(y_j^*)^2\mu_j \quad \textit{and} \quad z_{n-m+1}^2 + \cdots + z_n^2 = \tfrac{1}{\sigma^2}\sum_{i=1}^{m}(\widehat{\eta}_i - \eta_i)^2\gamma_i$$

$$\begin{aligned}\|\mathbf{z}\|^2 = \mathbf{z}'\mathbf{z} &= z_1^2 + \cdots + z_{n-m}^2 + z_{n-m+1}^2 + \cdots + z_n^2 \\ &= \frac{1}{\sigma^2}\mathbf{y}'\mathbf{M}\mathbf{y} + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{N}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\ &= \frac{1}{\sigma^2}\|\mathbf{y} - E\{\mathbf{y}\}\|^2 = \frac{1}{\sigma^2}(\mathbf{y} - E\{\mathbf{y}\})'(\mathbf{y} - E\{\mathbf{y}\}).\end{aligned}$$

Obviously, *the eigenspace synthesis* of the matrices $\mathbf{N} = \mathbf{A}'\mathbf{A}$ and $\mathbf{M} = \mathbf{I}_n - -\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ has guided us to the proper structure synthesis of the *generalized Helmert transformation*.

The fourth action item

Fourth, the norm decomposition unable us to split the *cumulative probability*

$$dF = f(y_1, \ldots, y_n)dy_1 \cdots dy_n$$

into the *pdf* of the *Helmert random variable* $x := z_1^2 + \cdots + z_{n-m}^2 = \sigma^{-2}(n - \mathrm{rk}\mathbf{A})\widehat{\sigma}^2 = \sigma^{-2}(n-m)\widehat{\sigma}^2$ and the *pdf* of the *difference random parameter vector* $z_{n-m+1}^2 + \cdots + z_n^2 = \sigma^{-2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})$.

$$dF = f(z_1, \ldots, z_{n-m}, z_{n-m+1}, \ldots, z_n)dz_1 \cdots dz_{n-m}dz_{n-m+1}dz_n$$

$$\boxed{\begin{aligned} f(z_1,\ldots,z_n) &= \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right) \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}}\exp\left(-\frac{1}{2}(z_1^2+\cdots+z_{n-m}^2)\right)\left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \\ &\quad\times\exp\left(-\frac{1}{2}(z_{n-m+1}^2+\cdots+z_n^2)\right). \end{aligned}}$$

Part A

$$x := r^2 = z_1^2+\cdots+z_{n-m}^2 \Rightarrow dx = 2(z_1 dz_1+\cdots+z_{n-m}dz_{n-m})$$

$$\boxed{\begin{gathered} z_1 = r\cos\phi_{n-m-1}\cos\phi_{n-m-2}\cdots\cos\phi_2\cos\phi_1 \\ z_2 = r\cos\phi_{n-m-1}\cos\phi_{n-m-2}\cdots\cos\phi_2\sin\phi_1 \\ \cdots \\ z_{n-m-1} = r\cos\phi_{n-m-1}\sin\phi_{n-m-2} \\ z_{n-m} = r\sin\phi_{n-m-1} \end{gathered}}$$

$$\begin{bmatrix} z_1 \\ \cdots \\ z_{n-m} \end{bmatrix} = \frac{1}{\sigma}Diag(\sqrt{\mu_1},\ldots,\sqrt{\mu_{n-m}})\mathbf{V}'_1\mathbf{y}$$

$$\mathbf{V} = [\mathbf{V}_1,\mathbf{V}_2],\quad \mathbf{V}'_1\mathbf{V}_1 = \mathbf{I}_{n-m},\quad \mathbf{V}'_2\mathbf{V}_2 = \mathbf{I}_m,\quad \mathbf{V}'_1\mathbf{V}_2 = 0$$

$$\mathbf{V}\mathbf{V}' = \mathbf{I}_n,\quad \mathbf{V}\in\mathbb{R}^{n\times n},\quad \mathbf{V}_1\in\mathbb{R}^{n\times(n-m)},\quad \mathbf{V}_2\in\mathbb{R}^{n\times m}\quad \textit{and}$$

$$\begin{bmatrix} z_{n-m+1} \\ \cdots \\ z_n \end{bmatrix} = \frac{1}{\sigma}Diag(\sqrt{\gamma_1},\ldots,\sqrt{\gamma_m})\mathbf{U}'(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})$$

altogether

$$\begin{bmatrix} z_1 \\ \cdots \\ z_{n-m} \\ z_{n-m+1} \\ \cdots \\ z_n \end{bmatrix} = \sigma^{-1}\begin{bmatrix} Diag(\sqrt{\mu_1},\ldots,\sqrt{\mu_{n-m}})\mathbf{V}'_1\mathbf{y} \\ Diag(\sqrt{\gamma_1},\ldots,\sqrt{\gamma_m})\mathbf{U}'(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) \end{bmatrix}.$$

The partitioned vector of the standard random variable z is associated with the norm $\|z_{n-m}\|^2$ and $\|z_m\|^2$, namely

$$
\begin{aligned}
&\|z_{n-m}\|^2 + \|z_m\|^2 \\
&\quad = z_1^2 + \cdots + z_{n-m}^2 + z_{n-m+1}^2 + \cdots + z_n^2 \\
&\quad = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{V_1}Diag(\sqrt{\mu_1}, \ldots, \sqrt{\mu_{n-m}})Diag(\sqrt{\mu_1}, \ldots, \sqrt{\mu_{n-m}})\mathbf{V}'{}_1\mathbf{y} \\
&\qquad + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\, Diag(\sqrt{\gamma_1}, \ldots, \sqrt{\gamma_m})\, Diag(\sqrt{\gamma_1}, \ldots, \sqrt{\gamma_m})\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\
&\quad = \frac{1}{\sigma^2}\mathbf{y}'\mathbf{V}_1 Diag(\mu_1, \ldots, \mu_{n-m})\mathbf{V}'{}_1\mathbf{y} + \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{U}\, Diag(\gamma_1, \ldots, \gamma_m)\mathbf{U}'(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})
\end{aligned}
$$

$$
\boxed{\begin{aligned}
&= dz_1 dz_2 \cdots dz_{n-m-1} dz_{n-m} = r^{n-m-1} dr (\cos\phi_{n-m-1})^{n-m-1}(\cos\phi_{n-m-2})^{n-m-2} \\
&\quad \cdots \cos^2\phi_3 \cos\phi_2 d\phi_{n-m-1} d\phi_{n-m-2} \cdots d\phi_3 d\phi_2 d\phi_1.
\end{aligned}}
$$

The representation of the local $(n-m)$ dimensional hypervolume element in terms of polar coordinates $(\phi_1, \phi_2, \ldots, \phi_{n-m-1}, r)$ has already been given by Lemma B.4. Here, we only transform the random variable r to *Helmert's* random variable x.

$$
x := r^2 : dx = 2r dr, \quad dr = \frac{dx}{2\sqrt{x}}, \quad r^{n-m-1} = x^{(n-m-1)/2}
$$

$$
\boxed{r^{n-m-1} dr = \frac{1}{2} x^{(n-m-1)/2} dx.}
$$

Part A concludes with the representation of the *left pdf* in terms of *Helmert's polar coordinates*

$$
\boxed{\begin{aligned}
dF_\ell &= \left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} \exp -\tfrac{1}{2}(z_1^2 + \cdots + z_{n-m}^2) dz_1 \cdots dz_{n-m} \\
&= \frac{1}{2}\left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{\frac{n-m-2}{2}} dx (\cos\phi_{n-m-1})^{n-m-1}(\cos\phi_{n-m-2})^{n-m-2} \\
&\quad \cdots \cos^2\phi_3 \cos\phi_2 d\phi_{n-m-1} d\phi_{n-m-2} \cdots d\phi_3 d\phi_2 d\phi_1.
\end{aligned}}
$$

Part B

Part B focuses on the representation of the *right pdf*, first in terms of the random variables $\widehat{\boldsymbol{\xi}}$, *second* in terms of the canonical random variables $\widehat{\eta}$.

$$
dF_r = \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \exp\left(-\frac{1}{2}(z_{n-m+1}^2 + \cdots + z_n^2)\right) dz_{n-m+1} \cdots dz_n
$$

$$
z_{n-m+1}^2 + \cdots + z_n^2 = \frac{1}{\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})
$$

$$
\boxed{dz_{n-m+1} \cdots dz_n = \frac{1}{\sigma^{m/2}} |\mathbf{A}'\mathbf{A}|^{1/2} d\widehat{\xi}_1 \cdots d\widehat{\xi}_m.}
$$

The computation of the local m-dimensional hyper volume element $dz_{n-m+1}\cdots dz_n$ has followed Corollary B.3 which is based upon the matrix of the metric $\sigma^{-2}\mathbf{A}'\mathbf{A}$. The *first* representation of the *right pdf* is given by

$$\boxed{\begin{aligned} dF_r &= \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \exp\left(-\frac{1}{2}(z_{n-m+1}^2+\cdots+z_n^2)\right) dz_{n-m+1}\cdots dz_n \\ &= \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{\sigma^{m/2}} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) d\widehat{\xi}_1\cdots d\widehat{\xi}_m. \end{aligned}}$$

Let us introduce the *canonical random variables* $(\widehat{\eta}_1,\ldots,\widehat{\eta}_m)$ which are generated by the correlating quadratic form $\|\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}}$.

$$(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) = (\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{U}\,Diag\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_m}\right)\mathbf{U}'(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}).$$

Here, we took advantage of the *eigenspace synthesis* of the matrix $\mathbf{A}'\mathbf{A} =: \mathbf{N}$ and $(\mathbf{A}'\mathbf{A})^{-1} =: \mathbf{N}^{-1}$. Such an inverse normal matrix is the representing *dispersion matrix* $D\{\widehat{\boldsymbol{\xi}}\} = (\mathbf{A}'\mathbf{A})^{-1}\sigma^2 = \mathbf{N}^{-1}\sigma^2$.

$$\begin{aligned} &\mathbf{U}\mathbf{U}' = \mathbf{I}_m \sim \mathbf{U} \in \mathrm{SO}(m) := \{\mathbf{U} \in^{n\times m} |\mathbf{U}\mathbf{U}' = \mathbf{I}_m, |\mathbf{U}| = +1\} \\ &\mathbf{N} := \mathbf{A}'\mathbf{A} = \mathbf{U}\,Diag(\gamma_1,\ldots,\gamma_m)\mathbf{U}' \quad versus \\ &\mathbf{N}^{-1} := (\mathbf{A}'\mathbf{A})^{-1} = \mathbf{U}\,Diag(\lambda_1,\ldots,\lambda_m)\mathbf{U}' \quad subject\ \ to \\ &\gamma_1 = \lambda_1^{-1},\ldots,\gamma_m = \lambda_m^{-1} \quad or \quad \lambda_1 = \gamma_1^{-1},\ldots,\lambda_m = \gamma_m^{-1} \\ &|\mathbf{A}'\mathbf{A}|^{1/2} = \sqrt{\gamma_1\cdots\gamma_m} = \frac{1}{\sqrt{\lambda_1\cdots\lambda_m}} \end{aligned}$$

$$\boxed{\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta} := \mathbf{U}'(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) \Leftrightarrow \widehat{\boldsymbol{\xi}}-\boldsymbol{\xi} := \mathbf{U}'(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})}$$

$$\|\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2_{\mathbf{A}'\mathbf{A}} =: (\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi}) = (\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})'Diag\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_m}\right)(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta}).$$

The local m-dimensional hypervolume element $d\widehat{\xi}_1\cdots d\widehat{\xi}_m$ is transformed to the local m-dimensional hypervolume element $d\widehat{\eta}_1\cdots d\widehat{\eta}_m$ by

$$d\widehat{\xi}_1\cdots d\widehat{\xi}_m = |\mathbf{U}|d\widehat{\eta}_1\cdots d\widehat{\eta}_m = d\widehat{\eta}_1\cdots d\widehat{\eta}_m.$$

Accordingly we have derived the *second* representation of the *right pdf* $f(\widehat{\boldsymbol{\eta}})$.

$$\begin{aligned} dF_r = &\left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{\sigma^{m/2}\sqrt{\lambda_1\cdots\lambda_m}} \\ &\times \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})'Diag\left(\frac{1}{\lambda_1},\ldots,y\frac{1}{\lambda_m}\right)(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\right) d\widehat{\eta}_1\cdots d\widehat{\eta}_m. \end{aligned}$$

Part C
Part C is an attempt to merge the *left and right pdf*

$$\boxed{\begin{aligned} dF &= dF_\ell dF_r = \frac{1}{2}\left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} dx\, d\omega_{n-m-1} \\ &* \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{\sigma^m} \exp\left(-\frac{1}{2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) d\widehat{\xi}_1 \cdots d\widehat{\xi}_m \quad \textit{or} \\ dF &= dF_\ell dF_r = \frac{1}{2}\left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} dx\, d\omega_{n-m-1} \\ &* \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{\sigma^m\sqrt{\lambda_1\cdots\lambda_m}} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})' Diag\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_m}\right)(\widehat{\boldsymbol{\eta}}-\boldsymbol{\eta})\right) d\widehat{\eta}_1 \cdots d\widehat{\eta}_m . \end{aligned}}$$

The local $(n-m-1)$-dimensional hypersurface element has been denoted by $d\omega_{n-m-1}$ according to Lemma B.4.

The fifth action item

Fifth, we are going to compute the *marginal pdf* of $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$.

$$dF_1 = f_1(\widehat{\boldsymbol{\xi}})\, d\widehat{\xi}_1 \cdots d\widehat{\xi}_m$$

as well as

$$dF_1 = f_1(\widehat{\boldsymbol{\eta}})\, d\widehat{\eta}_1 \cdots d\widehat{\eta}_m$$

include the *first marginal pdf* $f_1(\widehat{\boldsymbol{\xi}})$ and $f_1(\widehat{\boldsymbol{\eta}})$, respectively.
The definition

$$\begin{aligned} f_1(\widehat{\boldsymbol{\xi}}) := \int_0^\infty dx \int d\omega_{n-m-1} \frac{1}{2}\left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} \\ * \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{(\sigma^2)^{m/2}} \exp\left(-\frac{1}{2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) \end{aligned}$$

subject to

$$\int_0^\infty dx \int d\omega_{n-m-1} \frac{1}{2}\left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} = 1$$

leads us to

$$\boxed{f_1(\widehat{\boldsymbol{\xi}}) = \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{\sigma^m} \exp\left(-\frac{1}{2}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}}-\boldsymbol{\xi})\right).}$$

Unfortunately, such a general *multivariate Gauss-Laplace normal distribution* cannot be tabulated. An alternative is offered by introducing canonical unknown parameters $\widehat{\boldsymbol{\eta}}$ as random variables.
The definition

$$f_1(\widehat{\boldsymbol{\eta}}) := \int_0^\infty dx \int d\omega_{n-m-1} \frac{1}{2} \left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2}$$

$$* \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{\sigma^m} (\lambda_1 \lambda_2 \cdots \lambda_{m-1} \lambda_m)^{-1/2}$$

$$\times \exp\left(-\frac{1}{2\sigma^2} (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})' Diag\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_m}\right) (\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})\right)$$

subject to

$$\int_0^\infty dx \int d\omega_{n-m-1} \frac{1}{2} \left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} = 1$$

alternatively leads us to

$$\boxed{\begin{gathered} f_1(\widehat{\boldsymbol{\eta}}) = \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{1}{\sigma^m} \frac{1}{\sqrt{\lambda_1 \cdots \lambda_m}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m \frac{(\widehat{\eta}_i - \eta)^2}{\lambda_i}\right) \\ f_1(\widehat{\eta}_1, \ldots, \widehat{\eta}_m) = f_1(\widehat{\eta}_1) \cdots f_1(\widehat{\eta}_m) \\ f_1(\widehat{\eta}_i) := \frac{1}{\sigma \sqrt{\lambda_i} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(\widehat{\eta}_i - \eta)^2}{\lambda_i}\right) \quad \forall i \in \{1, \ldots, m\}. \end{gathered}}$$

Obviously the transformed random variables $(\widehat{\eta}_1, \ldots, y\widehat{\eta}_m)$ BLUUE of $(\eta_1, \ldots, \eta_m)$ are mutually independent and *Gauss-Laplace normal.*

The sixth action item

Sixth, we shall compute the *marginal pdf* of *Helmert's random variable* $x = (n - \text{rk}\mathbf{A})\widehat{\sigma}^2/\sigma^2 = (n - m)\widehat{\sigma}^2/\sigma^2$, $\widehat{\sigma}^2$ BIQUUE σ^2,

$$dF_2 = f_2(x)dx$$

includes the second marginal pdf $f_2(x)$.
The definition

$$f_2(x) := \int d\omega_{n-m-1} \frac{1}{2} \left(\frac{1}{2\pi}\right)^{\frac{n-m}{2}} x^{(n-m-2)/2} * \int_{-\infty}^{+\infty} d\widehat{\xi}_1 \cdots \int_{-\infty}^{+\infty} d\widehat{\xi}_m \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{\sigma^m}$$

$$\times \exp\left(-\frac{1}{2\sigma^2} (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \mathbf{A}' \mathbf{A} (\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right)$$

subject to

$$\omega_{n-m-1} = \int d\omega_{n-m-1} = \frac{2 * \pi^{(n-m-1)/2}}{\Gamma(\frac{n-m-1}{2})},$$

according to Lemma B.4

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} d\widehat{\xi}_1 \cdots d\widehat{\xi}_m \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \frac{|\mathbf{A}'\mathbf{A}|^{1/2}}{\sigma^m} \exp\left(-\frac{1}{2\sigma^2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\mathbf{A}'\mathbf{A}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right)$$

$$= \int_{-\infty}^{+\infty} dz_1 \cdots \int_{-\infty}^{+\infty} dz_m \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} \exp\left(-\frac{1}{2}(z_1^2 + \cdots + z_m^2)\right)$$

leads us to

$$p := n - \mathrm{rk}\mathbf{A} = n - m$$

$$\sqrt{\pi}\Gamma\left(\frac{n-m-1}{2}\right) = \Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-m-1}{2}\right) = \Gamma\left(\frac{n-m}{2}\right) = \Gamma\left(\frac{p}{2}\right)$$

$$\boxed{f_2(x) = \frac{1}{2^{p/2}\Gamma(p/2)} x^{\frac{p}{2}-1} \exp\left(-\frac{1}{2}x\right),}$$

namely the standard pdf of the normalised sample variance, known as *Helmert's Chi Square pdf* χ_p^2 with $p = n - \mathrm{rk}\mathbf{A} = n - m$ "*degree of freedom*". If you substitute $x = (n - rk\mathbf{A})\widehat{\sigma}^2/\sigma^2 = (n-m)\widehat{\sigma}^2/\sigma^2$, $dx = (n - \mathrm{rk}\mathbf{A})\sigma^{-2}d\widehat{\sigma}^2 = (n-m)\sigma^{-2}d\widehat{\sigma}^2$ we arrive at the pdf of the sample variance $\widehat{\sigma}^2$, in particular

$$\boxed{\begin{aligned} dF_2 &= f_2(\widehat{\sigma}^2)d\widehat{\sigma}^2 \\ f_2(\widehat{\sigma}^2) &= \frac{1}{\sigma^r 2^{p/2}\Gamma(p/2)} p^{p/2}\widehat{\sigma}^{p-2} \exp\left(-\frac{1}{2}p\frac{\widehat{\sigma}^2}{\sigma^2}\right). \end{aligned}}$$

Here is our proof's end.

Theorem B.12. (marginal probability distributions, special linear Gauss–Markov model with datum defect):

$$\begin{array}{lll} & & \mathbf{A} \in^{n\times m}, \quad r := rk\mathbf{A} < \min\{n, m\} \\ E\{\mathbf{y}\} = \mathbf{A}\boldsymbol{\xi} & & \\ & \textit{subject to} & E\{\mathbf{y}\} \in \mathcal{R}(\mathbf{A}) \\ D\{\mathbf{y}\} = \mathbf{V}\sigma^2 & & \\ & & \mathbf{V} \in^{n\times n}, \quad rk\mathbf{V} = n \end{array}$$

defines a *special linear Gauss–Markov model with datum defect* of fixed effects $\boldsymbol{\xi} \in \mathbb{R}^m$ and a positive definite variance-covariance matrix $D\{\mathbf{y}\} = \boldsymbol{\Sigma}_y$ of multivariate Gauss-Laplace distributed observations $\mathbf{y} := [y_1, \ldots, y_n]'$.

$$\widehat{\boldsymbol{\xi}} = \mathbf{A}^+\mathbf{y} \quad \textit{subject to} \quad \begin{bmatrix} \widehat{\boldsymbol{\xi}} = \mathbf{L}\mathbf{y} \text{ ``\textit{linear}''} \\ \|\mathbf{L}\mathbf{A} - \mathbf{I}_m\|^2 = \min \text{ ``\textit{minimum bias}''} \\ trD\{\widehat{\boldsymbol{\xi}}\} = tr\,\mathbf{L}\boldsymbol{\Sigma}_y\mathbf{L}' = \min \text{ ``\textit{best}''} \quad \textit{and} \end{bmatrix}$$

$$\widehat{\sigma}^2 = \frac{1}{n - \mathrm{rk}\mathbf{A}}(\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\xi}})'\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \mathbf{A}\widehat{\boldsymbol{\xi}}) \quad \textit{subject to} \quad \begin{bmatrix} E\{\widehat{\sigma}^2\} = \sigma^2 \\ D\{\widehat{\sigma}^2\} = \dfrac{2\sigma^4}{n - \mathrm{rk}\mathbf{A}} \end{bmatrix}$$

identify $\widehat{\boldsymbol{\xi}}$ BLIMBE of $\boldsymbol{\xi}$ and $\widehat{\sigma}^2$ BIQUUE of σ^2.

Part A

The *cumulative pdf* of the multivariate *Gauss-Laplace probability distribution* of the observation vector $\mathbf{y} = [y_1, \ldots, y_n]' \in \mathbb{Y}$

$$\begin{aligned} & f(\mathbf{y}|E\{\mathbf{y}\}, \quad D\{\mathbf{y}\} = \boldsymbol{\Sigma}_y)dy_1 \cdots dy_n \\ & \quad = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\boldsymbol{\Sigma}_y|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - E\{\mathbf{y}\})'\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - E\{\mathbf{y}\})\right) dy_1 \cdots dy_n \end{aligned}$$

is transformed by

$$\begin{aligned} \boldsymbol{\Sigma}_y &= \mathbf{W}'Diag(\sigma_1, \ldots, \sigma_n)\mathbf{W} = \mathbf{W}'Diag(\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n})\,Diag(\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n})\mathbf{W} \\ \boldsymbol{\Sigma}_y^{1/2} &:= Diag(\sqrt{\sigma_1}, \ldots, \sqrt{\sigma_n})\mathbf{W} \\ \boldsymbol{\Sigma}_y &= (\boldsymbol{\Sigma}_y^{1/2})'\boldsymbol{\Sigma}_y^{1/2} \quad \textit{versus} \\ \boldsymbol{\Sigma}_y^{-1} &= \mathbf{W}'Diag\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_n}\right)\mathbf{W} \\ &= \mathbf{W}'Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right)Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right)\mathbf{W} \\ \boldsymbol{\Sigma}^{-1/2} &:= Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right)\mathbf{W} \\ \boldsymbol{\Sigma}^{-1} &= (\boldsymbol{\Sigma}_y^{-1/2})'\boldsymbol{\Sigma}_y^{-1/2} \end{aligned}$$

subject to the orthogonality condition

$$\mathbf{W}\mathbf{W}' = \mathbf{I}_n$$

$$\|\mathbf{y} - E\{\mathbf{y}\}\|^2_{\boldsymbol{\Sigma}_y^{-1}} := (\mathbf{y} - E\{\mathbf{y}\})'\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - E\{\mathbf{y}\})$$

$$= (\mathbf{y} - E\{\mathbf{y}\})'\mathbf{W}'Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right) Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right)$$

$$\times \mathbf{W}(\mathbf{y} - E\{\mathbf{y}\})'$$

$$= (\mathbf{y}^* - E\{\mathbf{y}^*\})'(\mathbf{y}^* - E\{\mathbf{y}^*\}) =: \|\mathbf{y}^* - E\{\mathbf{y}^*\}\|^2_{\mathbf{I}_n}$$

subject to the star or canonical coordinates

$$\mathbf{y}^* = Diag\left(\frac{1}{\sqrt{\sigma_1}}, \ldots, \frac{1}{\sqrt{\sigma_n}}\right)\mathbf{W}\mathbf{y} = \boldsymbol{\Sigma}_y^{-1/2}\mathbf{y}$$

$$f(\mathbf{y}|E\{\mathbf{y}^*\}, D\{\mathbf{y}^*\} = \mathbf{I}_n)dy_1^* \cdots dy_n^*$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\mathbf{y}^* - E\{\mathbf{y}^*\})'(\mathbf{y}^* - E\{\mathbf{y}^*\})\right) dy_1^* \cdots dy_n^*$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\|\mathbf{y}^* - E\{\mathbf{y}^*\}\|^2\right) dy_1^* \cdots dy_n^*$$

into the canonical Gauss-Laplace pdf.

Part B

The marginal pdf of $\widehat{\boldsymbol{\xi}}$ BLUUE of $\boldsymbol{\xi}$, is represented by

(1st version)

$$dF_1 = f_1(\widehat{\boldsymbol{\xi}})d\xi_1 \cdots d\xi_n.$$

B-8 Multidimensional Variance Analysis, Sampling from the Multivariate Gauss–Laplace Normal Distribution

Let $x_1, \ldots, x_N$ denote a random sample from as k-variate *Gauss-Laplace normal distribution* with μ as the mean vector and Σ as the *variance-covariance matrix*, such that

$$\text{(B 370)} \quad \mathbf{x}_i := (\mathbf{X}_{i,1}, \ldots, \mathbf{X}_{i,k})'$$

for $i \in \{1, \ldots, N\}$. Let $\mathbf{X}^{-1} := (\mathbf{x}_1, \ldots, \mathbf{x}_N)$. The *joint density* of the random sample is generated by the *product of N multivariate normal densities*. Let us denote the *parameter space* by $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\Sigma}$ positive definite$\} =: \boldsymbol{\Omega}$. Here we describe estimation of the parameters from a *multivariate Gauss-Laplace normal distribution* and describe properties of these estimates. In addition, we care for inference for a *simple, multiple, and partial correlation* coefficients based on the normal random sample. Of course, we have to proof that we can assume a multivariate *Gauss-Laplace normal distribution*. We add some methods for assessing *Gauss-Laplace normality* and describe *suitable transformation* to normality.

B-81 Distribution of Sample Mean and Variance-Covariance

By the representation of the *sample mean* of the random sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ by (B371) and the sample variance-covariance matrix by (B372)–(B375), namely the $k \times k$ dimensional matrix $\mathbf{S}_N/(N-1) = \{\mathbf{S}_{jl}\}$ for all $j, l \in \{1, \ldots, k\}$ subject to $\mathbf{X} = \sum_{i=1}^{N} X_{i,j}/N$. The sample variance-covariance matrix has k *variances* $\mathbf{S}_{jj}$ for $j \in \{1, \ldots, k\}$ and $k(k-1)/2$ possibly distinct *covariances*. $\mathbf{S}_{jl}$ for $j, l, j, l \in \{1, \ldots, k\}$. The *generalized sample variance* is the scalar quality $|\mathbf{S}_N/(N-1)|$ which determines the degree of *"peakedness"* of the joint density of $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and is a normal summary measure of variability in the sample.

Table B.1 Sample mean, sample variance-covariance matrix

Sample mean

$$\mathbf{x}_N^\wedge := \sum_{i=1}^{N} \mathbf{x}_i \qquad \text{(B371)}$$

Sample variance-covariance matrix

$$\mathbf{S}_N := \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{x}_N^\wedge)(\mathbf{x}_i - \mathbf{x}_N^\wedge)\prime = \sum_{i=1}^{N} \mathbf{x}_i x_i' - N\mathbf{x}_i - \mathbf{x}_N^\wedge \mathbf{x}_i - \mathbf{x}_N^{\wedge\prime} \qquad \text{(B372)}$$

$$\mathbf{S}_N/(N-a) =: \{\mathbf{S}_{jl}\} \text{ for } j, l \in \{1, \ldots, k\} \qquad \text{(B373)}$$

$$\mathbf{S}_{jl} := \sum_{i=1}^{N} (\mathbf{X}_{i,j} - \mathbf{X}_j^\wedge)(\mathbf{X}_{il} - \mathbf{X}_l)/(N-1) \qquad \text{(B374)}$$

Subject to

$$\mathbf{X}_j^\wedge = \sum_{i=1}^{N} \mathbf{X}_{i,j} N \qquad \text{(B375)}$$

Given a univariate random sample $\mathbf{X}_1, \ldots, \mathbf{X}_N$ from a $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ population, that the sample mean $\mathbf{X}^\wedge$ has a normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2/N$, the static $(N-1)\mathbf{S}^2/\boldsymbol{\sigma}^2$ is distributed as a $\boldsymbol{\chi}^2$ variable and the *two distributions are independent*. The corresponding distributional result for the k-variable situations is given by our previous result. We first define a multivariate distribution called the *Wishart distribution* which is derived from the multivariate *Gauss-Laplace normal distribution* as its *sampling distribution of the sample* statistics

$$\sum_{i=1}^{N} (\mathbf{X}_i - \mathbf{X}_N^\wedge)(\mathbf{X}_i - \mathbf{X}_N^\wedge)' \qquad \text{(B 376)}$$

The *Wishart distribution* is a multivariate extension of the chi-square distribution.

Definition B.18. Wishart distribution

A random k-dimensional matrix $\mathbf{W}$ is said to follow a *k-dimensional noncentral Wishart distribution*. $\mathbf{W}_k(\boldsymbol{\Sigma}, m, \mathbf{x})$ with *m degrees of freedom*, parameter $\boldsymbol{\Sigma}$, and noncentrality parameter $\boldsymbol{\lambda} := \boldsymbol{\mu}' \sum^{-1} \boldsymbol{\mu}'/2$ if $\mathbf{W}$ can be represented as, $\mathbf{W} := \sum_{j=1}^{m} \mathbf{x}_i \mathbf{x}_j$ where $\mathbf{x}_j$ for all $j \in \{1, \ldots, m\}$ are $i.i.i.N_k(\mu, \sum)$ vectors. Provide $m.k$ the density function of $\mathbf{W}$ is

$$f(\mathbf{W}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{|\mathbf{W}|^{(m-k-1)/2} \exp\{tr(-\boldsymbol{\Sigma}^{-}1\mathbf{W}/2)\}}{2^{km/2}|\boldsymbol{\Sigma}|^{m/2}\Gamma_k(m/2)} \qquad \text{(B 377)}$$

where $\Gamma_k(m/2) = \pi^{k(k-1)/4} \Pi_{j=1}^{k} \Gamma(\frac{1}{2}(m + 1 - j))$ is the *multivariate Gamma functions*. If $\boldsymbol{\lambda} = 0$, we say that $\mathbf{W}$ follows a *central Wishart distribution* $\mathbf{W}_k(\boldsymbol{\Sigma}, m)$; since $\boldsymbol{\Sigma}$ is p.d., it is clear that $\boldsymbol{\lambda} = 0$ if and only if $\boldsymbol{\mu} = 0$. The additivity property of Wishart matrices states that if $\mathbf{W}_j \overset{ind}{\sim} \mathbf{W}_k(\boldsymbol{\Sigma}, m_j)$, $j = 1, \ldots, J$, then $\sum_{j=1}^{J} \mathbf{W}_j \sim \mathbf{W}_k\big(\boldsymbol{\Sigma}, \sum_{j=1}^{J} m_j\big)$.

End of Definition B.18: Wishart distribution

Intermezzo

Before we proceed we must introduce results about the *Fourier transform of a random vector*, namely called *"characteristic function"* being generalized in *Appendix C* to the *"characteristic functional"* for random functions, namely *stochastic processes*. We will introduce some lemmas and definitions.

Definition B.19. Characteristic function of a random vector

The characteristic function of a random vector $\mathbf{X}$ is $\mathbf{E}\{\exp i t'\mathbf{x}\}$ define for every vector t.

Definition B.20. Decomposition of a complex-valued function

Let the complex-valued function $g(\mathbf{x})$ be written as

$$g(\mathbf{x}) = g_1(\mathbf{x}) + i g_2(\mathbf{x}) \qquad \text{(B 378)}$$

where $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ are *real-valued*. Then the expected value of $g(\mathbf{X})$ is

$$\mathbf{E}\{g(\mathbf{X})\} = \mathbf{E}\{g_1(\mathbf{x})\} + i\mathbf{E}\{g_2(\mathbf{x})\} \qquad \text{(B 379)}$$

in particular,

$$\mathbf{E}\{\exp(i t'\mathbf{x})\} = \mathbf{E}\{\cos(t'\mathbf{x})\} + i\mathbf{E}\{\sin(t'\mathbf{x})\} \qquad \text{(B 380)}$$

Lemma B.18. Product rule for independent variables

Let $\mathbf{X}' := (\mathbf{X}^{(1)\prime}\mathbf{X}^{(2)\prime})$. If $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent and $g(\mathbf{x}) = g(1)(\mathbf{x}^{(1)})g(2)(\mathbf{x}^{(2)})$, then

$$\mathbf{E}\{g(\mathbf{X})\} = \mathbf{E}\{g^{(1)}(\mathbf{X}^{(1)})\}\mathbf{E}\{g^{(2)}(\mathbf{X}^{(2)})\} \qquad \text{(B 381)}$$

Lemma B.19. Components of independent distribution data

If the component of $\mathbf{X}$ are independently distributed, then

$$\mathbf{E}\{\exp(it'\mathbf{X}_N^{\wedge})\} = \prod_{j=1}^{N} \mathbf{E}\{it_j\mathbf{X}j\} \qquad \text{(B 382)}$$

Theorem B.20. Characteristic function of a vector-valued random variant of type Gauss-Laplace distribution

The characteristic function of $\mathbf{X}$ which is *Gauss-Laplace distribution* according to $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$\mathbf{E}\{\exp(it'\mathbf{X})\} = \exp\left\{t'\boldsymbol{\mu} + \frac{1}{2}t'\boldsymbol{\Sigma}t\right\} \qquad \text{(B 383)}$$

for any real vector t.

Theorem B.21. Characteristic function $\phi(t)$

If the random vector $\mathbf{X}$ has the density $f(\mathbf{x})$ and the characteristic function $f(t)$, then

$$f(\mathbf{x}) = \frac{1}{(2\pi)\boldsymbol{\gamma}} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp(-it'\mathbf{x})\boldsymbol{\phi}^N(t)\mathrm{d}\boldsymbol{\omega}_1 \dots \mathrm{d}\boldsymbol{\omega}_{\boldsymbol{\gamma}} \qquad \text{(B 384)}$$

and

$$f^N(\boldsymbol{\omega}) = \mathbf{E}\{+i\boldsymbol{\omega}'\mathbf{X}\} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp(+i\boldsymbol{\omega}'\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x}_1 \dots \mathrm{d}\mathbf{x}_{\boldsymbol{\gamma}} \qquad \text{(B 385)}$$

illustrating *synthesis and analysis*

Here ends our intermezzo

$$f^N(\boldsymbol{\omega}) = \mathbf{E}\{+i\boldsymbol{\omega}'\mathbf{X}\} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \exp(+i\boldsymbol{\omega}'\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x}_1 \dots \mathrm{d}\mathbf{x}_{\boldsymbol{\gamma}} \qquad \text{(B 386)}$$

Results: Distribution of $\mathbf{x}_N^\wedge$ and $\mathbf{S}_N$

Let $\mathbf{x}_1, \dots, \mathbf{x}_N$ denote a *random sample* from the *Gauss-Laplace normal distribution* $N_k(\boldsymbol{\mu}, \boldsymbol{\sigma})$.

(a) The distribution of $\mathbf{x}_N^\wedge$ is $N_k(\boldsymbol{\mu\sigma}/N)$
(b) For $N \geq 2$, $\mathbf{S}_N$ follows a *Wishart distribution* $\mathbf{W}_k(\boldsymbol{\Sigma}'_1 N - 1)$
(c) $\mathbf{x}_N^\wedge$ and $\mathbf{S}_N$ are *independently distributed.*

Proof

$$\mathbf{E}\{\exp(it'\mathbf{x}_N^\wedge)\} = \prod_{j=1}^{N} \mathbf{E}\{\exp(it'_j\mathbf{x}_j)/N\} = \exp(t'\boldsymbol{\mu} + \frac{1}{2}t'(\boldsymbol{\Sigma}/N)t \qquad \text{(B 387)}$$

for $N = 2$ it follows

$$\mathbf{S}_2 = \left[\mathbf{x}_1 - \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right]\left[\mathbf{x}_1 - \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right]' + \left[\mathbf{x}_2 - \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)\right]\left[\mathbf{x}_2 - \frac{1}{2}(\mathbf{x}_1 + x_2)\right]'$$

$$\Rightarrow S_2 = \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)' \qquad \text{(B 388)}$$

Obviously $(\mathbf{x}_1 - \mathbf{x}_2)\sqrt{2}$ is distributed according to $N_k(0, \boldsymbol{\Sigma})$ and $\mathbf{S}_2$ according to $\mathbf{W}_k(\boldsymbol{\Sigma}, 1)$

$$\begin{aligned}\mathbf{S}_3 = &\left[\mathbf{x}_1 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]\left[\mathbf{x}_1 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]' \\ &+ \left[\mathbf{x}_2 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]\left[\mathbf{x}_2 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]' \\ &+ \left[\mathbf{x}_3 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]\left[\mathbf{x}_3 - \frac{1}{3}(\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3)\right]' \qquad \text{(B 389)}\end{aligned}$$

$$\mathbf{S}_3 = \mathbf{S}_2 + \frac{2}{3}(\mathbf{x}_3 - \mathbf{x}_2^\wedge)(\mathbf{x}_3 - \mathbf{x}_2^\wedge)' \qquad \text{(B 390)}$$

subject to

$$\mathbf{x}_2^\wedge = \frac{1}{3}(2\mathbf{x}_2^\wedge + x_3) \qquad \text{(B 391)}$$

in general

$$\mathbf{x}_m^{\wedge} = \frac{1}{m+1}(m x_m^{\wedge} + \mathbf{x}_{m+1}) \quad \textit{and}$$

$$\mathbf{S}_{m+1} = \mathbf{S}_m + \frac{m}{m+1}(\mathbf{x}_{m+1} - \mathbf{x}_m^{\wedge})(\mathbf{x}_{m+1} - \mathbf{x}_m^{\wedge})' \qquad \text{(B 392)}$$

B-82 Distribution Related to Correlation Coefficients

Since $\mathbf{x}_i$ for all $i \in \{1, \ldots, m+1\}$ are mutually independent and therefore $\left(\mathbf{x}_{m+1}^{\wedge}, \mathbf{S}_m\right)$ are functions only of $\mathbf{x}_i$ for $i \in \{1, \ldots, m\}$, it follows that $\left(\mathbf{x}_{m+1}^{\wedge}, \mathbf{S}_m\right)$ are independent of $\mathbf{x}_{m+1}^{\wedge}$.
By the *hypothesis of induction* we enjoy the *independence* of $\mathbf{x}_m^{\wedge}$ and $\mathbf{S}_m$

$$\mathbf{x}_{m+1}^{\wedge}, \mathbf{S}_m \quad \textit{and} \quad \mathbf{x}_{m+1} \qquad \text{(B 393)}$$

are mutually independent. Thus, $\mathbf{S}_m$ is independent of $\left(\mathbf{x}_m^{\wedge}, \mathbf{x}_{m+1}\right)$ and therefore independent of $\mathbf{x}_{m+1} - \mathbf{x}_m^{\wedge}$ which follows a $N_k\left(\mathbf{0}, \frac{m+1}{m}\Sigma\right) \frac{m+1}{m}\left(\mathbf{x}_{m+1} - \mathbf{x}_m^{\wedge}\right)(\mathbf{x}_{m+1} - \mathbf{x}_m^{\wedge})'$ are distributed as $\mathbf{W}_k\,(\Sigma, 1)$. Note that

$$\boxed{cov\left(\mathbf{x}_i - \mathbf{x}_m^{\wedge}, \mathbf{x}_n^{\wedge}\right) = 0} \qquad \text{(B 394)}$$

Which implies that $\left[\left(\mathbf{x}_1 - \mathbf{x}_N^{\wedge}\right), \ldots, \left(\mathbf{x}_N - \mathbf{x}_N^{\wedge}\right)'\right]'$ is uncorrelated and *therefore independent of* $\mathbf{x}_N^{\wedge}$. In consequence, $\mathbf{S}_n$ is distributed independently of $\mathbf{x}_N^{\wedge}$.

Results: Maximum likelihood estimation if $\boldsymbol{\mu}$ and $\sum$

Based on a random sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from a $N_k\,(\boldsymbol{\mu}, \Sigma)$ distribution, the maximum likelihood estimates of $\boldsymbol{\mu}$ and $\sum$ are

$$\widehat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i = \mathbf{x}_N^{\wedge}, \qquad \text{(B 395)}$$

and

$$\widehat{\Sigma}_{ML} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}_i - \mathbf{x}_N^{\wedge}\right)\left(\mathbf{x}_i - \overline{\mathbf{X}}_N^{\wedge}\right)' = \frac{N-1}{N}\mathbf{S}_N \qquad \text{(B 396)}$$

$\boxed{\text{Proof}}$

The likelihood function $\mathbf{L}\,(\boldsymbol{\mu}, \Sigma; \mathbf{x}_1, \ldots, \mathbf{x}_N)$ has the form shown on the right side. The MLE's of $\boldsymbol{\mu}$ and Σ are denoted by $\widehat{\mu}_{ML}$ and $\widehat{\Sigma}_{ML}$, and are the values that maximize $\mathbf{L}(\boldsymbol{\mu}, \Sigma; \mathbf{x}_1, \ldots, \mathbf{x}_N)$. Since Σ^{-1} is p.d., the distance $\left(\mathbf{x}_N^{\wedge} - \boldsymbol{\mu}\right)' \Sigma^{-1} \left(\mathbf{x}_N^{\wedge} - \boldsymbol{\mu}\right)$ in the exponent of the likelihood function is positive unless $\boldsymbol{\mu} = \overline{\mathbf{X}}_N$. The MLE of $\boldsymbol{\mu}$ is then $\mathbf{x}_N^{\wedge}$, since it is the value of $\boldsymbol{\mu}$ that maximizes the likelihood function. We next maximize the following function with respect to Σ:

$$L\left(\widehat{\mu}_{ML}, \Sigma\right) = \frac{\exp\left\{-\frac{1}{2} tr\left[\Sigma^{-1} \sum_{i=1}^{N} \left(\mathbf{x}_i - \mathbf{x}_N^{\wedge}\right)\left(\mathbf{x}_i - \overline{\mathbf{X}}_N^{\wedge}\right)'\right]\right\}}{(2\pi)^{Nk/2} |\Sigma|^{N/2}} \qquad \text{(B 397)}$$

With $A = \Sigma$, $B = \sum_{i=1}^{N} \left(\mathbf{x}_i - \mathbf{x}_N^{\wedge}\right)\left(\mathbf{x}_i - \overline{\mathbf{X}}_N^{\wedge}\right)'$, and $b = N/2$, we can show that $\widehat{\Sigma}_{ML} = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{x}_i - \mathbf{x}_N^{\wedge}\right)\left(\mathbf{x}_i - \overline{\mathbf{X}}_N^{\wedge}\right)'$ maximizes $L\left(\widehat{\mu}_{ML}, \Sigma\right)$. The maximum likelihood is

$$L\left(\widehat{\mu}_{ML}, \widehat{\Sigma}_{ML}\right) = \frac{exp - \frac{Nk}{2}}{(2\pi)^{Nk/2} |\widehat{\Sigma}_{ML}|^{N/2'}} \qquad \text{(B 397)}$$

which completes the proof.

Corollary: Invariance property of MLE

Using the invariance property of MLE's which states that the MLE of a function $g(\theta)$ of a parameter θ is $g\left(\widehat{\theta}_{ML}\right)$, we see that

1. The MLE of the function $\mu\prime\Sigma^{-1}\mu$ is $\widehat{\mu}'_{ML}\widehat{\Sigma}^{-1}_{ML}\widehat{\mu}_{ML}$,
2. The MLE of $\sqrt{\sigma_{lj}}$, the (l,j)th element of $\sum$ is given by $\sqrt{\widehat{\sigma}_{lj}}$, the square root of the MLE of the(l,j)th entry in $\widehat{\sum}_{ML}$

Although the estimator $\widehat{\mu}_{ML}$ is an unbiased estimator of μ, $\widehat{\Sigma}_{ML}$ is a biased estimator of Σ. analogous to the *univariate case*, an unbiased estimator of Σ is $\widehat{\Sigma} = \frac{1}{N-1}\mathbf{S}_N$. The unbiased estimators of μ abd Σ denoted by $\widehat{\mu}$ and $\widehat{\Sigma}$ respectively are *complete sufficient statistics*.
Of course, you have realized that we gave no space to derive the *Wishart distribution*. Instead we refer to his works in Wishart (1928), Wishart (1931), Wishart (1955) as well as Wishart and Bartlett (1932). the general theory of *multivariate statistical analysis* including the *Gauss-Laplace multivariate normal distribution*, the estimation of the mean vector and the variance-covariance matrix the distribution of *sample correlation coefficients*, the generalized t^2 statistics and the distribution of the *sample covariance matrix* can be taken from the excellent review of Anderson (1958) including *413 references*. finally we advice the readers to consider the works of Koch (1988a), Koch (1988b), pp. 160–173

B.8 Distributions related to correlation coefficients

Traditionally we divide *simple, multiple and partial correlations*. Here we restrict our study to the analysis of the *simple correlation based on iid bivariate Gauss-Laplace normal samples*.

$$\left(\mathbf{X_{i,j}}, \mathbf{X_{i,l}}\right) \quad \textit{for all } i \in \{1, \ldots, N\} \qquad \text{(B 399)}$$

where $\mathbf{X_{i,j}}$ and $\mathbf{X_{i,l}}$ denote the j^{th} and l^{th} components of the k - 1 dimensional vector x. Suppose $E\left(\mathbf{X_{i,j}}\right) = \mu_j$, $E\left(\mathbf{X_{i,l}}\right) = \mu_l$, $var\left(\mathbf{X_{i,j}}\right) = \sigma_j^z$, $var\left(\mathbf{X_{i,l}}\right) = \sigma_l^z$ and $corr\left(\mathbf{X_{i,j}}, \mathbf{X_{i,l}}\right) = \rho_{jl}$.
The exact sampling distribution of ρ was first derived by Fisher (1955) who *showed that for* $\rho = 0$ the statistic $\sqrt{N-2}\rho^\wedge\sqrt{1-\rho}$ is distributed as t with (N-2) degrees of freedom. He introduced the Z-transformation,

$$\mathbf{Z} := \frac{1}{2}\left(1 + \rho^\wedge/1 - \rho^\wedge\right) \qquad \text{(B 400)}$$

and proved that even for relative small samples it is approximately Gauss-Laplace normally distributed, and that as N increases the variance of the Z-transformation *rapidly becomes nearly independent of* ρ converging to $1/(N-3)$. In 1938, *F. N. David* (1938) published *tables of the exact sampling distribution of* $\rho^\wedge$ for $\rho = 0(0.1)0.9$, $N = B(1)25, 50, 100, 200, 400$, $\rho^\wedge = -1(0.01)+1$, and investigated the accuracy of the approximate distributions suggested by R.H. Fisher. She found approximate formulas for the mean/variance of Z: They were *"extraordinarily accurate even for the low values of N provided* $|\rho|$ *is not too near unity"*. Hotdling (1953) intensively studied the *mathematical properties of the distribution function of* $\rho^\wedge$ and the *suggested improvements* of the Z-transformation in terms of a hypergeometric function. Beside discussing methods of interpolating the density he also *calculated moments of* $\rho^\wedge$. From the vast reference list of the subject we like to mention the contribution of Soper et al. (1955), Kraemer (1973), Wilcox (1997) and Green (1952) as well as Khan (1994).

Example B: Correlation coefficient, "iid" Gauss-Laplace two dimensional normal

We start from the two dimensional Gauss-Laplace normal observations (x_i, y_i) for N-independent identical distributed data ("iid'') estimated by BLUUE and BIQUUE. By a special decomposition we shift the *sample distribution* into two normal distributions relating to
$(\widehat{\mu}_1, \widehat{\mu}_2)$ *on the one* side *and* to $\left(\widehat{\sigma}_1^2\widehat{\sigma}_2^2\widehat{\sigma}_{12}\right)$ or $\left(\widehat{\sigma}_1^2\widehat{\sigma}_2^2\rho_{12}\right)$ *on the other side.*
$\widehat{\rho}_{12}$ indicates the *sample correlation coefficient* via the *characteristic function* based on *Fourier-Stieltjes integration* we derive the sampling distribution of the correlation coefficient $\widehat{\rho}_{12}$ (ρ_{12}) using *tables of David (1954).*
"N independently, identically distributed observations : iid"
"Gauss-Laplace normal, two-dimensional observations: (x_i, y_i) *for all* $i \in \{1, \ldots, N\}$*"*

"Probability density"

$$f\,(x,y)=\frac{1}{2\pi}\frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp-\frac{1}{2\,(1-\rho^2)}$$
$$\times\left[\frac{(x-\mu_1)}{\sigma 1^2}+2\frac{\rho\,(xi\mu_1)\,(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right] \qquad \text{(B 31)}$$

$$f\,(x_1,\ldots,x_N,y_1,\ldots,y_N)=\prod_{i=1}^{N} f\,(x_i,y_i)=f\,(x_1,y_1)\cdots(x_N,y_N)$$
$$=\left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)^N$$
$$\exp\left(-\frac{1}{2\,(1-\rho^2)}\sum_{i=1}^{N}\left(\frac{(x_i-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_i-\mu_1)\,(y_i-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-i-mu_2)^2}{\sigma_2^2}\right)\right)$$
$$\text{(B 32)}$$

$$BLUUE:\mu_1^{\wedge}=\frac{1}{N}\sum_{i=1}^{N}x_i=\|'\mathbf{x}/N \qquad \text{(B 33)}$$

$$BLUUE:\mu_2^{\wedge}=\frac{1}{N}\sum_{i=1}^{N}y_i=\|'\mathbf{y}/N \qquad \text{(B 34)}$$

BIQUUE:

$$\sigma_1^{\wedge 2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_1^{\wedge}\right)^2=\left(\mathbf{x}-\|\mu_1^{\wedge}\right)'\left(\mathbf{x}-\|\mu_1^{\wedge}\right)/\,(N-1) \qquad \text{(B 35)}$$

$$\sigma_2^{\wedge 2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(y_i-\mu_1^{\wedge}\right)^2=\left(\mathbf{y}-\|\mu_1^{\wedge}\right)'\left(\mathbf{y}-\|\mu_2^{\wedge}\right)/\,(N-1) \qquad \text{(B 36)}$$

$$\sigma_{12}^{\wedge}=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_1^{\wedge})(y_i-\mu_2^{\wedge})=(\mathbf{x}-\|\mu_1^{\wedge})'(\mathbf{y}-\|\mu_2^{\wedge})/(N-1) \qquad \text{(B 37)}$$

"Decomposition"

$$\sum_{i=1}^{N}(x_i-\mu_1)^2=\sum_{i=1}^{N}[(x_i-\widehat{\mu}_1)(\mu_1-\mu_1^{\wedge})]^2=(\mathbf{N}-1)\sigma_1^{\wedge 2}+\mathbf{N}(\mu_1-\mu_1^{\wedge})^2 \tag{B39}$$

$$\sum_{i=1}^{N}(y_i-\mu_2)^2=\sum_{i=1}^{N}[(y_i-\widehat{\mu}_2)(\mu_2-\mu_2^{\wedge})]^2=(\mathbf{N}-1)\sigma_2^{\wedge 2}+\mathbf{N}(\mu_2-\mu_2^{\wedge})^2 \tag{B40}$$

$$\sum_{i=1}^{N}(x_i-\mu_2)(y_i-\mu_2)=(\mathbf{N}-1)\rho_{12}^{\wedge}\sigma_1^{\wedge}\sigma_2^{\wedge}+\mathbf{N}(\mu_1-\mu_1^{\wedge})(\mu_2-\mu_2^{\wedge}) \tag{B41}$$

$$\begin{aligned}
f(x_1,\ldots,x_N,y_1,\ldots,y_N) &= \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)^N \\
&* \exp\left(-\frac{N}{2(1-\rho^2)}\left[\frac{(\mu_1-\widehat{\mu}_1)^2}{\sigma_1^2}-\frac{2\rho_{12}(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)}{\sigma_1\sigma_2}+\frac{(\mu_2-\widehat{\mu}_2)^2}{\sigma_2^2}\right]\right) \\
&* \exp\left(-\frac{N-1}{2(1-\rho_{12}^2)}\left[\frac{\widehat{\sigma}_1^2}{\sigma_1^2}-\frac{2\rho_{12}\widehat{\rho}_{12}\widehat{\sigma}_1\widehat{\sigma}_2}{\sigma_1\sigma_2}+\frac{\widehat{\sigma}_2^2}{\sigma_2^2}\right]\right)
\end{aligned} \tag{B42}$$

$\boxed{\text{“Product of two normal distributions”}}$

First part, marginal distribution

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\right)\exp\left(-\frac{N}{2(1-\rho_{12}^2)}\left[\frac{(\mu_1-\widehat{\mu}_1)^2}{\sigma_1^2}-\frac{2\rho_{12}(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)}{\sigma_1\sigma_2}+\frac{(\mu_2-\widehat{\mu}_2)^2}{\sigma_2^2}\right]d\widehat{\mu}_1 d\widehat{\mu}_2\right) \tag{B43}$$

$\boxed{\text{“Fourier-Stieltjes analysis”}}$

$$f^{\sim}(\omega)=\int(\exp i\omega x)f(x)dx \Leftrightarrow \tag{B44}$$

$$\Leftrightarrow f(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}(\exp i\omega x)f(x))f^{\wedge}(\omega)d\omega \tag{B45}$$

First part of the marginal normal distribution

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}}}\int\int_{-\infty}^{+\infty} d\widehat{\mu}_1 d\widehat{\mu}_2 \exp - \frac{N}{2(1-\rho_{12})^2}\frac{1-it_1}{\sigma_1^2}(\mu_1-\widehat{\mu}_1)^2$$
$$* \exp + \frac{N}{2(1-\rho_{12}^2)}2\rho_{12}\frac{1-it_2}{\sigma_1\sigma_2}(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)^2$$
$$* \exp - \frac{N}{2(1-\rho_{12})^2}\frac{1-it_3}{\sigma_2^2}(\mu_2-\widehat{\mu}_2)^2 \qquad \text{(B46)}$$

"abbreviations"

$$\boxed{\mu_1 := \frac{N}{2(1-\rho_{12}^2)}\frac{(\mu_1-\widehat{\mu}_1)^2}{\sigma_1^2} = \mu_1(\widehat{\mu}_1)} \qquad \text{(B47)}$$

$$\exp + it_1\mu_1 = \exp it_1 \frac{N}{2(1-\rho_{12}^2)}\frac{(\mu_1-\widehat{\mu}_1)^2}{\sigma_1^2}$$

$$\boxed{\mu_2 := \frac{N}{2(1-\rho_{12}^2)}2\rho_{12}\frac{(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)}{\sigma_1\sigma_2} = \mu_2(\widehat{\mu}_1,\widehat{\mu}_2)} \qquad \text{(B48)}$$

$$\exp + it_2\mu_2 = \exp + it_2 \frac{N}{2(1-\rho_{12}^2)}2\rho_{12}\frac{(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)}{\sigma_1\sigma_2}$$

"First part of the marginal normal distribution"

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}}\int\int_{-\infty}^{+\infty} \exp - \frac{N}{2(1-\rho_{12}^2)}\Bigg[\frac{1-i\omega_1}{\sigma_1^2}(\mu_1-\widehat{\mu}_1)^2$$
$$- \frac{2\rho_{12}(1+i\omega_2)}{\sigma_1\sigma_2}(\mu_1-\widehat{\mu}_1)(\mu_2-\widehat{\mu}_2)$$
$$+ \frac{1-i\omega_3}{\sigma_2^2}(\mu_2-\widehat{\mu}_2)^2\Bigg] d\widehat{\mu}_1 d\widehat{mu}_2$$
$$= \frac{(1-\rho_{12}^2)^{1/2}}{[(1-i\omega_1)(1-i\omega_3)-\rho_{12}^2(1+i\omega_2)^2]^{1/2}} \qquad \text{(B49)}$$

"Check yourself the integral $\iint \exp[-\alpha x^2 + \beta xy - \gamma y^2]dxdy$"

"Second part of the marginal normal distribution"

$$\frac{(1-\rho_{12}^2)^{\frac{N-1}{2}}}{[(1-it_1)(1-it_3)-\rho_{12}^2(1+it_2)]^{\frac{N-1}{2}}}\exp\left(\frac{N-1}{2(1-\rho_{12}^2)}\left[\frac{\widehat{\sigma}_1^2}{\sigma_1^2}-\frac{2\rho_{12}\widehat{\rho}_{12}\widehat{\sigma}_1\widehat{\sigma}_2}{\sigma_1\sigma_2}+\frac{\widehat{\sigma}_2^2}{\sigma_2^2}\right]\right) \tag{B50}$$

$$\begin{aligned} & f(x_1,\ldots,n_N,y_1,\ldots,y_N)dx_1\ldots dx_N dy_1\ldots dy_N \\ & \quad \Rightarrow f(x_1,y_1)\ldots f(x_N,y_N)dx_1 dy_1\ldots dx_N dy_N \\ & f(\widehat{\mu}_1,\widehat{\mu}_2)f(\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\sigma}_{12})d\widehat{\mu}_1 d\widehat{\mu}_2 d\widehat{\sigma}_1 d\widehat{\sigma}_2 d\widehat{\sigma}_{12} \end{aligned} \tag{B51}$$

$$f(x_1,\ldots,x_N,y_1,\ldots,y_N)=p_1(\widehat{\mu}_1,\widehat{\mu}_2)p_2(\widehat{\sigma}_1,\widehat{\sigma}_2,\widehat{\sigma}_{12})$$

"first"

$$\begin{aligned} & f_1^N(t_1,t_2,t_3)\leftarrow f_1(\widehat{\mu}_1,\widehat{\mu}_2) \\ & p_1^N(t_1,t_2,t_3)=\frac{(1-\rho_{12}^2)^{\frac{N-1}{2}}}{[(1-it_1)(1-it_3)-\rho_{12}^2(1+it_2)^2]} \end{aligned} \tag{B52}$$

"second"

$$p(g_1,g_2,g_3)=\frac{(1-\rho_{12}^2)^{\frac{N-1}{2}}}{(2n)^3}\iiint_{-\infty}^{+\infty}\frac{\exp(-i\omega_1\mu_1-i\omega)_2\mu_2-i\omega_3\mu_3)}{[(1-it_1)(1-it_3)-\rho_{12}^2(1+it_2)^2]} * d\omega_1 d\omega_2 d\omega_3 \tag{B53}$$

$$p(\widehat{\rho}_{12})=\frac{(1-\rho_{12}^2)^{\frac{N-1}{2}}}{\pi(N-3)!}(1-\widehat{\rho}_{12}^2)^{\frac{N-4}{2}}\frac{d^{N-2}}{d(\widehat{\rho}_{12}\rho)}^{N-2}\left(\frac{\arccos(-\rho_{12}\widehat{\sigma}_{12})}{\sqrt{1-\rho_{12}^2\widehat{\rho}_{12}^2}}\right) \tag{B.1}$$

End of Lemma Example

As a specific example, we consider the confidence interval of the confidence coefficient 0.95 on the correlation of 0.782 observed by a sample of ten. Using Graph 4 of David (1954) we find the two limits are 0.34 and 0.98 . This result leads us to the interval $0.34 < \rho_{12} < 0.98$ (Anderson 1958, p 72).

Lemma B (probability interval for the correlation coefficient, functional determinant: David 1954, pp. xxxv–xxxviii):

$$\left|\frac{\partial(x_1,\ldots,x_N,y_1,\ldots,y_N)}{\partial(\widehat{\mu}_1,\widehat{\mu}_2;\upsilon_1,\ldots,\upsilon_{N-1},\nu_1,\ldots,\nu_{N-1})}\right|=N$$

subject to the Helmert transformation

$$
\begin{aligned}
x_1 &= \widehat{\mu}_1 + \frac{1}{\sqrt{2\cdot 1}}\upsilon_1 + \frac{1}{\sqrt{3\cdot 2}}\upsilon_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\upsilon_{N-1} \\
x_2 &= \widehat{\mu}_1 - \frac{1}{\sqrt{2\cdot 1}}\upsilon_1 + \frac{1}{\sqrt{3\cdot 2}}\upsilon_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\upsilon_{N-1} \\
x_3 &= \widehat{\mu}_1 \qquad\qquad\quad - \frac{2}{\sqrt{3\cdot 2}}\upsilon_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\upsilon_{N-1} \\
&\qquad\qquad\qquad \cdots \\
x_N &= \widehat{\mu}_1 \qquad\qquad\qquad\qquad\qquad\qquad - \frac{1}{\sqrt{N(N-1)}}\upsilon_{N-1} \\
y_1 &= \widehat{\mu}_2 + \frac{1}{\sqrt{2\cdot 1}}\nu_1 + \frac{1}{\sqrt{3\cdot 2}}\nu_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\nu_{N-1} \\
y_2 &= \widehat{\mu}_2 - \frac{1}{\sqrt{2\cdot 1}}\nu_1 + \frac{1}{\sqrt{3\cdot 2}}\nu_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\nu_{N-1} \\
y_3 &= \widehat{\mu}_2 \qquad\qquad\quad - \frac{2}{\sqrt{3\cdot 2}}\nu_2 + \cdots + \frac{1}{\sqrt{N(N-1)}}\nu_{N-1} \\
&\qquad\qquad\qquad \cdots \\
y_N &= \widehat{\mu}_2 \qquad\qquad\qquad\qquad\qquad\qquad - \frac{1}{\sqrt{N(N-1)}}\nu_{N-1}
\end{aligned}
$$

$$
p(\widehat{\mu}_1, \widehat{\mu}_2, \upsilon_j, \nu_j) = p(x_i, y, i)\left|\frac{\partial(x_i, y_i)}{\partial(\widehat{\mu}_1, \widehat{\mu}_2, \upsilon_j, \nu_j)}\right|
$$

$$
\textit{for all } j = 1, 2, \ldots, N-1; \quad i = 1, 2, \ldots, N
$$

Appendix C
Statistical Notions, Random Events and Stochastic Processes

Definitions and lemmas relating to statistics, random events as well as stochastic processes are given, neglecting their proofs. *First*, we review statistical moments of a probability distribution of *random vectors* and list the *Gauss-Laplace normal distribution* by introducing the notion of a quasi-normal distribution. As the end of Appendix C1 we take reference to two lemmas about the Gauss-Laplace 3σ rule, namely the *Gauss-Laplace inequality* and the *Vysochainskii-Potunin inequality*. Appendix C2 reviews the *spatial linear error propagation* as well as the *general nonlinear error propagation* based on $y = g(x)$ introducing the moments of second, third and fourth order. The special role *Jacobi matrix* as well as the *Hesse matrix* is classified. Appendix C3 reviews useful identities like $E\{yy' \otimes j\}$ and $E\{yy' \otimes yy'\}$ as well as $\psi = E\{(y - E\{y\})(y - E\{y\})' \otimes (y - E\{y\})\}$ relating to the *matrix of obliquity* and $\Gamma := E\{(y - E\{y\})(y - E\{y\})' \otimes (y - E\{y\})(y - E\{y\})'\} + \ldots$ relating to the *matrix of courtosis*.

Appendix C.4 newly introduces *scalar-valued stochastic processes of one parameter* enriched by Appendix C.5 reviewing *characteristics* of a one parameter stochastic processes like *non-stationary, stationary* and *cryodic* stochastic processes. *Extensive simple examples* are presented in Appendix C.6 including ARIMA processes namely. Appendix C.7 introduce the important *WIENER process* of type ORNSTEIN- UHLENBECK, WIENER process with drift integrated WIENER process. The *spectral analysis* of one parameter is enriched by 14 examples like the *DIRAC representation, spectral densities, band limited white noise*. An extensive introduction into *scalar-, vector-, and tensor-valued stochastic* processes in Appendix C.9 refers to the *characteristic functional* the *moment representations of stochastic processes for scalar-valued and vector-valued qualities, for statistical homogeneous and isotropic fields of multi-point systems*. The Appendix is specified by Example C.9.1 on the topic of two-dimensional Euclidean networks under the *postulates of homogeneity and isotropy in the statistical sense*, by Example C.9.2 on criterion matrices for absolute coordinates in a space time astronomic frame of reference, for instance represented in *scalar-valued spherical harmonic*, by Example C.9.3 on *space gravity spectroscopy* illustrating the benefits of

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2,
© Springer-Verlag Berlin Heidelberg 2012

TAYLOR-KARMAN structured criterion matrices, by Example C.9.4 on *Nonlocal prediction* and Example C.9.5 on *Nonlocal Time Series Analysis*.

C-1 Moments of a Probability Distribution, the Gauss–Laplace Normal Distribution and the Quasi-Normal Distribution

First, we define the *moments of a probability distribution of a vector valued random function* and explain the notion of a *Gauss-Laplace normal distribution* and its moments. Especially we define the terminology of a *quasi - Gauss-Laplace normal distribution*.

Definition C.1. (statistical moments of a probability distribution):

(1) The expectation or *first moment of a continuous random vector* $[\mathbf{X}_i]$ for all $i = 1, \ldots, n$ of a probability density $f(x_1, \ldots x_n)$ is defined as the mean vector $\boldsymbol{\mu} := [\mu_1, \ldots, \mu_n]$ of $E\{\mathbf{X}_i\} = \boldsymbol{\mu}_i$

$$\boldsymbol{\mu}_i := E\{\mathbf{X}_i\} = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} x_i f(x_1, \ldots, x_n) dx_1 \ldots dx_n. \tag{C1}$$

The *first moment* related to the random vector $[\mathbf{e}_i] := [\mathbf{X}_i - E\{\mathbf{X}_i\}]$ is called *first central moment*

$$\pi_i := E\{\mathbf{X}_i - \boldsymbol{\mu}_i\} = E\{\mathbf{X}_i\} - \boldsymbol{\mu}_i = 0. \tag{C2}$$

(2) The *second moment of a continuous random vector* $[\mathbf{X}_i]$ for all $i = 1, \ldots, n$ of a probability density $f(x_1, \ldots, x_n)$ is the *mean matrix*

$$\boldsymbol{\mu}_{ij} := E\{\mathbf{X}_i \mathbf{X}_j\} = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \mathbf{x}_i \mathbf{x}_j f(x_1, \ldots, x_n) dx_1 \cdots dx_n. \tag{C3}$$

The *second moment* related to the random vector $[\mathbf{e}_i] := [\mathbf{X}_i - E\{\mathbf{X}_i\}]$ is called *variance - covariance matrix* or *dispersion matrix* $[\sigma_{ij}]$, especially *variance* or dispersion σ^2 for $i = j$ or *covariance* for $i \neq j$:

$$\pi_{ii} = \sigma_i^2 = V\{\mathbf{X}_i\} = D\{\mathbf{X}_i\} := E\{(\mathbf{X}_i - \boldsymbol{\mu}_i)^2\} \tag{C4}$$

$$\pi_{ij} = \sigma_{ij} = C\{\mathbf{X}_i, \mathbf{X}_j\} := E\{(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)\} \tag{C5}$$

$$D\{\mathbf{x}\} = [\sigma_{ij}] = [C\{\mathbf{X}_i, \mathbf{X}_j\}] = E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})'\}. \tag{C6}$$

$\mathbf{x} := [\mathbf{X}_1, \ldots, \mathbf{X}_n]'$ is a collection of the $n \times 1$ random vector. The random variables $\mathbf{X}_i, \mathbf{X}_j$ for $i \neq j$ are called with respect to the *central moment of second order uncorrelated* if $\sigma_{ij} = 0 \text{for} i \neq j$. (3) The *third central moment* with respect to the random vector $[\mathbf{e}_i] := [\mathbf{X}_i - E\{\mathbf{X}_i\}]$ defined by

$$\pi_{ijk} := E\{\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k\} = E\{(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)(\mathbf{X}_k - \boldsymbol{\mu}_k)\} \tag{C7}$$

contains for $i = j = k$ the *vector of obliquity* with the components

$$\boldsymbol{\Psi}_i := E\{\mathbf{e}_i^3\} \tag{C8}$$

Uncorrelation with respect to the central moment up to *third order* is defined by

$$E\{\mathbf{e}_{i_1}^{n_1}\mathbf{e}_{i_2}^{n_2}\mathbf{e}_{i_3}^{n_3}\} = \prod_{j=1}^{3} E\{\mathbf{e}_{i_j}^{n_j}\} \quad \left[\begin{array}{c} \forall 1 \le i_1 \neq i_2 \neq i_3 \le n \\ \text{and} \\ 0 \le n_1 + n_2 + n_3 \le 3. \end{array}\right. \tag{C9}$$

(4) The *fourth central moment* relative to the random vector $[\mathbf{e}_i] := [\mathbf{X}_i - E\{\mathbf{X}_i\}]$

$$\pi_{ijk\ell} := E\{\mathbf{e}_i\mathbf{e}_j\mathbf{e}_k\mathbf{e}_\ell\} = E\{(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)(\mathbf{X}_k - \boldsymbol{\mu}_k)(\mathbf{X}_\ell - \boldsymbol{\mu}_\ell)\} \tag{C10}$$

leads for $i_1 = i_2 = i_3 = i_4$ to the *vector of curtosis* with the components

$$\boldsymbol{\gamma}_i := E\{\mathbf{e}_i^4\} - 3\{\sigma_i^2\}^2 \tag{C11}$$

Uncorrelation with respect the *central moments up the fourth order* is defined by

$$E\{\mathbf{e}_{i_1}^{n_1}\mathbf{e}_{i_2}^{n_2}\mathbf{e}_{i_3}^{n_3}\mathbf{e}_{i_4}^{n_4}\} = \prod_{j=1}^{4} E\{\mathbf{e}_{i_j}^{n_j}\} \quad \left[\begin{array}{c} \forall 1 \le i_1 \neq i_2 \neq i_3 \neq i_4 \le n \\ und \\ 0 \le n_1 + n_2 + n_3 + n_4 \le 4\,. \end{array}\right. \tag{C12}$$

(5) The *central moments of the nth order* relative to the random vector $[\mathbf{e}_i] := [\mathbf{X}_i - E\{\mathbf{X}_i\}]$ are defined by

$$\pi_{i_1 \ldots i_n} := E\{\mathbf{e}_{i_1} \ldots \mathbf{e}_{i_n}\} = E\{(\mathbf{X}_{i_1} - \mu_{i_1}) \ldots (\mathbf{X}_{i_n} - \mu_{i_n})\} \tag{C13}$$

A special distribution is the *Gauss-Laplace normal distribution* of random vectors in $\mathbb{R}^n$. Note that alternative distributions *on manifolds* exist in large numbers, for instance the *von Mises distribution* on $\mathbb{S}^2$ or the *Fisher distribution* on $\mathbb{S}^3$ of Chap. 7.

Definition C.2. (Gauss-Laplace normal distribution):

An $n \times 1$ random vector $\mathbf{x} := [\mathbf{X}_1, \ldots, \mathbf{X}_n]'$ is a Gauss-Laplace normal distribution if its probability density $f(x_1, \ldots x_i)$ has the representation

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp[-\frac{1}{2}(\mathbf{x} - E\{\mathbf{x}\})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - E\{\mathbf{x}\})] \tag{C14}$$

Symbolically we can write

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{C15}$$

with the *moment of first order* or *mean vector* $\boldsymbol{\mu} := E\{\mathbf{x}\}$ and the *central moment of second order or variance – covariance matrix* $\boldsymbol{\Sigma} := E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})'\}$. The moments of the Gauss-Laplace normal distribution are given next.

Lemma C.3. (moments of the Gauss-Laplace normal distribution):

Let the $n \times 1$ random vector $\mathbf{x} := [\mathbf{X}_1, \ldots, \mathbf{X}_n]'$ follow a *Gauss-Laplace normal distribution. Then all central moments of odd order disappear and the central moments of even order are product sums of the central moments of second order exclusively*.

$$E\{\mathbf{e}_{i_1} \ldots \mathbf{e}_{i_n}\} = 0 \forall n = 2m + 1, m = 1, \ldots, \infty \tag{C16}$$

$$E\{\mathbf{e}_{i_1} \ldots \mathbf{e}_{i_n}\} = \text{fct}(\sigma^2_{i_1}, \sigma_{i_1 i_2}, \ldots, \sigma^2_{i_2}, \ldots, \sigma^2_{i_n}) \forall n = 2m, m = 1, \ldots, \infty \tag{C17}$$

$$\pi_{ij} = \sigma_{ij}, \pi_{ii} = \sigma_i^2, \tag{C18}$$

$$\pi_{ijk} = 0, \tag{C19}$$

$$\pi_{ijk\ell} = \pi_{ij}\pi_{k\ell} + \pi_{ik}\pi_{j\ell} + \pi_{i\ell}\pi_{jk}, \tag{C20}$$

$$\pi_{iijj} = \pi_{ii}\pi_{jj} + 2\pi_{ij}\pi_{ij} = \sigma_i^2\sigma_j^2 + 2\sigma_{ij}^2, \pi_{iiii} = 3(\sigma_i^2)^2 \tag{C21}$$

$$\pi_{ijk\ell m} = 0, \tag{C22}$$

$$\begin{aligned}\pi_{i_1 i_2 \ldots i_{2m-2} i_{2m-1} i_{2m}} &= \pi_{i_1 i_2 \ldots i_{2m-2}} \pi_{i_{2m-1} i_{2m}} + \pi_{i_1 i_2 \ldots i_{2m-3} i_{2m-1}} \pi_{i_{2m-2} i_{2m}} + \ldots \\ &+ \pi_{i_2 i_3 \ldots i_{2m-1}} \pi_{i_1 i_{2m}}.\end{aligned} \tag{C23}$$

The vector of obliquity $\boldsymbol{\psi} := [\psi_1, \ldots, \psi_m]'$ and the vector of curtosis $\boldsymbol{\gamma} := [\gamma_1, \ldots, \gamma_m]'$ vanish.
A *weaker assumption* compared to the *Gauss-Laplace normal distribution* is the assumption that the central moments *up to the order four* are of the form (C18)–(C21). Thus we allow for a larger class of distributions which have a similar structure compared to (C18)–(C21).

Definition C.4. (quasi-Gauss-Laplace normal distribution):

A random vector x is *quasi-Gauss-Laplace normally distributed* if it has a continuous symmetric probability distribution $f(\mathbf{x})$ which allows a representation of its central moments up to the order four of type (C18)–(C21).
Of special importance is the computation of *error bounds* for the *Gauss-Laplace normal distribution*, for instance. As an example, we have the case called the 3σ

rule which states that the probability for the random variable for **X** falling away, from its mean by more than three *standard deviations* (SDs) is at most 5%,

$$P\{|\mathbf{X}-\boldsymbol{\mu}| \geq 3\sigma\} \leq \frac{4}{81} < 0.05. \tag{C24}$$

Another example is the *Gauss-Laplace inequality, that bounds the probability for the deviation from the mode* ν.

Lemma C.5. (Gauss inequality):

The expected squared deviation from the mode $\boldsymbol{\nu}$ is

$$P\{|\mathbf{X}-\boldsymbol{\nu}| \geq r\} \leq \frac{4\tau^2}{9r^2} \text{for all} r \geq \sqrt{4/3}\,\tau \tag{C25}$$

$$P\{|\mathbf{X}-\boldsymbol{\nu}| \geq r\} \leq 1-(r/\sqrt{3}\tau) \text{for all} r \leq \sqrt{4/3}\,\tau \tag{C26}$$

subject to $\tau^2 := E\{(\mathbf{X}-\boldsymbol{\nu})^2\}$.
Alternatively, we take advantage of the Vysochanskii–Potunin inequality.

Lemma C.6. (Vysochanskii–Potunin inequality):

The expected squared deviation from an arbitrary point $\alpha \in$ is

$$P\{|\mathbf{X}-\alpha| \geq r\} \leq \frac{4\rho^2}{9r^2} \text{for all} r \geq \sqrt{8/3}\rho \tag{C27}$$

$$P\{|\mathbf{X}-\alpha| \geq r\} \leq \frac{4\rho^2}{3r^2} - \frac{1}{3} \text{for all} r \leq \sqrt{8/3}\rho \tag{C28}$$

subject to $\rho^2 := E\{(\mathbf{X}-\alpha)^2\}$.
References about the two inequalities are Gauss (1823), Pukelsheim (1994).

C-2 Error Propagation

At the beginning we note some properties of operators *"expectation* E*"* and *"dispersion* D*"*. Those derivatives can be taken from Pukelsheim (1993) and Teuniseen (1989b, 1990). Afterwards we review the special and general, in particular *nonlinear error propagation*.

Lemma C.7. (expectation operators $E\{\mathbf{X}\}$):

E is defined as a *linear operator* in the space of random variables in $\mathbb{R}^n$, also called *expectation operator*. For arbitrary constants $\alpha, \beta, \delta \in \mathbb{R}$ there holds the identity

$$E\{\alpha \mathbf{X}_i + \beta \mathbf{X}_i + \delta\}. \tag{C29}$$

Let **A** and **B** be two $m \times n$ and $m \times \ell$ matrices and $\boldsymbol{\delta}$ an $m \times 1$ vector of constants $\mathbf{x} := [\mathbf{X}_i, \ldots, \mathbf{X}_n]'$ and $\mathbf{y} := [\mathbf{Y}_i, \ldots, \mathbf{Y}_n]'$ two $n \times 1$ and $\ell \times 1$ random vectors such that

$$E\{\mathbf{AX} + \mathbf{BX} + \boldsymbol{\delta}\} = \mathbf{A}E\{\mathbf{x}\} + \mathbf{B}E\{\mathbf{y}\} + \delta \tag{C30}$$

holds. The expectation operator E is *not multiplicative* that is

$$E\{\mathbf{X}_i \mathbf{X}_j\} = E\{\mathbf{X}_i\}E\{\mathbf{X}_j\} + C\{\mathbf{X}_i, \mathbf{X}_j\} \neq E\{\mathbf{X}_i\}E\{\mathbf{X}_j\}, \tag{C31}$$

if $\mathbf{X}_i$ and $\mathbf{X}_j$ are correlated.

Lemma C.8. (special error propagation):

Let y be an $n \times 1$ dimensional random vector which depends *linear* of the $m \times 1$ dimensional random vector x by means of a constant $n \times m$ dimensional matrix A and of a constant dimensional vector $\boldsymbol{\delta}$ of the form $\mathbf{y} := \mathbf{Ax} + \boldsymbol{\delta}$. Then hold the "*error propagation law*"

$$\boxed{D\{\mathbf{y}\} = D\{\mathbf{Ax} + \delta\} = \mathbf{A}D\{\mathbf{x}\}\mathbf{A}'.} \tag{C32}$$

The dispersion function D is *not addition*, in consequence a nonlinear operator that is

$$D\{\mathbf{X}_i + \mathbf{X}_j\} = D\{\mathbf{X}_i\} + D\{\mathbf{X}_j\} + 2C\{\mathbf{X}_i, \mathbf{X}_j\} \neq D\{\mathbf{X}_i\} + D\{\mathbf{X}_j\}, \tag{C33}$$

if $\mathbf{X}_i$ and $\mathbf{X}_j$ are correlated.
The "*special error propagation law*" holds for a *linear transformation* $\mathbf{x} \to \mathbf{y} = \mathbf{Ax} + \delta$. The "*general nonlinear error propagation*" will be presented by Lemmas C.7 and C.8. The detailed proofs are taken from Chap. 8, Examples.

Corollary C.9. ("nonlinear error propagation"):

Let $y = g(x)$ be a scalar valued function between one random variable x and one random variable y. g(x) is assumed to allow a Taylor expansion around the *fixed approximation point* ξ_0 :

$$\begin{aligned} g(x) &= g(\xi_0) + g'(\xi_0)(x - \xi_0) + \tfrac{1}{2}g''(\xi_0)(x - \xi_0)^2 + \mathcal{O}(3) \\ &= \gamma_0 + \gamma_1(x - \xi_0) + \gamma_2 g(\xi_0)^2 + \mathcal{O}(3)\,. \end{aligned} \tag{C34}$$

Then the *expectation and dispersion identities* hold

$$E\{y\} = g(\mu_x) + \frac{1}{2}g''(\xi_0)\sigma_x^2 + \mathcal{O}^*(3) = g(\mu_x) + \gamma_2\sigma_x^2 + \mathcal{O}^*(3), \tag{C35}$$

$$\begin{aligned} D\{y\} = {} & \gamma_1^2\sigma_x^2 + 4\sigma_x^2[\gamma_1\gamma_2(\mu_x - \xi_0) + \gamma_2^2(\mu_x - \xi_0)^2] - \gamma_2^2\sigma_x^4 \\ & + E\{(x-\mu_x)^3\}[2\gamma_1\gamma_2 + 4\gamma_2^2(\mu_x - \xi_0)] + E\{(x-\mu_x)^4\}\gamma_2^2 + \mathcal{O}^*(3). \end{aligned} \tag{C36}$$

For the *special case* of a *fixed approximation point* $\boldsymbol{\xi}_0$ chosen to coincide with mean value $\mu_x = \xi_0$ and x being *quasi - Gauss-Laplace normal distributed* we arrive at the identities

$$E\{y\} = g(\mu_x) + \frac{1}{2}g''(\mu_x)\sigma_x^2 + \mathcal{O}^*(4), \tag{C37}$$

$$D\{y\} = [g'(\mu)]^2\sigma_x^2 + \frac{1}{2}[g''(\mu_x)]^2\sigma_x^4 + \mathcal{O}^*(3). \tag{C38}$$

The representation (C37) and (C38) characterize the *nonlinear error propagation* which is in general dependent of the central moments of the order two and higher, especially of the *obliquity* and the *curtosis*.

Lemma C.10. ("nonlinear error propagation"):

Let $\mathbf{y} = \mathbf{f}(\mathbf{x})$ be a vector-valued function between the $m \times 1$ random vector x and the $n \times 1$ random vector $\mathbf{y}$. $\mathbf{g}(\mathbf{x})$ is assumed to allow a *Taylor expansion* around the $m \times 1$ *fixed approximation vector* $\boldsymbol{\xi}_0$:

$$\boxed{\begin{aligned} \mathbf{g}(\mathbf{x}) &= \mathbf{g}(\boldsymbol{\xi}_0) + \mathbf{g}'(\boldsymbol{\xi}_0)(\mathbf{x} - \boldsymbol{\xi}_0) + \tfrac{1}{2}\mathbf{g}''(\boldsymbol{\xi}_0)[(\mathbf{x} - \boldsymbol{\xi}_0) \otimes (\mathbf{x} - \boldsymbol{\xi}_0)] + \mathcal{O}^*(3) \\ &= \boldsymbol{\gamma}_0 + \mathbf{J}(\mathbf{x} - \boldsymbol{\xi}_0) + \tfrac{1}{2}\mathbf{H}[(\mathbf{x} - \boldsymbol{\xi}_0) \otimes (\mathbf{x} - \boldsymbol{\xi}_0) + \mathcal{O}(3). \end{aligned}} \tag{C39}$$

With the $n \times m$ *Jacobi matrix* $\mathbf{J} := [\mathbf{J}_j \mathbf{g}_i(\boldsymbol{\xi}_0)]$ and the $n \times m^2$ *Hesse matrix* $\mathbf{H} := [\mathrm{vec}\mathbf{H}_1, \ldots, \mathrm{vec}\mathbf{H}_n]'$,

$$\boxed{\mathbf{H}_i := [\partial_j \partial_k \mathbf{g}_i(\boldsymbol{\xi}_0)](i = 1, \ldots, n; j, k = 1, \ldots, m)} \tag{C40}$$

there hold the following expectation and dispersion identities ("*nonlinear error propagation*")

$$E\{\mathbf{y}\} = \boldsymbol{\mu}_y = \mathbf{g}(\boldsymbol{\mu}_x) + \frac{1}{2}\mathbf{H}\mathrm{vec}\boldsymbol{\Sigma} + \mathcal{O}^*(3) \tag{C41}$$

$$\boxed{\begin{aligned}D\{\mathbf{y}\} = \boldsymbol{\Sigma}_{\mathbf{y}} = J\Sigma J' &+ \frac{1}{2}\mathbf{J}\cdot[\boldsymbol{\Sigma}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)' + (\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)'\otimes\boldsymbol{\Sigma}]\mathbf{H}' \\
&+ \frac{1}{2}\mathbf{H}[\boldsymbol{\Sigma}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0) + (\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)\otimes\boldsymbol{\Sigma}]\mathbf{J}' \\
&+ \frac{1}{4}\mathbf{H}[\boldsymbol{\Sigma}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)' + (\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)'\otimes\boldsymbol{\Sigma} \\
&+(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)\otimes\boldsymbol{\Sigma}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)' + (\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)'\otimes\boldsymbol{\Sigma}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)]\mathbf{H}' \\
&+\frac{1}{2}\mathbf{J}\cdot E\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)'\}\mathbf{H}' \\
&+\frac{1}{2}\mathbf{H}\cdot E\{(\mathbf{x}-\boldsymbol{\mu}_x)\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\}\mathbf{J}' \\
&+\frac{1}{4}\mathbf{H}\cdot[E\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)\}\cdot(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)' \\
&+(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)\cdot E\{(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\} \\
&+(\mathbf{I}\otimes(\boldsymbol{\mu}_x-\boldsymbol{\xi}_0))\cdot E\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)'\} \\
&+E\{(\mathbf{x}-\boldsymbol{\mu}_x)\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\}\cdot((\boldsymbol{\mu}_x-\boldsymbol{\xi}_0)')\otimes\mathbf{I})]\mathbf{H}' \\
&+\frac{1}{4}\mathbf{H}\cdot\{(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\} \\
&-\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})']\mathbf{H}' + \mathcal{O}^*(3).\end{aligned}} \tag{C42}$$

In the *special case* that the fixed approximations vector $\boldsymbol{\xi}_0$ coincides to the mean vector $\boldsymbol{\mu}_{\mathbf{x}} = \boldsymbol{\xi}_0$ and the random vector x is quasi-Gauss-Laplace normally distributed the following identities hold:

$$\boxed{\begin{aligned}E\{\mathbf{y}\} &= \boldsymbol{\mu}_{\mathbf{y}} = \mathbf{g}(\boldsymbol{\mu}_{\mathbf{x}}) + \tfrac{1}{2}\mathbf{H}\mathrm{vec}\boldsymbol{\Sigma} + \mathcal{O}^*(4) \qquad (C-43)\\
D\{\mathbf{y}\} &= \boldsymbol{\Sigma}_{\mathbf{y}} = J\Sigma J' - \tfrac{1}{4}\mathbf{H}\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})'\mathbf{H}' \\
&\quad+\frac{1}{4}\mathbf{H}\cdot E\{(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})'\otimes(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})'\}\cdot\mathbf{H}' + \mathcal{O}^*(3) \\
&= \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' + \frac{1}{4}\mathbf{H}[\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma} + E\{(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})'\otimes\boldsymbol{\Sigma}\otimes(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{x}})\}]\cdot\mathbf{H}' + \mathcal{O}^*(3).\end{aligned}} \tag{C44}$$

C-3 Useful Identities

Notable identities about higher order moments are the following.

Lemma C.11. (identities: higher order moments):

(a) *Kronecker-Zehfuss products* #1

$$E\{\mathbf{y}\mathbf{y}'\} = E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\} + E\{\mathbf{y}\}E\{\mathbf{y}\}' \tag{C45}$$

$$\begin{aligned} E\{\mathbf{y}\mathbf{y}'\otimes\mathbf{y}\} &= E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\otimes(\mathbf{y}-E\{\mathbf{y}\})\} \\ &\quad + E\{\mathbf{y}\mathbf{y}'\}\otimes E\{\mathbf{y}\} + E\{\mathbf{y}\otimes E\{\mathbf{y}\}'\otimes\mathbf{y}\} \\ &\quad + E\{\mathbf{y}\}\otimes E\{\mathbf{y}\mathbf{y}'\} - 2E\{\mathbf{y}\}\otimes E\{\mathbf{y}\}'\otimes E\{\mathbf{y}\} \end{aligned} \tag{C46}$$

$$\begin{aligned} E\{\mathbf{y}\mathbf{y}'\otimes\mathbf{y}\mathbf{y}'\} &= G - \boldsymbol{\Psi}\otimes E\{\mathbf{y}\}' - E\{\mathbf{y}\}'\otimes\boldsymbol{\Psi} \\ &\quad -\boldsymbol{\Psi}'\otimes E\{\mathbf{y}\} - E\{\mathbf{y}\}\otimes\boldsymbol{\Psi}' + E\{\mathbf{y}\mathbf{y}'\}\otimes E\{\mathbf{y}\mathbf{y}'\} \\ &\quad + E\{\mathbf{y}'\otimes E\{\mathbf{y}\mathbf{y}'\}\otimes\mathbf{y}\} + E\{\mathbf{y}\otimes\mathbf{y}\}E\{\mathbf{y}\otimes\mathbf{y}\}' \\ &\quad -2E\{\mathbf{y}\}E\{\mathbf{y}\}'\otimes E\{\mathbf{y}\}E\{\mathbf{y}\}'. \end{aligned} \tag{C47}$$

(b) *Kronecker products* #2

$$\boldsymbol{\Psi} := E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\otimes(\mathbf{y}-E\{\mathbf{y}\})\} \tag{C48}$$

$$\begin{aligned} G := {} & E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\otimes(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\} \\ & -E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\otimes E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\} \\ & -E\{(\mathbf{y}-E\{\mathbf{y}\})'\otimes E\{(\mathbf{y}-E\{\mathbf{y}\})(\mathbf{y}-E\{\mathbf{y}\})'\}\otimes(\mathbf{y}-E\{\mathbf{y}\})\} \\ & -E\{(\mathbf{y}-E\{\mathbf{y}\})\otimes(\mathbf{y}-E\{\mathbf{y}\})\}E\{(\mathbf{y}-E\{\mathbf{y}\})'\otimes(\mathbf{y}-E\{\mathbf{y}\})'\}. \end{aligned} \tag{C49}$$

The $n^2 \times n$ matrix $\boldsymbol{\Psi}$ contains the components of *obliquity*, the $n^2 \times n^2$ matrix G the components of *curtosis* relative to the $n \times 1$ central random vector $\mathbf{y} - E\{\mathbf{y}\}$.

(c) *Covariances between linear and quadratic forms*

$$\begin{aligned} & C\{\mathbf{F}_1\mathbf{y}+d_1, \mathbf{F}_2\mathbf{y}+d_2\} := E\{(\mathbf{F}_1\mathbf{y}+d_1-E\{\mathbf{F}_1\mathbf{y}+d_1\})(\mathbf{F}_2\mathbf{y}+d_2 \\ & \quad -E\{\mathbf{F}_2\mathbf{y}+d_2\})'\} = \mathbf{F}_1\boldsymbol{\Sigma}F_2 \end{aligned} \tag{C50}$$

(linear error propagation)

$$\begin{aligned} & \mathbf{C}\{\mathbf{F}\mathbf{y}+d, \mathbf{y}'\mathbf{H}\mathbf{y}\} := E\{(\mathbf{F}\mathbf{y}+d-E\{\mathbf{F}\mathbf{y}+d\})(\mathbf{y}'\mathbf{H}\mathbf{y}-E\{\mathbf{y}'\mathbf{H}\mathbf{y}\})\} \\ & \quad = \mathbf{F}[E\{\mathbf{y}\mathbf{y}'\otimes\mathbf{y}'\} - E\{\mathbf{y}\}E\{\mathbf{y}'\otimes\mathbf{y}'\}]\mathrm{vec}\mathbf{H} \\ & \quad = \frac{1}{2}\mathbf{F}\boldsymbol{\Psi}'\mathrm{vec}(\mathbf{H}+\mathbf{H}') + \mathbf{F}\boldsymbol{\Sigma}(\mathbf{H}+\mathbf{H}')E\{\mathbf{y}\} \end{aligned} \tag{C51}$$

$$\begin{aligned}
&C\{\mathbf{y}'\mathbf{G}\mathbf{y}, \mathbf{y}'\mathbf{H}\mathbf{y}\} := E\{(\mathbf{y}'\mathbf{G}\mathbf{y} - E\{\mathbf{y}'\mathbf{G}\mathbf{y}\})(\mathbf{y}'\mathbf{H}\mathbf{y} - \mathbf{E}\{\mathbf{y}'\mathbf{H}\mathbf{y}\})\} \\
&\quad = (\text{vec}\mathbf{G})'[E\{\mathbf{y}\mathbf{y}' \otimes \mathbf{y}\mathbf{y}'\} - E\{\mathbf{y} \otimes \mathbf{y}\}E\{\mathbf{y}' \otimes \mathbf{y}'\}]\text{vec}\mathbf{H} \\
&\quad = \frac{1}{4}[\text{vec}(\mathbf{G} + \mathbf{G}')]' G \,\text{vec}(\mathbf{H} + \mathbf{H}') \\
&\qquad -\frac{1}{2}[\text{vec}(\mathbf{G} + \mathbf{G}')]'(\boldsymbol{\Psi}' \otimes E\{\mathbf{y}\} + \boldsymbol{\Psi} \otimes E\{\mathbf{y}\}')\text{vec}(\mathbf{H} + \mathbf{H}') \\
&\qquad +\frac{1}{2}\text{tr}[(\mathbf{G} + \mathbf{G}')\boldsymbol{\Sigma}(\mathbf{H} + \mathbf{H}')\boldsymbol{\Sigma}] + E\{\mathbf{y}\}'(\mathbf{G} + \mathbf{G}')\boldsymbol{\Sigma}(\mathbf{H} + \mathbf{H}')E\{\mathbf{y}\}.
\end{aligned} \tag{C52}$$

(d) *quasi-Gauss-Laplace-normally distributed data*

$$C\{\mathbf{F}_1\mathbf{y} + \delta, \mathbf{F}_2\mathbf{y} + \delta_2\} = \mathbf{F}_1\boldsymbol{\Sigma} F_2 \tag{C53}$$

(independent from any distribution)

$$C\{\mathbf{F}\mathbf{y} + \boldsymbol{\delta}, \mathbf{y}'\mathbf{H}\mathbf{y}\} = \boldsymbol{F}\Sigma(\mathbf{H} + \mathbf{H}')E\{\mathbf{y}\} \tag{C54}$$

$$\begin{aligned}
C\{\mathbf{y}'\mathbf{G}\mathbf{y}, \mathbf{y}'\mathbf{H}\mathbf{y}\} = &\frac{1}{2}\text{tr}[(\mathbf{G} + \mathbf{G}')\boldsymbol{\Sigma}(\mathbf{H} + \mathbf{H}')\boldsymbol{\Sigma}] \\
&+E\{\mathbf{y}\}'(\mathbf{G} + \mathbf{G}')\boldsymbol{\Sigma}(\mathbf{H} + \mathbf{H}')E\{\mathbf{y}\}.
\end{aligned} \tag{C55}$$

C-4 Scalar – Valued Stochastic Processes of One Parameter

We have previously reviewed the moments of a probability distribution up to order four. The deviation was based on integration from minus infinity to plus infinity, A different concept is needed if we work with *distributions on manifolds*. For instance, we are confronted with the problem to find a distributions for data on a circle or a sphere or on a hypersphere. As mentioned before, *directional measurements*, also called angular observations or longitudinal data are *not* Gauss-Laplace distributed, but enjoy a *von Mises-Fisher distribution* in $\mathbb{S}^\gamma \subset \mathbb{R}^{\rho+1}$

Example C.4.1: p=1 (von Mises, 1918)

$$f(\Lambda|\mu,\kappa) = [2\pi I_0(\kappa)]^{-1}\exp[\kappa\cos(\Lambda - \mu_\Lambda)] \tag{C56}$$

Example C.4.2: p=2, (Fisher, 1953)

$$
\begin{aligned}
f(\Lambda, \Phi \parallel \mu_\Lambda, \mu_\Phi, \kappa) &= \frac{\kappa}{4\pi \sinh \kappa} \exp[\cos \Phi \cos \mu_\Phi \cos(\gamma - \mu_\gamma) + \sin \Phi \sin \mu_\Phi] \\
&= \frac{\kappa}{4\pi \sinh \kappa} \exp \kappa < \mu | \mathbf{X} >
\end{aligned}
\tag{C57}
$$

Here, we have introduced the *circular normal distribution* $CN(\mu, \kappa)$ parameterized by the *mean direction* $\mu(0 \leq \mu \leq 2\pi)$ and by the *concentration parameter* $\kappa(\kappa > 0)$, the reciprocal of a dispersion measure. The *spherical normal distribution* $CN(\mu_\gamma, \mu_\Phi, \kappa)$ is parameterize by the "*longitudinal direction*, lateral mean direction (μ_γ, μ_Φ)" and the concentration parameter κ, the reciprocal dispersion measure. We note that the p-spherical normal distribution is an element of the *exponential class*.

Most important examples of the *circular normal (R. von Mises)* distributions are *phase observation* within the *Global Positioning System* (*"Global Problem Solver"*) according to Cai et al.(2007) and the *spherical normal* (Fisher) distribution for *spherical harmonic analysis-synthesis* (*Fourier-Legendre functions*) describing the gravitational field within gravities, electrostatics and magnetostaties according to Fisher et al. (1981), Kent (1983), Mardia (1972), Man (2005), Mc Fadden and Jones (1981) and Roberts and Ursell (1960).

There have been made studies to consider an *ellipsoidal reference* to analyze the proper reference field of celestial bodies like the Earth, the Moon, or any other planetary body. The *ellipsoid-of-revolution* is the *equilibrium figure* of type MacLaurin which represents to first order of the *Earth*, the *Moon*, *Mars* or *other planets* even the *Jacobi triaxial reference figure* is a better approximation following, for instance Chandra Sekhar (1969). We leave the question: what is the characteristic distribution of a ellipsoid-of-revolution or a triaxial ellipsoid open!

A random event $\mathbf{X}$ is the result of a random experiment under a given condition. If these conditions change it might have an influence on the result of random experiment, namely to the probability distribution. If we consider the varying conditions on the random quantities $\mathbf{X} = \mathbf{X}(t)$, we arrive at the random effects *which depend on the deterministic parameter t*. We illustrate the new detailed definition of *one parameter stochastic process*.

Example C.4.1

We measure the temperature of a point continuously by a sensor for 3 years. Naturally we model the sensor recording by a graph shown in Fig. C.1. Actually we find a random variable which is function of the *continuous parameter* "time t", namely $\mathbf{X}(t)$. Every year we receive a temperature profile which is a deterministic function of time $t : x = x(t), 0 \leq t \leq 1$. In this form $x(t)$ is understood as *realization* of the random variable $\mathbf{X}(t)$. Neglecting the *influence of climate changes* the function $x = x(t)$ is *cyclic*. We introduce the generalized random experiment "*continuous measurement of temperature over one year*" and refer it to $\{\mathbf{X}(t) | 0 \leq t \leq t\}$. Actually one could represent climate changes by a *trend function*. The character of a generalized random experiment is very obvious since for closely measuring points t_i and t_{i+1} range in the second up to the minute.

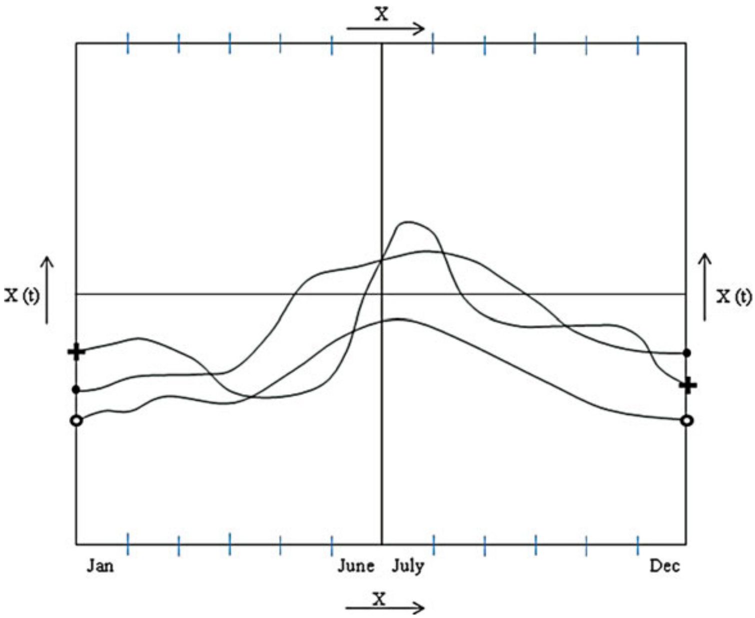


Fig. C.1 Graph of the function f(t) over a period of 1 year or 12 months. Two years recording: *trajectories of the stochastic process*

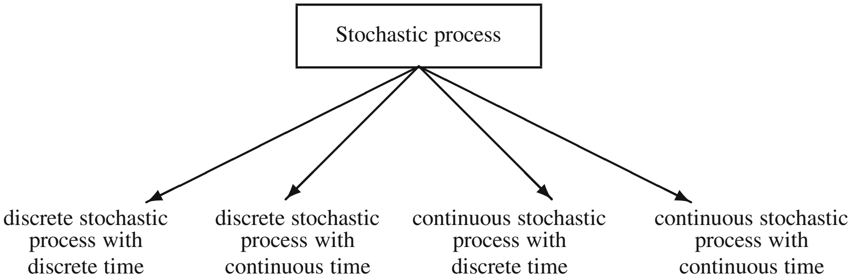
Example C.4.2

Alternatively we consider the *length* L=60 mm of *Nylon wires* and the design condition of a variance of 0.5 mm produced by *one machine* under equal conditions. If we analyzed the *dependence of the diameter of a function of the distance t* from the initial point, we receive for any measurement of the wire a function $x = x(t), 0 \leq t \leq L$. These functions differ from wire to wire, but these is no significant difference between the functional different wires. Indeed we assume changes to the constant condition of the wire. The *random variations* of the diameter are in the range of ± 0.02 mm and caused by the experimental conditions. Let us refer to the random wire diameter as a function of the *one parameter* t, we are able to define by a "*continuous measurement of the various diameter of a wire as a dependence of the initial point*" the generalized random experiment and formalize it by $\{\mathbf{X}(t), 0 \leq t \leq L\}$.
In contrast to random experiments which resulted in real numbers, such as generalized random experiment generates *random functions*, also called *stochastic process or random process*. We have called our parameter "t" derived from the term "*time*", but it can be any parameter for instance "position", "placement","*pressure*" etc. Finally we are able to define a *stochastic process of one parameter*.

Definition C.4.2: Stochastic process

A stochastic process of one parameter $t \in T$, the *parameter space*, and the state $z \in Z$, the *state space*, is defined as the set of random quantities $\{X(t), t \in T\}$.

If the set of parameter is *finite* or *measurable infinite*, we refer to a *stochastic process with discrete time*. Such processes are summarized as a sequence of random quantities $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$. If the parameter $t \in \mathbf{T}$ is interval, we refer to *stochastic process with continuous time*. A stochastic process $\{\mathbf{X}(t), t \in \mathbf{T}\}$ is called *discrete* if the *state space* z is a finite or measurable infinite. In contrast we refer to a *continuous stochastic process* if he *state space* is an interval.



If we consider various realization of $X(t)$ for all $t \in T$, we call them *trajectory* of the stochastic process. For a *discrete stochastic process* with *continuous time* the trajectories are "*step functions*"

C-5 Characteristic of One Parameter Stochastic Processes

The most important function in connection with stochastic process is the *distribution function* of $\mathbf{X}_t, t \in \mathbb{T}$, namely $F(x) = P(\mathbf{X}_t \leq x)$. But with the set of the distributions $\{F_t(x), t \in \mathbb{T}\}$ a stochastic ic process is *not* uniquely determined. A *complete characterization* of a stochastic process it is necessary to have information on all n-type $\{t_1, t_2, \cdots, t_{n-1}, t_n\}$ for the discrete set $N := \{1, 2, \cdots, n-1, n\}$ with $t_i \in \mathbb{T}$ and the information of the distribution function of random vectors $\{\mathbf{X}(t_1), \mathbf{X}(t_2), \cdots, \mathbf{X}(t_{n-1}), \mathbf{X}(t_n)\}$, namely

$F_{t_1, t_2, \cdots t_{n-1}, t_n}(x_1, x_2, \cdots, x_{n-1}, x_n) := P(\mathbf{X}(t_1) \leq x_1, \cdots, \mathbf{X}(t_n) \leq x_n)$
The set of *distribution functions* will be introduced by Table C.5.1 as well as the trend function, the covariance function, the variance function, the correlation function and the symmetry condition.

Table C.5.1 Characteristics of one parameter stochastic process

"Distribution functions of an n-dimensional random vector"

$$F_{t_1,t_2,\cdots t_{n-1},t_n}(x_1, x_2, \cdots, x_{n-1}, x_n)$$
$$:= P\{\mathbf{X}(t_1) \leq x_1, \mathbf{X}(t_2) \leq x_2, \cdots, \mathbf{X}(t_{n-1}) \leq x_{n-1}, \mathbf{X}(t_n) \leq x_n\} \quad \text{(C58)}$$

"Trend function"

$$E\{\mathbf{X}_t\} = \mu(t) \text{ for } t \in \mathbb{T} \quad \text{(C59)}$$
$$\mu(t) = \int_{-\infty}^{+\infty} x f_t(x) dx \quad \text{(C60)}$$

"Covariance function and stochastic process"

$$Cov(s, t) := Cov\{\mathbf{X}_s, \mathbf{X}_t\}$$
$$= E\{[\mathbf{X}_s - E\{\mathbf{X}_s\}][\mathbf{X}_t - E\{\mathbf{X}_t\}]\}$$
$$\text{for all } s, t \in T \quad \text{(C61)}$$

$$Cov(s, t) = E\{\mathbf{X}_s, \mathbf{X}_t\} - \mu(s)\mu(t) \quad \text{(C62)}$$

"Variance function"

$$Var\{\mathbf{X}_s\} =: Cov(t, t) \quad \text{(C63)}$$

"Correlation function"

$$\rho(s, t) =: Cov\{\mathbf{X}_s, \mathbf{X}_t\} \div (\sqrt{Var\{\mathbf{X}_s\}}\sqrt{Var\{\mathbf{X}_t\}}) \quad \text{(C64)}$$

"Symmetry"

$$Cov(s, t) =: Cov(t, s) \quad \text{(C65)}$$
$$\lim_{|t-s|\to\infty} Cov(s, t) = \lim_{|t-s|} \rho(s, t) = 0 \quad \text{(C66)}$$

End of Table C.5.2

Of special importance are those stochastic processes which do *not* depend on the absolute values of t_i, but are functions of the distances between the points. They are considered as functions of *the relative positions*.

Definition 5.1 (Stationary in the narrow sense)

A stochastic process $\{\mathbf{X}_t, t \in \mathbb{T}\}$ is *stationary in the narrow sense* if for all $n \in \{1, 2, \cdots\}$ for arbitrary subject to $t_i \in \mathbb{T}$ and $t_i + h \in \mathbb{T}$ holds

$$F_{t_1,t_2,\cdots t_{n-1},t_n}(x_1, x_2, \cdots, x_{n-1}, x_n) = F_{t_1+h,t_2+h,\cdots t_{n-1}+h,t_n+h}(x_1, x_2, \cdots, x_{n-1}, x_n) \tag{C67}$$

A *special property* is for

$$\mu(t) = E\{\mathbf{X}(t)\} = \mu(const) \tag{C68}$$

$$Var\{\mathbf{X}(t)\} = (const) \tag{C69}$$

End of Definition 5.1

Special case $n = 2$, for instance, $t_1 = 0, t_2 = t - s, h = s$ will lead us to

$$F_{0,t-s}(X_1 x_2) = F_{s,t}(x_1 x_2) = F(\tau) \text{ for } \tau := t - s \tag{C70}$$

$$Cov(s, t) = Cov(s, s + \tau) = Cov(\tau) \text{ for } \tau := t - s \tag{C71}$$

$$\text{“Symmetry”:} Cov(\tau) = Cov(-\tau) = Cov(|\tau|) \tag{C72}$$

$$\lim_{|\tau| \to \infty} Cov(\tau) = 0 \tag{C73}$$

Another generalization is achieving of the treat stochastic process ”*of higher order*”

Definition 5.2: (stationary in the wider sense):

A stochastic process of second order is stationary in the wider sense if the conditions

1. $\mu(t) = \texttt{const}$ and
2. $Cov(|\tau|) = Cov\{\mathbf{X}_s, \mathbf{X}_t\}$

End of Definition 5.2

Besides the condition of stationary another support property of a stochastic process is the set of conditions which hold for *increments of stochastic process* $\{\mathbf{X}(t), t \in \mathbb{T}\}$ in the interval $[t_1, t_2]$ for $t_1 < t_2$, *namely the random difference*

$$\mathbf{X}_{t_2} - \mathbf{X}_{t_1}$$

In general such increments may be *negative*. The concept of *incremental stochastic processes* will lead us later to the concept of *Kolmogov’s structure functions* which is very popular in the applied sciences.

Definition 5.3: (Stationary increments):

A stochastic process $\{\mathbf{X}(t), t \in \mathbb{T}\}$ is called *homogenous* or *stationary incremental*, if the increments $\{\mathbf{X}(t_2 + \tau)\} - \{\mathbf{X}(t_1 + \tau)\}$ for all τ have identical probability distributions for any fixed t_1 and t_2

End of Definition 5.3

The great advantage of the incremental stochastic process in applications has the origin in the fact *that a stationary stochastic process is not necessary.*

Definition 5.4: (Stochastic processes with incremental properties):

A stochastic process $\{\mathbf{X}(t), t \in \mathbb{T}\}$ has independent increment for all $n = 3, 4, \cdots$ and values $\{t_1, t_2, \cdots, t_{n-1}, t_n\}$ subject to the *order* $\{t_1 < t_2 < \cdots < t_{n-1} < t_n\}$ and $t_i \in \mathbb{T}$ with increments

$$\mathbf{X}_{t_2} - \mathbf{X}_{t_1}, \mathbf{X}_{t_3} - \mathbf{X}_{t_2}, \cdots, \mathbf{X}_{t_{n-1}} - \mathbf{X}_{t_{n-2}}, \mathbf{X}_{t_n} - \mathbf{X}_{t_{n-1}} \tag{C74}$$

being independent.

End of Definition 5.4

Try to consider that the point t_{n-1} is placed *in the future* the point t_n at *present time*, is general for $n > 1$ all points $\{t_1, t_2, \cdots, t_{n-1}, t_n\}$ *in the past*. The event at t_{n+1} for a Markov process depends only on the events at t_n:

Definition 5.5: (Simple Markov process):

A stochastic process $\{\mathbf{X}(t)|t \in \mathbb{T}\}$ has the property of a *Markov process* is for all $n \in \{1, 2, \cdots, \}$ if for all $n \in \{1, 2, \cdots\}(n+1)$ tuple $\{t_1, t_2, \cdots, t_{n-1}, t_n\}$ *ordered according to* $\{t_1 < t_2 < \cdots < t_{n-1} < t_n\}$ the probability relation

$$\begin{aligned} P\{\mathbf{X}_{t_{n+1}} \in \mathbb{Z}_{n+1}|\mathbf{X}(t_n) \in \mathbb{Z}_n, \mathbf{X}(t_{n-1}) \in \mathbb{Z}_{n-1}, \cdots, \mathbf{X}(t_1) \in \mathbb{Z}_1\} \\ = P\{\mathbf{X}(t_{n-1}) \in \mathbb{Z}_{n-1}|\mathbf{X}(t_n) \in \mathbb{Z}_n\} \end{aligned} \tag{C75}$$

End of Definition 5.5

Stochastic processes with the property of a *simple Markov process* is characterized by *independent increments*.

Lemma 5.6: Markov process

A *Markov process* is in narrow sense stationary if and only if the one-dimensional probability is *independent of the time* such that

$$F(x) = P\{\mathbf{X}_t \leq t\} = x \text{ for all } t \in \mathbb{T} \tag{C76}$$

End of Lemma 5.6

Markov process with a finite state space are called *Markov chains*, otherwise they are called *continuous Markov chains*.

Definition 5.7: (Continuity in the squared mean):

A stochastic process of second order $\{\mathbf{X}(t)|t \in \mathbb{T}\}$ located in the point t_0 is in the *squared mean continuous* if

$$\lim_{h \to 0} E([\mathbf{X}(t_0 + h) - \mathbf{X}(t_0)]^2) = 0 \tag{C77}$$

The stochastic process $\{\mathbf{X}(t)|t \in \mathbb{T}\}$ is in the domain $\mathbb{T}_0 = \mathbb{T}$ in the *quadratic mean continuous* if the stochastic process has this property *for all* $t \in \mathbb{T}_0$

End of Definition 5.7

Lemma 5.8: Squared mean and its covariance function

A stochastic process of second order $\{\mathbf{X}(t)|t \in \mathbb{T}\}$ is continuous in the point t_0 if its *covariance function* $Cov(s,t)$ is characterized by $(s,t) = (t_0 t_0)$ and its *trend function* $\mu(t)$ is continuous in the point $t = t_0$.

End of Lemma 5.8

C-6 Simple Examples of One Parameter Stochastic Processes

Simple examples of one parameter stochastic process are introduced divided in two types of stochastic processes, first with *continuous time* and second with *discrete time*:

(i) Process with linear realizations
(ii) Cosine oscillations with random amplitude
(iii) Cosine oscillations with random amplitude and random phase
(iv) Superposition of two uncorrelated random functions
(v) Impulse modulation
(vi) Randomly retarded pulse modulation
(vii) Random sequences with discrete signals
(viii) Random sequences of moving averages of order n: $MA(n)$
(ix) Random sequences of moving averages of order n: $M\mathbf{A}(n)$
(x) Random sequences of unlimited order
(xi) Autoregressive sequences of first order $AR(1)$
(xii) Autoregressive sequences of first order r: $AR(r)$
(xiii) ARMA (r,s) models.

We present here *simple examples of stochastic processes*. Statements with respect to stationarity relate always to *stationarity in the wide sense*.

Ex 1: Processes with Linear Scalisations

Examples of stochastic processes with continuous time

Example 6.1 Process with linear realizations

Many linear models can be characterized by the *simple linear model* $\mathbf{X}(t) = Vt + W$, V and W are *random effects* with the mean $E\{\mathbf{X}(t)|t \geq 0\}$ called *trend functions* and the variance-covariance function $E\{[\mathbf{X}(s) - E\{\mathbf{X}(s)\}][\mathbf{X}(t) - E\{\mathbf{X}(t)\}] = E\{[Vs + W][Vt + w]\} - \mu(s)\mu(t) =: \Sigma(s,t)$.
A first result is the variance-covariance function oriented to our model.

$$\boxed{\Sigma(s,t) = st\ Var(V) + (s+t)\ Cov(V,W) + Var(W)} \tag{C78}$$

A second result is the special case W = constant. In such a case we find the variance-covariance function $\Sigma(s,t) = st\ Var(V)$

End of Example 6.1

Ex 2: Cosine Oscillations with Random Amplitudes

Example 6.2 Cosine oscillation with random amplitude

Here we assume the *cosine model* $\mathbf{X}(t) = A \cos \omega t$, where A is a non-negative random effect, the amplitude subject to $E\{A\} < \infty$. Here we realize the *trend function* of oscillations of type $E\{\mathbf{X}(t)\} = E\{A\} \cos \omega t$. The variance-covariance function for such a simple model is given by $E\{[\mathbf{X}(s) - E\{\mathbf{X}(s)\}][\mathbf{X}(t) - E\{\mathbf{X}(t)\}] = E\{[A \cos \omega s][A \cos \omega t]\} - \mu(s)\mu(t) = [E\{A\}^2 - E\{A\}]^2 (\cos \omega s)(\cos \omega t) =: \Sigma(s,t)$ Let us denote the variance $\sigma^2 = Var(A)$ such that we receive the variance-covariance function $\Sigma(s,t)$, namely

$$\boxed{\Sigma(s,t) = \sigma^2 (\cos \omega s)(\cos \omega t)} \tag{C79}$$

End of Example 6.2

Ex 3: Cosine Oscillations with Random Amplitudes and Phase

Example 6.3 Cosine oscillation with random amplitude and random phase

This time we assume the cosine model $\mathbf{X}(t) = A \cos(\omega t + \Phi)$, where A is a non-negative *random effect* and finite variance. The *phase* is a *random effect*, equally distributed $\Phi \in [0, 2\pi]$, The two random effects A and Φ are independently from each other in the statistical sense. Such a stochastic process $\{\mathbf{X}(t) \in (-\infty, +\infty)\}$ can be considered as the outputs of the larger number of similar oscillators, arbitrarily chosen when applied at different time instances. Because of

$$E\{\cos(\omega t+\Phi)\} = \frac{1}{2\pi}\int_0^{2\pi}\cos(\omega t+\Phi)d\Phi$$
$$= \frac{1}{2\pi}[\sin(\omega t+\Phi)]_0^{2\pi} = 0 \tag{C80}$$

We *experience a zero trend function*. According the variance-covariance function we may represent by

$$\Sigma(s,t) = E\{[A\cos(\omega s+\Phi)][A\cos(\omega t+\Phi)]\}d\Phi$$
$$= E\{A^2\}\frac{1}{2\pi}\int_0^{2\pi}\cos(\omega s+\Phi)\cos(\omega t+\Phi)d\Phi$$
$$= E\{A^2\}\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\{\cos(t-s)\cos[\omega(s+t)+2\Phi]\}d\Phi \tag{C81}$$

The first term characterizes a *constant*, while the second term approaches zero such the resulting variance-covariance function is proportional to the difference $\tau =: t-s$

$$\boxed{\Sigma(\tau) = \Sigma(t-s) = \frac{1}{2}E\{A^2\}\cos\omega t} \tag{C82}$$

Such a process is *stationary*.

End of Example 6.3

Ex 4: Superposition of Two Uncorrelated Random Function

Example 6.4 Superposition of two uncorrelated random functions

Let A and B be *two uncorrelated random functions* characterized by $E\{A\} = E\{B\} = 0$ and $var(A) = var(B) = \sigma^2 < \infty$, The derived stochastic process is defined by $\{\mathbf{X}(t), t \in (-\infty, +\infty)\}$, namely by

$$\mathbf{X}(t) = A\cos\omega t + B\sin\omega t \tag{C83}$$

Because of $var(\mathbf{X}(t)) = \sigma^2 < \infty$ we have an example of random process of second order. Due to

$$E\{\mathbf{X}(t)\} = E\{A\}\cos\omega t + E\{B\}\sin\omega t = 0 \tag{C84}$$

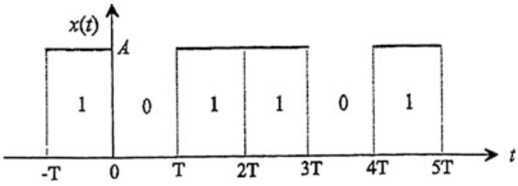


Fig. C.2 Pulse demodulation of a binary system

We are able to compute the variance-covariance function $\Sigma(s,t) = E\{\mathbf{X}(t)\mathbf{X}(s)\}$ assuming $E\{A, B\} = E\{A\}E\{B\} = 0$ due to uncorrelation

$$\begin{aligned}\Sigma(s,t) &= E\{A^2 \cos\omega s \cos\omega t + B^2 \sin\omega s \sin\omega t\} \\ &\quad + E\{AB \cos\omega s \sin\omega t + \sin\omega s \cos\omega t\} \\ &= \sigma^2(\cos\omega s \cos\omega t + \sin\omega s \sin\omega t) \\ &\quad + E\{AB\}(\cos\omega s \sin\omega t + \sin\omega s \cos\omega t) \qquad \text{(C85)}\end{aligned}$$

$$\Sigma(s,t) = \sigma^2 \cos\omega t \text{ for } \tau =: t - s \qquad \text{(C86)}$$

Again we gain a stochastic process which depends only on the difference $t - s =: \tau$ which is *stationary*.

End of Example 6.4

Ex 5: Pulse Modulation

Example 6.5 Pulse modulation

A source generates at each time interval T time units of independent from each other a sign “1” or “0” with the probability 1-p or γ. The transfer of a “1” or “0” appears in this way that T time units produce a pulse with the amplitude A. In this way they generate a random signal, namely a stochastic process $\{\mathbf{X}(t), t \in (-\infty, +\infty)\}$, with this property:

$$\mathbf{X}(t) = \begin{cases} 0 \;\; \textit{with probability} \;\; \gamma \\ A \;\; \textit{with probability} \;\; (1-\gamma). \end{cases}$$

We start from the partial sequence, for instance

$$\cdots 1\ 0\ 1\ 1\ 0\ 1 \cdots,$$

illustrated by Fig. C.2. We have chosen the initial value $t = 0$ in such a way that we start the transmission of a message.
Finally we review the *trend function* of the process as a *constant* and the *covariance function as a non stationary process*.

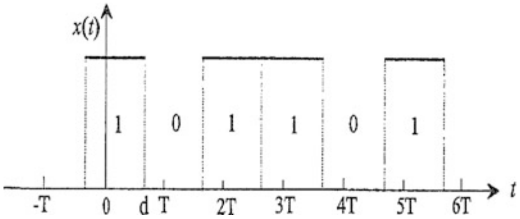


Fig. C.3 Randomly retarded pulse demodulation of a binary system

$$E\{\mathbf{X}(t)\} = AP\{\mathbf{X}(t) = A\} + OP\{\mathbf{X}(t) = 0\} = A(1-p) \quad \text{(C87)}$$

$$nT \le s, t < (n+1)T, \quad n = 0, \pm 1, \pm 2, \cdots$$

$$\begin{aligned} E\{\mathbf{X}(s), \mathbf{X}(t)\} &= E\{\mathbf{X}(s), \mathbf{X}(t)|\mathbf{X}(s) = A\}P\{\mathbf{X}(s)|\mathbf{X}(s) = A\} \\ &\quad + E\{\mathbf{X}(s), \mathbf{X}(t)|\mathbf{X}(s) = 0\}P\{\mathbf{X}(s)|\mathbf{X}(s) = 0\} \\ &= A^2(1-p) \end{aligned} \quad \text{(C88)}$$

For $m \neq n$, the process $\mathbf{X}(s)$ and $\mathbf{X}(t)$ are *independent* due to $mT \le s \le (m+1)T$ *and* $nT \le t < (n+1)T$. Accordingly the *variance-covariance* function is *not stationary*:

$$\begin{aligned} \Sigma(s,t) &= Cov(s,t) \\ &= \begin{bmatrix} A^2 p(1-p) \; \textit{for all} \;\; nT \le t < (n+1)T, \; \textit{for all} \;\; n = 0, \pm 1, \pm 2, \cdots \\ 0 \qquad\qquad\qquad \textit{otherwise}. \end{bmatrix} \end{aligned} \quad \text{(C89)}$$

End of Example 6.5

Ex 6: Random Retarded Pulse Modulation

Example 6.6: Randomly retarded pulse modulation

With regard to our previous example of a stochastic process $\{\mathbf{X}(t), t \in (-\infty, +\infty)\}$, we consider another stochastic process $\{\mathbf{Y}(t) = \{\mathbf{X}(t-1)\}$ subject to $t \in (-\infty, +\infty)$ assuming that $D \in [0, T]$ is a *random variable* of a trajectory $\{\mathbf{X}(t), t \in (-\infty, +\infty)\}$ shifted to the right producing the other stochastic process $\{\mathbf{Y}(t), t \in (-\infty, +\infty)\}$. We illustrate the transfer by a random number $\cdots 101101 \cdots$ under the constraint $D = 0$. The *trend function* of such a process is characterized by $E\{\mathbf{X}(t-D)\} = A(1-p)$ as we have seen before.
$\mathbf{X}(s)$ and $\mathbf{X}(t)$ are independent if and only if when the inequality $|t-s| > T$ holds *or* if s and t at the points $nT + D$ for $n = 0, \pm 1, \pm 2, \cdots$ are disjunct. Let us call the random event by B. The *complementary event* is denoted by $\overline{B}$. The related

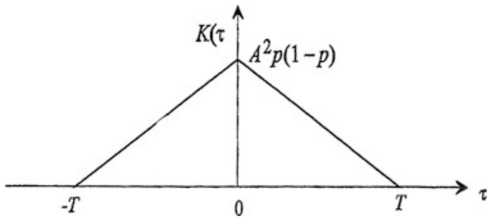


Fig. C.4 Covariance function of a randomly retarded pulse modulation

probability values are called $P(B) = (t-s)/T$ and $P(\overline{B}) = 1-(t-s)/T$. For all $|t-s| \leq T$ to variance-covariance function is represented by

$$\begin{aligned}\Sigma(s,t) &= E\{\mathbf{X}(s), \mathbf{X}(t)|B\}P\{B\} + E\{\mathbf{X}(s), \mathbf{X}(t)|\overline{B}\} - E\{\mathbf{X}(s)\}E\{\mathbf{X}(t)\} \\ &= E\{\mathbf{X}(s), \mathbf{X}(t)|B\}P\{B\} + E\{[\mathbf{X}(s)]^2\}P\{\overline{B}\} - E\{\mathbf{X}(s)\}E\{\mathbf{X}(t)\} \\ &= [A(1-p)]^2|t-s|/T + A^2(1-p)(1-|t-s|/T) - [A(1-p)]^2\end{aligned} \tag{C90}$$

$$\Sigma(\tau := t-s) = \begin{bmatrix} A^2p(1-p)(1-|\tau|/T)\ for|\tau| \leq T \\ 0 \qquad\qquad\qquad\qquad otherwise. \end{bmatrix} \tag{C91}$$

The random process of second order $\{\mathbf{Y}(t), t \in (-\infty, +\infty)\}$ is therefore stationary illustrated by its variance covariance-function of Fig. C.4

Example of stochastic processes with discrete time

Stochastic processes with discrete time scale can be understood as a *random sequence*. Here we restrict us to *stationary random sequences* as a first order model. They play a dominant role in signal transform and signal transfer.

End of Example 6.6

Ex 7: Random Sequences with Discrete Signals

Example C6.7 Random sequences with discrete signals

Let $\{\cdots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \cdots\}$ be a *random sequence of uncorrelated identical distributed random events characterized by*

$$E\{\mathbf{X}_i\} = 0, \quad Var\{\mathbf{X}_i\} = \sigma^2 < \infty, \quad i \in \{0, \pm 1, \pm 2, \cdots$$

Due to the restricted range of variances we model a *random process of second order*, its *trend function is zero*.

$$E\{\mathbf{X}_i\} = 0, \ \ for \ \ all \ \ i \in \{0, \pm 1, \pm 2, \cdots$$

$$cov(\tau) = \begin{bmatrix} \sigma^2 \, for \ \ all \ \ \tau = 0; \\ 0 \ \ for \ \ \tau \neq 0. \end{bmatrix} \tag{C92}$$

Its covariance function with odd numbers s and t is characterized by the form $Cov(s,t) = E\{\mathbf{X}_s\mathbf{X}_t\}$ such that $Cov(s,t) = E\{\mathbf{X}_s\}E\{\mathbf{X}_t\} = 0$ for $s \neq t$ and $Cov(s,t) = E\{\mathbf{X}^2\} = \sigma^2$ for $s = t$. Obviously beside the *property of stationarity* of random sequences it is important the *property of independent increments*.

End of Example 6.7

Ex 8: Random Sequences of Moving Averages of Order n : $MA(n)$

Example C.6.8 Random sequences of moving averages of order n: MA(n)

A random signal may be given by

$$\mathbf{Y}_t = \sum_{i=0}^{n} c_i \mathbf{X}_{t-1} for \, allt \in \{0, \pm 1, \pm 2, \cdots\}$$

where $\{\mathbf{X}_i\}$ is a random sequence of independent effects with parameter. The natural number n and the sequence of real restricted real numbers $c_0, c_1, \cdots, c_{n-1}, c_n$ are given. The construction of the $\mathbf{Y}_t$ are influenced by the momentary date $\mathbf{X}_t$ as well as n previous values $\{\cdots, \mathbf{X}_{-2}, \mathbf{X}_{-1}, \mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \cdots\}$. Such a relation is called *Principle of moving averages* characterized by *the process of second order* $\mathbf{X}_t, t \in \{0, \pm 1, \pm 2, \cdots\}$.

$$Var\{\mathbf{Y}_t\} = \sigma^2 \sum_{i=0}^{n} c_i c_i < \infty \ \ for \ \ all \ \ t \in \{0, \pm 1, \pm 2, \cdots\} \tag{C93}$$

"trend function"

$$E\{\mathbf{X}_t\} = 0 \ \ for \ \ all \ \ t \in \{0, \pm 1, \pm 2, \cdots\} \tag{C 93}$$

"covariance function"

$$cov(s,t) = E\{\mathbf{Y}_s\mathbf{Y}_t\} = E\left\{\left[\sum_{i=0}^{n} c_i \mathbf{X}_{s-i}\right][\sum_{j=0}^{n} c_j \mathbf{X}_{t-j}]\right\}$$

$$= E\left\{\left[\sum i = 0^n \sum_{j=0}^{n} c_i c_j \mathbf{X}_{s-i}\mathbf{X}_{t-j}\right]\right\} \tag{C94}$$

"if $|t-s| > n$, *then* $E\{\mathbf{X}_{s-i}\mathbf{X}_{t-j}\} = 0$ subject to $s-i \neq t-j$",

"if $s-i = t-j$, *then*"

$$cov(s,t) = E\{\sum_{0\le i\le n, 0\le |t-s|-i\le n}^{n} c_i c_{|t-s|+i}\mathbf{X}^2_{s-i}\}$$

$$= \sigma^2 \sum_{i=0}^{n-|t-s|} c_i c_{|t-s|+i} \tag{C95}$$

As a summary, the covariance function $cov(s,t) = cov(t)$ subject to $\tau := t-s$ is *as a sequence of moving average stationary*

$$cov(\tau) = \begin{bmatrix} \sigma^2(c_0c_{|\tau|} + c_1c_{|\tau|+1}\cdots + c_{n-|\tau|}c_n) & for\ \ 0 \le |\tau| \le n; \\ 0 & for\ \ |\sigma| > n. \end{bmatrix} \tag{C96}$$

End of Example C 6.8

Ex 9: Random Sequences with Constant Coefficients of Moving Averages of Order n : $MA(n)$

Example C.6.9 Random sequences of moving averages of order n: MA(n)

A special case if the coefficient c_i in our previous example is the setup $c = 1/(n+1)$ for $i \in \{0, 1, \cdots, n-1, n\}$. The sequence of moving average of order n can be defined by

$$\mathbf{Y}_t = \frac{1}{n+1}\sum_{i=0}^{n}\mathbf{X}_{t-i} \quad for \quad all \quad t \in \{0, \pm 1, \pm 2, \cdots\} \tag{C97}$$

$$cov(\tau) = \begin{bmatrix} \frac{\sigma^2}{n+1}\left(1 - \frac{|\tau|}{n+1}\right) & for\ \ 0 \le |\tau| \le n; \\ 0 & otherwise. \end{bmatrix} \tag{C98}$$

"MA" summarizes "*moving average*" of order n.

End of Example C 6.9

Ex 10: Random Sequences of Unlimited Order

Example C.6.10 Random sequences of unlimited order

Let be given the random sequence of unlimited order of type

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} c_i \mathbf{X}_{t-i} \quad for \quad all \quad t \in \{0, \pm 1, \pm 2, \cdots\} \tag{C99}$$

for real numbers c_i. *In order to guarantee the convergence of the series we assume*

$$\sum_{i=0}^{\infty} c_i < \infty \tag{C100}$$

its related covariance function can be represented

$$Cov(\tau, t) = \Sigma(\tau) = \sigma^2 \sum_{i=0}^{\infty} c_i c_{|\tau|+i} \quad for \quad all \quad \tau = \{0, \pm 1, \pm 2, \cdots\} \tag{C101}$$

$$Var(\mathbf{Y}_t) = \Sigma(0) = \lim_{J \to \infty} \sigma^2 \sum_{i=0}^{J} c_i \quad for \quad all \quad t = \{0, \pm 1, \pm 2, \cdots\} \tag{C102}$$

$$\mathbf{Y}_t = \Sigma(0) = \lim_{i \to -\infty, +\infty} c_i \mathbf{X}_{t-i} \quad for \quad all \quad t = \{0, \pm 1, \pm 2, \cdots\}$$

$$Cov(s, t) = \Sigma(\tau) = \sigma^2 \lim_{J \in [-\infty, +\infty]} c_i c_{|\tau|+i}$$

The result is *a two-sided sequence of moving averages*

End of Example C 6.10

Ex 11: Autoregression Sequence of First Order: $AR(1)$

Example C.6.11 Autoregressive sequences of first order: AR(1)

The starting point is the *autoregressive sequence of first order* AR(1) of type

$$\mathbf{Y}_t = a\mathbf{Y}_{t-1} + b\mathbf{X}_t \quad for \quad all \quad t = \{0, \pm 1, \pm 2, \cdots\} \tag{C103}$$

limited by $|a| < 1$ and b as well as the random sequence $\{\mathbf{X}\}$. The *instant state* $\mathbf{Y}_t$ depends only of the previous state $\mathbf{Y}_{t-1}$ as well as a random disturbance $\mathbf{X}_t$ subject to $\{\mathbf{X}\} = 0$ and $E\{(\mathbf{X}_t - E\{X_t\})^2\} = b^2\sigma^2$. Applying n times of the series we gain

$$\mathbf{Y}_t = a^n \mathbf{Y}_{t-n} + b \sum_{i=0}^{n-1} a^i \mathbf{X}_{t-i} \tag{C104}$$

$$\lim_{n\to\infty} a^n = 0 \quad \text{(C 105)} \quad \Rightarrow \quad \mathbf{Y}_t = b \sum_{i=0}^{\infty} a^i \mathbf{X}_{t-\tau} \tag{C105}$$

"Special Case"

$$c_i =: ba^i : \lim_{J\to\infty} \sum_{i=0}^{J} \left(ba^i\right)^2 = b^2 \lim_{J\to\infty} \sum_{i=0}^{J} a^{2i} = \frac{b^2}{1-a^2} \quad forJ < \infty \tag{C106}$$

For an *autoregressive sequence of first order of a stationary random process* we arrive a the *characteristic covariance function*

$$Cov(\tau) = (b\sigma)^2 \lim_{J\to\infty} \sum_{i=0}^{J} a^i a^{|\tau|+i} = a^{|\tau|}(b\sigma)^2 \lim_{J\to\infty} \sum_{i=0}^{J} a^{2i} \tag{C107}$$

$$\boxed{Cov(\tau) = \frac{(b\sigma)^2}{1-a^2} \quad a^{|\tau|} \text{ for } \tau = 0, \pm 1, \pm 2, \cdots} \tag{C108}$$

End of Example C 6.11

Ex 12: Autoregression Sequence of First Order r: $AR(r)$

Example C 6.12 autoregressive sequences of order r: AR(r)

An *autoregressive sequence of order r* called "AR(r)" is the random sequence $\mathbf{Y}_t, t = 0, \pm 1, \pm 2, \cdots$ over the set of real numbers $a_1, a_2, \cdots, a_{r-1}, a_r$ namely

$$\boxed{\mathbf{Y}_t + a_1 \mathbf{Y}_{t-1} + a_2 \mathbf{Y}_{t-2} + \cdots + a_{r-1} \mathbf{Y}_{t-(r-1)} + a_r \mathbf{Y}_{t-r} = b\mathbf{X}_t} \tag{C109}$$

The set $\mathbf{X}_i$ is a random sequence of parameters

Is the setup $\mathbf{Y}_t = \lim_{J\to\infty} \sum_{i=1}^{J} c_t \mathbf{X}_{t-1}$ *for* $\sum_{i=0}^{\infty} c_i < \infty$ *for a stationary sequence transformable in an autoregressive sequence of order r ?*

For solving this problem we have to *solve* for the relation of constants

$$\begin{aligned} c_0 &= b \\ c_1 + a_1 c_0 &= 0 \\ c_2 + a_1 c_1 + a_2 c_0 &= 0 \\ c_r + a_1 c_{r-1} + \cdots + a_r c_0 &= 0 \\ c_i + a_1 c_{i-1} + \cdots + a_r c_{i-r} &= 0 \end{aligned} \tag{C110}$$

A non-trivial solution of the equation leads us to the *algebraic equation*

$$y^r = a_1 y^{r-1} + \cdots + a_{r-1} y + a_r = 0 \tag{C111}$$

$$\boxed{\text{Example: } y^2 + a_1 y + a_2 = 0}$$

For an *autoregressive sequence of second order*, namely r=1, we study the equation of quadratic order, namely by the *eigen values* λ_1 and λ_2. *For the case* $\lambda_1 \neq \lambda_2$ we compute the *covariance function*.

$$\boxed{Cov(\tau) = Cov(0)\frac{(1-\lambda_1^2)\lambda_2^{|\tau|+1} - (1-\lambda_2^2)\lambda_1^{|\tau|+1}}{(\lambda_2 - \lambda_1)(1-\lambda_1\lambda_2)}} \tag{C112}$$

Alternatively, for the case $\lambda_1 = \lambda_2 = \lambda$ we calculate the covariance function:

$$\boxed{Cov(\tau) = Cov(0)\left(1 + \frac{1-\lambda^2}{1+\lambda^2}|\tau|\right)\lambda^{|\tau|} \quad for \quad \tau = 0, \pm 1, \pm 2, \cdots} \tag{C113}$$

subject to

$$Cov(0) = Var\{\mathbf{Y}_t\}$$

$$Cov(0) = \frac{1 + a_2}{(1-a_2)[(1+a_2)^2 - a_1^2]} \tag{C114}$$

If λ_1 and λ_2 are *complex*, then there exist λ and ω *subject to*

$$\lambda_1 = \lambda \;\; exp(i\;\omega), \;\; \lambda_2 = \lambda \;\; exp(-i\;\omega) \tag{C115}$$

$$Cov(\tau) = Cov(0)\alpha\lambda^{|\tau|}\sin(\omega|\tau| + \beta) \quad for \quad \tau = 0, \pm 1, \pm 2, \cdots \tag{C116}$$

$$\alpha := \frac{1}{\sin\beta}, \quad \beta = \arctan\left(\frac{1+\lambda^2}{1-\lambda^2}\tan\omega\right) \tag{C117}$$

Example: Numerical example

Let us introduce the *numerical example* for the *autoregressive sequence*

$$\mathbf{Y}_t - 0.6\,\mathbf{Y}_{t-1} + 0.05\,\mathbf{Y}_{t-2} = 2\mathbf{X}_t \quad for \quad t = 0, \pm 1, \pm 2, \cdots \tag{C118}$$

subject to

$$Var\{\mathbf{X}_t\} = 1 \tag{C119}$$

Obviously the *impact of the term at the time* instant t-2 on **Y** is *small*, at least *compared to term* on the time instant t-1. The related *algebraic equation of second order* can be represented by

$$\mathbf{Y}^2 - 0.6\,y + 0.05 = 0 \tag{C120}$$

Their solution is $\lambda_1 = 0.1$ and $\lambda_2 = 0.5$. Indeed compared to one, they are small than one. The solution leads to a *stationary autoregressive sequence* characterized by the *covariance functions*

$$Cov(\tau) = 7.017 * (0.5)^{|\tau|} - 1.063 * (0.1)^{|\tau|} \tag{C121}$$

subject to

$$\tau = 0, \pm 1, \pm 2, \cdots$$
$$Cov(0) = Var\{\mathbf{Y}_t\} = 5.954$$

More details can be taken from Andel (1984)

End of Example C.6.12

Ex 13: ARMA (r,s) Models

Example C.6.13 ARMA(r,s) models

As a combination of random sequences of moving averages *and* of auto regressive sequences we introduce the *ARMA(r,s) of the order*

$$\boxed{\left(\mathbf{Y}_t + a_1\mathbf{Y}_{t-1} + \cdots + a_{r-1}\mathbf{Y}_{t-(r-1)} + a_r\mathbf{Y}_{t-r} = b_0\mathbf{X}_t + b_1\mathbf{X}_{t-1} + \cdots + b_{s-1}\mathbf{X}_{t-(s-1)} + b_s\mathbf{X}_{t-s}\right)} \tag{C122}$$

In practice, ARMA model we used to model time series as a *sequence of real numbers determined by observing* the state variables in discrete time instants. Very often, time series are *not* realizations of *stationary random sequences* of type $\{\mathbf{Y}_t, t = 0, \pm 1, \pm 2, \cdots\}$ with *constant trend functions* $E\{\mathbf{Y}_t\} = \mu(t)$, but the *transformation*

$$\mathbf{Z}(t) = E\{\{\mathbf{Y}(t)\} - E\{\mathbf{Y}(t)\}\} \tag{C123}$$

reduced mean

$$E\{\mathbf{Z}_i\} = E[\{\mathbf{Y}_t\} - E\{\mathbf{Y}_t\}] = 0 \tag{C124}$$

reduced covariance matrix

$$D\{\mathbf{Z}\} = E[\{\mathbf{Y}_{t_1}\} - E\{\mathbf{Y}_{t_1}\}] = [\mathbf{Y}_{t_2} E\{\mathbf{Y}_{t_2}\}] = cov(t_1, t_2) \tag{C125}$$

leads *approximately* to proper fit of ARMA models. They are called *trend reduced* moving average interregressive sequences being *"nearly stationary"*. *Very efficient software exist nowadays.*

End of Example C.6.13: ARMA(r,s)

C-7 Wiener Processes

First, we define a *Wiener Process* by three axioms, namely by *stationary increments* of a random function, together with path continuity. The relation to a *random walk* is outlined. Especially the probability density of an *ordered n-dimensional Wiener Process* is reviewed. Secondly, special Wiener A Process of type (a) Ornstein-Uhlenbeck (b) WIENER Process with drift and (c) integrated WIENER Process are introduced

C-71 Definition of the Wiener Processes

The English botanician *R. Brown* published in the year 1828 the *random movement* of microscopic small organic and inorganic particles suspended in liquids illustrated in Fig. C.3. We observe *a totally irregular motion of particles*, nowadays called *Brownian motion*. Its first analytical treatment, we owe Bachelier (1900) and

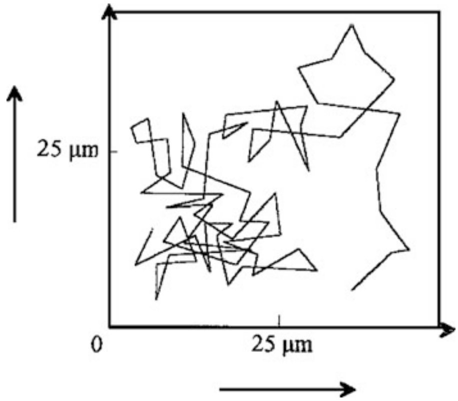


Fig. C.5 Trajectory of a two dimensional Brownian motion of 30 s distance (Perrin 1916)

Einstein (1905). Both found that a *two-dimensional Gauss-Laplace distribution* can model the *typical Brownian motion*. They spoke a *chaotic movement of microscopic small particles*. Wiener (1923) built up his theory of *stochastic process with independent increments* to model this *Brownian motion* we will define first.

Definition C.7.1: Wiener Process

A stochastic process $\{\mathbf{X}(t), t \geq 0\}$ with continuous time and the state space $\mathbf{Z} = (-\infty, +\infty)$ is called Wiener Process when it can be described as a path-continuous process by the following axioms.

(i) $\mathbf{X}(0) = 0$ (C 126)
(ii) $\{\mathbf{X}(t), t \geq 0\}$ has *stationary increments* (C 127)
(iii) for all $t > 0$, $\mathbf{X}(t)$ is *Gauss-Laplace normally distributed* (C 128)

End of Definition C.7.1: Wiener

Because of the *stationarity of the increments*, the difference $\mathbf{X}(t) - \mathbf{X}(s)$ for all $s, t \geq 0$ are *Gauss-Laplace normally distributed* with expectation $E\{\mathbf{X}(t) - \mathbf{X}(s)\} = 0$ and variance $\sigma^2|t - s|$. Since the increments are independent the increments are a *Markov-Wiener process. If* $\sigma = 1$, *then* we call the process *"standard Wiener"*.

Continuity and differentiability

$$E\{|\mathbf{X}(t) - \mathbf{X}(s)|^2\} = Var\{\mathbf{X}(t) - \mathbf{X}(t)\} = \sigma^2|t - s| \tag{C129}$$

$$\lim_{h\to 0} E\{|\mathbf{X}(t + h) - \mathbf{X}(t)|^2\} = \lim_{h\to} \sigma^2|h| = 0 \tag{C130}$$

"In the mean square sense the Wiener Process is continuous"

Wiener Process and the random walk

There is a intimate contact of *Wiener Process* and *"a random walk"* of a particle. The *random walk* of a particle can be described by

$$\mathbf{X}(t) = (\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_{[t/\Delta t]})\Delta x \tag{C131}$$

"In the time interval starting by $x = 0$, a particle moves in the time difference Δt by Δx as a length unit to the left and the right *with probability* 1/2".

$$\mathbf{X}(t) = \left[\begin{array}{l} +1 \text{ if the jump is to the right} \\ -1 \text{ if the jump is to the left} \end{array}\right. \tag{C132}$$

$\boxed{\text{"}t/\Delta t\text{" is the largest even number smaller or equal "}t/\Delta t\text{"}}$

$$P(\mathbf{X}_t = 1) = P(\mathbf{X}_t = -1) = 1/2 \tag{C133}$$

$E\{\mathbf{X}_i\} = 0$ and $Var\{\mathbf{X}_i\} = 1 \Rightarrow$

$\Rightarrow E\{\mathbf{X}(t)\} = 0$ and $Var\{\mathbf{X}(t)\} = (\Delta x)^2[t/\Delta t]$

$$\boxed{\Delta t \to 0 \Rightarrow E\{\mathbf{X}(t)\} = 0,\ Var\{\mathbf{X}(t)\} = \sigma^2 t}$$

$\boxed{\text{Multidimensional distribution and conditional probability}}$

Let $\{\mathbf{X}(t)\}$ for $t \geq 0$ be a *Wiener Process*. We ask the question what is distribution of the sum of independent Gauss-Laplace normally distributed incremental qualities of type

$$\mathbf{X}(t_i) = \mathbf{X}(t_1) + (\mathbf{X}(t_2)\mathbf{X}(t_1)) + \cdots + (\mathbf{X}(t_i)\mathbf{X}(t_{i-1}))$$

subject to

$$i = 2, 3, \cdots, n-1, n$$

The result will lead us to the *Special Wiener Process* of type *Gauss Process*

Lemma C.7.2: Probability density of an n-dimensional Wiener Process

Let us assume $0 < t_1 < t_2 < \cdots < t_{n-1} < t_n$ for an ordered n-dimensional *Wiener Process* of the random vectors $\{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}(t_{n-1}), \mathbf{X}(t_n)\}$ characterized by *independent Gauss-Laplace normally distributed increments of a Wiener Process*. Then the *probability density* has the property

$$f_{t_1,t_2,\cdots,t_{n-1},t_n}(\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}(t_{n-1}),\mathbf{X}(t_n))$$
$$= f_{t_1}(\mathbf{X}_1)f_{t_2-t_1}(\mathbf{X}_2-\mathbf{X}_1)\cdots f_{t_{n-1}-t_{n-2}}(\mathbf{X}_{n-1}-\mathbf{X}_{n-2})f_{t_n-t_{n-1}}(\mathbf{X}_n-\mathbf{X}_{n-1}) \tag{C134}$$

and has the structure

$$f_{t_1,t_2,\cdots,t_{n-1},t_n}(\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}(t_{n-1}),\mathbf{X}(t_n))$$
$$= \exp\{-\frac{1}{2}[\frac{x_1^2}{t_1}+\frac{(x_2-x_1)^2}{t_2-t_1}+\cdots+\frac{(x_{n-1}-x_{n-2})^2}{t_{n-1}-t_{n-2}}+\frac{(x_n-x_{n-1})^2}{t_n-t_{n-1}}]\}$$
$$\div(2\pi)^{n/2}\sigma^n\sqrt{t_1(t_2-t_1)\cdots(t_{n-1}-t_{n-2})(t_n-t_{n-1})} \tag{C135}$$

End of Lemma C.7.2: n-dimensional Wiener Process

Proof of Lemma 7.2

For the proof we depart from the one dimensional *probability density* of $\{\mathbf{X}_t\}$ of type *Wiener Process*.

First assumption

$$f_t(x) = -\exp[-x^2/(2\sigma^2 t)] \div \sqrt{2\pi t}\sigma \tag{C136}$$

Second assumption

Let $f_{s,t}(x_1x_2)$ the probability density of the random vector $\{\mathbf{X}(s)\mathbf{X}(t)\}$ subject to the order $0 < s < t$ characterized by

$$f_{s,t}(x_1x_2)dx_1dx_2 = P(\mathbf{X}(s) = x_1, \mathbf{X}(t) = x_2)dx_1dx_2 \tag{C137}$$
$$P(\mathbf{X}(s) = x_1, \mathbf{X}(t) = x_2) = P(\mathbf{X}(s) = x_1, \mathbf{X}(t) - \mathbf{X}(s) = x_2 - x_1)dx_1dx_2 \tag{C138}$$
$$f_{s,t}(x_1x_2)dx_1dx_2 = P(\mathbf{X}(s) = x_1)P(\mathbf{X}(t) - \mathbf{X}(s) = x_2 - x_1)dx_1dx_2 \tag{C139}$$

"independent increments"

$$f_{s,t}(x_1x_2) = f_s(x_1)f_t(x_2-x_1) \tag{C140}$$

Third assumption

$$f_{s,t}(x_1x_2) = exp\{-[tx_1^2 - 2sx_1x_2 + sx_1^2]\} \div 2\pi\sigma^2\sqrt{s(t-s)} \tag{C141}$$

"*correlation coefficient, covariance function*"

$$\rho s, t = +\sqrt{s/t} \tag{C142}$$

$$Cov\{\mathbf{X}(s), \mathbf{X}(t)\} = \sigma^2 s \text{ for } 0 < s \leq t$$

$$Cov\{\mathbf{X}(s), \mathbf{X}(t) - \mathbf{X}(s)\} = 0 \tag{C143}$$

$$\begin{aligned} Cov\{\mathbf{X}(s), \mathbf{X}(t)\} &= Cov\{\mathbf{X}(s), \mathbf{X}(s) + \mathbf{X}(t) - \mathbf{X}(s)\} \\ &= Cov\{\mathbf{X}(s), \mathbf{X}(s)\} + Cov\{\mathbf{X}(s), \mathbf{X}(t) - \mathbf{X}(s)\} \\ &= Cov\{\mathbf{X}(s), \mathbf{X}(s)\} = Var\{\mathbf{X}(s)\} = \sigma^2 \end{aligned} \tag{C144}$$

"*conditional probability density*"

Condition: $\mathbf{X}(t) = b$

$$f_{\mathbf{X}(s)}(x|\mathbf{X}(t) = b) = f_{s,t}(x, t) \div f_t(b) \tag{C145}$$

$$\boxed{f_{\mathbf{X}(s)}(x|\mathbf{X}(t) = b) = exp\{-(x - \frac{s}{t}b)^2 \div \sqrt{2\pi\frac{s}{t}(t-s)\sigma}\}} \tag{C146}$$

$$E\{\mathbf{X}(s)|\mathbf{X}(t) = b\} = \frac{s}{t}b, \quad Var\{\mathbf{X}(s)|\mathbf{X}(t) = b\} = \sigma^2\frac{s}{t}(t-s) \tag{C147}$$

$$\max Var\{\mathbf{X}(s)|\mathbf{X}(t) = b\} \Leftrightarrow s = t/2 \tag{C148}$$

Finally, a *generalization* from $f_t(t)$, $f_{s,t}(s, t)$ to $f_{t_1 t_2 \cdots t_{n-1} t_n}$ leads us to *Lemma C 7.2*

End of proof for Lemma C.7.2

At this end we will define the *Gauss process*.

Definition C.7.3: Gauss process

A stochastic process $\mathbf{X}(t), t \in \mathbb{T}$ is a *Gauss Process* of type (C135), if for arbitrary vectors $(t_1, t_2, \cdots, t_{n-1}, t_n)$ ordered as $t_1 < t_2 < \cdots < t_{n-1} < t_n$ and $n = 1, 2, \cdots$ are random vectors $\{\mathbf{X}(t_1), \mathbf{X}(t_2), \cdots, \mathbf{X}(t_{n-1}), \mathbf{X}(t_n)\}$ of an n-dimensional *Gauss-Laplace normal distribution*.

End of Definition C.7.3: Gauss process

C-72 Special Wiener Processes: Ornstein–Uhlenbeck, Wiener Processes with Drift, Integral Wiener Processes

Here we only introduce special *Wiener Process* of type

(i) Ornstein-Uhlenbeck
(ii) *Wiener Process with drift* and
(iii) *integrated Wiener Process*

It has to be remembered that the trajectories of a *Wiener Process* are *nowhere differentiable*. Particles in motion have an infinite velocity when modeled a *Wiener Process*. In order to overcome this difficulty, *Ornstein and Uhlenbeck* developed a stochastic model for modeling the velocity of particles, an important issue in *Continuum mechanics*.

Definition 7.3: Ornstein-Uhlenbeck Process

Let $\{\mathbf{X}(t), t \geq 0\}$ be a *Wiener Process* with the parameter σ. Then we call

$$V(t) = [exp(-\alpha t)]\mathbf{X}(exp(\alpha t)) \tag{C149}$$

subject to $\alpha > 0$, a stochastic process $\{V(t), -\infty < t < +\infty\}$ an *Ornstein-Uhlenbeck Process*.

End of Definition: Ornstein-Uhlenbeck Process

What are the characteristic properties of the Ornstein-Uhlenbeck Process?

Property 1: Probability density

$$f_{V(t)}(x) = [exp(-x^2/2\sigma^2)] \div (\sqrt{2\pi}\sigma) \tag{C150}$$

Property 1: Expectations

$$E\{V(t)\} = 0, \quad Var\{V(t)\} = \sigma^2 \tag{C151}$$

$$Cov(s,t) = \sigma^2[exp(-2\alpha(t-s))] \text{ for } s \leq t \tag{C152}$$

Proof for the covariance function

$$\begin{aligned} Cov(s,t) &= Cov(V(s), V(t)) = E\{V(s)V(t)\} \\ &= exp[-\alpha(s+t)]E\{\mathbf{X}(exp2\alpha s)\mathbf{X}(exp2\alpha t)\} \\ &= exp[-\alpha(s+t)]Cov(\mathbf{X}(exp2\alpha s)\mathbf{X}(exp2\alpha t)) \\ &= exp[-\alpha(s+t)]\sigma^2[exp(2\alpha(t-s))] \end{aligned} \tag{C153}$$

It has to be mentioned that the *Ornstein-Uhlenbeck Process* is *stationary in the wider sense*. As a *Gauss Process* it is also *stationary in the narrow sense*. In summary, the *Ornstein-Uhlenbeck Process* is an *instationary Wiener Process* which is stationary process due to the *action of time transformation and normalization*. In comparison to the *Wiener Process*, the *Ornstein-Uhlenbeck* has the following differences

1. The *Ornstein-Uhlenbeck Process* has *no* independent increments
2. The trajectories of an *Ornstein-Uhlenbeck Process* are *in the quadratic mean differentiable*

Wiener Process with drift

Another example for the transformation of an *instationary process to a stationary process* is the *Wiener Process with drift*.

Definition 7.4: Wiener Process with drift

A stochastic process $\{\mathbf{W}(t), t \geq 0\}$ with independent increments is called a *Wiener Process with drift*, if the following properties hold:

(i)
$$W(0) = 0 \tag{C154}$$

(ii) Each increment has the property $\mathbf{W}(t) - W(s)$ is Gauss-Laplace normally distributed with second order statistics

(iii)
$$E\{\mathbf{W}(t)\} = \mu(t - s) \tag{C155}$$

(iv)
$$Var\{\mathbf{W}(t)\} = \sigma^2|t - s| \tag{C156}$$

End of Definition 7.4: drifted Wiener Process

An equivalent formulation is next:

"$\{\mathbf{W}(t), t \geq 0\}$ is a Wiener Process with drift, if it has the structure

$$\mathbf{W}(t) = \mu t + \mathbf{X}(t) \tag{C157}$$

where $\{\mathbf{X}(t), t \geq 0\}$ is a Wiener Process subject to $\sigma^2 = Var\{\mathbf{X}(1)\}$ and μ an arbitrary constant, called as drift parameter".

Obviously, a *Wiener Process with drift* is generated by the *super position* of type "additive" of a *Wiener Process* with a *deterministically growing decreasing part*. The deterministic part is a *trend function* of type $\mu(t) = \mu t$.
The one dimension probability density distribution of a "Wiener Process with drift" can be characterized by

$$f_{\mathbf{W}(t)}(x) = exp[-(x - \mu t)^2/(2\sigma)] \div [\sqrt{2\pi t}\sigma], \quad -\infty, x, +\infty \tag{C158}$$

In practice, drift parameters are daily experience. We restrict ourselves with only one example.

Example: Geometric Wiener Process with drift

Let us introduce a stochastic process of type $\{\mathbf{Z}(t), t \geq 0\}$ characterized by

$$\mathbf{Z}(t) = exp\mathbf{W}(t) \tag{C159}$$

called "*geometric Wiener Process with drift*". Since the model $\mathbf{W}(t)$ has by definition a *Gauss-Laplace normal distribution* the expectation of $(\mathbf{Z}(t))$ equals the *moment generating function* of $\mathbf{W}(t)$. For instance,

$$E\{exp\mathbf{W}(t)\} = exp\{st(\mu + 1/2\sigma^2 s)\} \tag{C160}$$

$$s = 1 :\Rightarrow E\{\mathbf{Z}(t)\} = exp[t(\mu + \sigma^2/2)] \tag{C161}$$

$s = 2:$

$$\Rightarrow E\{\mathbf{Z}^2(t)\} = exp[2t(\mu + \sigma^2)] \tag{C162}$$

$$\Rightarrow Var\{\mathbf{Z}(t)\} = exp[t(2\mu + \sigma^2)][exp(t\sigma^2) - 1] \tag{C163}$$

Integrated Wiener Process

By transforming of a Wiener Process we receive a stochastic process which is very important in the analysis. We will summarize these transformations in three cases in our

Lemma 7.5: Elementary transformations os a Wiener Process

Let $\{\mathbf{X}(t), t \geq 0\}$ be a *standard Wiener Process* such that the following stochastic processes are as well standard Wiener Processes:

(i)

$$\{\mathbf{U}(t), t \geq 0\} \text{ subject to } \mathbf{U}(t) = c\mathbf{X}(t/c^2), c > 0 \tag{C164}$$

(ii)

$$\{\mathbf{V}(t), t \geq 0\} \text{ subject to } \mathbf{V}(t) = \mathbf{X}(t+h) - \mathbf{X}(h), h > 0 \tag{C165}$$

(iii) $\{\mathbf{V}(t), t \geq 0\}$ subject to

$$\mathbf{W}(t) = \begin{bmatrix} \mathbf{X}(t) & \text{for all } t > 0 \\ 0 & \text{for all } t > 0 \end{bmatrix} \tag{C166}$$

End of Lemma 7.5

We skip the proof. Instead let $\{\mathbf{X}(t), t \geq 0\}$ be a Wiener Process: Accordingly "near all" trajectories $x = x(t)$ are continuous in the squared mean. In consequence, there exist the integrals $u(t) = \int_0^t x(y)dy$ for "nearly all" trajectories $x = x(t)$. These integrals are *realizations of the random integral.*

$$\boxed{u(t) = \int_0^t x(y)dy} \tag{C167}$$

$u(t)$ is the symbol for the integral acting on "nearly all" trajectories of a *Wiener Process* subject to the probability distribution $\{\mathbf{U}(t), t \geq 0\}$ defines a stochastic process called *integrated Wiener Process*.

Stochastic integration

Assume the Riemenn integration for $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = 1$ subject to $\Delta t_i = t_i - t_{i-1}$ for forward computations and $i = 1, 2, \cdots, n-1, n$

$$u(t) = \lim_{n \to \infty, \Delta t_i \to 0} \{\sum_{i=1}^{n} [\mathbf{X}(t_i) - \mathbf{X}(t_{i-1})] \Delta t_i \tag{C168}$$

$u(t)$ is the lines of a sum of *independent Gauss-Laplace normally distributed* random variables as well *Gauss-Laplace normally distributed*: An *integrated Wiener Process* is being analyzed by

$\boxed{(a)\ \text{Trend function:}}$

$$E\{\int_0^t \mathbf{X}(y)dy\} = \int_0^t E\{\mathbf{X}(y)\}dy = E(t) = 0 \tag{C169}$$

$\boxed{(b)\ \text{Covariance function:}}$

$$\begin{aligned} Cov\{u(s), u(t)\} &= E\{\int_0^s \mathbf{X}(z)dz \int_0^t \mathbf{X}(y)dy\} = \int_0^s \int_0^t E\{\mathbf{X}(z)\mathbf{X}(y)\}dydz \\ &= \sigma^2 \int_0^s \int_0^t \min(y, z)dydz = \sigma^2 \int_0^s \int_0^y zdzdy + \sigma^2 \int_0^s \int_y^t y\, dz\, dy \\ &= \sigma^2 \int_0^s \frac{1}{z} y^2 dy + \sigma^2 \int_0^s y(t-y)dy \end{aligned} \tag{C170}$$

$$Cov\{u(s), u(t)\} = \frac{\sigma^2}{6}(3t - s)s^2 \text{ for } s \leq t \tag{C171}$$

$$Var\{u(t)\} = \frac{\sigma^2}{3}t^3 \text{ for } s = t \tag{C172}$$

An integrated wiener Process is not stationary, in general. But it can be proven that the *stochastic process* $\mathbf{V}(t) := u(t+\tau) - u(t)$ for $\mathbf{V}(t) \in \{\mathbf{V}(t), t \geq 0\}$ is stationary for all $\tau > 0$.

White noise

A stochastic process, $\mathbf{Z}(t) \in \{\mathbf{Z}(t), t \geq 0\}$ subject to

$$\mathbf{Z}(t) := d\mathbf{X}(t) \div dt = \mathbf{X}(t) \tag{C173}$$

or

$$d\mathbf{X}(t) := \mathbf{Z}(t)dt \tag{C174}$$

cannot be computed by standard differentiation: We have to introduce *stochastic integration, for instance*

$$\int_a^b f(t)d\mathbf{X}(t) = \lim_{n \to \infty, \Delta t_i \to 0} \{\sum_{i=1}^{n} f(t_{i-1})[\mathbf{X}(t_i) - \mathbf{X}(t_{i-1})]\} \tag{C175}$$

$$\int_a^b f(t)d\mathbf{X}(t) = f(b)\mathbf{X}(b) - f(a)\mathbf{X}(a) - \int_a^b \mathbf{X}(t)df(t) \tag{C176}$$

The representation reminds us to "*partial integration*". As the limit of the sum of *independent Gauss-Laplace normally distributed random variables* we have to compute the *expectation* of type

$$E\{\int_a^b f(t)d\mathbf{X}(t)\} = 0 \tag{C177}$$

$$
\begin{aligned}
&Var\{\sum_{i=1}^{n} f(t_{i-1})[\mathbf{X}(t_i) - \mathbf{X}(t_{i-1})]\\
&= \sum_{i=1}^{n} f^2(t_{i-1}) Var\{\mathbf{X}(t_i) - \mathbf{X}(t_{i-1})\}\\
&= \sigma^2 \sum_{i=1}^{n} f^2(t_{i-1})(t_i - t_{i-1}) = \sigma^2 \sum_{i=1}^{n} f^2(t_{i-1}) \Delta t_i
\end{aligned}
\tag{C178}
$$

The limit approaches the variance of the stochastic integral of type

$$
\boxed{Var\{\int_a^b f(t) d\mathbf{X}(t)\} = \sigma^2 \int_a^b f^2(t) dt}
\tag{C179}
$$

This result leads to the definition of *White Noise*

Definition C.7.6: White Noise

A stochastic process $\{\mathbf{Z}(t), t \geq 0\}$ is called "*White Noise*" if for any arbitrary interval $[a, b]$ of continuously differentiable functions $f(t)$, namely

$$
\int_a^b f(t) d\mathbf{Z}(t) = f(b)\mathbf{X}(b) - f(a)\mathbf{X}(a) - \int_a^b \mathbf{X}(t) df(t)
\tag{C180}
$$

where $\{\mathbf{X}(t), t \geq 0\}$ is a Wiener Process.

End of Definition C.7.6: White Noise

?Why we call stochastic process "*White Noise*"?

The origin of the question is the definition of the "*generalized derivative*". To this end we define "*the Dirac Delta function*" as a "*generalized function*"

$$
\delta(t) := \lim_{h \to 0} \begin{bmatrix} 1/h & \text{for } -h/2 \leq t \leq +h/2 \\ 0 & \text{otherwise} \end{bmatrix}
\tag{C181}
$$

or

$$
\delta(t) := \begin{bmatrix} \infty & \text{for } t = 0 \\ 0 & \text{otherwise} \end{bmatrix}
\tag{C182}
$$

$\boxed{\text{Properties of Dirac's Delta function}}$

(i)

$$\int_{-\infty}^{+\infty} f(t)\delta(t - t_0)dt = f(t_0) \tag{C183}$$

(ii) *Heaviside functions*

$$H(t) := \begin{bmatrix} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{bmatrix} \tag{C184}$$

"*The Dirac's Delta function is the formal derivative of the Heavide function*"

$$\delta(t) = \partial H(t)/\partial t \tag{C185}$$

(iii) *Covariance function*

$$\begin{aligned} Cov\{\mathbf{Z}(s), \mathbf{Z}(t)\} &= Cov[\partial \mathbf{X}(s)/\partial s, \partial \mathbf{X}(t)/\partial t \\ &= \frac{\partial}{\partial s}\frac{\partial}{\partial t} Cov\{\mathbf{X}(s), \mathbf{X}(t)\} = \frac{\partial}{\partial s}\frac{\partial}{\partial t} \min(s, t) \\ &= \frac{\partial}{\partial s} H(s - t) \end{aligned} \tag{C186}$$

$$\boxed{Cov\{\mathbf{Z}(s), \mathbf{Z}(t)\} = \delta(s - t)} \tag{C187}$$

the covariance function is a measure of the statistically independence between $\mathbf{Z}(s)$ and $\mathbf{Z}(t)$, have a function of the small distances between the point s and the point t, namely $|s - t|$.

"*How can we illustrate the Dirac function?*"
Let $\{N(t), t \geq 0\}$ be an *arbitrary numbering process* following our Definition C.7.6. The instants of the process at the points $T_1, T_2, \cdots$, admit the representation in terms of the *Heavide function* $N'(t)$ in terms of the *Dirac function*:

$$N(t) = \sum_{i=1}^{\infty} H(t - T_i), \qquad \partial N(t)/\partial t = \sum_{i=1}^{\infty} \delta(t - T_i) \tag{C188}$$

Figure C.6 an illustration of the *numbering process*.
The theory of stochastic integration has led us to the concept of generalized function, namely the Dirac function and Heaviside function, and to the notion of *White Noise*. More general notions like the concept of *Colored Noise* and finally to the notion of its stochastic processes.

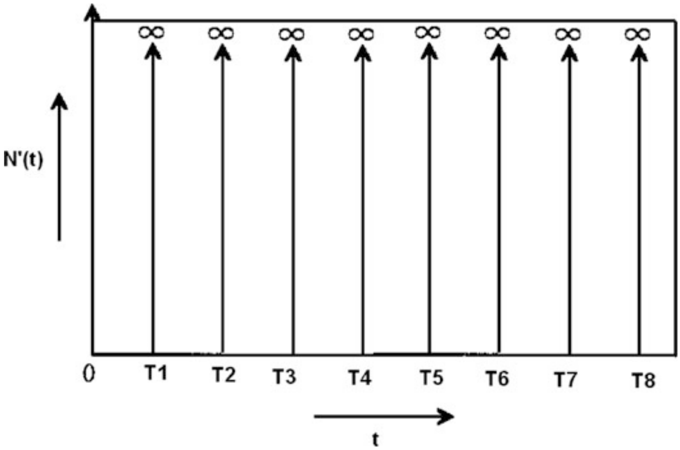


Fig. C.6 Dirac numbering system

C-8 Special Analysis of One Parameter Stationary Stochastic Process

C.8: Spectral analysis of one parameter stationary stochastic processes

Statistic or stochastic processes are conventionally divided into three classes we have already touched:

Class one	Class two	Class three
general instationary process	stationary process	ergodic stationary process

In real life, of course, we have live the property that all *natural process are instationary*. It is the task of modeling real world process, for instance by a first order trend model or by taking increments of higher order to come more closely to the case of stationary processes for *higher order increments*. We will come back to this problem.

For formal reasons we introduce *first* real-valued and *complex-valued* stochastic process. Second we study the structure of *process with a discrete spectrum*. At the end, we introduce stochastic process with *a continuous spectrum enriched by examples*.

C-81 Foundations: Ergodic and Stationary Processes

C.8.1: Foundations: Ergodic and stationary processes

We move on towards *complex-valued stochastic processes* defined by the following axioms:

$$
\begin{aligned}
&(a) && X(t) = Y(t) + iZ(t) \ \text{ subject to } \ i = \sqrt{-1} && \text{(C189)}\\
&(b) && \text{“Real”}: \ \{Y(t),\ t \in [-\infty, +\infty]\} &&\\
&(c) && \text{“Imaginary”}: \ \{Z(t),\ t \in [-\infty, +\infty]\} &&\\
& && \bar{Z} = a - b, \ \ |z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2} \ \ \text{“absolute value”} && \text{(C190)}\\
&(d) && \text{“Trend function”}: E\{X(t)\}E\{Y(t)\} + iE\{Z(t)\} &&\\
&(e) && \text{“Covariance function”}: &&\\
& && \operatorname{Cov}\{X(s), X(t)\} = E\{[X(s) - E\{X(s)\}][X(t) - E\{X(t)\}]\} && \text{(C191)}\\
&(f) && \textit{Stationary in the wider sense}: &&\\
& && (vi1) \ \ E\{X(t)\} = \mu(t) = \mu \ \ \text{(constant)} && \text{(C192)}\\
& && (vi2) \ \ \operatorname{Cov}\{X(s), X(t)\} = \operatorname{Cov}(o; s - t), \tau := t - s && \text{(C193)}\\
&(g) && \textit{Complex-valued stochastic process of second order:} &&\\
& && E\{|X(t)|^2\} < \infty \ \ forall \ \ t \in [-\infty, +\infty] && \text{(C194)}\\
& && && \text{(C195)}
\end{aligned}
$$

Ergodicity

A stochastic process $E\{X(t), \quad t \in [-\infty, +\infty]$ is *stationary in the narrow sense* if for all trajectories $x(t) = y(t) + iz(t)$ there hold

$$\boxed{E\{X(t)\} = E\{X(t_0)\} =: \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{+T} x(t)dt \ \text{ for } \ t_0 \in \mathbb{R}} \tag{C196}$$

In this integral relation, in order to compute the expectation value $E\{x(t_0)\}$ we use only all the information *which is located in one realization of the stochastic process*. In contrast to the general case, we have used *up to zero* the information of *all realizations* of the stochastic process. In stationary ergodic process we used only *one instant*. An *excellent approximation* is the presentation.

$$E\{X(t_0)\} = \mu = \lim_{N\to\infty} \frac{1}{N} \sum_{R=1}^{N} x_k(t_0) \tag{C197}$$

Such a definition has a simple physical interpretation: The mean of a fixed instant equals the mean of a long time interval: *The mean location is identical to the time* *average*. In essence, This is the property of ergodic stationary processes. Analogue we can define the covariance function for a such processes namely

$$\mathrm{Cov}(\tau) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{+T} [x(t) - \mu\sqrt{x(t+\tau) - \mu}dt \tag{C198}$$

The treatment of ergodic stationary processes is reasonable for cases in which we have *only one time like realization of the stochastic process*. Of course *increasing the length of the interval* $[-T, +T]$ *improves the estimation of* $E\{X(t_0)\} = \mu$.

Here we also take reference the *Euler's representations*:

$$\exp \pm ix = \cos x \pm i \sin x \tag{C199}$$

$$\sin x = \frac{1}{2i}[\exp(ix) - \exp(-ix)] \tag{C200}$$

$$\cos(x) = \frac{1}{2}[\exp(ix) + \exp(-ix)] \tag{C201}$$

C-82 Processes with Discrete Spectrum

C.8.2: Processes with discrete spectrum

We start with general structure of *stationary processes* with discrete spectrum: *we assume that all stationary process are ergodic*.

Ex 1: Special Stationary Ergodic Processes

Example 8.1: Special stationary ergodic processes

Let $\{X(t), \ \ t \geq 0\}$ be a stochastic process structure by

$$lrX(t) = X\Omega(t), X \text{ a complex random variable} \tag{C202}$$

$$E\{X\} = 0, \ \ E(|X|^2) < \infty, \ \ E\{X(t)\overline{X(t+\tau)}\} = \Omega(t)\overline{\Omega(t+\tau)}E\{X^2\} \tag{C203}$$

$$\tau = 0: \ \ \Omega(t)\overline{\Omega(t)} = |\Omega(t)|^2 = \text{ constant} \tag{C204}$$

$$\text{“}ansatz\text{”}: \ \Omega(t) = |\Omega(t) \exp[i\omega(t)] \text{ for } \omega(t) \text{ real function} \tag{C205}$$

$$\text{“}ansatz\text{”}: \ \omega(t+\tau) - \omega(t) \neq f(t) : d[\omega(t+\tau) - \omega(t)]/dt = 0 \tag{C206}$$

$$\text{or } d\omega(t)/dt = \text{ constant} \tag{C207}$$

$$\text{“}ansatz\text{”} : \ \omega(t) \text{ is a function of two constants, ; } \omega \text{ and } \phi : \tag{C208}$$

$$\Omega(t)|\Omega(t)|\exp[i(\omega+\phi)] \tag{C209}$$

End of Example 8.1: Special stationary ergodic processes

Lemma 8.1: decomposition of a complex stationary ergodic process

A stochastic process $\{X(t),\ t \geq 0\}$ of the structure $X(t) = X\Omega(t)$ is stationary in the wider sense if and only if

$$X(t) = X\exp(i\omega t) \tag{C210}$$

subject to $E\{X\} = 0,\ E\{|x|^2\} < \infty,\ s =: E\{|X|^2\},\ \mathrm{Cov}(\tau) = s\exp(-i\omega\tau)$

End Lemma 8.1

The real part of a complex stochastic process describes a *cosine oscillation with a random amplitude and phase*

$$y(t) = a\cos(\omega t + \phi) \tag{C211}$$

The parameter ω is called *frequency*.

We generalize to two stationary processes as a linear combination: we refer to X_1 and X_2 as two complex random variables with zero expectation and two constant parameter $\{\omega_1, \omega_1^\circ = 0,\ \omega_2, \omega_2^\circ = 0\}$. We assume that the two complex random functions X_1 and X_2 are uncorrelated such that $E\{X_1X_2\} = 0$ or $E\{X_1 - X_2\} = 0$. Our intention is to get the form of the covariance function:

$$X(t) = X_1\exp(i\omega_1 t) + X_2\exp(i\omega_2 t) \tag{C212}$$

$$\begin{aligned} &cov\{X(t)X(t+\tau)\} \\ &= E\{[X_1\exp(i\omega_1 t) + X_2\exp(i\omega_2 t)][\bar{X}_1\exp(-i\omega_1 t) + \bar{X}_2\exp(-i\omega_2 t)]\} \\ &= E\{[X_1\bar{X}_2\exp(-i\omega_1\tau) + X_1\bar{X}_2\exp(i(\omega_1-\omega_2)t - i\omega_2\tau)]\} \\ &\quad + E\{[X_2\bar{X}_1\exp(-i\omega_2\tau) + X_2\bar{X}_2\exp(i(\omega_2-\omega_1)t - i\omega_1\tau)]\} \end{aligned} \tag{C213}$$

$$s_1 =: E\{|X_1|^2\},\ \ s_2 =: E\{|X_2|^2\} \tag{C214}$$

$$Cov(\tau) = s_1\exp(-i\omega_1\tau) + s_2\exp(-i\omega_2\tau) \tag{C215}$$

Ex 2: Special Correlation Function of a Stationary Stochastic Processes

Example 8.2: Special correlation function of a stationary stochastic process

A special complex stationary stochastic process may be real-valued. For instance we define

$$X_1 = \frac{1}{2}(A + iB) \quad \text{and} \quad X_2 =: \bar{X}_1 = \frac{1}{2}(A - \delta B) \tag{C216}$$

A and B are *two real-valued functions*. In consequence of our representation $X(t) = X_1 \exp(i\omega_1 t) + X_2 \exp(i\omega_2 t)$ we specialize here to

$$X(t) = A \cos \omega t - B \sin \omega t \tag{C217}$$

A, B are *uncorrelated*. We received the *covariance function*

$$Cov(\tau) = 2s \cos \omega\tau \tag{C218}$$

subject to $s := E\{|X_1|^2\} = E\{|X_2|^2\}$

End of Example 8.2: Special correlation function

For pairwise uncorrelated stationary stochastic processes the general ansatz of type $X(t) = \sum_{k=1}^{n} \exp(i\omega_k t)$ subject to $\omega_j \neq \omega_k$ for $j \neq k$ and $i, j = 1, 2, \ldots, n-1, n$ and $E\{X_k\} = 0$ will lead us to the *covariance function*

$$Cov(\tau) = \sum_{k=1}^{n} s_k(i\omega_k\tau) \quad \text{and} \quad s_k = E\{|X_k|^2\} \quad \text{for} \quad k \in \{1, 2, \ldots, n-1, n\} \tag{C219}$$

in particular

$$Cov(\tau = 0) = E\{|X(t)|^2\} = \sum_{k=1}^{n} s_k \tag{C220}$$

In summary, the oscillation $X(t)$ is an *additive superposition of a harmonic oscillation*. This average energy per time unit is the same of *circle average energies per time unit*, an appropriate interpretation of our last representation.
In order to achieve a *real-valued function*, we have specialize to

(a) an *even* number n,
(b) and *a pair* of values X_k being *conjugate complex*.

Let the set $\{\omega_1, \omega_2, \ldots\}$ define the *spectrum*. We have to divide the various spectrum into two classes: *small band processes* and *broad band processes*, an entry into *filter theory* with many application in all sciences.

Ex 3: Dirac Representation of a Covariance Function

Example 8.3: Dirac representation of covariance function

We use the infinitely generalized *covariance function* for application of a *Dirac function representation*.

$$\mathrm{Cov}(\tau) = \lim_{K\to\infty} \sum_{k=1}^{K} s_k \int_{-\infty}^{+\infty} \exp(i\omega\tau)s(\omega-\omega_k)d\omega \tag{C221}$$

$$\mathrm{Cov}(\tau) = \int_{-\infty}^{+\infty} \exp(i\omega\tau)s(\omega)d\omega \tag{C222}$$

subject to

$$s(\omega) = \lim_{K\to\infty} \sum_{k=1}^{K} s_k\delta(\omega-\omega_k) \tag{C223}$$

The *generalized function* $s(\omega$ is denoted by *"spectral density"* of a stationary stochastic process, formally equivalent to the *Fourier transform.*

End of Example 8.3: spectral density

C-83 Processes with Continuous Spectrum

C.8.3 Processes with continuous spectrum

Here we present *first* the spectral *decomposition of the covariance function* of a stochastic process *with continuous spectrum.* Second we present *Kolmogorov's spectral decomposition theorem.*

C8-31 Spectral decomposition of a covariance function

Let $\{X(t),\ t \in \mathbb{R}\}$ be a complex-valued stationary stochastic process described by its covariance function. Then there exists a real-valued function $S(\omega)$ such that $\mathrm{cov}(\tau)$ enjoys the representation.

$$\mathrm{Cov}(\tau) = \int_{-\infty}^{+\infty} \exp(i\omega\tau)dS(\omega) \tag{C224}$$

called the spectral function of the process. Note the domain $t \in [-\infty, +\infty]$

$$\mathrm{Cov}(0) = S(\infty) - S(-\infty) = E\{|X|^2\} < \infty \tag{C225}$$

The *spectral function* is determined only up to an additive constant c. Most commonly c is chosen to $S(-\infty) = 0$. If there exists the first derivative $S(\omega) = dS(\omega)/d\omega$, we find the *spectral density of the stochastic process*: The covariance function of this stationary process is the Fourier transform of the spectral density. Since $S(\omega)$ has to be a decreasing function we find also its property.

$$\mathrm{Cov}(\tau) = \int_{-\infty}^{+\infty} \exp(i\omega\tau)s(\omega)d\omega \tag{C226}$$

$$s(\omega) \geq 0, \tag{C227}$$

$$\int_{-\infty}^{+\infty} s(\omega)d\omega < \infty \tag{C228}$$

The set $\{\omega, s(\omega) > 0\}$ defines the continuous spectrum of the stationary stochastic process. The term

$$\mathrm{Cov}(0) = S(\infty) - S(-\infty) = \int_{-\infty}^{+\infty} s(\omega)d\omega \tag{C229}$$

is called average power of the spectrum. If we describe the stationary stochastic process X_t by harmonic oscillations with random amplitude and random phase, we are able to express

$$\exp(i\omega t) = \exp[it(\omega + 2\pi k)] \text{ for all } k \in \{1, 2, \ldots\} \text{ or } k \in [-\pi, +\pi] \tag{C230}$$

$$\mathrm{Cov}(\tau) = \int_{-\pi}^{+\pi} \exp(i\tau\omega)s(\omega)d\omega \text{ for } \tau = 0, \pm 1, \pm 2, \ldots \tag{C231}$$

If $\int_{-\infty}^{+\infty} \mathrm{Cov}(\tau)d\tau < \infty$ for stationary stochastic processes, holds, we are able to introduce the *inverse Fourier transform* by

$$s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-i\omega\tau)\mathrm{Cov}(\tau)d\tau \tag{C232}$$

$$s(\omega_2) - s(\omega_1) = \frac{i}{2\pi} \int_{-\infty}^{+\infty} \frac{\exp(-\omega_2\tau) - \exp(-i\omega_1\tau)}{\tau} \mathrm{Cov}(\tau) d\tau \quad \text{(C233)}$$

Ex 4: Fourier Spectral Density

Example 8.4: Fourier spectral density

Assume real-valued constants a and c$|a| < 1$. Here we will analyze the covariance function for an autoregressive stationary sequence of first order and compute its spectral density.

$$\begin{aligned} s(\omega) &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \mathrm{Cov}(\tau) \exp(-i\tau\omega) \\ &= \frac{e}{2\pi} \left[\sum_{\tau=-\infty}^{-1} a^{-\tau} \exp(-i\tau\omega) + \sum_{\tau=0}^{\infty} a^{\tau} \exp(-i\tau\omega) \right] \\ &= \frac{c}{2\pi} \left[\sum_{\tau=1}^{\infty} a^{\tau} \exp(i\tau\omega) + \sum_{\tau=0}^{\infty} a^{\tau} \exp(-i\tau\omega) \right] \end{aligned} \quad \text{(C234)}$$

$$s(\omega) = \frac{c}{2\pi} \left[\frac{a\exp(i\omega)}{1 - a\exp(i\omega)} + \frac{1}{1 - a\exp(i\omega)} \right] \quad \text{(C235)}$$

End of Example: Fourier spectral density

Assume a *real-valued stochastic process* whose covariance function $\mathrm{Cov}(\tau) = \mathrm{Cov}(-\tau)$, namely $\mathrm{Cov}(\tau) = [\mathrm{Cov}(\tau) + \mathrm{Cov}(-\tau)]/2$, can be represented by

$$\mathrm{Cov}(\tau) = \int_{-\infty}^{+\infty} \cos(\omega\tau) s(\omega) d\omega \quad \text{(C236)}$$

$$\mathrm{Cov}(\tau) = 2 \int_{-\infty}^{+\infty} \cos(\omega\tau) s(\omega) d\omega \quad (\cos(\omega\tau) = \cos(-\omega\tau)) \quad \text{(C237)}$$

$$s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \cos(\omega\tau)\mathrm{Cov}(\tau)d\tau \tag{C238}$$

$$s(\omega) = \frac{1}{\pi} \int_{0}^{+\infty} \cos(\omega\tau)\mathrm{Cov}(\tau)d\tau \tag{C239}$$

Correlation time

In practice the notion of "*correlation time*" or "*correlation length*" is very appropriate:

$$\tau_0 := \frac{1}{\mathrm{Cov}(0)} \int_{0}^{+\infty} \mathrm{Cov}(\tau)d\tau \tag{C240}$$

or

$$\tau_0 := \frac{\pi s(0)}{2 \int_{0}^{+\infty} s(\omega)d\omega} \tag{C241}$$

For $|\tau| \leq \tau_0$ there is a unstable *correlation or statistical dependence* between $X(t)$ and $X(t) + \tau$. This *correlation decreases* for growing $|\tau|, |\tau| > \tau_0$.

Ex 5: Stationary Infinite Random Sequence

Example 8.5: Stationary infinite random sequence

The infinite random sequence $\{\cdots, -X_1, X_0, +X_1, \cdots\}$ may be structured according to $X_t = a_t X$ subject to $t = 0, \pm 1, \pm 2, \cdots$: The *complex random variable* is structured according to $E\{X\} = 0$ as well as $\mathrm{Var}\{X\} = \sigma^2$. The random sequence $\{\cdots, -X_1, X_0, +X_1, \cdots\}$ is the *wider sense stationary* if and only if the constants ω within the set-up.

$$a_t = \exp(it\omega) \quad \text{for } t = 0, \pm 1, \pm 2, \cdots \tag{C242}$$

are represented. In the *stationary case*, the harmonic oscillations X_t will be given by the *random amplitude* and the *random phase* as well as a *constant frequency* ω. For such a case a set-up of the *covariance-function* will be calculated.

$$\exp(it\omega) = \exp it(\omega + 2k\pi) \quad \text{for } t \in, \pm 1, \pm 2, \cdots \tag{C243}$$

$$\mathrm{Cov}(\tau) = \int_{-\pi}^{+\pi} \exp(i t \omega) s(\omega) d\omega \quad \text{for } \tau \in, \pm 1, \pm 2, \cdots \qquad \Leftrightarrow \tag{C244}$$

$\Leftrightarrow$

$$s(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{+\infty} \exp(-i t \omega) \mathrm{Cov}(\tau) \tag{C245}$$

If we specialize to X *real-valued* and uncorrelated we achieve a *pure random sequence* subject to

$$\boxed{\mathrm{Cov}(\tau) = \sigma^2 \quad \Leftrightarrow \quad s(\omega) = \sigma^2/(2\pi)} \tag{C246}$$

End of Example 8.5: random sequence

Ex 6: Spectral Density of a Random Telegraph Signal

Example 8.6: spectral density of a random telegraph signal

Given the covariance function, we intend to find a representation of the *spectral density for a random telegraph signal*.

$$\mathrm{Cov}(\tau) = a \exp(-b|\tau|) \quad \text{for } a > 0, b > 0 \tag{C247}$$

$$\begin{aligned} s(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-i \omega \tau) a \exp(-b|\tau|) d\omega \\ &= \frac{a}{2\pi} \left\{ \int_{-\infty}^{0} \exp(-i \omega \tau) d\tau + \int_{0}^{+\infty} \exp[-(b - i\omega)\tau] d\tau \right\} \\ &= \frac{a}{2\pi} \left\{ \frac{1}{b - i\omega} + \frac{1}{b + i\omega} \right\} \end{aligned} \tag{C248}$$

$$\boxed{s(\omega) = \frac{ab}{\pi(\omega^2 + b^2)}} \tag{C249}$$

The *correlation time* is calculated to $\tau_0 = 1/b$! this result is illustrated by Figs. C.7 and C.8, namely by the spectral structure of the covariance function and its spectral density

End of Example: spectral density telegraph signal

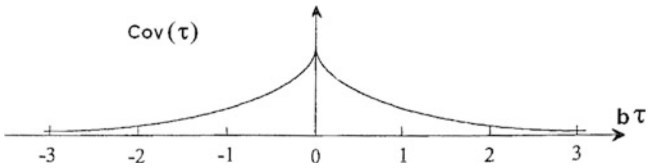


Fig. C.7 Covariance of Example 8.6

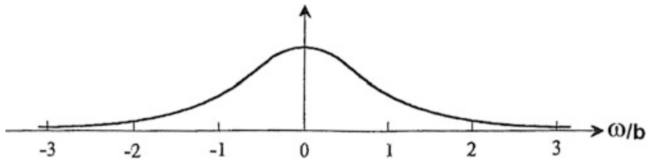


Fig. C.8 Spectral density of Example 8.6

Ex 7: Spectral Representation of a Pulse Modulation

Example 8.7: Spectral density of a pulse modulation

We start with the *covariance function* of a *pulse modulation* and study its *spectral density*.

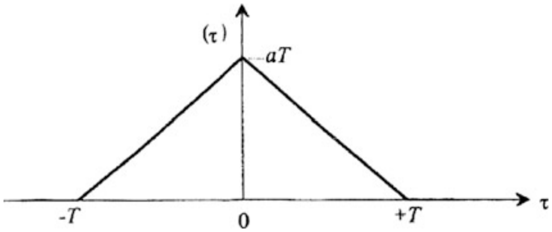
$$\mathrm{Cov}(\tau) = \begin{bmatrix} a(T - |\tau|) & \text{for } |\tau| \le T \\ 0 & \text{for } |\tau| > T \end{bmatrix} \tag{C250}$$

$$\begin{aligned} s(\omega) &= \frac{a}{2\pi} \int_{-T}^{+T} [\exp(-i\omega\tau)](T - |\tau|)d\tau \\ &= \frac{a}{2\pi} \left\{ T \int_{-T}^{T} \exp(-i\omega\tau)d\tau - \int_{0}^{+T} \tau \exp(i\omega\tau)d\tau - \int_{0}^{+T} \tau \exp(-i\omega\tau)d\tau \right\} \\ &= \frac{a}{2\pi} \left\{ \frac{2T}{\omega} \sin(\omega\tau) - 2 \int_{0}^{+T} \tau \cos(\omega\tau)d\tau \right\} \end{aligned} \tag{C251}$$

$$\boxed{s(\omega) = \frac{a}{\pi} \frac{1 - \cos \omega T}{\omega^2}} \tag{C252}$$

Figures C.7 and C.8 illustrate the *covariance function* and the *spectral density* of a *pulse modulation*. A slight modification of our example proves that it is *not automatic for stationarity*.

Fig. C.9 Covariance of Example 8.7



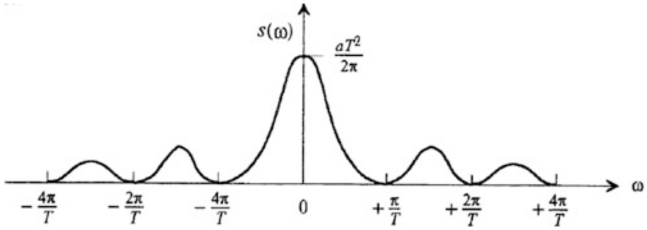


Fig. C.10 Spectral density of Example 8.7

$$\mathrm{Cov}(\tau) = \begin{bmatrix} a(T-\tau^2) & \text{for } |\tau| \leq 0 \\ & \\ 0 & \text{for } |\tau| < 0 \end{bmatrix} \quad \text{for } a > 0, T > 0 \tag{C253}$$

$s(\omega)$ *cannot* be the spectral density of a stationary stochastic process.

Alternatively, we like to resent the covariance function $\mathrm{Cov}(\tau)$ be a Dirac delta-function $\delta(\tau)$. In this case we find

$$\delta(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-i\omega\tau) d\tau \tag{C254}$$

$$\boxed{s(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-i\omega\tau)\delta(\tau) d\tau = \frac{1}{2\pi}} \tag{C255}$$

by formal inversion.

End of Example 8.7: pulse modulation

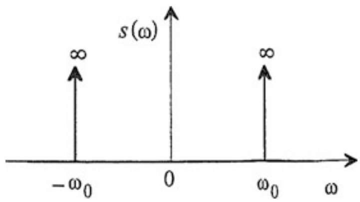


Fig. C.11 Spectral density of a harmonic function

Ex 8: Spectral Density and Spectral Function of a Harmonic Oscillation

Example 8.8: Spectral density and spectral function of a harmonic oscillation

Given a *harmonic oscillation* with a given *covariance function*, we will derive the structure of the *spectral density function* as well as the spectral function.

$$\mathrm{Cov}(\tau) = a\cos\omega_0\tau \tag{C256}$$

$$\begin{aligned} s(\omega) &= \frac{a}{2\pi}\int_{-\infty}^{+\infty}\exp(-i\omega\tau)\cos(\omega_0\tau)d\tau \\ &= \frac{a}{4\pi}\int_{-\infty}^{\infty}\exp(-i\omega\tau)[\exp(-i\omega_0\tau)-\exp(-i\omega_0\tau)]d\tau \\ &= \frac{a}{4\pi}\left\{\int_{-\infty}^{\infty}\exp[i(\omega_0+\omega)\tau]+\int_{-\infty}^{\infty}\exp[-i(\omega_0-\omega)\tau]\right\}d\tau \end{aligned} \tag{C257}$$

$$\boxed{s(\omega) = \frac{a}{2}\{\delta(\omega_0-\omega)+\delta(\omega_0+\omega)\}} \tag{C258}$$

$$\boxed{S(\omega) = \begin{bmatrix} 0 & \text{for } \omega < -\omega_0 \\ a/2 & \text{for } -\omega_0 < \omega \le -\omega_0 \\ a & \text{for } \omega > \omega_0 \end{bmatrix} \quad \text{step function}} \tag{C259}$$

We illustrate the spectral density as well as the spectral function of a harmonic oscillation by Fig. C.11 as well as Fig. C.12. It has to be emphasized that all three functions $\mathrm{Cov}(\tau)$, $s(\omega)$ and $S(\omega)$ are *stationary functions*.

End of Example 8.8: Spectral function of harmonic oscillation

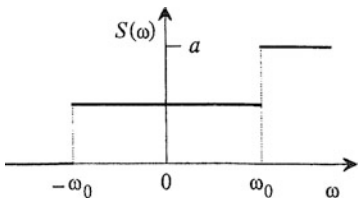


Fig. C.12 Spectral function of type step function for a harmonic oscillation

Ex 9: Spectral Density Function of a Damped Harmonic Oscillation

Example 8.9: Spectral density functions of a damped harmonic oscillation

Given the *covariance function* of a *damped harmonic oscillation*, we want to find the *spectral density* of this *stationary stochastic process*,

$$\mathrm{Cov}(\tau) = a\exp(-b|\tau|)\cos\omega_0\tau \tag{C260}$$

$$\begin{aligned} s(\omega) &= \frac{a}{\pi}\int_0^{\infty}\exp(-b\tau)\cos(\omega\tau)\cos(\omega_0\tau)d\tau \\ &= \frac{a}{2\pi}\int_0^{\infty}\exp(-b\tau)[\cos(\omega-\omega_0)\tau+\cos(\omega_0+\omega)\tau]d\tau \end{aligned} \tag{C261}$$

$$\boxed{s(\omega) = \frac{ab}{2\pi}\left\{\frac{1}{b^2+(\omega-\omega_0)^2}+\frac{1}{b^2+(\omega+\omega_0)^2}\right\}} \tag{C262}$$

ω_0 is called the *eigen function* of the stochastic process of type stationary. A typical example is the *fading of a radio signal*.

End of Example 8.9: Damped harmonic oscillation

Ex 10: Spectral Density of White Noise

Example 8.10: Spectral density of white noise

Perviously we introduced the concept of "*white noise*" as a *real-valued stationary process with continuous time of type Wiener process*. Given its *covariance function*, we are interested in the *spectral density function*.

$$\mathrm{Cov}(\tau) = 2\pi s_0\delta(\tau) \tag{C263}$$

$$s(\omega) = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \exp(-i\omega\tau) 2\pi s_0 \delta(\tau) d\tau = S_0 \tag{C264}$$

A comparative formulation of a white noise process is the following:

White noise is a real-valued process whose spectral density is a constant

Comment

the spectral density of a white noise process has only the property $s(\omega) \geq 0$, but not the property $\int\limits_{-\infty}^{\infty} s(\omega) d\omega < \infty$. Accordingly, a white noise process with an average power spectrum fulfils

$$s(\omega) d\omega = \infty \tag{C265}$$

does exist only in theory not in practice. A comparison is "*the mass point*" or "*the point mass*", which also exist only "*in theory*". Nevertheless it is *reasonable approximation of any reality*.
A stationary process can always be considered as a *excellent approximation of white noise* if the covariance between $Z(t)$ and $Z(t + \tau)$ for growing values $|\tau|$ extremely fast goes to zero.

Example:

Let us denote the *variations of the absolute value of the force* by $Z(t)$ acting on a particle of a liquid medium at the time t and leading to a *Brownian motion*. Per second, the resulting force leads to *approximately* 10^{21} *collisions* with other particles. Therefore the stochastic variables $Z(t)$ and $Z(t + \tau)$ are stochastically independent ("uncorrelated") if $|\tau|$ is of the *order of* 10^{-18} *seconds*. If we assume a covariance function of type $\text{Cov}(\tau) = \exp(-b|\tau|)$ for $b > 0$, the number b has to be larger than 10^{-17} seconds.

End of Example 8.10: White noise

Ex 11: Band Limit of White Noise

Example 8.11: Band limited white noise

A stationary stochastic process with a constant spectral density function $S(\omega) = s_0$ for $\omega \in \{-\infty, +\infty\}$ does *not* in practice. But a stationary stochastic process with the *spectral density*

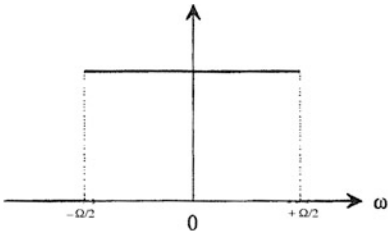


Fig. C.13 Spectral density of band limited white noise process

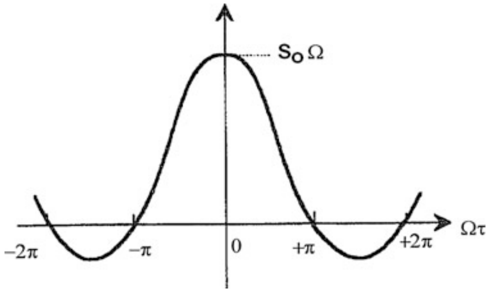


Fig. C.14 Covariance function of a band limited white noise process

$$s(\omega) = \begin{bmatrix} s_0 & \text{for } -\varOmega/1 \leq \omega \leq +\varOmega/2 \\ 0 & \text{otherwise} \end{bmatrix} \tag{C266}$$

exists

$$\operatorname{Cov}(\tau) = \int\limits_{-\varOmega/2}^{+\varOmega/2} \exp(i\omega\tau)s_o = 2s_0 \frac{\sin(\varOmega\tau)/2}{\tau} \tag{C267}$$

The average power of the process is proportional to Cov $(0) = s_0\varOmega$. The parameter $\varOmega$ is called the *band width of the process*. A *white noise process* is generated by the *limit of a band limited white noise* illustrated by Figs. C.13 and C.14.

End of Example 8.11: band limited white noise

C-84 Spectral Decomposition of the Mean and Variance-Covariance Function

C.8.4: Spectral decomposition of the expectation and the variance-covariance function

The decomposition of the *spectrum of a stationary process* $\{Y(x), x \in \mathbb{R}\}$ subject to $\mathbb{R} \in \{-\infty, +\infty\}$ *applied to orthogonal increments of various order* has many representations, if they do *not* overlap in the intervals

$$[x_1, x_2], [x_3, x_4] \text{ or } [x_1, x_2, x_3], [x_4, x_5, x_6] \text{ etc}$$

in the formulations of double, triple etc. differences. An example is

$$E\{[Y(x_2) - Y(x_1)][\overline{Y(x_4) - Y(x_4)}]\} \tag{C268}$$

$$E\{[Y(x_3) - 2Y(x_2) + Y(x_1)][\overline{Y(x_6) - 2Y(x_5) + Y(x_4)}]\} \tag{C269}$$

Accordingly, *a real-valued stochastic process with independent increments* whose trend function $E\{Y(x)\} = 0$ is zero, has *orthogonal increments*.

Example 8.12: Spatial best linear uniformly unbiased prediction: spatial BLUUP

Spatial BLUUP depends on the exact knowledge of proper covariance function. In general, the variance-covariance function *depends on three parameters* $\{\delta_1, \delta_2, \delta_3\}$ called { rugged, sill, range}

$$\mathrm{Cov}(\tau) = (\|\mathbf{x}_1 - \mathbf{x}_2\|) \tag{C270}$$

We assume stationarity and isotropy for the mean function as well as for the variance-covariance function. (In the next section we define isotropy)

$$\text{Model } : Z(\mathbf{x}) - f(\mathbf{x})'\beta + e(\mathbf{x}) \tag{C271}$$

$$\text{Subject to } E\{e(\mathbf{x})\} = 0$$

Postulate one: " Mean square continuity "

$$\lim_{\mathbf{y}\to\mathbf{x}} E\{\mathbf{Z}(\mathbf{y}) - \mathbf{Z}(\mathbf{x})\} = 0 \tag{C272}$$

Postulate two: Mean square differentiability

$$\lim_{h\to 0} \frac{Z(\| \times \| + h_n) - Z(\| \times \|)}{h_n} \underset{n\to\infty}{\longrightarrow} 0 \tag{C273}$$

If we assume "*mean square differentiability*", *then* Z' has a variance-covariance function. If "*mean square differentiability*" holds, it is also differentiated *in any dimension*. Here we introduce various classes of differential calculus applied to variance-covariance functions.

(i) DIGGLE chaos:

$$Cov(h) = c\exp(-a|h|^{\gamma}) \tag{C274}$$

(ii) MATERN class:

$$Cov(h) = c(\alpha|h|^{\nu}\mathcal{H}(a|h|)) \tag{C275}$$

" $\mathcal{H}$" denotes the "modified Bessel function of order ν"

(iii) Whittle class:
spectral density of type

$$c(\alpha^2 + \omega^2)^{-\nu-\alpha/2} \tag{C276}$$

(iv) Stochastic process on curved spaces like the sphere or the ellipsoid: no Matern classes, the characic functions have a *discrete spectrum*, for instance in terms of *harmonic functions*.

End of Example 8.12: spatial BLUUP

Let us give some references: Brown et al. (1994), Brus (2007), Bueso et al. (1999), Christenson (1991), Diggle and Lophaven (2006), Fedorov and Muller (2007), Groeningen et al. (1999), Kleijnen (2004), Muller et al. (2004), Muller and Pazman (2003), Müller (2005), Omre (1987), Omre and Halvorsen (1989), Pazman and Muller (2001), Pilz (1991), Pilz et al. (1997), Pilz and Spock (2006), PilzSpock(2008), Spöck (1997), Spock(2008), Stein (1999), Yaglom (1987), Zhu and Stein (2006).

Up to now we assumed a stochastic process of type $\{X(t), t \in \mathbb{R}\}, \mathbb{R} = \{-\infty, +\infty\}$ as a complex stationary process of *second order*. Then there exist a stochastic process of second order $\{U(\omega), \omega \in \mathbb{R}\}$ with *orthogonal increments which can be represented by*

$$X(t) = \int_{-\infty}^{+\infty} \exp(i\omega t) dU(\omega) \tag{C277}$$

The process $\{U(\omega), \omega \in \mathbb{R}\}$ of the process $\{X(t), t \in \mathbb{R}\}$ is called the related spectral process. Its unknown additive constant we fix by the postulate

$$P\{U(-\infty) = 0\} = 1 \tag{C278}$$

subject to

$$E\{U(\omega)\} =: 0 \tag{C279}$$

$$E\{[|U(\omega)|^2]\} - S(\omega) \tag{C280}$$

$$E\{[|dU(\omega)|^2]\} - dS(\omega) \tag{C281}$$

Our generalization was introduced by *A. N. Kolmogorov*: Compare this representation with our previous model

$$X(t) = \lim_{k \to \infty} \sum_{k=1}^{K} \exp(i\omega t) X_k \tag{C282}$$

and

$$X_t = \int_{-\infty}^{+\infty} \exp(i\omega t) dU(\omega) \tag{C283}$$

The *orthogonality* of the *Kolmogorov stochastic process* $\{U(t), t \geq 0\}$ corresponds in case of a *discrete spectrum* to the concept of *uncorrelation*. Convergence in the squared mean of type

$$\boxed{\int_{-\infty}^{+\infty} \exp(i\omega t) dU(\omega) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \lim_{\substack{n \to +\infty \\ \Delta\omega_k \to 0}} \sum_{k=1}^{n} \exp(i\omega_{k-1} t)[U(\omega_k) - U(\omega_{k-1})] \tag{C284}}$$

emphasizes the *various weighting functions* of frequencies: The frequency ω_{k-1} in the range $[\omega_{k-1}, \omega_k]$ receives the random weight $U(\omega_k) - U(\omega_{k-1})$. A analogue version of the *inversion process* leads to

$$\boxed{U(\omega_2) - U(\omega_1) = \frac{i}{2\pi} \int_{-\infty}^{+\infty} \frac{\exp(i\omega_2 t) - \exp(i\omega_1 t)}{t} X(t) dt \tag{C285}}$$

In practice, *Kolmogorov stochastic processes* are applied in *linear filters* for prognosis or prediction.

C-9 Scalar-, Vector-, and Tensor Valued Stochastic Processes of Multi-Parameter Systems

Appendix C.9: Scalar-, vector- and tensor-valued stochastic processes of multi-parameter systems

We have experienced the generalization of a scalar-valued stochastic process due to *A.N. Kolmogorov*. Another generalization due to him is the concept of the *structure functions* for more general vector-valued and tensor-valued for *increments of various order* as well as their *power spectrum* or *multi-parameter systems*.

First we introduce the notion of the *characteristic functional* replacing the notion of the *characteristic function* for random function. *Second*, the *moment representation* for stochastic processes is outlined. Third we generalize the notion from scalar-valued and vector-valued stochastic processes to *tensor-valued stochastic processes* of multi-point parameter systems. In particular, we introduce the notion of *homogeneous and isotropic stochastic fields* and extent the *Taylor-Karman* structured criterion matrices to *distributions of the sphere*.

C-91 Characteristic Functional

Let us begin with the *definition of the characteristic functional*.

$$\widetilde{Y}_{\mathbf{x}}[\theta] = E\{\exp(iY(\theta))\} = E\{i\} \int_a^b \theta(\mathbf{x})Y(\mathbf{x})d\mathbf{x} \tag{C286}$$

$\widetilde{Y}(\theta)$ as the *characteristic functional* elates uniquely to the random function $Y(\mathbf{x})$ of a fixed *complex number*. If the characteristic functional of *a specific random function* is known we are able to *derive the probability density distribution* $f_{X_1,\ldots,X_N}(Y_X,\ldots,Y_{X_N})$.

Example

If a scalar-valued function in a three dimensional field $\mathbf{x} := (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ exists define its *characteristic functional* by

$$\widetilde{Y}_Y[\theta] := E\{\exp(i \iiint_{-\infty}^{+\infty} Y(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\theta(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3))d\mathbf{x}_1 d\mathbf{x}_2 \tag{C287}$$

The *domain of integration* is very often chosen to be spherical coordinates of ellipsoidal coordinates (u, v, w) for $u \in \{0, 2\pi\}, v \in \{-\pi/2 < v < +\pi/2\}$ and $w \in \{0 < w < \infty\}$ as a *first chart*. A second chart is needed to cover totally the intersection surfaces, for instance by *meta-longitude*, *meta-latitude*.

End of Example

Probability distribution for random functions and the characteristic functional

As a generalization of the *characteristic function* for random events, the concept of the *characteristic functionals* for *random functions* or *stochastic processes* has been developed. Here we aim at reviewing the theory. We apply the concept of random functions for multi-functions of type *scalar-valued and vector-valued functions* which have been first applied in the *statistical theory turbulence*.

what is the characteristic functional?

We summarize in Table C.9.1 the *detailed* notion of a *scalar-valued as well as a vector-valued characteristic functional* illustrated by an example by introducing the "*Fourier-Stieljes functional*".

Table C.9.1 Probability distribution of a random function and the characteristic functional

"Scalar-valued characteristic functionals"

"Random function over an interval [a,b] with N-variables"

$\{Y(\mathbf{x}), \mathbf{x} \in \mathbb{R}^N\}$ is called a random function if the *density function*

$$f_{x_1,x_2,\ldots,x_{N-1},x_N}[Y(x_1), Y(x_2), \ldots, Y(x_{N-1}), Y(x_N)] \tag{C288}$$

exist at $(x_1, x_2, \ldots, x_{N-1}, x_N)$ points.

"Fourier representation of the distribution function"

$$\begin{aligned} \widetilde{Y}_{x_1,\ldots,x_N}(\theta_1, \ldots, \theta_N) &:= E\{\exp[i(\textstyle\sum_{k=1}^N \Theta_k Y_k)]\} \\ &= \textstyle\int_{-\infty}^{+\infty} \cdot \int \exp\{i \sum_{k=1}^N \Theta_k Y_k\} f_{x_1,\ldots,x_N}(Y_1, \ldots, Y_N) dY_1 \ldots dY_N \end{aligned} \tag{C289}$$

subject

$$Y_1 := Y(x_1), \ldots, Y_k := Y(x_k) \tag{C290}$$

"Existence condition"

$$Y[\theta_1, \ldots, \theta_N] = \int_a^b \theta(x) dx \tag{C291}$$

"Vector-valued characteristic functionals"

"at any point P of the domain D place a set of *base vectors* $\mathbf{e}^\alpha(P_i)$ at fixed points P_i for $\alpha = 1, 2, \ldots, n-1, n, \quad i = 1, 2, \ldots, N-1, N$.

$$Y_i^\alpha =, Y(P-i)|\mathbf{e}^\alpha(P_i) > \tag{C292}$$

"Vector-valued characteristic functionals"

$$\widetilde{\mathbf{Y}}_{\mathbf{Y}}[\theta] : \theta(x, y, z) = \textstyle\sum_{\alpha=1}^n \theta_\alpha(P_i)\mathbf{Y}(P_i) \tag{C293}$$

End of Table C.9.1: characteristic functional

C-92 The Moment Representation of Stochastic Processes for Scalar Valued and Vector Valued Quantities

The moment representation of stochastic processes is the standard tool of statistical analysis. *If* a random function is given by its *probability function*, *then* we find a unique representation of its moments. But the *inverse process* of a given moment representation converted in its probability function is *not unique*. Therefore we present only the direct representation of moments for a given probability function. *At first* we derive for a given *scalar-valued* probability function its moment representation in Table C.9.2. In Table C.9.3 we generalize this result for stochastic processes which are *vector-valued*. An *example* is presented for second order statistics, namely for the *Gauss-Laplace* normal distribution and its first and second order moments.

Table C.9.2 Moment representation of stochastic processes for scalar quantities

"scalar-valued stochastic processes"

"moment representation"

$Y(x), x \in \mathbb{R}^N$ is a *random function*, for instance applied in turbulence theory.
Define the set of random function $\{x_1, x_2, \ldots, x_{N-1}, x_N\} \ni \mathbb{R}$
with the *probability distribution* $f(x_1, x_2, \ldots, x_{N-1}, x_N)$.
Then the *moment representation* is the polynomial

$$\boxed{\begin{aligned} &M^{p_1, p_2, \ldots, \gamma_{N-1}, \gamma_N} := E\{x_1, x_2, \ldots, x_{N-1}, x_N\} \\ &= \iint_{-\infty}^{+\infty} \ldots \iint x_1^{\gamma_1} x_2^{\gamma_2} \ldots x_{N-1}^{\gamma_{N-1}} x_N^{\gamma_N} f\{x_1, x_2, \ldots, x_{N-1}, x_N\} \\ &\times dx_1 dx_2 \ldots dx_{N-1} dx_N \end{aligned}} \tag{C294}$$

(nonnegative even numbers) $p_1, \ldots, p_N : \gamma_1, \gamma_2, \ldots, p_{N-1}, p_N \in \mathbb{N}_N$

$$[p] := \gamma_1 + p_2 + \cdots + p_{N-1} + \gamma_N \tag{C295}$$

is called the order of the moment representation

Table C.9.2 Continued

"central moment representation"

$$\{x_1 - E\{x_1\}, x_2 - E\{x_2\}, \ldots, x_{N-1} - E\{x_{N-1}\}, x_N - E\{x_N\}\} \quad \text{(C296)}$$

$$B^{\gamma_1\gamma_2\ldots p_{N-1},p_N} := E\{(x_1 - E\{x_1\})^{\gamma_1}, (x_2 - E\{x_2\})^{\gamma_2}, \ldots, (x_{N-1} - E\{x_{N-1}\})^{\gamma_{N-1}}, (x_N - E\{x_N\})^{\gamma_N}\} \quad \text{(C297)}$$

"Characteristic function"

$$\Phi(\theta_1, \ldots, \theta_N) := \iint_{-\infty}^{+\infty} \cdots \iint \exp\{i \sum_k \theta_k x_k\} f\{x_1, x_2, \ldots, x_{N-1}, x_N\} dx_1 dx_2 \ldots dx_{N-1} dx_N \quad \text{(C298)}$$

"differentiate at point $\theta_1 = 0, \ldots, \theta_N(0)$"

$$(i)^{[\gamma]} M^{\gamma_1 p_2\ldots p_{N-1},p_N} = \left[\frac{\partial^{[\gamma]}}{\partial\theta_1^{p_1}\ldots\partial\theta_N^{p_N}} \Phi(\theta_1, \ldots, \theta_N)\right]_{\theta_1=0,\ldots,\theta_N=0} \quad \text{(C299)}$$

"semi-invariants"

$$\Psi := \ln \Phi \quad \text{(C300)}$$

$$S^{\gamma_1,\ldots,\gamma_N} := (i)^{[\gamma]}(-1)^{[\gamma]} \left[\frac{\partial^{[\gamma]}}{\partial\theta_1^{\gamma_1}\partial\theta_2^{\gamma_2}\ldots\partial\theta_{N-1}^{\gamma_{N-1}}\partial\theta_N^{p_N}} \Psi(\theta_1, \ldots, \theta_N)\right]_{\theta_1=0,\ldots,\theta_N=0} \quad \text{(C301)}$$

"transformation of semi invariants to moments of order"

$$\begin{gathered} S_i = M_1, \quad S_2 = M_2 - M_1^2 = B_2, \\ S_3 = M_3 - 3M_2M_1 + 2M_1^3 = B_3 \\ S_4 = B_1 - 3B_2^2, \quad S_5 = B_5 - 10B_3B_2 \end{gathered} \quad \text{(C302)}$$

End of Table C.9.2: Moment representation of stochastic processes

Table C.9.3 Moment representation of stochastic processes for vector-valued quantities

"Vector-valued stochastic processes"

$\boxed{\text{“moment representation”}}$

"multi-linear functions of a vector field Gauss-Laplace distributed"

$$\tilde{\mathbf{Y}}_{\mathbf{Y}}[\theta] : \theta(\mathbf{x}) = \sum_{\alpha=1}^{n} \theta_\alpha(P_{\mathbf{x}})\mathbf{Y}^\alpha(P_i) \tag{C303}$$

for a set of base vectors $e^\alpha(P_i)$ at fixed points
$P_\alpha, \quad \alpha = 1, 2, \ldots, n-1, n; \quad i = 1, 2, \ldots, N-1, N$

"one-point n-dimensional Gauss-Laplace distribution"

$$\mathbf{Y}(P_\alpha) \sim \mathbf{f}(P_\alpha)$$
$$\mathbf{f}(P_N) = C \exp\left(\frac{1}{2}\sum_{i,j}^{N} g_{ij}(P_\alpha)(\mathbf{Y}^i_\alpha - E\{\mathbf{Y}^i\})(\mathbf{Y}^j_\alpha - E\{\mathbf{Y}^j_\alpha\})\right) \tag{C304}$$

"gauge"

$$\iint\limits_{-\infty}^{+\infty} \ldots \iint f(P_\alpha) dY_1 dY_2 \ldots dY_{N-1} dY_N := 1 \tag{C305}$$

"two-point n-dimensional Gauss-Laplace distribution: covariance function"

$$B^{\alpha,\beta}(P_\alpha, P_\beta) = E\{[Y^\alpha(P_\alpha) - E\{Y^\alpha\}][Y^\beta(P_\beta) - E\{Y^\beta\}]\} \tag{C306}$$

$$Y(P_\alpha), Y(P_\beta) \sim f(P_\alpha, P_\beta) \tag{C307}$$

$$f(P_\alpha, P_\beta) = D \exp\left(-\frac{1}{2}\sum_{i,j=1}^{N} g_{ij}(P_\alpha, P_\beta)(Y^i_\alpha - E\{Y^i_\alpha\})(Y^i_\beta - E\{Y^i_\beta\})\right) \tag{C308}$$

Example : *Gauss-Laplace normal distribution, character's functional*

$\boxed{\text{“ansatz”}}$

Table C.9.3 Continued

$$A(\theta) = E\{X(\theta)\} := E\{\int_a^b \theta(Y)Y(x)\}dx = \int_a^b E\{Y(x)\}dx \tag{C309}$$

$$B(\theta) = \int_a^b \int_a^b B(x_1, x_2)\theta(x_1)\theta(x_2)dx_1 dx_2 \tag{C310}$$

"multi-point functions"

$$\widetilde{Y}(Y)[\theta(x)] := \exp\{i \int_a^b \theta(x)E\{Y(x)\}dx - \tfrac{1}{2}\int_a^b \int_a^b B(x_1, x_2)\theta(x_1)\theta(x_2)dx_1 dx_2 \tag{C311}$$

in general: multi-point functional series
Fourier-Stieljes functional series
Gauss-Laplace normal distribution functional:
first and second order moments

An example is the central moment of fourth order of type two-points
subject to a Gauss-Laplace normal distribution:

$$\begin{aligned} B^{\alpha\beta,\gamma^\delta}(P_1, P_2) &= B^{\alpha,\beta}(P_1)B^{\gamma^\delta}(P_2)+ \\ +B^{\alpha,\gamma}(P_1, P_2)B^{\beta\delta}(P_1, P_2) &+ B^{\alpha\delta}(P_1, P_2)B^{\beta\delta}(P_1, P_2) \end{aligned} \tag{C312}$$

general rule by *JSSERLIS*: on a formula for the product moment coefficient in any number of variables *Biometrika* 12(1918) No. 138

$\boxed{\text{First and Second order moments of a probability density}}$
$\boxed{\text{and the uncertainty relation}}$

$$\int_a^b [\mathbf{Y}(\mathbf{x}) - E\{\mathbf{Y}(\mathbf{x})]\theta_0(\mathbf{x})d\mathbf{x} = 0 \tag{C313}$$

$$\int_a^b \int_a^b B(x_1, x_2)\theta(x_1)\theta(x_2)dx_1 dx_2 = 0 \tag{C314}$$

C-93 Tensor-Valued Statistical Homogeneous and Isotropic Field of Multi-Point Systems

First, we introduce the ideal structure of a variance-covariance matrix of multi-point systems in an Euclidean space called

homogenous and isotropic in the statistical sense.

Such a notion generalized the *concept of stationary functions* so far introduced. Throughout we use the terminology "*signal*" and "*signal analysis*".
Let us begin with the notion of an *ideal structure* of the variance-covariance matrix of scalar value called signal, a multi-point function. An example is the gravity force as a *one-point function*, namely the distance between two-points as a *two-point function* or the angle between to a reference point and two targets as a *three-point function*. Here, we are interested in a criterion matrix of *vector-vector signals* which are multi-point functions. An example is the placement vector between network points determined by an adjustment process, namely a *vector-valued one-point function* proportional to the difference vector between network points, a *vector-valued two-point function*.
Let $S(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha)$ be a scalar-valued function between placement vectors $\mathbf{X}_1, \cdots, \mathbf{X}_\alpha$ with in a three dimensional Euclidean space. The *variance-covariance* of the *scalar-valued signal* at placement vectors $\mathbf{X}_1, \cdots, \mathbf{X}_\alpha$ and $\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta$ is defined by

$$\begin{aligned}\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) &:= E\{[S(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha) - E\{S(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha)\}] \\ &\times [S(\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) - E\{S(\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)\}]\}\end{aligned} \tag{C315}$$

For the special case $\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta$ the multi-point function agrees to $\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ to the *variance* in the other case to the *covariance*.

Definition C.9.1: Homogeneity and isotropy of a scalar-valued multi-point function

The scalar-valued multi-point function $\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ is called *homogenous in the statistical sense*. if it is *translational invariant*, namely

$$\sum(\mathbf{X}_1 + t, \cdots, \mathbf{X}_\alpha + t, \mathbf{X}_{\alpha+1} + t, \cdots, \mathbf{X}_\beta + t) = \sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \tag{C316}$$

t is called the *translational vector*. The function $\sum$ is called the *isotropic*, if it is *rotational invariant* in the sense of

$$\sum(\mathbf{R}\mathbf{X}_1, \cdots, \mathbf{R}\mathbf{X}_\alpha, \mathbf{R}\mathbf{X}_{\alpha+1}, \cdots, \mathbf{R}\mathbf{X}_\beta) = \sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \tag{C317}$$

R is called the *rotational matrix*.

End of Definition C.9.1: homogeneity and isotropy

The following *representation theorems* apply:

Lemma C.9.2: homogeneity and isotropy of a scalar-valued multi-point function

The scalar-valued multi-point function is *homogenous* in the statistical sense if and only if it is a *function of the difference vectors* $\mathbf{X}_2 - \mathbf{X}_1, \cdots, \mathbf{X}_\beta - \mathbf{X}_{\beta-1}$:

$$\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) = \sum(\mathbf{X}_2 - \mathbf{X}_1, \cdots, \mathbf{X}_{\alpha+1} - \mathbf{X}_\alpha, \cdots, \mathbf{X}_\beta - \mathbf{X}_{\beta-1}) \tag{C318}$$

This function is *isotropic* in the statistical sense,

$$\begin{aligned} &\sum (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \\ &= \sum(|\mathbf{X}_1|, < \mathbf{X}_1, \mathbf{X}_2 >, < \mathbf{X}_1, \mathbf{X}_3 >, \cdots, < \mathbf{X}_{\beta-1}, \mathbf{X}_\beta >, |\mathbf{X}_\beta|) \end{aligned} \tag{C319}$$

is a function of all *scalar products* including the length of the vectors $\mathbf{X}_1$, $\cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta$.

The function $\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ is *homogenous and isotropic* in the statistical sense if

$$\begin{aligned} &\sum \quad (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \\ &= \quad \sum(|\mathbf{X}_2 - \mathbf{X}_1|, < \mathbf{X}_2 - \mathbf{X}_1 | \mathbf{X}_3 - \mathbf{X}_2 >, \cdots, \\ &, \cdots, < \mathbf{X}_\beta - \mathbf{X}_{\beta-1} | \mathbf{X}_{\beta-1} - \mathbf{X}_{\beta-2} >, |\mathbf{X}_\beta - \mathbf{X}_{\beta-1}|) \end{aligned} \tag{C320}$$

is a function of all *scalar products* including the length of the difference vectors $\mathbf{X}_2 - \mathbf{X}_1, \cdots, \mathbf{X}_{\alpha+1} - \mathbf{X}_\alpha, \mathbf{X}_\beta - \mathbf{X}_{\beta-1}$.

End of Lemma C.9.2: homogeneity and isotropy

As a basic result we take advantage of the property that a scalar-valued function $\sum$ of type multi-points is *homogenous and isotropic in the statistical sense* if

$$\boxed{\begin{aligned} &\sum (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \\ &= \sum(|\mathbf{X}_2 - \mathbf{X}_1|, |\mathbf{X}_3 - \mathbf{X}_1|, \cdots, |\mathbf{X}_\beta - \mathbf{X}_{\beta-2}|, |\mathbf{X}_\beta - \mathbf{X}_{\beta-1}|) \end{aligned}} \tag{C321}$$

is a function of length of all difference vectors $\mathbf{X}_\gamma - \mathbf{X}_\delta$ for all $1 \leq \delta < \gamma \leq 1$. For a proof we need the decomposition

$$
\begin{aligned}
< \mathbf{X}, \mathbf{Y} > &= \frac{1}{2}(|\mathbf{X}|^2 + |\mathbf{Y}|^2 - |\mathbf{X} - \mathbf{Y}|^2) \\
&= \frac{1}{2}(|\mathbf{X} + \mathbf{Y}|^2 - |\mathbf{X} - \mathbf{Y}|^2)
\end{aligned}
\tag{C322}
$$

Example: four-point function, var-cov function

Choose a scalar-values two-point function as the distance $S(\mathbf{X}_1, \mathbf{X}_2)$ between two points $\mathbf{X}_1$ and $\mathbf{X}_2$. The criterion matrix of a variance-covariance function under the postulate of *homogeneity and isotropy in the statistical sense* $\sum(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ as a *four-point function described by* $\beta(\beta - 1)/2 = 6$ *quantities*, for instance

$$
\begin{aligned}
&\sum (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \\
&= \sum(|\mathbf{X}_2 - \mathbf{X}_1|, |\mathbf{X}_3 - \mathbf{X}_1|, |\mathbf{X}_4 - \mathbf{X}_1|, |\mathbf{X}_3 - \mathbf{X}_2|, |\mathbf{X}_4 - \mathbf{X}_2|, |\mathbf{X}_4 - \mathbf{X}_3|)
\end{aligned}
\tag{C323}
$$

End of example: var-cov function, 4 points

Another example is a scalar-valued three-point function of the type *angular observation* $A(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ between a fixed point $\mathbf{X}_1$ and the two moving points $\mathbf{X}_2$ and $\mathbf{X}_3$.
The situation of *vector-valued functions* between placement vectors $\mathbf{X}_1, \cdots, \mathbf{X}_\alpha$ *in a three-dimensional Euclidian space* based on the representation

$$
\mathbf{S}(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha) = \sum_{i=1}^{3} \mathbf{S}_i (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha)\mathbf{e}^i (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha) \tag{C324}
$$

is slightly more complicated. the Euclidean space is spanned by the *base vectors* $\{e^1, e^2, e^3\}$ fixed to a reference point. The variance-covariance of a vector signals at placements $\mathbf{X}_1, \cdots, \mathbf{X}_\alpha$ and $\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta$ is defined as the second order tensor

$$
\sum := \sum_{i,j}(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)\mathbf{e}^i (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha) \otimes \mathbf{e}^j (\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \tag{C325}
$$

when we have applied the sum notion convention over repeated indices.

$$
\begin{aligned}
&\textstyle\sum_{i,j} (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) := E\{[S_i (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha) - E\{S_i (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha)\}] \\
&\quad \times [S_j (\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) - E\{S_j (\mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)\}]\}
\end{aligned}
\tag{C326}
$$

We remark that all *Latin indices* run $1, 2, 3$ *but* all *Greek indices* label the number of multi-points. For the case $\mathbf{X}_1 = \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\alpha = \mathbf{X}_\beta$ we denote the multi-point function $\sum_{i,j} (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ as the *local variance* covariance matrix *for the general case nonlocal covariance*

Definition C.9.3: Homogeneity and isotropy in the statistical sense of a tensor valued multi-point function

The multi point tensor valued function $\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ is called *homogenous in the statistical sense*, if it is *translational invariant* in the sense of

$$\sum(\mathbf{X}_1 + \mathbf{t}, \cdots, \mathbf{X}_\alpha + \mathbf{t}, \mathbf{X}_{\alpha+1} + \mathbf{t}, \cdots, \mathbf{X}_\beta + \mathbf{t}) = \sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \tag{C327}$$

Where $\mathbf{t}$ denotes the **t***ranslational vector*
In contrast the multi-point tensor-valued function $\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)$ is called *isotropic in the statistical sense* if it is *rotational invariant* in the sense of

$$\sum(\mathbf{R}\mathbf{X}_1, \cdots, \mathbf{R}\mathbf{X}_\alpha, \mathbf{R}\mathbf{X}_{\alpha+1}, \cdots, \mathbf{R}\mathbf{X}_\beta) = \mathbf{R}\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta)\mathbf{R} \tag{C328}$$

where $\mathbf{R}$ denotes the rotation matrix element of the $SO(3)$ group

End of tensor-valued multi-point functions

of key importance are the following representations for tensor-valued multi-point homogenous and isotropic functions

Lemma C.9.4: homogeneity and isotropy in the statistical sense of tensor-valued symmetric multi-point functions

A second order symmetric tensor-valued multi-point function is *homogenous* if,

$$\sum(\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) = \sum(\mathbf{X}_2 - \mathbf{X}_1, \cdots, \mathbf{X}_{\alpha+1} - \mathbf{X}_\alpha, \cdots, \mathbf{X}_\beta - \mathbf{X}_{\beta-1}) \tag{C329}$$

is a function of difference vector $(\mathbf{X}_2 - \mathbf{X}_1, \cdots, \mathbf{X}_{\alpha+1} - \mathbf{X}_\alpha, \cdots, \mathbf{X}_\beta - \mathbf{X}_{\beta-1})$. This function is isotropic if

$$\begin{aligned}
&\sum (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \\
&= \sigma_0(|\mathbf{X}_1|, <\mathbf{X}_1, \mathbf{X}_2>, <\mathbf{X}_1, \mathbf{X}_3>, \cdots, <\mathbf{X}_{\beta-1}, \mathbf{X}_\beta>, |\mathbf{X}_\beta|)\delta_{ij}\mathbf{e}^i \otimes \mathbf{e}^j \\
&\quad + \sum\nolimits_{\gamma=1}^{\beta}\sum\nolimits_{\mu\gamma}^{\beta} \sigma_{\gamma\mu}(|\mathbf{X}_1|, <\mathbf{X}_1, \mathbf{X}_2>, <\mathbf{X}_1, \mathbf{X}_3>, \cdots, <\mathbf{X}_{\beta-1}, \mathbf{X}_\beta>, |\mathbf{X}_\beta|) \\
&\quad \cdot[(\mathbf{X}_\gamma \otimes \mathbf{X}_\mu) + (\mathbf{X}_\mu \otimes \mathbf{X}_\gamma)]
\end{aligned} \tag{C330}$$

for a certain isotropic scalar-valued function σ_0 and $\sigma_{\gamma\mu}$ for all $1 \leq \gamma \leq \mu \leq \beta$

Finally such a function is *homogenous and isotropic in the statistical sense*, if and only if

$$
\begin{aligned}
&\sum (\mathbf{X}_1, \cdots, \mathbf{X}_\alpha, \mathbf{X}_{\alpha+1}, \cdots, \mathbf{X}_\beta) \\
&= \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|, |\mathbf{X}_3 - \mathbf{X}_1|, \cdots, |\mathbf{X}_\beta - \mathbf{X}_{\beta-2}|, |\mathbf{X}_\beta - \mathbf{X}_{\beta-1}|)\delta_{ij}(\mathbf{e}^i \otimes \mathbf{e}^j) \\
&\quad + \sum\nolimits_{\gamma=2}^{\beta} \sum\nolimits_{\mu\gamma}^{\beta} \sigma_{\gamma\mu}(|\mathbf{X}_2 - \mathbf{X}_1|, |\mathbf{X}_3 - \mathbf{X}_1|, \cdots, |\mathbf{X}_\beta - \mathbf{X}_{\beta-1}|) \\
&\quad \cdot [(\mathbf{X}_\gamma - \mathbf{X}_{\gamma-1}) \otimes (\mathbf{X}_\mu - \mathbf{X}_{\mu-1}) + (\mathbf{X}_\mu - \mathbf{X}_{\mu-1}) \otimes (\mathbf{X}_\gamma - \mathbf{X}_{\gamma-1})] \qquad \text{(C331)}
\end{aligned}
$$

holds for a certain homogenous and isotropic scalar-valued functions σ_0 and $\sigma_{\gamma\mu}$ for all $2 \leq \gamma \leq \mu \leq \beta$

End of Lemma C.9.4: tensor-valued function var-cov functions

For the applications the *Taylor-Karman structure* for the case $\beta = 2$ is most important:

Corollary C.9.5: $\beta = 2$, *Taylor-Karman structure* for two-point tensor function homogeneity and isotropy

A symmetric second order two-point tensor-function is *homogenous and isotropic* in the statistical sense if and only if

$$
\begin{aligned}
&\Sigma\ (\mathbf{X}_1, \mathbf{X}_2) \\
&= \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij}(\mathbf{e}^i \otimes \mathbf{e}^j) \\
&\quad + \sigma_{22}(|\mathbf{X}_2 - \mathbf{X}_1|)[(|\mathbf{X}_1 - \mathbf{X}_2|) \otimes (|\mathbf{X}_2 - \mathbf{X}_1|)] \\
&= \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij} + \sigma_{22}(|\mathbf{X}_2 - \mathbf{X}_1|) < \mathbf{X}_2 - \mathbf{X}_1, \mathbf{e}^i >< \mathbf{X}_2 - \mathbf{X}_1, \mathbf{e}^j > \\
&= \Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)(\mathbf{e}^i \otimes \mathbf{e}^j) \qquad \text{(C332)}
\end{aligned}
$$

(summation convention)

and

$$
\begin{aligned}
\Sigma_{ij} &= \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij} + [\sigma_{22}(|\mathbf{X}_2 - \mathbf{X}_1|)|\mathbf{X}_2 - \mathbf{X}_1|^2 + \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|)] \\
&\quad - \sigma_0(|\mathbf{X}_2 - \mathbf{X}_1|)\frac{\Delta \mathbf{X}_i \Delta \mathbf{X}_i}{|\mathbf{X}_2 - \mathbf{X}_1|^2} \qquad \text{(C333)}
\end{aligned}
$$

$$\Sigma_{ij}(\mathbf{X}_1\mathbf{X}_2) = \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij} + \{[\Sigma_l(|\mathbf{X}_2 - \mathbf{X}_1|) - \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)]\frac{\Delta\mathbf{X}_i\,\Delta\mathbf{X}_i}{|\mathbf{X}_2 - \mathbf{X}_1|^2} \tag{C334}$$

Σ_m and Σ_l are the *homogenous and isotropic scalar-values functions*

End of Corollary C.9.5: $\beta = 2$, Taylor-Karman structure

The representation of the symmetric second-order two-point tensor-valued function under the postulate of homogeneity and isotropy has been analyzed by *G. J. Taylor* (1935) and *T. Karman* (1937) dealing with problems of *Hydrodynamics*. It was introduced into Geodetic sciences by *E. Grafarend*(1970, 1972). Most often the homogenous and isotropic scalar-valued function Σ_m and Σ_l are called

longitudinal and lateral

correlation function. It has been *C.C. Wang* who wrote *first* the symmetric second order two-point tensor-valued function being *homogenous and isotropic* if

$$\Sigma(\mathbf{X}_1\mathbf{X}_2) = a_1(\mathbf{X}_2 - \mathbf{X}_1)\delta_{ij}[e^i\mathbf{X}_1 \otimes e^j\mathbf{X}_2] + a_2(\mathbf{X}_2 - \mathbf{X}_1)[(\mathbf{X}_2 - \mathbf{X}_1) \otimes (\mathbf{X}_2 - \mathbf{X}_1)] \tag{C335}$$

holds. This representations refers to $a_1(\mathbf{X}_2 - \mathbf{X}_1) = b_1(|\mathbf{X}_2 - \mathbf{X}_1|)$ and $a_2(\mathbf{X}_2 - \mathbf{X}_1) = b_2(|\mathbf{X}_2 - \mathbf{X}_1|)$ according to *C. C. Wang* (1970 p, 214/215) as *scalar-valued isotropic functions*. In case we decompose the *difference vector*

$$\begin{aligned} \mathbf{X}_2 - \mathbf{X}_1 &= < \mathbf{X}_2 - \mathbf{X}_1 | e^i(\mathbf{X}_1) >< \mathbf{X}_2 - \mathbf{X}_1, e^j(\mathbf{X}_2) > [e^i\mathbf{X}_1 \otimes e^j\mathbf{X}_2] = \\ &= \Delta\mathbf{X}_i\,\Delta\mathbf{X}_j[e^i\mathbf{X}_1 \otimes e^j\mathbf{X}_2] \end{aligned} \tag{C336}$$

$$\Sigma_{ij}(\mathbf{X}_1\mathbf{X}_2) = b_1(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij} + b_2(|\mathbf{X}_2 - \mathbf{X}_1|)\Delta\mathbf{X}_i\,\Delta\mathbf{X}_j$$

for any choice of the basis e^{i_1}, e^{j_2}. A *special choice* is to put parallel to the basic vector $\mathbf{X}_2 - \mathbf{X}_1$ such that

$$\Sigma_{11} = b_1(|\mathbf{X}_2 - \mathbf{X}_1|) + b_2(|\mathbf{X}_2 - \mathbf{X}_1|)|\mathbf{X}_2 - \mathbf{X}_1|^2 \tag{C337}$$

$$\Sigma_{22} = \Sigma_{22} = b_1(|\mathbf{X}_2 - \mathbf{X}_1|) \tag{C338}$$

The base vector e^2 and e^3 are *orthogonal* to the "*longitudinal direction*" $\mathbf{X}_2 - \mathbf{X}_1$. For such a reason

$$b_1(|\mathbf{X}_2 - \mathbf{X}_1|) := \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|) \tag{C339}$$

and

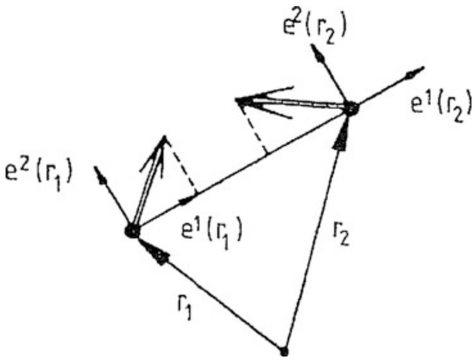


Fig. C15 Longitudinal and lateral correlation: Two-point function

$$b_2(|\mathbf{X}_2 - \mathbf{X}_1|) := \frac{1}{|\mathbf{X}_2 - \mathbf{X}_1|^2}\{\Sigma_l(|\mathbf{X}_2 - \mathbf{X}_1|) - \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)\} \tag{C340}$$

of the scalar-valued isotropic functions $\Sigma_l(|\mathbf{X}_2 - \mathbf{X}_1|)$ and $\Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)$. The result is illustrated in Fig. C.9.1

A *two-dimensional Euclidean example* is the vector-valued one-point function of a network point. Let

$$\Sigma = \begin{bmatrix} \sigma_{x_1x_1} & \sigma_{x_1y_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} \\ \sigma_{y_1x_1} & \sigma_{y_1y_1} & \sigma_{y_1x_2} & \sigma_{y_1y_2} \\ \sigma_{x_2x_1} & \sigma_{x_2y_1} & \sigma_{x_2x_2} & \sigma_{x_2y_2} \\ \sigma_{y_2x_1} & \sigma_{y_2y_1} & \sigma_{y_2x_2} & \sigma_{y_2y_2} \end{bmatrix} \tag{C341}$$

be the variance-covariance matrix of the two-dimensional network of two points with Cartesian coordinates. The variance-covariance matrix will be characterized by the variance $\sigma_{x_1x_1} = \sigma^2_{x_1}$ of the X-coordinate of the first point, the covariance $\sigma_{x_1y_1}$ of the x, y coordinates of the first point and the second point and so on. The *discrete variance-covariance matrix*

$$\Sigma = E\{[\mathbf{X} - E\{\mathbf{X}\}][\mathbf{X} - E\{\mathbf{X}\}]\} \tag{C342}$$

or

$$\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2) = E\{[x_i(\mathbf{X}_1) - E\{x_i\mathbf{X}_1\}][x_j - E\{x_j\mathbf{X}_2\}]\} \tag{C343}$$

or

$$x_i(\mathbf{X}_1) = \{x_1\mathbf{X}_1, x_2\mathbf{X}_1\} =: \{x_1, y_1\}$$
$$x_j(\mathbf{X}_2) = \{x_1\mathbf{X}_2, x_2\mathbf{X}_2\} =: \{x_2, y_2\}$$

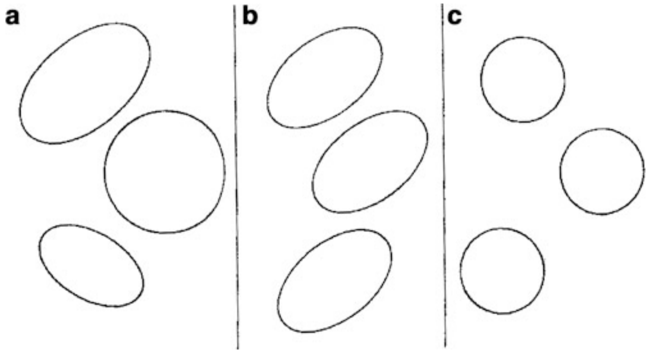


Fig. C16 Homogenous and isotropic dispersion situation: (**a**) general, (**b**)homogenous and (**c**) homogenous-isotropic

are the estimated coordinates of the *Cartesian vectors* of the first network point $\mathbf{X}_1$ or the second network point $\mathbf{X}_2$. The two-point functions $\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)$ are the coordinates labeling the tensor of second order. The *local variances and covariances* are usually illustrated by *local error ellipses*:

$$\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2 = \mathbf{X}_1) \tag{C344}$$

and

$$\Sigma_{ij}(\mathbf{X}_1 = \mathbf{X}_2, \mathbf{X}_2), \qquad \begin{bmatrix} \sigma_{x_1x_1} & \sigma_{x_1y_1} \\ \sigma_{y_1x_1} & \sigma_{y_1y_1} \end{bmatrix} \text{ and } \begin{bmatrix} \sigma_{x_2x_2} & \sigma_{x_2y_2} \\ \sigma_{y_2x_2} & \sigma_{y_2y_2} \end{bmatrix} \tag{C345}$$

The *nonlocal* covariances are contained in the function $\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)$ for the case $\mathbf{X}_1 \neq \mathbf{X}_2$, in particular

$$\begin{bmatrix} \sigma_{x_1x_2} & \sigma_{x_1y_2} \\ \sigma_{y_1x_1} & \sigma_{y_1y_2} \end{bmatrix} \tag{C346}$$

Figure C.9.2 illustrates the postulates of *homogeneity and isotropy* for *variance-covariance matrices of* cartesian coordinates left, we illustrate the *general dispersion situation* in *two-dimensional networks, of course* those *local dispersion ellipses*. *In the center*, the homogenous dispersion situation: At each network points, due the *translation invariance* all *dispersion ellipses are equal.* In the right side, the *homogenous and isotropic dispersion situation of type Taylor-Karman is* finally illustrated.

"All dispersion ellipses are equal and circular"

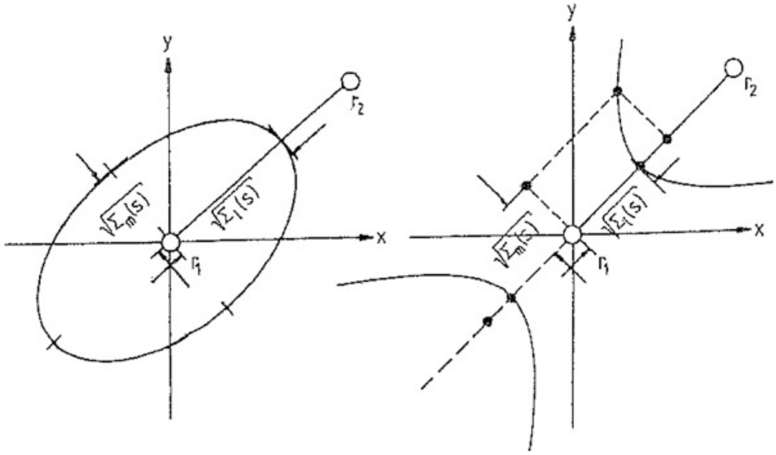


Fig. C17 Correlation figures of type (**a**) ellipse and (**b**) hyperboloid

Finally we illustrate"*longitudinal*" and "*lateral*" within homogenous and isotropic networks by Fig. C16, namely the characteristic figures of type (a) *ellipse* and (b) *hyperboloid*. Those *characteristic*
functions Σ_l and Σ_m are described the *eigen values* of the variance-covariance matrix. *If* the eigenvalues are *positive, then* the correlation figure is an *ellipse*. But if the *two eigen values change sign* (either "+" or "−" or "−" or "+") the characteristic function has to be illustrated by a *hyperboloid*.

Ex 1: Two-Dimensional Euclidean Networks Under the Postulates of Homogeneous and Isotropy in the Statistical Sense

Example C.9.1: Two-dimensional Euclidean networks under the postulates of homogeneity and isotropy in the statistical sense

Two-dimensional network in a Euclidean space are presented in terms of the postulates homogeneity and isotropy in the statistical sense. At the beginning, we assume a variance-covariance matrix of absolute cartesian net coordinates of type homogenous and isotropic, in particular with *Taylor Karman structure*. Given such a dispersion matrix, we *derive* the associated variance-covariance matrix of *coordinate differences*, a special case of the *Kolmogorov structure function*, *and directions, angles and distances*. In order to write more simple formulae, we assume quantities of type $E\{\mathbf{Y}\} = 0$

Definition C.9.6: Variance covariance matrix of Cartesian coordinate, coordinate differences, azimuths, angular observations and distances in two-dimension Euclidean networks

Let $\delta x_i(\mathbf{X}_1)$ be absolute Cartesian coordinates $\delta x_i(\mathbf{X}_1) - \delta x_i(\mathbf{X}_2)$ relative cartesian coordinates or *coordinate differences*, $\delta\alpha(\mathbf{X}_1, \mathbf{X}_2)$ the *azimuths* of a line observation point $\mathbf{X}_1$ to a target point $\mathbf{X}_2$, the difference $\delta\alpha(\mathbf{X}_1, \mathbf{X}_2)$ to $\delta\alpha\mathbf{X}_1, \mathbf{X}_3)$ as *angular observations* or azimuth differences and $\delta s(\mathbf{X}_1\mathbf{X}_2)$ the distances from the point $\mathbf{X}_1$ to the point $\mathbf{X}_2$. Then

$$\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2) = E\{\delta x_i(\mathbf{X}_1)\delta x_j(\mathbf{X}_2)\} \tag{C347}$$

$$\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) = E\{[\delta x_i(\mathbf{X}_1) - \delta x_i(\mathbf{X}_2)][\delta x_j(\mathbf{X}_3) - \delta x_j(\mathbf{X}_4)]\} \tag{C348}$$

$$\Lambda(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = E\{[\delta\alpha(\mathbf{X}_1, \mathbf{X}_2)\delta\alpha(\mathbf{X}_3, \mathbf{X}_4)]\} \tag{C349}$$

$$\begin{aligned}\Lambda(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_6) &= E\{[\delta\alpha(\mathbf{X}_1, \mathbf{X}_2) - \delta\alpha(\mathbf{X}_1, \mathbf{X}_3)]\\ &\quad \times [\delta\alpha(\mathbf{X}_4, \mathbf{X}_5) - \delta\alpha(\mathbf{X}_4, \mathbf{X}_6)]\}\end{aligned}$$

$$\Pi(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) = E\{\delta s(\mathbf{X}_1, \mathbf{X}_2) - \delta s(\mathbf{X}_3, \mathbf{X}_4)\} \tag{C350}$$

are the derived variance-covariance matrices for the *increments* $\mathbf{Y}(x_1, \cdots, x_i) - y(x_1, \cdots, x_i) = \delta y(x_1, \cdots, x_i)\{\delta x_i.\delta x_i - \delta x_j, \delta\alpha, \delta\alpha_i - \delta\alpha_j, \delta s\}$ with respect to a model y.

End of Definition C: derived var-cov matrices

Let the variance-covariance matrix $\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)$ of absolute Cartesian coordinates of the homogenous and isotropic type the characteristic isotropic functions Σ_l and Σ_m given. Then we derive the variance-covariance of

Lemma C.9.7: Representation of derived variance-covariance matrices of *Taylor-Karman* structured variance-covariance of absolute Cartesian coordinates.

Let the variance-covariance matrix of absolute Cartesian $\sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)$ be isotropic and homogenous in terms of the *Taylor-Karman* structure be given. Then the derived variance-covariance matrices are structured according to

$$\begin{aligned}\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) &= \Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_3) - \Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_4)\\ &\quad -\Sigma_{ij}(\mathbf{X}_2, \mathbf{X}_3) - \Sigma_{ij}(\mathbf{X}_2, \mathbf{X}_4)\end{aligned} \tag{C351}$$

$$\Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) = a_i(\mathbf{X}_1, \mathbf{X}_2)a_j(\mathbf{X}_3, \mathbf{X}_4)\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \tag{C352}$$

$$\begin{aligned}\Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5, \mathbf{X}_6) &= \Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_5) - \Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_4, \mathbf{X}_6)\\ &\quad -\Delta(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5) + \Delta(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_6)\end{aligned} \tag{C353}$$

$$\Pi(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) = b_i(\mathbf{X}_1, \mathbf{X}_2)b_j(\mathbf{X}_3, \mathbf{X}_4)\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \tag{C354}$$

subject to

$$\begin{aligned}\Sigma_{ij}(\mathbf{X}_\beta, \mathbf{X}_\alpha) &= \Sigma_m(|\mathbf{X}_\beta - \mathbf{X}_\alpha|)\delta_{ij} + [\Sigma_l(|\mathbf{X}_\beta - \mathbf{X}_\alpha|) \\ &\quad -\Sigma_m(|\mathbf{X}_\beta - \mathbf{X}_\alpha|)]b_i(\mathbf{X}_\beta, \mathbf{X}_\alpha)b_j(\mathbf{X}_\beta, \mathbf{X}_\alpha) \qquad \text{(C355)}\end{aligned}$$

$$a_1(\mathbf{X}_\alpha, \mathbf{X}_\beta) := -[\dot{y}(\mathbf{X}_\alpha) - \dot{y}(\mathbf{X}_\beta)] \div [\dot{s}(\mathbf{X}_\beta, \mathbf{X}_\alpha)] \qquad \text{(C356)}$$

$$a_2(\mathbf{X}_\alpha, \mathbf{X}_\beta) := +[\dot{x}(\mathbf{X}_\alpha) - \dot{x}(\mathbf{X}_\beta)] \div [\dot{s}^2(x_\alpha, x_\beta)] \qquad \text{(C357)}$$

$$b_1(\mathbf{X}_\alpha, \mathbf{X}_\beta) := +[\dot{x}(\mathbf{X}_\alpha) - \dot{x}(\mathbf{X}_\beta)] \div [\dot{s}^2(x_\alpha, x_\beta)] = \cos\dot{\alpha}_{\beta\alpha} \qquad \text{(C358)}$$

$$b_2(\mathbf{X}_\alpha, \mathbf{X}_\beta) := +[\dot{y}(\mathbf{X}_\alpha) - \dot{y}(\mathbf{X}_\beta)] \div [\dot{s}^2(x_\alpha, x_\beta)] = \sin\dot{\alpha}_{\beta\alpha} \qquad \text{(C359)}$$

End of Lemma C.9.7: Taylor-Karman derived structures

For the proof, we apply summation convention with respect to two identical indices. In addition, the derivations are based on

$$\delta\alpha(\mathbf{X}_1, \mathbf{X}_2) = a_1(\mathbf{X}_1, \mathbf{X}_2)[\delta x(\mathbf{X}_1) - \delta x(\mathbf{X}_2)] + a_2(\mathbf{X}_1, \mathbf{X}_2)[\delta y(\mathbf{X}_1) - \delta y(\mathbf{X}_2)] \qquad \text{(C360)}$$

$$\delta s(\mathbf{X}_1, \mathbf{X}_2) = b_1(\mathbf{X}_1, \mathbf{X}_2)[\delta x(\mathbf{X}_1) - \delta x(\mathbf{X}_2)] + b_2(\mathbf{X}_1, \mathbf{X}_2)[\delta y(\mathbf{X}_1) - \delta y(\mathbf{X}_2)] \qquad \text{(C361)}$$

In order to be informed about local variance-covariance we need special results about the situation $\mathbf{X}_1 = \mathbf{X}_3, \mathbf{X}_2 = \mathbf{X}_4$:

Lemma C.9.8: derived *local* variance-covariance matrices of a Taylor-Karman structure variance-covariance matrix of absolute Cartesian coordinates

$$\begin{aligned}\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2) &= 2[\Sigma_m(0)\delta_{ij}(|\mathbf{X}_2 - \mathbf{X}_1|)] \\ &= \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|)\delta_{ij} + [\Phi_l(|\mathbf{X}_2 - \mathbf{X}_1|)] \\ &\quad - \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|)]b_i(\mathbf{X}_1, \mathbf{X}_2)b_j(\mathbf{X}_1, \mathbf{X}_2) \qquad \text{(C362)}\end{aligned}$$

$$\Phi_l(|\mathbf{X}_2 - \mathbf{X}_1|) = 2[\Sigma_l(0) - \Sigma_l(|\mathbf{X}_2 - \mathbf{X}_1|)] \qquad \text{(C363)}$$

$$\Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) = 2[\Sigma_m(0) - \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)] \qquad \text{(C364)}$$

$$\begin{aligned}\Delta(\mathbf{X}_1, \mathbf{X}_2) &= 2a_i(\mathbf{X}_1, \mathbf{X}_2)a_j(\mathbf{X}_1, \mathbf{X}_2)[\Sigma_m(0)\delta_{ij} - \Sigma(\mathbf{X}_1, \mathbf{X}_2)] \\ &= [\Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|)] \div [|\mathbf{X}_2 - \mathbf{X}_1|^2] \qquad \text{(C365)}\end{aligned}$$

$$
\begin{aligned}
\Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) &= \Delta(\mathbf{X}_1, \mathbf{X}_2) - 2\Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_1, \mathbf{X}_3) + \Delta(\mathbf{X}_1, \mathbf{X}_3) \\
&= \frac{\Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|)}{|\mathbf{X}_2 - \mathbf{X}_1|^2} + \frac{\Phi(|\mathbf{X}_3 - \mathbf{X}_1|)}{|\mathbf{X}_3 - \mathbf{X}_1|^2} - a_i(\mathbf{X}_1, \mathbf{X}_2) a_j(\mathbf{X}_1, \mathbf{X}_3) \\
&\quad * [\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2) - \Phi_{ij}(\mathbf{X}_2, \mathbf{X}_3) - \Phi_{ij}(\mathbf{X}_1, \mathbf{X}_3)] \\
&= \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) \frac{< \mathbf{X}_3 - \mathbf{X}_2, \mathbf{X}_3 - \mathbf{X}_1}{|\mathbf{X}_2 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_1|^2} \\
&\quad - \Phi_m(|\mathbf{X}_3 - \mathbf{X}_2|) \frac{< \mathbf{X}_2 - \mathbf{X}_1, \mathbf{X}_3 - \mathbf{X}_2 >}{|\mathbf{X}_2 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_2|^2} \\
&\quad + \Phi_m(|\mathbf{X}_3 - \mathbf{X}_2|) \frac{< \mathbf{X}_2 - \mathbf{X}_1, \mathbf{X}_3 - \mathbf{X}_1 >}{|\mathbf{X}_2 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_1|^2} \\
&\quad + [\Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) - \Phi(|\mathbf{X}_3 - \mathbf{X}_2|)] \\
&\quad * \frac{[|\mathbf{X}_2 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_1|^2 - < \mathbf{X}_2 - \mathbf{X}_1, \mathbf{X}_3 - \mathbf{X}_1 >]}{[|\mathbf{X}_2 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_1|^2 |\mathbf{X}_3 - \mathbf{X}_2|^2} \qquad \text{(C366)}
\end{aligned}
$$

$$
\Pi(\mathbf{X}_1, \mathbf{X}_2) - 2b_i(\mathbf{X}_1, \mathbf{X}_2) | \Sigma_m(0)\delta_{ij} - \Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2) = \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) \qquad \text{(C367)}
$$

End of Lemma C 9.8: Derived local var-cov matrices

The results can be interpreted as following.

First, the scalar-valued functions $\Delta(\mathbf{X}_1, \mathbf{X}_2), \Delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ and $\Pi(\mathbf{X}_1, \mathbf{X}_2)$ are *homogenous and isotropic*. *Second*, of the *local variance-covariance matrix of Cartesian coordinate differences* has the structure of *nonlocal covariance matrix of absolute Cartesian coordinates* as long as we depart from a *Taylor-Karman matrix*: *Third*, not every homogenous and isotropic *local* variance-covariance matrix of *Cartesian coordinate differences* produces derived homogenous and isotropic functions as we illustrated in our next example.

Of course, *not* every homogenous and isotropic local variance-covariance matrix of type *coordinate difference* has the structure of being derived by the transformation of absolute coordinates into coordinate difference. For instance, for the special case $\mathbf{X}_1 = \mathbf{X}_2$ the variance-covariance matrix $\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2)$ due to $\Sigma_m(0) = \Sigma_l(0)$ can be characterized by *circles*. The limits

$$
\Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_1) = \Sigma_{ij}(\mathbf{X}_1, \lim_{\mathbf{X}_2 \to \mathbf{X}_1} \mathbf{X}_2) = \lim_{\mathbf{X}_2 \to \mathbf{X}_1} \Sigma_{ij}(\mathbf{X}_1, \mathbf{X}_2) \qquad \text{(C368)}
$$

are fixed for any direction of approximation. *However*, due to $\Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) \neq \Phi_l(|\mathbf{X}_2 - \mathbf{X}_1|)$ lead us to *symmetry breaking in a two-dimensional EUCLIDEAN* *space*: Though we have started from a homogenous and isotropic variance-covariance matrix of *absolute Cartesian coordinates into* relative coordinates or coordinate differences the variance-covariance matrix $\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ is *not* homogenous and isotropic, a result mentioned first by *W. Baarda* (1977): *The error eclipses are not circular*. In a *three dimensional Euclidean space*. In contrast, we receive *rotational symmetric ellipsoids* based on the difference vector $\mathbf{X}_2 - \mathbf{X}_1$ with respect to the axis of rotation and circles as intersection surfaces orthogonal to the vector $\mathbf{X}_2 - \mathbf{X}_1$.

In order to summarize our results we present

Lemma C.9.9: Circular two-point functions

The image of the two-point function $\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2) = \Phi(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_1, \mathbf{X}_2)$ is a circle if and only if $\Phi_l(|\mathbf{X}_2 - \mathbf{X}_1|) = \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|)$ or $\Sigma_l(|\mathbf{X}_2 - \mathbf{X}_1|) = \sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)$

End of Lemma C.9.9: Circular two-point functions

The result is illustrated by an example: Assume *network points* with a distance function $|\mathbf{X}_2 - \mathbf{X}_1| = 1$. In this distance the longitudinal *and* the lateral correlation functions are assumed to the values 1/2 *and* 1/4. Then there holds

$$\Sigma_l(0) = \Sigma_m(0) = 1 \tag{C369}$$

$$\Sigma_l = \frac{1}{2}, \quad \Sigma_m = \frac{1}{4} \tag{C370}$$

$$\Phi_l = 1, \quad \Phi_m = \frac{3}{2}, \quad \Delta(1) = \frac{3}{2}, \quad \Pi(1) = 1 \tag{C371}$$

Lemma C.9.10: Derived local variance-covariance matrix for a variance-covariance matrix of absolute Cartesian coordinate of *Special Taylor-Karman structure* $\Sigma_l = \Sigma_m$

Let the variance-covariance matrix of absolute Cartesian coordinates with *Special Taylor-Karman structure* $\Sigma_l = \Sigma_m$ for homogenous and isotropic signals given. Then the derived variance-covariance matrices are structured accordingly to

$$\begin{aligned}\Phi_{ij}(\mathbf{X}_1, \mathbf{X}_2) &= 2[\Sigma_m(0) - \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)]\delta_{ij} \\ &= \Phi(|\mathbf{X}_2 - \mathbf{X}_1|)]\delta_{ij}\end{aligned} \tag{C372}$$

$$\Phi_l(|\mathbf{X}_2 - \mathbf{X}_1|) = \Phi_m(|\mathbf{X}_2 - \mathbf{X}_1|) = 2[\Sigma_m(0) - \Sigma_m(|\mathbf{X}_2 - \mathbf{X}_1|)] \tag{C373}$$

$$\begin{aligned}\Delta(\mathbf{X}_1,\mathbf{X}_2) &= \frac{2}{|\mathbf{X}_2-\mathbf{X}_1|^2}[\Sigma_m(0)-\Sigma_m(|\mathbf{X}_2-\mathbf{X}_1|)] \\ &= \frac{1}{|\mathbf{X}_2-\mathbf{X}_1|}\Phi_m(|\mathbf{X}_2-\mathbf{X}_1|) \end{aligned} \tag{C374}$$

$$\begin{aligned}\Delta(\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3) &= \frac{1}{|\mathbf{X}_2-\mathbf{X}_1|^2|\mathbf{X}_3-\mathbf{X}_1|^2} \\ &\quad *[\Phi_m(|\mathbf{X}_2-\mathbf{X}_1|) < \mathbf{X}_3-\mathbf{X}_2|\mathbf{X}_3-\mathbf{X}_1 > - \\ &\quad - \Phi_m(|\mathbf{X}_3-\mathbf{X}_1|) < \mathbf{X}_2-\mathbf{X}_1, \mathbf{X}_3-\mathbf{X}_2 > \\ &\quad + \Phi_m(|\mathbf{X}_3-\mathbf{X}_2|) < \mathbf{X}_3-\mathbf{X}_1, \mathbf{X}_2-\mathbf{X}_1 >] \end{aligned} \tag{C375}$$

$$\Pi(\mathbf{X}_1,\mathbf{X}_2) = 2[\Sigma_m(0)-\Sigma_m(|\mathbf{X}_2-\mathbf{X}_1|)] = \Phi_m(|\mathbf{X}_2-\mathbf{X}_1|) \tag{C376}$$

End of Lemma C 9.10: Special Taylor-Karman matrix

Our example is characterized now by a homogenous and isotropic variance-covariance matrix of *Cartesian absolute coordinates* which are the *gradient of a homogenous and isotropic scalar-valued function*. In addition, we assume that the scalar valued function is an element of a *two-dimensional Markov process of first order*. Its variance-covariance matrix is characterized by normalized longitudinal and lateral correlation function which possess the distance function $s := |\mathbf{X}_2 - \mathbf{X}_1|$ between network points d the characteristic *distance*, K_0 and K_1 the modified Bessel function of the second kind of zero and first order. The *characteristic distance* d is the *only variable or degree of freedom*. Within the parameter d we depend on our experience with geodetic networks. Due to the property $\Sigma_l(0) = \Sigma_l(0)$ we call the *longitudinal and lateral normalized*. In order to find the scale for the variance-covariance matrix we have to multiply by the variance σ^2. The *modified Bessel function* K_0 and K_1 are tabulated in Table C18 in the range 0.1(0.1)8 where we followed *M. Abramowtz and I. Stegun* (1970 p. 378-379).

In Fig. C18 we present the characteristic functions of type absolute coordinates $\Sigma_l(X) = -4x^{-2} + 2K_0(X) + 4x^{-1}K_1(X) + 2xK_1(X)$ and $\Sigma_m(X) = +4x^{-2} - 2K_0(X) - 2K_0(X) - 4x^{-1}K_1(X)$. Next is the representation of the derived functions $\Phi_l(X_) = \Pi(X) = 2\Sigma_m(0) - 2\Sigma_l(X)$, $\Phi_m(X) = 2\Sigma_m(0) - 2\Sigma_m(X)$ and $\Delta(X) = 2x^{-2}|\Sigma_m(0) - \Sigma_m(X)| = x^{-2}\Phi_m(X)$. Figures C19 and C20 illustrate the various variance-covariance functions.

An explicit example is a two-dimensional network in Fig. C21. Let point one be the center of a fixed (x,y) Cartesian coordinate system. In consequence, we measure nine points by distance observations assuming the distance between point one and point two is unity, namely

$$1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}$$

The related characteristic functions for a characteristic distance d=1 are approximately given in Table C18. The variance-covariance matrix of absolute Cartesian

(x)	$K_0(x)$	$K_1(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_5(x)$
0	∞	∞					
0,1	2,42704	9,85361	—	—	—	—	—
0,2	1,75269	4,77601	0,93598	0,97442	0,05116	0,12803	1,27900
0,3	1,37240	3,05598	0,88034	0,95324	0,09351	0,23931	1,03901
0,4	1,11448	2,18435	0,81994	0,92754	0,14492	0,36012	0,90575
0,5	0,92441	1,65640	0,75642	0,89998	0,20004	0,48716	0,80016
0,6	0,77754	1,30282	0,69282	0,87056	0,25887	0,61436	0,71909
0,7	0,66052	1,05023	0,62941	0,84091	0,31818	0,74118	0,64934
0,8	0,56533	0,86174	0,56814	0,81064	0,37872	0,86371	0,59175
0,9	0,48679	0,71660	0,51008	0,77980	0,44039	0,97985	0,54370
1,0	0,42097	0,60193	0,45352	0,75034	0,49932	1,09296	0,49932
1,1	0,36552	0,50972	0,40017	0,72122	0,55756	1,19967	0,46080
1,2	0,31856	0,43464	0,35128	0,69186	0,61628	1,29744	0,42798
1,3	0,27819	0,37259	0,30468	0,66405	0,67189	1,39064	0,39757
1,4	0,24366	0,32091	0,26194	0,63661	0,72678	1,47613	0,37081
1,5	0,21380	0,27740	0,22176	0,61044	0,77911	1,55649	0,34627
1,6	0,18796	0,24063	0,18501	0,58501	0,82999	1,62998	0,32421
1,7	0,16550	0,20936	0,15135	0,56047	0,87906	1,69729	0,30417
1,8	0,14593	0,18262	0,12055	0,53689	0,92623	1,75891	0,28587
1,9	0,12885	0,15966	0,09250	0,51421	0,97159	1,81500	0,26914
2,0	0,11390	0,13986	0,06696	0,49248	1,01504	1,86608	0,25376
2,1	0,10078	0,12274	0,04383	0,47168	1,05664	1,91234	0,23960
2,2	0,08927	0,10790	0,02304	0,45172	1,09665	1,95393	0,22656
2,3	0,07914	0,09499	0,00429	0,43266	1,13467	1,99142	0,21449
2,4	0,07021	0,08372	−0,01264	0,41449	1,17102	2,02527	0,20330
2,5	0,06234	0,07389	−0,02765	0,39710	1,20581	2,05529	0,19293
2,6	0,05540	0,06528	−0,04103	0,38049	1,23903	2,08206	0,18329
2,7	0,04926	0,05774	−0,05284	0,36464	1,27073	2,10568	0,17431
2,8	0,04383	0,05111	−0,06331	0,34953	1,30094	2,12663	0,16594
2,9	0,03900	0,04529	−0,07247	0,33516	1,32969	2,14495	0,15811
3,0	0,03475	0,04015	−0,08051	0,32141	1,35718	2,16102	0,15080
3,1	0,03096	0,03564	−0,08736	0,30833	1,38335	2,17472	0,14395
3,2	0,02760	0,03164	−0,09338	0,29588	1,40825	2,18676	0,13752
3,3	0,02461	0,02812	−0,09841	0,28400	1,43199	2,19683	0,13150
3,4	0,02196	0,02500	−0,10269	0,27269	1,45462	2,20538	0,12583
3,5	0,01960	0,02224	−0,10623	0,26191	1,47617	2,21247	0,12050
3,6	0,01750	0,01980	−0,10908	0,25164	1,49672	2,21816	0,11549
3,7	0,01563	0,01763	−0,11140	0,24186	1,51627	2,22281	0,11076
3,8	0,01397	0,01571	−0,11314	0,23253	1,53494	2,22627	0,10630
3,9	0,01248	0,01400	−0,11447	0,22367	1,55267	2,22893	0,10208
4,0	0,01116	0,01248	−0,11536	0,21520	0,56960	2,23072	0,09810
4,1	0,00998	0,01114	−0,11578	0,20713	1,5875	2,23155	0,09433
4,2	0,00893	0,00994	−0,11593	0,19943	1,60114	2,23187	0,09077
4,3	0,00799	0,00887	−0,11582	0,19210	1,61580	2,23164	0,08739
4,4	0,00715	0,00792	−0,11542	0,18511	1,62978	2,23083	0,08418
4,5	0,00640	0,00708	−0,11472	0,17844	1,64312	2,22944	0,08114
4,6	0,00573	0,00633	−0,11384	0,17207	1,65586	2,22767	0,07825
4,7	0,00513	0,00565	−0,11290	0,16601	1,66798	2,22580	0,07551
4,8	0,00460	0,00505	−0,11172	0,16020	1,67959	2,22345	0,07290
4,9	0,00412	0,00452	−0,11037	0,15467	1,69067	2,22074	0,07042
5,0	0,00369	0,00404	−0,10899	0,14939	1,70122	2,217[illegible]8	0,06805
5,1	0,00331	0,00362	−0,10740	0,14433	1,71134	2,21481	0,06580
5,2	0,00297	0,00324	−0,10580	0,13950	1,72101	2,21160	0,06365
5,3	0,00266	0,00290	−0,10415	0,13489	1,73022	2,20830	0,06160
5,4	0,00238	0,00260	−0,10241	0,13049	1,73902	2,20482	0,05964
5,5	0,00214	0,00233	−0,10063	0,12626	1,74749	2,20125	0,05777
5,6	0,00192	0,00208	−0,09893	0,12223	1,75555	2,19786	0,05598
5,7	0,00172	0,00187	−0,09704	0,11836	1,76327	2,19409	0,05427
5,8	0,00154	0,00167	−0,09530	0,11467	1,77065	2,19060	0,05264
5,9	0,00139	0,00150	−0,09341	0,11111	1,77777	2,18683	0,05107
6,0	0,00124	0,00134	−0,09166	0,10774	1,78452	2,18332	0,04957
6,1	0,00112	0,00120	−0,08983	0,10447	1,79106	2,17966	0,04813
6,2	0,00100	0,00108	−0,08797	0,10136	1,79728	2,17594	0,04676
6,3	0,00090	0,00097	−0,08614	0,09837	1,80327	2,17229	0,04543
6,4	0,00081	0,00087	−0,08436	0,09549	1,80902	2,16871	0,04417
6,5	0,00073	0,00078	−0,08259	0,09273	1,81453	2,16519	0,04295
6,6	0,00065	0,00070	−0,08086	0,09010	1,81979	2,16173	0,04178
6,7	0,00058	0,00063	−0,07913	0,08757	1,82486	2,15826	0,00654
6,8	0,00053	0,00056	−0,07750	0,08512	1,82977	2,15500	0,03957
6,9	0,00047	0,00051	−0,07574	0,08278	1,83444	2,15148	0,03853
7,0	0,00042	0,00045	−0,07424	0,08054	1,83893	2,14847	0,03753
7,1	0,00038	0,00041	−0,07254	0,07836	1,84328	2,14507	0,03657
7,2	0,00034	0,00037	−0,07095	0,07627	1,84745	2,14189	0,03564
7,3	0,00031	0,00033	−0,06944	0,07426	1,85148	2,13888	0,03474
7,4	0,00028	0,00030	−0,06788	0,07232	1,85535	2,13577	0,03388
7,5	0,00025	0,00027	−0,06642	0,07047	1,85907	2,13283	0,03305
7,6	0,00022	0,00024	−0,06504	0,06869	1,86263	2,13008	0,03225
7,7	0,00020	0,00021	−0,06372	0,06696	1,86609	2,12744	0,03147
7,8	0,00018	0,00019	−0,06232	0,06529	1,86942	2,12465	0,03073
7,9	0,00016	0,00017	−0,06100	0,06369	1,87263	2,12200	0,03001
8,0	0,00015	0,00016	−0,05956	0,06212	1,87576	2,11912	0,02931

Fig. C18 Modified Bessel function $K_0(X)$ and $K_1(X)$ longitudinal and lateral function of absolute Cartesian coordinates $f_1(X) =: \Sigma_l(X)$ and $f_2(X) =: \Sigma_m(X)$, longitudinal and lateral correlation of relative Cartesian coordinates $f_3 := \Phi_m(X)$ and $f_4 := \Phi_l(X)$, variances of directions $f_5 := \Delta(X)$ and distances $\Pi(X) =: \Delta(X)$ of a two-dimensional conservation Markov process of first order

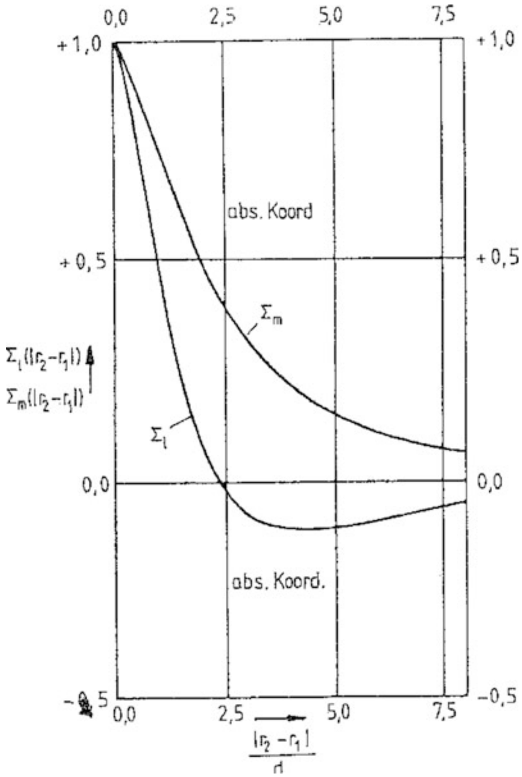


Fig. C19 Characteristic functions σ_l and Σ_m for *Taylor-Karman* structured varaince-covariance matrix of *Cartesian coordinates*

coordinates is relative Cartesian coordinates or coordinate differences is computed with respect to $\sigma^2 = 1.09$, $\sigma_{y_1-y_2} = 0.50$, $\sigma_{x_1-x_2,y_1-y_2} = 0$

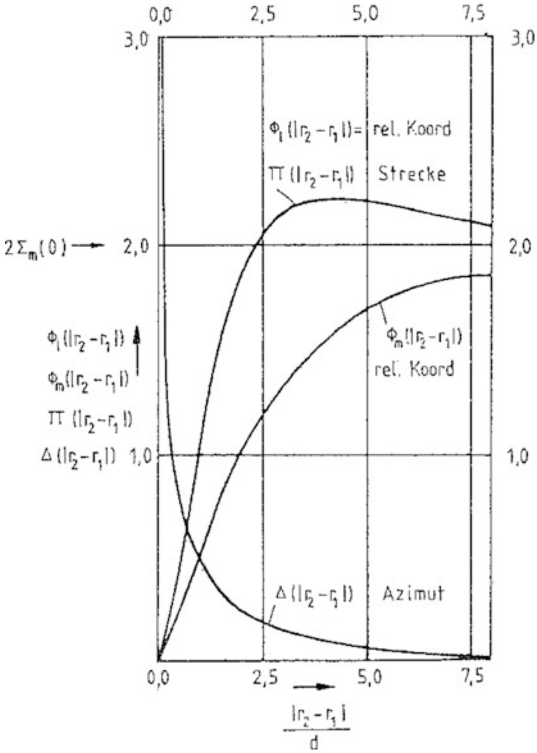
We added by purpose according to Fig. C19 the characteristic functions for the *special homogenous and isotropic variance-covariance matrix* given by $\Sigma_l = \Sigma_m$. The associated variance-covariance matrix for absolute and relative Cartesian coordinates is given in Tables C25 and C26, structured by

$$\Sigma_l = \frac{4d^2}{s^2} + 2K_0(\frac{s}{d}) + \frac{4d}{s}K_1(\frac{s}{d}) + 2(\frac{s}{d})K_1(\frac{s}{d}) \tag{C377}$$

$$\Sigma_m = \frac{4d^2}{s^2} + 2K_0(\frac{s}{d}) + \frac{4d}{s}K_1(\frac{s}{d}) \tag{C378}$$

Finally we present the derived variance-covariance matrices for relative Cartesian coordinates, direction s, direction differences namely angles and distances, for the *special case* $\Sigma_l = \Sigma_m$.

Fig. C20 Derived variance-covariance functions for *Taylor-Karman* structured variance-covariance matrix of *Cartesian coordinates*



$$
\begin{aligned}
\Phi_l(s) &= \Pi(s) = 2\Sigma_m(0) - 2\Sigma_m(s), \\
&= 2 + \left(\frac{8d^2}{s^2}\right) - 4K_0\left(\frac{s}{d}\right) - \left(\frac{8d}{s}\right)K_1\left(\frac{s}{d}\right) - 4\left(\frac{s}{d}\right)K_1\left(\frac{s}{d}\right)
\end{aligned} \tag{C379}
$$

$$
\begin{aligned}
\Phi_m(s) &= \Pi(s) = 2\Sigma_m(0) - 2\Sigma_m(s), \\
&= 2 - \left(\frac{8d^2}{s^2}\right) - 4K_0\left(\frac{s}{d}\right) + \left(\frac{8d}{s}\right)K_1\left(\frac{s}{d}\right),
\end{aligned} \tag{C380}
$$

$$
\begin{aligned}
\Delta(s) &= \Pi(s) = \frac{2}{s^2}\left[\Sigma_m(0) - \Sigma_m(s)\right], \\
&= \frac{2}{s^2}\left[1 - \left(\frac{4d^2}{s^2}\right) - 2K_0\left(\frac{s}{d}\right) + 4\left(\frac{s}{d}\right)K_1\left(\frac{s}{d}\right)\right]
\end{aligned} \tag{C381}
$$

Example 9.1 is based on our contribution *E. Grafarend and B. Schaffrin* (1979) on the subject of criterion matrices of type *homogenous and isotropic*. Other statistical properties like *anisotropic* or *non-isotropic* correlation functions were treated by *G. K. Batchelor* (1946), *S. Chandrasekhar* (1950) and E. Grafarend (1971, 1972) on the subject of *axisymmetric turbulence*.

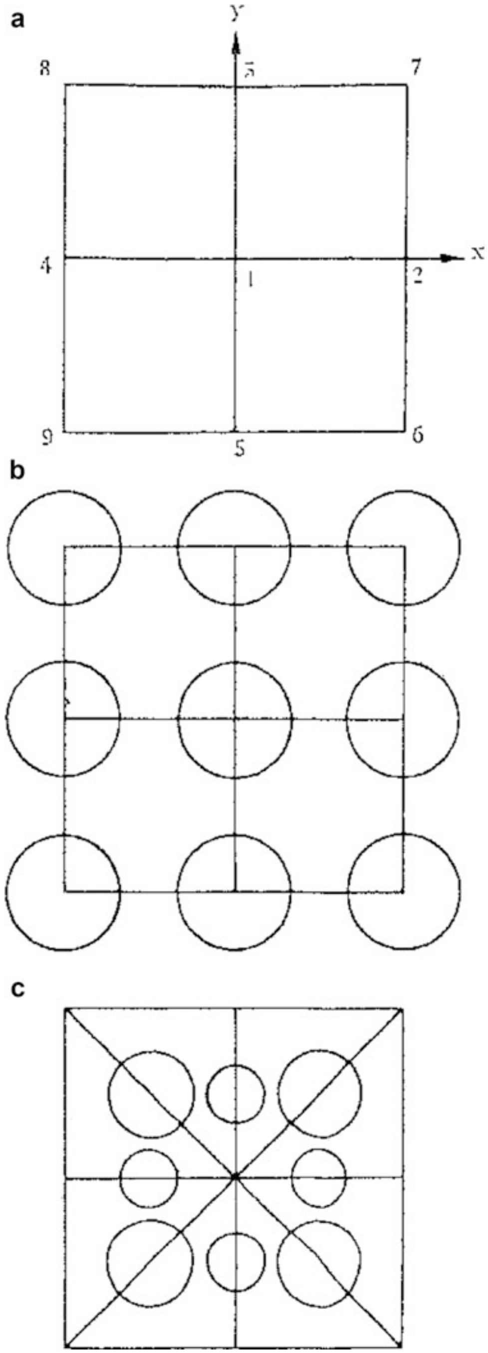


Fig. C21 Two dimensional network

Fig. C22 Characteristic functions of a network example

s/d	Σ_l	Σ_m	Φ_l	Φ_m	Π	Δ
1	0,45	0,75	1,09	0,50	1,09	0,50
$\sqrt{2}$	0,26	0,63	1,47	0,72	1,47	0,37
2	0,07	0,49	1,87	1,02	1,87	0,25
$\sqrt{5}$	0,02	0,45	1,97	1,11	1,97	0,21
$\sqrt{8}$	−0,06	0,35	2,13	1,30	2,13	0,16

Fig. C23 Cartesian coordinate data for a network example

Point	x	y
1	0	0
2	1	0
3	0	1
4	−1	0
5	0	−1
6	1	−1
7	1	1
8	−1	1
9	−1	−1

Of course there is a rich list of references for "criterion matrices". We list only few of them.

J. E. Alberta (1974), *A. B. Amiri* (2004), *W. Baarda* (1973, 1977), *K. Bill* (1985), *R. Bill et. al.* (1986), *Bischoff et. al.* (2006), *G. Boedecker* (1977), *F. Bossler et. al.* (1973), *F. Comte et. al.* (2002), *P. Cross et. al.* (1978), *G. Gaspari et. al.* (1999), *E. Grafarend* (1971, 1972, 1975, 1976), *E. Grafarend, F. Krumm and B. Schaffrin* (1985), *E. Grafarend and B. Schaffrin* (1979), *T. Karman* (1937), *T. Karman et. al.* (1939), *R. Kelm* (1976), *R. M. Kerr* (1985), *S. Kida* (1989), *F. Krumm et. al* (1982, 1986), *S Meier* (1976, 1977, 1978, 1981), *S. Meier and W. Keller* (1990), *J. van Mierlo* (1982), *M. Molenaar* (1982), *A. B. Obuchow* (1958), *T. Reubelt et. al.* (2003), *H. P. Robertson* (1940), *B. Schaffrin and E. Grafarend* (1977, 1981), *B. Schaffrin and E. Grafarend* (1982), *B. Schaffrin et. al.* (1977), *G. Schmitt* (1977, 1978 i, ii), *J. F. Schoeberg* (1938), *J. P. Shkarofky* (1968), *V. I. Tatarski* (1961), *A. I. Taylor* (1935, 1938), *C. C. Wang* (1970), *H. Wimmer* (1972, 1978), *M. J. Yadrenko* (1983) and *A. A. Yaglom* (1987).

Ex 2: Criterion Matrices for Absolute Coordinates in a Space Time Holonomic Frame of Reference, for Instance Represented in Scalar-Valued Spherical Harmonics

Example C.9.2: Criterion matrices for absolute coordinates in a space time holonomic frame of reference, for instance represented in scalar-valued spherical harmonics

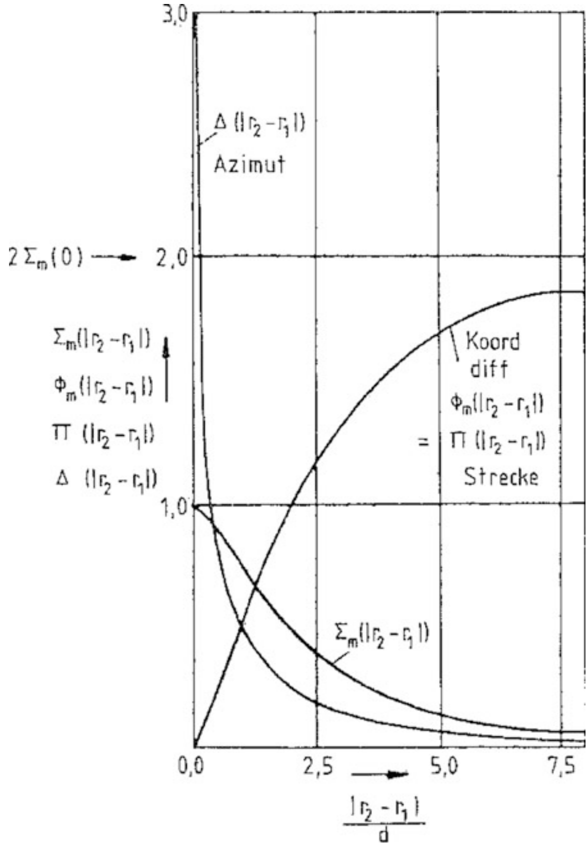


Fig. C24 Derived variance-covariance functions for a special Taylor-Karman structure variance-covariance matrix of type $\Sigma_l = \Sigma_m$ for Cartesian coordinates

Conventional geodetic networks have been adjusted by methods of *Total Least Squares* and optimized in different reference frames. The advantage of a rigourously three dimensional adjustment in a uniform reference frame is obvious, especially under the influence of subdtic geodetic networks within various *Global Positioning Systems* (GPS, *Global Problem Solver*). Numerical results in a three dimensional, *Total Least Squares* approach are presented for instance by *K. Bergbauer* et. al. (1979) *K. Euler* et. al. (1982), *E. Grafarend* (1991), *G. Hein and H. Landau* (1983), and by *W. Jorge and H. G. Wenzel* (1973).

The *diagnosis* of a geodetic network is rather complicated since the *measure of quality*, namely the *variance-covariance matrix* of three dimensional coordinates of Cartesian type contains $3g(3g+1)/2$ *independent elements*. The integer number g accounts for the *number of ground stations*.

$\boxed{\text{Example: g = 100}}$

Fig. C25 Homogenous isotropic normalized variance-covariance matrix of absolute coordinates from a network example

	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4	x_5	y_5	x_6	y_6	x_7	y_7	x_8	y_8	x_9	y_9
x_1	1	0	0,45	0	0,75	0	0,45	0	0,75	0	0,45	0,18	0,18	−0,18	0,45	0,18	0,45	−0,18
y_1		1	0	0,75	0	0,45	0	0,75	0	0,45	0,18	0,45	−0,18	0,45	0,18	0,45	−0,18	0,45
x_2			1	0	0,45	0,18	0,07	0	0,45	−0,18	0,75	0	0,75	0	0,11	0,18	0,11	0,18
y_2				1	0,18	0,45	0	0,63	−0,18	0,45	0	0,45	0,18	0,36	−1,18	0,36	−0,18	0,36
x_3					1	0	0,45	−0,18	0,49	0	0,36	0,18	0,45	0	0,45	0	0,36	−0,18
y_3						1	−0,18	0,44	0	0,07	0,18	0,11	0	0,75	0	0,75	−0,18	0,11
x_4							1	0	0,24	0,18	0,11	0,18	0,11	−0,18	0,75	0	0,75	0
y_4								1	0,09	0,44	0,18	0,36	−0,18	0,36	0	0,45	0	0,45
x_5									1	0	0,45	0	0,36	−0,18	0,36	0,18	0,45	0
y_5										1	0	0,75	−0,18	0,11	0,75	0,11	0	0,75
x_6											1	0	0,49	0	0,15	0,20	0,07	0
y_6												1	0	0,07	0,20	0,15	0	0
x_7													1	0	0,07	0	0,15	−0,20
y_7														1	0	0,49	−0,20	0,15
x_8															1	0	0,49	0
y_8																1	0	0,07
x_9																	1	0
y_9																		1

	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4	x_5	y_5	x_6	y_6	x_7	y_7	x_8	y_8	x_9	y_9
x_1	1	0	0,75	0	0,75	0	0,75	0	0,75	0	0,63	0	0,63	0	0,63	0	0,63	0
y_1		1	0	0,75	0	0,75	0	0,75	0	0,75	0	0,63	0	0,63	0	0,63	0	0,63
x_2			1	0	0,63	0	0,49	0	0,63	0	0,75	0	0,75	0	0,45	0	0,45	0
y_2				1	0	0,63	0	0,49	0	0,63	0	0,75	0	0,75	0	0,45	0	0,45
x_3					1	0	0,63	0	0,49	0	0,45	0	0,75	0	0,75	0	0,45	0
y_3						1	0	0,63	0	0,49	0	0,45	0	0,75	0	0,75	0	0,45
x_4							1	0	0,63	0	0,45	0	0,45	0	0,75	0	0,75	0
y_4								1	0	0,63	0	0,45	0	0,45	0	0,75	0	0,75
x_5									1	0	0,75	0	0,45	0	0,45	0	0,75	0
y_5										1	0	0,75	0	0,45	0	0,45	0	0,75
x_6											1	0	0,49	0	0,35	0	0,49	0
y_6												1	0	0,49	0	0,35	0	0,49
x_7													1	0	0,49	0	0,35	0
y_7														1	0	0,49	0	0,35
x_8															1	0	0,49	0
y_8																1	0	0,49
x_9																	1	0
y_9																		1

Fig. C26 Special homogenous isotropic normalized variance-covariance matrix of *absolute coordinates* for a network example case $\Sigma_l = \Sigma_m$

	x_1-x_2	y_1-y_2	x_1-x_3	y_1-y_3	x_1-x_4	y_1-y_4	x_1-x_5	y_1-y_5	x_1-x_6	y_1-y_6	x_1-x_7	y_1-y_7	x_1-x_8	y_1-y_8	x_1-x_9	y_1-y_9
x_1-x_2	0,50	0	0,13	0	0	0	0,13	0	0,37	0	0,37	0	0,07	0	0,07	0
y_1-y_2		0,50	0	0,13	0	0	0	0,13	0	0,37	0	0,37	0	0,07	0	0,07
x_1-x_3			0,50	0	0,13	0	0	0	0,07	0	0,37	0	0,37	0	0,07	0
y_1-y_3				0,50		0,13	0	0	0	0,07	0	0,37	0	0,37	0	0,07
x_1-x_4					0,50	0	0,13	0	0,07	0	0,07	0	0,37	0	0,37	0
y_1-y_4						0,50	0	0,13	0	0,07	0	0,07	0	0,37	0	0,37
x_1-x_5							0,50	0	0,37	0	0,07	0	0,07	0	0,37	0
y_1-y_5								0,50	0	0,37	0	0,07	0	0,07	0	0,37
x_1-x_6									0,72	0	0,23	0	0,09	0	0,23	0
y_1-y_6										0,72	0	0,23	0	0,09	0	0,23
x_1-x_7											0,72	0	0,23	0	0,09	0
y_1-y_7												0,72	0	0,23	0	0,09
x_1-x_8													0,72	0	0,23	0
y_1-y_8														0,72	0	0,23
x_1-x_9															0,72	0
y_1-y_9																0,72

Fig. C27 Special homogenous isotropic normalized variance-covariance matrix of *relative coordinates* for a network example case $\Sigma_l = \Sigma_m$

The number of independent elements with in the coordinate variance-covariance matrix *amounts to 45,150.*

$\boxed{\text{Example: g = 1,000}}$

The number of independent elements within the coordinate variance-covariance matrix *amounts to 4,501,500.*

$\boxed{\text{Example: g = 5,000 (GPS)}}$

The number of independent elements within the coordinate variance-covariance matrix amounts to 125,507,500 !
In a definition, the same number of unknown variance-covariance elements of vertical deflections and gravity disturbances.

The open big problem

In light of the great number of elements of the variance-covariance matrix it is an open question how to achieve *any measure of quality* of geodetic network based on GPS. We think the only way to get away from this number of millions of unknowns is to *design an idealized* variance-covariance matrix again called *criterion matrix*. Comparing the *real* situation of the variance-covariance with an *idealized* one, we receive an *objective measure of quality of the geodetic network*.

Here we are aiming for designing a *three-dimensional criterion matrices* for absolute coordinates in a *holonomic frame of reference* and for functionals of the "*disturbing potential*". In a *first section* we extensively discuss the notion of criterion matrix on a surface of a moving frame of type *homogenous and isotropic on the sphere* which is a *generalization of the classical Taylor-Karman structure* for a vectorized stochastic process. The *vector values stochastic process on the sphere is found*. In detail we review the representation of *the criterion matrix in scalar- and vector-valued spherical harmonics*. In *section two* we present the criterion matrices for *functionals of the disturbing potential*. The property of harmonicity is taken into account for the case that a representation in terms of *scalar-valued spherical harmonics exists*. At the end, we present a numerical example.

Example C.9.21: Criterion matrices for absolute coordinates in a space time holonomic frame if reference

The variance-covariance matrix of absolute Cartesian coordinates in the fixed orthonormal frame of equatorial type, holonomic in spacetime, is transformed into a moving frame, e.g. of spherical type. In the moving frame *longitudinal and lateral characteristic functions* are derived. The postulates of *spherical* homogenity and *spherical* isotropy are reviewed extending to the classical *Taylor-Karman structure* with respect to three dimensional Cartesian coordinates into three dimensions. Finally the *spherical harmonic representation* of the coordinate criterion matrix is introduced.

Transformation of the coordinate criterion matrix from a fixed into a moving frame of reference

$$\boxed{\mathbf{F}^{\bullet} \to \mathbf{e}^{*} :}$$

We depart from the fixed orthonormal frame $\{\underline{E}_{1\bullet}, \underline{E}_{2\bullet}, \underline{E}_{3\bullet}\}$ which we transform into a moving one $\{\underline{e}_{1*}, \underline{e}_{2*}, \underline{e}_{3*}, \}$ by

$$\boxed{\mathbf{F}^{\bullet} \to \mathbf{e}^{*} = \mathbf{R}_E(\lambda, \phi, 0)\mathbf{F}^{\bullet}} \tag{C382}$$

$\mathbf{R}_E(\lambda, \phi, 0) := \mathbf{R}_3(0)\mathbf{R}_2(\frac{\pi}{2} - \phi)\mathbf{R}_3(\lambda)$ is the three dimensional euler rotation matrix including geometric longitude λ and geometric latitude ϕ. As an example assume a representation of a placement vector $\underset{\sim}{x}$ by spherical coordinates (λ, ϕ, r) such that

$$\mathbf{x} = \mathbf{F}_{1\bullet} r \cos\lambda \cos\phi + \mathbf{F}_{2\bullet} r \sin\lambda \cos\phi + \mathbf{F}_{3\bullet} r \sin\phi \tag{C383}$$

$$\begin{aligned} \mathbf{e}_{1*} &= -\mathbf{x}_{,\phi} \div \|\mathbf{x}_{,\phi}\| = -\mathbf{e}_{\phi} \\ \mathbf{e}_{2*} &= -\mathbf{x}_{,\lambda} \div \|\mathbf{x}_{,\lambda}\| = -\mathbf{e}_{\lambda} \\ \mathbf{e}_{3*} &= -\mathbf{x}_{,r} \div \|\mathbf{x}_{,r}\| = -\mathbf{e}_{r} \end{aligned} \tag{C384}$$

$\mathbf{x}_{,\phi}$ denotes $\delta x / \delta \phi$ etc.

$$\boxed{\mathbf{e}^{*} \to \mathbf{f}^{*}}$$

Now let us rotate the orthonormal frame $\{\mathbf{e}_{1*}, \mathbf{e}_{2*}, \mathbf{e}_{3*},\}$ into the orthonormal frame $\{\mathbf{f}_{1*}, \mathbf{f}_{2*}, \mathbf{f}_{3*},\}$ by means of the South azimuth α

$$\begin{aligned} \mathbf{e}^{*} \to \mathbf{f}^{*} = \mathbf{R}_3(\alpha) e^{*} &\sim \begin{bmatrix} \mathbf{f}_{1*} \\ \mathbf{f}_{2*} \\ \mathbf{f}_{3*} \end{bmatrix} \\ &= \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{e}_{1*} \\ \mathbf{e}_{2*} \\ \mathbf{e}_{3*} \end{bmatrix} \end{aligned} \tag{C385}$$

The transformation $\mathbf{e}^{*} \to \mathbf{f}^{*}$ on the unit sphere orientates $\mathbf{f}_{1*}$ into the direction of the great circle connecting the points (λ, ϕ, r) and (λ', ϕ', r') For this reason the unit vector $\mathbf{f}_{1*}$ will be called *"longitudinal direction"*. In contrast, $\mathbf{f}_{2*}$ points in direction orthogonal to $\mathbf{f}_{1*}$, but still in the tangential plane of the unit sphere. Thus we call the unit vector $\mathbf{f}_{2*}$ *"lateral direction"*. Obviously $\mathbf{f}_{3*} = \mathbf{e}_{3*}$ holds

$$\mathbf{F}^{\bullet} \to \mathbf{g} \to \mathbf{f}^{*}$$

There is an alternative way to connect the triads $\mathbf{F}^{\bullet}$ and $\mathbf{f}^{*}$. Consider the *spherical* position of a point in the "orbit" of the great circle passing through the points $\mathbf{x}, \mathbf{x}'$, respectively. For a given set of *"Keplerian elements"* $\{\Omega, i, \omega\}$ - Ω the right ascension of the ascending node, i the inclination and ω the angle measured in the "orbit" - we transform the fixed orthonormal frame $\{\mathbf{F}_{1\bullet}, \mathbf{F}_{2\bullet}, \mathbf{F}_{3\bullet}\}$ into the orthonormal "great circle moving frame" $\{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3\} = \{\mathbf{f}_{3*}, \mathbf{f}_{1*}, \mathbf{f}_{2*},\}$ by

$$\mathbf{F}^{\bullet} \to \mathbf{g} = \mathbf{R}_3(\omega)\mathbf{R}_1(i)\mathbf{R}_3(\Omega)\mathbf{F} \tag{C386}$$

$$\mathbf{g} \to \mathbf{F}^{\bullet} = \mathbf{R}_3(-\Omega)\mathbf{R}_1(-i)\mathbf{R}_3(-\omega)\mathbf{F} \tag{C387}$$

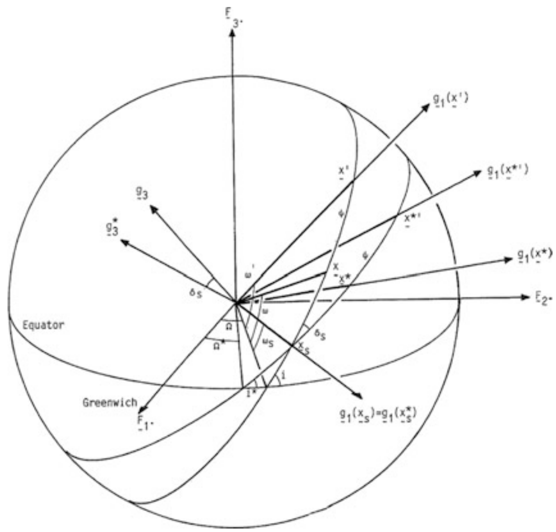


Fig. C28 Moving and fixed frame of reference on the sphere : two point correlation function

or explicitly

$$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix} = \begin{bmatrix} \cos\omega\cos\Omega - \sin\omega\cos i\sin\Omega & \cos\omega\sin\Omega + \sin\omega\cos i\cos\Omega & +\sin\omega\sin i \\ -\sin\omega\cos\Omega - \cos\omega\cos i\sin\Omega & -\sin\omega\sin\Omega + \cos\omega\cos i\cos\Omega & \cos\omega\sin i \\ \sin i\sin\Omega & -\sin i\cos\Omega & \cos i \end{bmatrix} \tag{C388}$$

$$\cdot \begin{bmatrix} \mathbf{F}_{1\bullet} \\ \mathbf{F}_{2\bullet} \\ \mathbf{F}_{3\bullet} \end{bmatrix} \tag{C389}$$

$$\mathbf{g}_3 = \sin i\sin\Omega\mathbf{F}_{1\bullet} - \sin i\cos\Omega\mathbf{F}_{2\bullet} + \cos i\mathbf{F}_{3\bullet} \tag{C390}$$

represents the "*great circle normal*" parameterized by the *longitude* $\frac{\pi}{2} - \Omega$ and the *latitude* $\frac{\pi}{2} - i$. $\mathbf{g}_1$ and $\mathbf{g}_2$ span the "*great circle plane*", namely $\mathbf{g}_1$ is directed towards the spherical point $\mathbf{x}$.

Figure C28 illustrates the various fames of type *"fixed"* and *"moving"*.

$\{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3\} = \{\mathbf{f}_{3*}, \mathbf{f}_{1*}, \mathbf{f}_{2*}, \}$, $\mathbf{g}_2$ is the vector product of $\mathbf{g}_3$ and $\mathbf{g}_1$

Transformation of variance-covariance matrices

Next we compute the Cartesian coordinates variance-covariance matrix.

$$\begin{aligned}\mathbf{\Sigma}(\mathbf{x},\mathbf{x}') &= \mathbf{\Sigma}_{i\bullet j\bullet}(\mathbf{x},\mathbf{x}')\{\mathbf{F}_{i\bullet}\otimes\mathbf{F}_{j\bullet}\}\\ \mathbf{\Sigma}(\mathbf{x},\mathbf{x}') &= \mathbf{\Sigma}_{i*j*}(\mathbf{x},\mathbf{x}')\{\mathbf{f}_{i*}(\mathbf{x})\otimes\mathbf{f}_{j*}(\mathbf{x}')\}\\ \mathbf{\Sigma}(\mathbf{x},\mathbf{x}') &= \mathbf{\Sigma}_{ij}(\mathbf{x},\mathbf{x}')\{\mathbf{g}_i(\mathbf{x})\otimes\mathbf{g}_j(\mathbf{x}')\}\end{aligned}\tag{C391}$$

where$i, j\varepsilon\{1,2,3\}$ holds

$$(\Sigma^{ij'}) = \begin{bmatrix}\Sigma_{\mathbf{x},\mathbf{x}'} & \Sigma_{\mathbf{x},\mathbf{y}'} & \Sigma_{\mathbf{x},\mathbf{z}'}\\ \Sigma_{\mathbf{y},\mathbf{x}'} & \Sigma_{\mathbf{y},\mathbf{y}'} & \Sigma_{\mathbf{y},\mathbf{z}'}\\ \Sigma_{\mathbf{z},\mathbf{x}'} & \Sigma_{\mathbf{z},\mathbf{y}'} & \Sigma_{\mathbf{z},\mathbf{z}'}\end{bmatrix}\tag{C392}$$

$$(\Sigma^{ij'}) = E\{[x^i - E\{x^i\}][X^{j'} - E\{X^{j'}\}]\} = E\{\varepsilon^i\varepsilon^{j'}\}\tag{C393}$$

$$\Sigma(\mathbf{x},\mathbf{x}') := E\{\varepsilon(\mathbf{x})\varepsilon^T(\mathbf{x}')\}\tag{C394}$$

Note that the variance-covariance matrix $\Sigma^{ij'}(\mathbf{x},\mathbf{x}')$ is a two-point tensor field paying attention to the fact that the coordinate errors $\varepsilon(\mathbf{x})$ and $\varepsilon(\mathbf{x})$ are functions of the point $\mathbf{x},\mathbf{x}'$, respectively. Now let us transform $\Sigma^\bullet \leftrightarrow \Sigma^* \leftrightarrow \Sigma$.

$$\begin{aligned}\Sigma^\bullet_{3\times 0} &= \mathbf{R}_E(\lambda,\phi,\alpha)\Sigma^*_{3\times 3}\mathbf{R}_E^T(\lambda',\phi',\alpha')\\ &= \mathbf{R}_3(-\Omega)\mathbf{R}_1(-i)\mathbf{R}_3(-\omega)\Sigma_{3\times 3}\mathbf{R}_3(-\omega')\mathbf{R}_1^T(-i')\mathbf{R}_3^T(-\Omega')\\ \Sigma^*_{3\times 0} &= \mathbf{R}_E(\lambda,\phi,\alpha)\Sigma^\bullet\mathbf{R}_E^T(\lambda',\phi',\alpha')\\ &= \mathbf{R}_E(\lambda,\phi,\alpha)\Sigma^*\mathbf{R}(-\omega,-i,-\Omega)\Sigma_{3\times 3}\mathbf{R}^T(-\omega',-i',-\Omega')\cdot\mathbf{R}_E^T(\lambda',\phi',\alpha')\\ \Sigma_{3\times 0} &= \mathbf{R}^T(-\omega',-i',-\Omega')\Sigma_{3\times 3}\mathbf{R}(-\omega,-i,-\Omega)\\ &= \mathbf{R}^T(-\omega',-i',-\Omega')\mathbf{R}_E^T(\lambda,\phi,\alpha)\Sigma^*_{3\times 3}\mathbf{R}_E(\lambda',\phi',\alpha')\cdot\mathbf{R}(-\omega',-i',-\Omega')\end{aligned}\tag{C395}$$

A detailed write-up of the variance-covariance matrix Σ^* is

$$(\Sigma^{ij'}) = \begin{bmatrix}\Sigma_{ll'} & \Sigma_{lm'} & \Sigma_{lr'}\\ \Sigma_{ml'} & \Sigma_{mm'} & \Sigma_{mr'}\\ \Sigma_{rl'} & \Sigma_{rm'} & \Sigma_{rr'}\end{bmatrix}\tag{C396}$$

indicating the *"longitudinal"* base vector $\mathbf{f}_{1*} = \mathbf{f}_{l*}$, *"lateral"* base vector $\mathbf{f}_{2*} = \mathbf{f}_{m*}$ and the *"radical"* base vector $\mathbf{f}_{3*} = \mathbf{f}_{r*} = \mathbf{e}_r$. In terms of spherical coordinates Σ^*, Σ, respectively, can be written

$$\Sigma^*_{3\times 0}(\mathbf{x},\mathbf{x}') = \Sigma^*(\lambda,\phi,r,\lambda',\phi',r')\tag{C397}$$

$$\Sigma_{3\times 0}(\mathbf{x},\mathbf{x}') = \Sigma(\omega,i,\Omega,r,\omega',i',\Omega',r')\tag{C398}$$

$$\begin{aligned}\Sigma_{ll'} &= E\{\varepsilon_l(\mathbf{x})\varepsilon_{l'}(\mathbf{x}')\},\\ &,\ \Sigma_{ll'}(\mathbf{x},\mathbf{x}')\{\mathbf{f}_l(\mathbf{x})\otimes\mathbf{f}_{l'}(\mathbf{x}')\}\end{aligned}$$

$$\begin{aligned}
\Sigma_{lm'} &= E\{\varepsilon_l(\mathbf{x})\varepsilon_{m'}(\mathbf{x}')\}, \\
&, \ \Sigma_{lm'}(\mathbf{x},\mathbf{x}')\{\mathbf{f}_l(\mathbf{x}) \otimes \mathbf{f}_{m'}(\mathbf{x}')\} \\
\Sigma_{lr'} &= E\{\varepsilon_l(\mathbf{x})\varepsilon_{r'}(\mathbf{x}')\}, \\
&, \ \Sigma_{lr'}(\mathbf{x},\mathbf{x}')\{\mathbf{f}_l(\mathbf{x}) \otimes \mathbf{f}_{r'}(\mathbf{x}')\}
\end{aligned} \tag{C399}$$

The characteristic function $\Sigma_{ll'}(\mathbf{x},\mathbf{x}')$ is called *longitudinal correlation function*, $\Sigma_{mm'}(\mathbf{x},\mathbf{x}')$ *lateral correlation function*, $\Sigma_{lm'}(\mathbf{x},\mathbf{x}')$ *cross correlation function* between the longitudinal error $\varepsilon_l(\mathbf{x})$ at the point $\mathbf{x}$ and the lateral error $\varepsilon_{m'}(\mathbf{x}')$ at the point $\mathbf{x}'$.

Spherically homogenous and spherically isotropic coordinate criterion matrices in a surface moving frame

terrestrial geodetic networks constitute points on the earths surface. Once we want to develop an idealized variance-covariance matrix of the coordinates of the network points, we have to reflect the restriction given by the closed surface. A *three dimensional criterion matrix* of Taylor-Karman type is therefore not suited since it assumes extension of a network in three dimension up to infinity. a more suitable concept of a coordinate criterion matrix starts from the representation where the radial part is treated separately from the longitudinal-lateral part. Again the nature of the coordinate criterion matrix as a *second order two-point tensor field* will help us to construct homogenous and isotropic representations. The spherically homogeneity-isotropy-postulate reflects the intuitive notion os an idealized variance-covariance matrix.

Definition (spherical homogeneity and spherical isotropy of a tensor-valued two-point function):

The two-point second order tensor function $\Sigma(\mathbf{x},\mathbf{x}')$ is called spherically homogenous, if it is translational invariant in the sense of

$$\mathbf{\Sigma}_{3\times3}(\omega+\tau, i, \Omega, r, \omega'+\tau, i', \Omega', r') = \mathbf{\Sigma}_{3\times3}(\omega, i, \Omega, r, \omega', i', \Omega', r') \tag{C400}$$

for all $\tau \in [0, 2\pi], i = i', \Omega = \Omega', r = r'$, where $\omega' = \omega + \psi$ holds. It is called spherically isotropic, if it is rotational invariant in the sense of

$$\begin{aligned}
&\Sigma_{3\times3}(\omega^* - \omega_s^*, i^*, \Omega^*, r, \omega^{*'} - \omega_s^*, i^{*'}, \Omega^{*'}, r') \\
&= R_3(\delta_s)\Sigma_{3\times3}(\omega - \omega_s, i, \Omega, r, \omega' - \omega_s, i', \Omega', r')R_3(\delta_s)
\end{aligned} \tag{C401}$$

for all $\delta\varepsilon[0, 2\pi], r = r'$, where $\omega' = \omega + \psi$, $\omega^{*'} = \omega^* + \psi$ holds. $R_3(\delta_s)$ indicates a local rotation around the intersection points $\mathbf{x}_s$ of the great circle through $\mathbf{x}, \mathbf{x}'$ and through $\mathbf{x}^*, \mathbf{x}^{*'}$, respectively. $\mathbf{x}_s$ has the common coordinates $(\Omega, i, \omega_s$ and $(\Omega^*, i*, \omega_s*$ with respect to the locally rotated frames.

End of Definition

Corollary *(spherical homogeneity and spherical isotropy of a tensor-valued two-point function)*:

The two-point second order tensor function $\Sigma(\mathbf{x}, \mathbf{x}')$ *is spherically homogenous, if it is a function of* (ψ, i, Ω, r) *only:*

$$\Sigma(\omega, i, \Omega, r, \omega', i', \Omega', r) = \Sigma(\omega, i, \Omega, r, \omega - \omega_s, \psi, r) \tag{C402}$$

It is spherically homogenous and isotropic, if it is a function of (ψ, r) *only*

$$\Sigma(\omega, i, \Omega, r, \omega', i', \Omega', r) = \Sigma(\psi, i, \Omega, r) \tag{C403}$$

such that

$$\begin{aligned} \Sigma(\mathbf{x}, \mathbf{x}') = \{&\Sigma_0(\psi, r)\delta^{ij} + \Sigma_{22}(\psi, r) < \mathbf{x}' - \mathbf{x}|\mathbf{f}_i > \cdot \\ &\cdot < \mathbf{x}' - \mathbf{x}|\mathbf{f}_i >\}(\mathbf{f}_i \otimes \mathbf{f}_j) \end{aligned} \tag{C404}$$

holds for some spherically isotropic and homogenous scalar-valued functions Σ_0 *and* Σ_{22}

End of Corollary

Let us interpret the definition of the corollary, especially with respect to Fig. C28. Two points $\mathbf{x}, \mathbf{x}'$ are connected to the sphere of radius $r = r' = r_0$ by a great circle such that $\psi = 2 \arcsin(\|\mathbf{x}, \mathbf{x}'\|/2r_0)$ is the spherical distance if these points.

Obviously a "translation" of the configuration $\mathbf{x}, \mathbf{x}'$ is generated by $\omega \to \omega + \tau, \omega \to \omega' + \tau$ where the location of $\mathbf{x}, \mathbf{x}'$ in the great circle / "orbit" plane is measured by ω, ω'. The "rotation" of the configuration $\mathbf{x}, \mathbf{x}'$ is more complicated. Assume therefore an alternative configuration $\mathbf{x}^*, \mathbf{x}^{*'}$ of two point connected by another great circle. *Note the spherical distances of* $\mathbf{x}, \mathbf{x}'$ and $\mathbf{x}^*, \mathbf{x}^{*'}$ *respectively, coincide*. The great circles intersect at $\mathbf{x}_s$, : The location of $\mathbf{X}_s$ within the plane $\mathbf{x}, \mathbf{x}'$ is described by ω_s with respect to the ascending node, while within the plane $\mathbf{x}^*, \mathbf{x}^{*'}$ by ω_s^*. the angle between the two great circles is denoted by δ_s.

The configuration $\mathbf{x}, \mathbf{x}'$ and $\mathbf{x}^*, \mathbf{x}^{*'}$ is described with respect to $\mathbf{x}^s$ by $\omega - \omega_s, i, \Omega, \omega' - \omega_s, i', \Omega'$ and $\omega^* - \omega_s^*, i^*, \Omega^*, \omega^{*'} - \omega_s^*, i^{*'}, \Omega^{*'}$ respectively. Now it should be obvious that the two configurations under the spherical isotropy postulate coincide once we rotate around $\mathbf{x}_s$ by the angle δ_s.

We expressed $\Sigma^\bullet$. given in the $F^\bullet$-frame in terms of the g-frame, namely by

$$\Sigma = R_3^T(-\omega)R_1^T(-i)R_3^T(-\Omega)\Sigma^\bullet R_3(-\Omega)R_1(-i)R_3(-\omega) \tag{C405}$$

Introducing $\omega = (\omega - \omega_s) + \omega_s$ we compute the variance-covariance matrix Σ

$$\Sigma = \mathbf{R}_3^T(\omega - \omega_s)\mathbf{R}_3^T(\omega_s)\mathbf{R}_1^T(-i)\mathbf{R}_3^T(-\Omega)\Sigma^\bullet\mathbf{R}_3(-\Omega')\mathbf{R}_1(-i')\mathbf{R}_3(\omega_s)\mathbf{R}_3(\omega' - \omega_s) \tag{C406}$$

$$\Sigma = \mathbf{R}_3^T(\omega - \omega_s)\Sigma_s\mathbf{R}_3(\omega' - \omega_s) \tag{C407}$$

in terms of the variance-covariance matrix Σ_s in the $\mathbf{g}(\mathbf{x}_s)$-frame. Now we are prepared to introduce

Corollary (spherical Taylor-Karman structure):

Denote the variance-covariance matrix $\Sigma^{i*j*'}$ in the f^*- and e^*-frame by

$$f^{\Sigma^{i*j*'}} = \begin{bmatrix} \Sigma_{ll'}(\psi, r) & 0 \\ 0 & \Sigma_{mm'}(\psi, r) \end{bmatrix} \sim f^{\Sigma^{2\times 2}} \tag{C408}$$

$$e^{\Sigma^{i*j*'}} = \begin{bmatrix} \Sigma_{11'} & \Sigma_{12'} \\ \Sigma_{21'} & \Sigma_{22'} \end{bmatrix} \sim e^{\Sigma^{2\times 2}} \tag{C409}$$

and transform $f^\Sigma \to e^\Sigma$ by

$$e^\Sigma = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} f^\Sigma = \begin{bmatrix} \cos\alpha' & \sin\alpha' \\ -\sin\alpha' & \cos\alpha' \end{bmatrix} \tag{C410}$$

Then

$$\begin{aligned}
\Sigma_{11'} &= \Sigma_{ll'}(\psi, r)\cos\alpha\cos\alpha' + \Sigma_{mm'}(\psi, r)\sin\alpha\sin\alpha' \\
\Sigma_{12'} &= \Sigma_{ll'}(\psi, r)\cos\alpha\sin\alpha' - \Sigma_{mm'}(\psi, r)\sin\alpha\cos\alpha' \\
\Sigma_{21'} &= \Sigma_{ll'}(\psi, r)\sin\alpha\cos\alpha' - \Sigma_{mm'}(\psi, r)\cos\alpha\sin\alpha' \\
\Sigma_{22'} &= \Sigma_{ll'}(\psi, r)\sin\alpha\sin\alpha' + \Sigma_{mm'}(\psi, r)\cos\alpha\cos\alpha'
\end{aligned} \tag{C411}$$

$$\Sigma_{lm'} = \Sigma_{ml'} = 0 \tag{C412}$$

holds for spherically homogenous and isotropic scalar-valued functions $\Sigma_{ll'}(\psi, r)$, $\Sigma_{mm'}(\psi, r)$ called longitudinal and lateral correlation functions on the sphere.

End of Corollary

The corollary extends results of *G.J. Taylor* (1935), *T. Karman* (1937) and C. Wang (1970 p. 214-215) and has been given for a special variance-covariance two-point tensor field of *potential type* by *H. Moritz* (1972 p. 111). Here it has been generalized for any vector-valued field. Note that incase of homogenity and isotropy the cross-correlation functions $\Sigma_{lm'}$ and $\Sigma_{ml'}$ may vanish.

The classical *planar Taylor-Karman structure* is constrained within.

Let us put $\alpha = \alpha'$ such that $\cos\alpha\cos\alpha' = \cos^2\alpha = 1 - \sin^2\alpha$, $\sin\alpha\sin\alpha' = \sin^2\alpha$ etc. can be rewritten

$$\begin{aligned}
\Sigma_{11'} &= \Sigma_{mm'}(\psi) + [\Sigma_{ll'}(\psi) - \Sigma_{mm'}(\psi)]\cos^2\alpha \\
\Sigma_{12'} &= [\Sigma_{ll'}(\psi) - \Sigma_{mm'}(\psi)]\sin\alpha\cos\alpha = \Sigma_{21'} \\
\Sigma_{22'} &= \Sigma_{mm'}(\psi) + [\Sigma_{ll'}(\psi) - \Sigma_{mm'}(\psi)]\sin^2\alpha
\end{aligned} \tag{C413}$$

which is the standard Taylor-Karman form

Representation of the coordinate criterion matrix in scalar spherical harmonics

Let us assume a coordinate variance-covariance matrix which is homogenous and isotropic in the sphere. With reference to (C412) we postulate a coordinate vector of *potential* type, in addition, namely

$$\varepsilon = \operatorname{grad} R \sim \varepsilon^i = \partial_i R \quad (i = 1, 2, 3) \tag{C414}$$

where $(\partial_i) = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ nd R the scalar potential. Now the variance-covariance matrix of absolute coordinates can be represented by

$$\Sigma^{ij'} = E\{\varepsilon^i(\mathbf{x})\varepsilon^{j'}(\mathbf{x}')\} = \partial_i\partial_{j'}E\{R(\mathbf{x})R(\mathbf{x}')\} \tag{C415}$$

where we used the commutator relation $\partial_i E = E\partial_i$ of differentiation and integration operators. Denote the scalar-valued variance-covariance function $E\{R(\mathbf{x})R(\mathbf{x}')\} =: K(\mathbf{x}, \mathbf{x}')$. Once we postulate homogeneity and isotropy of two-point function $K(\mathbf{x}, \mathbf{x}')$ on the sphere we arrive at

$$K(\mathbf{x}, \mathbf{x}') = K(r, r', \psi) \tag{C416}$$

For a harmonic coordinate error potential, $\partial_i\partial_i R = 0$, we find for the covariance function $K(r, r', \psi)$

$$\partial_i\partial_i K(r, r', \psi) = E\{\partial_i\partial_i R(\mathbf{x})R(\mathbf{x}')\} = 0 \tag{C417}$$

$$\partial_{j'}\partial_{j'} K(r, r', \psi) = E\{R(\mathbf{x})\partial_{j'}\partial_{j'} R(\mathbf{x}')\} = 0 \tag{C418}$$

$\boxed{\textit{Case 1}\ (\|\mathbf{x}\| = \|\mathbf{x}'\|)}$

On a sphere $(\|\mathbf{x}\| = \|\mathbf{x}'\|) = r_0$ the scalar-valued variance-covariance function $K(\psi)$ which fulfils (C414 - C 417) may be represented by

$$K(\psi) = \sum_{n=0}^{\infty} k_n P_n(\cos\psi) \tag{C419}$$

where P_n are the Legendre polynomials. The constants k_n depend, of course, on the choice of r_0

$\boxed{\textit{Case 2 } (\|\mathbf{x}\| \neq \|\mathbf{x}'\|)}$

Denote $(\|\mathbf{x}\| = r, \|\mathbf{x}'\| = r')$. the scalar-valued variance-covariance function $K(r, r', \psi)$ which fulfils (C414 - C 417) may be represented by

$$\begin{aligned} K &= K(r, r', \psi) \\ &= \sum_{n=0}^{\infty} {}^{1}k_n \left(\frac{r_0^2}{rr'}\right)^{n+1} P_n(\cos\psi) + \sum_{n=0}^{\infty} {}^{2}k_n \left(\frac{rr'}{r_0}\right)^{n} P_n(\cos\psi) \end{aligned} \tag{C420}$$

where the coefficients ${}^{1}k_n, {}^{2}k_n$ represent the influence of the *decreasing and increasing solution,* by Legendre polynomials with respect to r, r', respectively. We can only hope that the covariances $K(r, r', \psi)$ *decrease* with respect to r, r', respectively in order to be allowed to discriminate ${}^{2}k_n = 0$. In the following we mae this assumption, thus representing

$$K(\mathbf{x}, \mathbf{x}') = K(r, r', \psi) = \sum_{n=0}^{\infty} k_n \left(\frac{r_0^2}{rr'}\right)^{n+1} P_n(\cos\psi) \tag{C421}$$

Next let us assume the coordinate variance-covariance matrix (C420) is of potential type. Then longitudinal and lateral correlation functions on the sphere if radius r can be represented by

$$\Sigma_{ll'}(\psi) = \frac{d^2}{d\psi^2} K(\psi) \tag{C422}$$

$$\Sigma_{mm'}(\psi) = \frac{1}{\sin\psi} \frac{1}{d\psi} K(\psi) \tag{C423}$$

$$\Sigma_{ll'}(\psi) = \frac{d}{d\psi} (\Sigma_{mm'} \sin \tag{C424}$$

if $\sigma(\mathbf{x}, \mathbf{x}')$ is homogenous and isotropic on the sphere.

End of Corollary

Note that the result $\Sigma_{ll'} = \frac{d}{d\psi}(\sin\psi \Sigma_{mm'})$ for the sphere corresponds to $\Sigma_{ll'} = \frac{d}{ds}(s\Sigma_{mm'})$ for a harmonic field in the plane.

$s = \sqrt{(x'-x)^2 + (y'-y)^2}$ is the Euclidean distance in the plane which corresponds to ψ at the distance on the sphere.

Corollary (longitudinal and lateral correlation functions on the sphere of harmonic type):

Let us assume that the coordinate variance-covariance matrix id homogenous, isotropic and harmonic on the sphere in the sense of the definition. Longitudinal and lateral correlation functions on the sphere can be represented by

$$\begin{aligned}\Sigma_{ll'}(\psi) = \sum_{n=0}^{\infty} k_n \Bigg\{ \left[\frac{(n^2+3n+2)\cos^2\psi}{\sin^2\psi} + (n+1) \right] . \\ \cdot P_n(\cos\psi) - [n^2+3n+2] \frac{2\cos\psi}{\sin^2\psi} P_{n+1}(\cos\psi) \Bigg\} \\ + [n^2+3n+2] \frac{1}{\sin^2\psi} P_{n+2}(\cos\psi)\}\end{aligned} \tag{C425}$$

$$\begin{aligned}\Sigma_{mm'}(\psi) = -\sum_{n=0} \infty k_n \frac{n+1}{\sin^2\psi} [\cos\psi P_n(\cos\psi) \\ - P_{n+1}(\cos\psi)\end{aligned} \tag{C426}$$

End of Corollary

For the proof of the above corollary we have made successive use of the recurrence relation $(x^2-1)dP_n(x)/dx = -(n+1)[xP_n(x) - P_{n+1}(x)]$, $x = \cos\psi, d/d\psi = (dx/d\psi)(d/dx) = -\sin\psi d/dx$

Representation of the coordinate criterion matrix in vector spherical harmonics

Previously we restricted the coordinate criterion matrix being homogenous and isotropic on the sphere to be of *potential type*. Here we generalize the coordinate criterion matrix to be of *potential and "turbulent" type*. Thus according to the Lame lemma – for more details see *A. Ben Menahem - S. J. Singh* (1981) - we represent the coordinate vector

$$\varepsilon = gradR + rot\mathbf{A} \tag{C427}$$

where we gauge the *vector potential* **A** by $div\mathbf{A} = 0$. In vector analysis $gradR$ is called the longitudinal vector field, $rot\mathbf{A}$ the transversal vector field. In order to base the analysis only on *scalar* potentials, the source-free vector field $rot\mathbf{A}$ is decomposed $\mathbf{S}+\mathbf{T}$ where **S** is the *poloidal* part while **T** is the *toroidal* part such that

$$\varepsilon = gradR + rotrot(\mathbf{e}_r\mathbf{S}) + rot(\mathbf{e}_r\mathbf{T}) \tag{C428}$$

(R, S, T) are the three scalar potentials we were looking for. The variance-covariance matrix of absolute coordinates can now be structured according to

$$
\begin{aligned}
\Sigma(\mathbf{x},\mathbf{x}') &= E\{\varepsilon(\mathbf{x})\} \\
&= grad \otimes grad' E\{R(\mathbf{x})R(\mathbf{x}')\} \\
&\quad + rotrot \otimes rot'rot'\mathbf{e}_r\mathbf{e}_{r'} E\{S(\mathbf{x})S(\mathbf{x}')\} \\
&\quad + rot \otimes rot' e_r e_{r'} E\{T(\mathbf{x})T(\mathbf{x}')\}
\end{aligned} \tag{C429}
$$

where we have assumed $E\{R(\mathbf{X})S(\mathbf{X}')\} = 0$, $E\{R(\mathbf{X})T\ (\mathbf{X}')\} = 0$, $E\{S(\mathbf{X})T(\mathbf{X}')\} = 0$ holds for a coordinate variance-covariance matrix $\Sigma(\mathbf{xx}')$ homogenous and isotropic on the sphere. Thus with respect to the more general representation is

$$
\boxed{
\begin{aligned}
\Sigma(\mathbf{x},\mathbf{x}') &= grad \otimes grad' K \\
&\quad + rotrot \otimes rot'rot'\mathbf{e}_r\mathbf{e}_{r'} L \\
&\quad + rot \otimes rot'\mathbf{e}_r\mathbf{e}_{r'} M
\end{aligned} \qquad \text{(C430)}
}
$$

where $K := E\{R(\mathbf{X})R(\mathbf{X}')\}, L := E\{S(\mathbf{X})S(\mathbf{X}')\}, M := E\{T(\mathbf{X})T(\mathbf{X}')\}$ are supposed to be the scalar-valued two point covariance functions depending on (r, r', ψ) for $(\Sigma(\mathbf{x},\mathbf{x}'))$ homogenous and isotropic on the sphere.

The representation of the criterion matrix $(\Sigma(\mathbf{x},\mathbf{x}'))$ in terms of vector spherical harmonics is achieved once we introduce the *spherical surface harmonics*.

$$
e_{nm}(\lambda\phi) = \begin{bmatrix}
\sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}} P_{nm}(\sin\phi)\cos m\lambda & \forall m > 0 \\
\sqrt{2(2n+1)} P_n(\sin\phi) & \forall m = 0 \\
\sqrt{2(2n+1)\frac{(n-|m|)!}{(n+|m|)!}} P_{n|m|}(\sin\phi)\sin|m|\lambda & \forall m < 0.
\end{bmatrix} \tag{C431}
$$

where $P_{nm}(\sin\phi)$ are the *associated Legendre functions*

$$
P_{nm}(x) = \frac{1}{2^n n!}(1 = x^2)^{m/2} \frac{d^{n+m}}{dx^{n+m}}(x^2-1)^n \tag{C432}
$$

Note that the quantum numbers run $n = 0, 1, \cdots, \infty; m = -n, -n+1, \cdots, 0, \cdots, n-1, n$. The base functions $e_{nm}(\lambda, \phi)$ are orthonormal in the sense

$$
\frac{1}{4\pi}\int_0^{2\pi} d\lambda \int_{-\pi/2}^{+\pi/2} d\phi\ \cos\phi\ e_{kl}e_{nm} = \delta_{kn}\ \delta_{lm} \tag{C433}
$$

Finally a vector field ε can be represented in terms of *vector spherical harmonics* by

$$\varepsilon = U_{nm}\mathbf{a}_{nm} + V_{nm}\mathbf{b}_{nm} + W_{nm}\mathbf{c}_{nm} \tag{C434}$$

$$\left[\begin{array}{l} \mathbf{a}_{nm} := \mathbf{e}_r \; e_{nm}(\lambda, \phi) \\ \\ \mathbf{b}_{nm} := r \;\; grad \;\; e_{nm}(\lambda, \phi) \\ \\ \mathbf{c}_{nm} := -\mathbf{x} \;\; grad \;\; e_{nm}(\lambda, \phi). \end{array}\right. \tag{C435}$$

For more details on vector spherical harmonics we refer to standard textbooks, here e.g. to *E. Grafarend* (1982).

$$\begin{aligned} \Sigma(\mathbf{x}, \mathbf{x}') &= E\{\varepsilon(\mathbf{x}) \otimes \varepsilon(\mathbf{x}')\} \\ &= U_{nmn'm'}\mathbf{a}_{nm} \otimes \mathbf{a}_{n'm'} \\ &\quad + V_{nmn'm'}\mathbf{b}_{nm} \otimes \mathbf{b}_{n'm'} \\ &\quad + W_{nmn'm'}\mathbf{c}_{nm} \otimes \mathbf{c}_{n'm'} \end{aligned} \tag{C436}$$

which is due to $< \mathbf{a}|\mathbf{b} >=< \mathbf{a}|\mathbf{c} >=< \mathbf{b}|\mathbf{c} >= 0$. $U_{nmn'm'}$ etc. are spherical covariance functions.

Example C 9.22: Towards criterion matrices for functionals of the distributing potential in a space-time holonomic frame if reference

The geodetic observational equations in their linearized form contain beside unknown *Cartesian coordinate corrections* in an orthonormal reference frame, being holonomic in space-time, the *disturbance of the potential and its derivatives as unknowns* . for a complete set-up of a criterion matrix of geodetic unknowns we have therefore present thoughts about a homogenous and isotropic variance-covariance matrix of the potential disturbance and its derivatives, The postulates of homogeneity and isotropy will lead again to the extension of an classical *Talyor-Karman structure*. finally the representation of the criterion matrix of potential related quantities in terms of *scalar* spherical harmonics is introduced. Since the potential disturbance is harmonic function, *vector* spherical harmonics are not needed.

We begin the transformation of the gravity criterion matrix from spherical into Cartesian coordinates.

With reference to the linearized observational equations the unknowns $(\widehat{\delta\lambda}_\gamma, \widehat{\delta\phi}_\gamma, \widehat{\delta\gamma})$ have to be equipped with a criterion matrix, reasonable to be homogenous and isotropic on the sphere. In order to take advantage of lemma and corollaries of the second chapter, we better transform the spherical coordinates of the disturbance gravity vector $\delta\gamma$ into Cartesian ones and then apply the variance-covariance "propagation" on the transformation formula.

$$\begin{aligned}
\delta\gamma_i := \Gamma_i - \gamma_i &= \begin{bmatrix} -\Gamma \cos\Lambda_\Gamma \cos\Phi_\Gamma \\ -\Gamma \sin\Lambda_\Gamma \cos\Phi_\Gamma \\ -\Gamma \sin\Phi_\Gamma \end{bmatrix} - \begin{bmatrix} -\gamma \cos\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\phi_\gamma \end{bmatrix} \\
&= \begin{bmatrix} -(\gamma+\delta\gamma)\cos(\lambda_\gamma+\delta\lambda_\gamma)\cos(\phi_\gamma+\delta\phi_\gamma) \\ -(\gamma+\delta\gamma)\sin(\lambda_\gamma+\delta\lambda_\gamma)\cos(\phi_\gamma+\delta\phi_\gamma) \\ -(\gamma+\delta\gamma)\sin(\phi_\gamma+\delta\phi_\gamma) \end{bmatrix} \\
&\quad - \begin{bmatrix} -\gamma \cos\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\phi_\gamma \end{bmatrix}
\end{aligned} \tag{C437}$$

$$\begin{aligned}
\delta\gamma_i &\doteq \begin{bmatrix} -(\gamma+\delta\gamma)(\cos\lambda_\gamma - \delta\lambda_\gamma \sin\lambda_\gamma)(\cos\phi_\gamma - \delta\phi_\gamma \sin\phi_\gamma) \\ -(\gamma+\delta\gamma)(\sin\lambda_\gamma + \delta\lambda_\gamma \cos\lambda_\gamma)(\cos\phi_\gamma - \delta\phi_\gamma \sin\phi_\gamma) \\ -(\gamma+\delta\gamma)(\sin\phi_\gamma + \delta\phi_\gamma \cos\phi_\gamma) \end{bmatrix} \\
&\quad - \begin{bmatrix} -\gamma \cos\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\lambda_\gamma \cos\phi_\gamma \\ -\gamma \sin\phi_\gamma \end{bmatrix}
\end{aligned} \tag{C438}$$

$$\delta\gamma_i \doteq \begin{bmatrix} -\delta\gamma \cos\lambda_\gamma \cos\phi_\gamma + \delta\lambda_\gamma \gamma \sin\lambda_\gamma \cos\phi_\gamma + \delta\phi_\gamma \gamma \cos\lambda_\gamma \sin\phi_\gamma \\ -\delta\gamma \sin\lambda_\gamma \cos\phi_\gamma - \delta\lambda_\gamma \gamma \cos\lambda_\gamma \cos\phi_\gamma - \delta\phi_\gamma \gamma \sin\lambda_\gamma \sin\phi_\gamma \\ -\delta\gamma \sin\phi_\gamma - \delta\phi_\gamma \gamma \cos\phi_\gamma \end{bmatrix} \tag{C439}$$

$$\begin{bmatrix} \delta\gamma_1 \\ \delta\gamma_2 \\ \delta\gamma_3 \end{bmatrix} \doteq \begin{bmatrix} +\gamma \sin\lambda_\gamma \cos\phi_\gamma & +\gamma \cos\lambda_\gamma \sin\phi_\gamma & -\cos\lambda_\gamma \cos\phi_\gamma \\ -\gamma \cos\lambda_\gamma \cos\phi_\gamma & -\gamma \sin\lambda_\gamma \sin\phi_\gamma & -\sin\lambda_\gamma \cos\phi_\gamma \\ 0 & -\gamma \cos\phi_\gamma & -\sin\phi_\gamma \end{bmatrix} \begin{bmatrix} \delta\lambda_\gamma \\ \delta\phi_\gamma \\ \delta\gamma \end{bmatrix} \tag{C440}$$

$$\begin{bmatrix} \delta\gamma_1 \\ \delta\gamma_2 \\ \delta\gamma_3 \end{bmatrix} \doteq \begin{bmatrix} -\gamma_2 & +\gamma_1\gamma_2(\gamma_1^2+\gamma_2^2)^{-1/2} & \gamma_1\gamma^{-1} \\ +\gamma_1 & -\gamma_2\gamma_3(\gamma_1^2+\gamma_2^2)^{-1/2} & \gamma_2\gamma^{-1} \\ 0 & -(\gamma_1^2+\gamma_2^2)^{-1/2} & \gamma_3\gamma^{-1} \end{bmatrix} \begin{bmatrix} \delta\lambda_\gamma \\ \delta\phi_\gamma \\ \delta\gamma \end{bmatrix} \tag{C441}$$

The variance-covariance matrix $D\{\widehat{\delta\gamma_i}(\mathbf{x}), \widehat{\delta\gamma_{j'}}(\mathbf{x}')\}$ as a function of the variance-covariance matrix $D\{(\widehat{\delta\lambda}_\gamma, \widehat{\delta\phi}_\gamma, \widehat{\delta\gamma})(\mathbf{x}), (\widehat{\delta\lambda}_\gamma, \widehat{\delta\phi}_\gamma, \widehat{\delta\gamma})(\mathbf{x}')\}$ based on the transformation (C 438) is given. finally the Cartesian variance-covariance matrix $D\{\widehat{\delta\gamma_i}(\mathbf{x}), \widehat{\delta\gamma_{j'}}(\mathbf{x}')\}$ in the *fixed frame* $F^\bullet$ should be transformed into the *moving* frame f^* according to (C381).

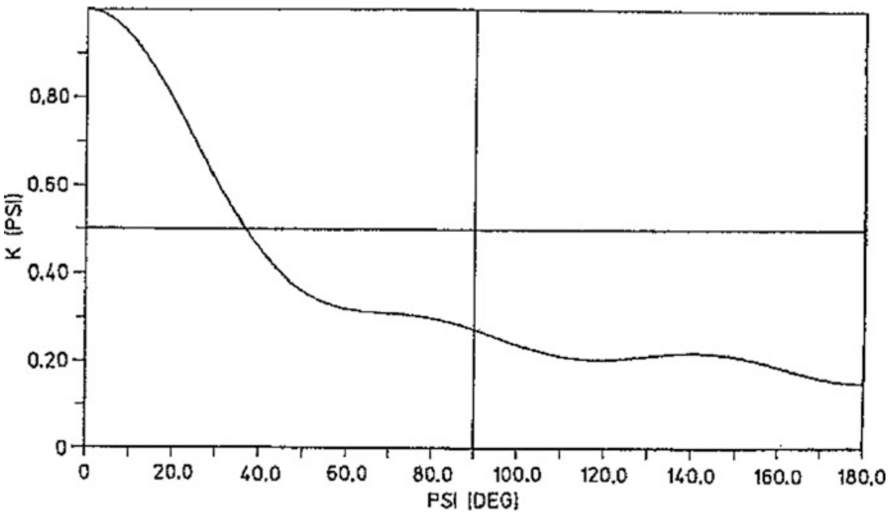


Fig. C29 Normalized covariance function $K(\psi) = \sum_{n=0}^{5} k_n P_n(\cos\psi)$ for $k_n = \frac{1}{n+2}$

Thus we refer to

$$\gamma^{\Sigma}_{ll'}, \gamma^{\Sigma}_{lm'}, \gamma^{\Sigma}_{lr'}, \gamma^{\Sigma}_{ml'}, \gamma^{\Sigma}_{mm'}, \cdots, \gamma^{\Sigma}_{rr'} \tag{C442}$$

as the *longitudinal* correlation function $\gamma^{\Sigma}_{ll'}$, the *lateral* correlation function $\gamma^{\Sigma}_{mm'}$ etc. of the gravity disturbance vector $\delta\gamma_i$.

In order to construct a *homogenous and isotropic criterion matrix* for functionals of the disturbing potential in a surface moving frame we refer to the introduction which applies here, too. Again we make use of the definition of homogeneity and isotropy of the tensor-valued two-point function $\gamma^{\Sigma(\mathbf{X},\mathbf{X}')}$ on the sphere, the basic lemma and the corollary for the spherical Taylor-Karman structure. Just by adding the left hand index γ on Σ within formula (C410) (C411), we arrive at the corresponding formulae for longitudinal, lateral and cross-correlation of the gravity disturbance vector $\delta\gamma$.

For the representation of the *criterion matrix for functionals of the disturbing potential* in *scalar* spherical harmonics we once more refer to (C417). All results can be transferred, especially (C 420 - C 425) once we write $K(r, r', \psi)$ as the scalar-valued two-point function $E\{[\widehat{\delta w} - E(\widehat{\delta w})]\mathbf{x}[\widehat{\delta w} - E(\widehat{\delta w})]\mathbf{x}'\}$ homogenous and isotropic on the sphere. Besides the *first-order functionals* of the potential δw and their variance-covariance matrices, higher order functionals, e.g. for gravity gradients, and their variance-covariance matrices can be easily obtained, a project for the future.

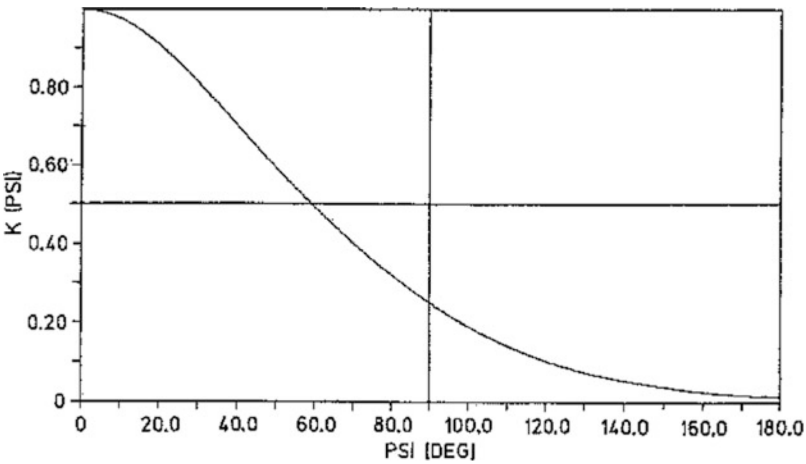


Fig. C30 Normalized covariance function $K(\psi) = \sum_{n=0}^{5} k_n P_n(\cos\psi)$ for $k_n = \frac{1}{(n+2)(n^2-2n+2)}$

A numerical example of the homogenous and isotropic two-point function $K(\mathbf{x}, \mathbf{x}', \psi)$ on the space

For demonstration purpose plots C29 - C31 show the normalized covariance functions $K(\mathbf{X}, \mathbf{X}') = K(r, r', \psi)$

$$= K(\psi) = \sum_{n=0}^{5} k_n P_n(\cos\psi) \ \text{ for } \ case \ \ 1$$

$(\|\mathbf{X}\| = \|mathbf\mathbf{X}'\| = r_0)$ and three different choices of the coefficients k_n. The coefficients are chosen in such a manner that their sum equals the unity (so as to normalize the covariance function) and the covariance function approaches zero as ψ tends to π. As an example for the construction of a criterion matrix, the coefficients k_n could be estimated for an ideal network and its variance-covariance matrix; that is derived covariance function $K(\mathbf{x}, \mathbf{x}')$.
We have taken a part of Appendix C 9.2 from our contribution *E. Grafarend, I. Krumm and B. Schaffrin* (Manuscripta geodetica 10 (1988) 3-22). Spherical homogeneity and isotropy in tensor-valued two-point functions have been introduced by *E. Grafarend* (1976) and *H. Moritz* (1972, 1978).

Ex 3: Space Gravity Spectroscopy: The Benefits of Taylor–Karman Structured Criterion Matrices

Example C.9.3. Space gravity spectroscopy: the benefits of Taylor-Karman structured criterion matrices (eg, P. Marinkovice et. al. (2003))

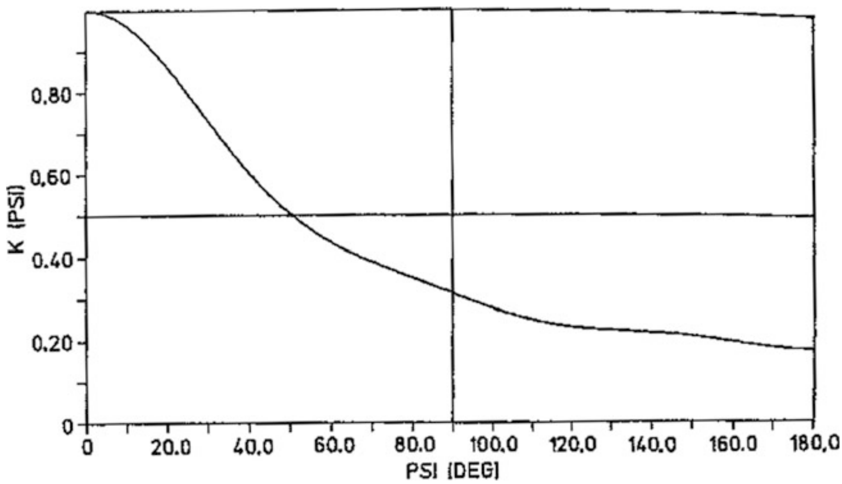


Fig. C31 Normalized covariance function $K(\psi) = \sum_{n=0}^{5} k_n P_n(\cos\psi)$ for $k_n = \frac{1}{n^2+3}$

As soon as the space spectroscopy was successfully performed, for instance by means of semi-continuous ephemeris of LEO-GPS tracked satellites, the problem of data validation appeared. It is for this purpose that a stochastic model for the homogenous isotropic analysis of measurements, obtained as "directly" measured values in LEO satellite missions (CHAMP, GRACE, GOCE), is studied. An isotropic analysis is represented by the homogenous distribution os measured values and the statistical properties of the model are calculated. In particular a correlation structure function is defined by the third order tensor (*Taylor-Karman tensor*), for the ensemble average of a set of incremental differences in measured components. Specifically, *Taylor-Karman correlation tensor* is calculated with the assumption that the analyzed random function is of a "potential type". The special class of *homogenous and isotropic correlation functions* is introduced. Finally, a successful application of the concept is presented in the case study CHAMP and a comparison between modeled and estimated correlations is performed.

Example C.9.31. Introduction

A significant problem of LEO satellites, both in geometry and gravity space, is the association of quality standards to *Cartesian ephemeris* in terms of variance-covariance matrix valued-functions. as a pessimistic measure of the quality standards os LEO satellite ephemeris, a three dimensional *Taylor-Karman structured* criterion matrix has been proposed, named in honor of Taylor (1938) and Karman (1938), the founders of the statistical theory of turbulence.

The concept of the *Taylor-Karman criterion matrices* was first introduced by *E. Grafarend* (1979) and subsequently further developed by *B. Schaffrin and E. Grafarend* (1982), *H. Wimmer* (1982), and *E. Grafarend* et. al. (1985, 1986). with this contribution we extend the application of the Taylor-Karman structured matrices to the third-dimension, namely to the long-arc orbit analysis.

If we assume the vector-valued stochastic process to be the gradient of a random scalar-valued potential, in particular *its longitudinal and lateral correlation function* or the "correlation function along-track and across-track", what would be the structure of a three-dimensional *Taylor-Karman variance-covariance matrix*. In order to answer this question, a three-dimensional correlation tensor and its decomposition in the case of homogeneity and isotropy is studied with the emphasis on $\mathbb{R}^3$. Additionally, we deal with a special class of homogenous and isotropic tensor-valued correlation functions. They are derived, analyzed and applied to the data validation process. The theoretical concept for the application of the previously discussed criterion matrix in the geometric and gravitational analysis of LEO satellite ephemeris is presented. Finally the case study CHAMP is used for the application of our theory concept followed by the results and conclusion.

Example C.9.32: Homogenous and isotropic variance-covariance tensor and correlation functions in a three-dimensional Euclidean space

Example C.9.321: Notions of homogeneity and isotropy

The notions of homogeneity and isotropy for functions on $\mathbb{R}^{\kappa}$ are briefly explained as the following. The general context for these two definitions involves the action of a transitive group of motions on a homogenous and space belongs to the extensive theory of *Lie groups* (*F. W. Warner* (1983), *A. M. Yaglom* (1987)). However, it is important to clarify the different notions of homogeneity and isotropy exist due to the variety of homogenous spaces and transitive group actions on these spaces. The terminology introduced associates the notion of homogeneity and isotropy with

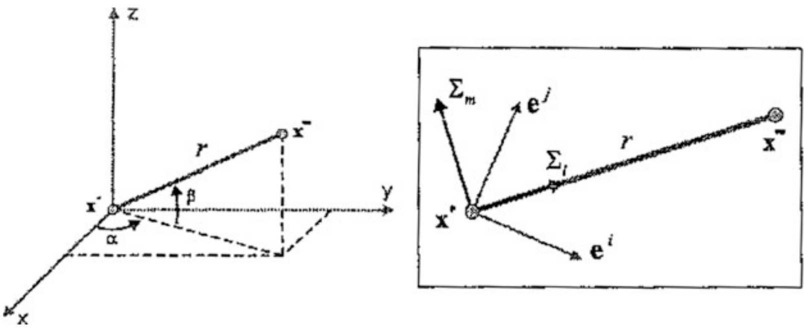


Fig. C32 Graphical representation of the longitudinal and lateral correlation functions the functions on $\mathbb{R}^{\ltimes}$ that are invariant under the translation group acting on $\mathbb{R}^{\ltimes}$. The notion of isotropy is defined for functions on $\mathbb{R}^{\ltimes}$ that are invariant under the orthogonal group acting on $\mathbb{R}^{\ltimes}$.

Table C.9.9: Tensor-valued correlation function of second-order

"two-point correlation function"

$$\Sigma(\mathbf{x}^*, \mathbf{x}^{**}) = \sum_{i,j=1}^{3} e^i e^j \Sigma_{ij}(\mathbf{x}^*, \mathbf{x}^{**}) = \sum_{i,j=1}^{3} \Sigma_{ij}(\mathbf{x}^*, \mathbf{x}^{**}) e^i e^j \tag{C443}$$

$$r := \|\mathbf{x}^*, \mathbf{x}^{**}\| \quad \text{and} \quad r := \mathbf{x}^*, \mathbf{x}^{**}$$

$$\Sigma_{ij}(r) = \Sigma_m(r)\delta_{ij} + [\Sigma_l(r) - \Sigma_m(r)]\frac{\Delta x_i \Delta x_j}{r^2}, \quad i, j \in \{1, 2, 3\} \tag{C444}$$

"Cartesian versus spherical coordinates"

$$\begin{aligned} \Delta x_1 = \Delta x &:= (x_1^{**} - x_1^*) = r \cos\beta \cos\alpha \\ \Delta x_2 = \Delta y &:= (x_2^{**} - x_2^*) = r \cos\beta \sin\alpha \\ \Delta x_3 = \Delta z &:= (x_3^{**} - x_3^*) = r \sin\beta \end{aligned} \tag{C445}$$

"continuity of a potential type"

$$\Sigma_l(r) = \frac{d[r\Sigma_m(r)]}{dr} = \Sigma_m(r) + r\frac{d\Sigma_m(r)}{dr} \tag{C446}$$

End of Table C.9.9

Example C.9.322: Homogenous and isotropic variance-covariance tensor (correlation tensor)

Taylor (1938) proved that he homogenous random field $X(\mathbf{t})$ correlation function $\Sigma(r) =< X(\mathbf{t}+\mathbf{r})X(\mathbf{t}) >$ depends only on the length $r = \|\mathbf{r}\|$ of the vector $\mathbf{r}$ and not on its direction, where $< ... >$ denoted the ensemble average. If the correlation function $\Sigma(r)$ of the homogenous random field $X(\mathbf{t})$ in $\mathbb{R}^{\ltimes}$ has the property, then the field $X(t)$ is said to be an isotropic random field in $\mathbb{R}^{\ltimes}$. The corresponding correlation function $\Sigma(r)$ is then called an isotropic correlation function in $\mathbb{R}^{\ltimes}$ (or an n-dimensional isotropic correlation function). Process, which satisfy the introduced postulates of homogeneity and isotropy, are said to be (widely) stationary (*M. J. Yadrenko* (1983)). Note that for an isotropic random field in $\mathbb{R}^{\ltimes}$, all directions in space are obviously equivalent.

The decomposition of a homogenous and isotropic variance-covariance tensor-valued function, shown in box 1, was introduced by von *Karman and Howarth*

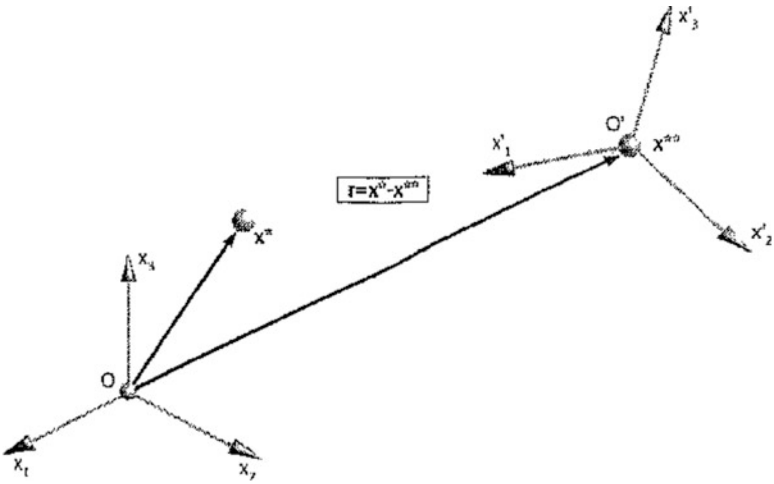


Fig. C33 Graphical representation of the correlation tensor transformation

(1938) by means of a more general and direct method than the one used by *G. J. Taylor*(1938). Additionally, *Robertson* (1940) refined and reviewed the *T. Karman and L. Howard* equation in the light of a classical invariant tensor theory.

The decomposition of $\Sigma_{ij}(\|\mathbf{x}^* - \mathbf{x}^{**}\|)$, with a special emphasis on $n=3$, is performed in terms of Cartesian coordinates and with respect to the orthonormal frame of reference $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3|0\}$ attached to the origin 0 of a three-dimensional Euclidean space. $\|\mathbf{x}^* - \mathbf{x}^{**}\|$ denotes the Euclidean distance r between the two points $\mathbf{x}^*$ and $\mathbf{x}^{**}$ of $\mathbb{E}^{\nVdash} := \{\mathbb{R}^{\nVdash}, \delta_{ij}$. *Longitudinal and lateral correlation functions* Σ_l and Σ_m, are the structural elements of such a homogenous and isotropic tensor-valued variance-covariance function which appear in the spherical tensor $\Sigma_m(r)\delta_{ij}$ as well as in the oriented tensor $[\Sigma_l(r) - \Sigma_m(r)]\Delta x_i \Delta x_j / r^2$ for all $i, j \in \{1, 2, 3\}$, see (C443) and Fig. C.32. δ_{ij} denotes the *Kronecker* or unit matrix, Δx_i the Cartesian coordinate difference between the points $\mathbf{x}^*$ and $\mathbf{x}^{**}$. These differences are also represented in terms of relative spherical coordinates (α, β, r), (C444). Finally the continuity condition of a potential type is formulated by (C445) which provides the unique relation between Σ_l and Σ_m (*G. J. Taylor* (1938); *H. M. Obuchow* (1958); *E. Grafarend* (1979)).
Due to its complexity, it is necessary to further elaborate on the previous equation set. The correlation tensor $\Sigma_{ij}(r)$ was transformed to a special coordinate system $O'x'_1x'_2x'_2$ in $\mathbb{R}^{\nVdash}$ instead of the initial set $Ox_1x_2x_3$. The new set $O'x'_1x'_2x'_3$ is selected in such a way so that its origin O' is shifted by the vector x^{**} with respect to the origin O, as illustrated in Fig. C15. This means that O' coincides with the terminal point of the vector $\mathbf{x}^{**}$ that refers to the initial coordinates, while the axis $O'x'_1$ lies along the vector $\mathbf{x}^* - \mathbf{x}^{**}$.

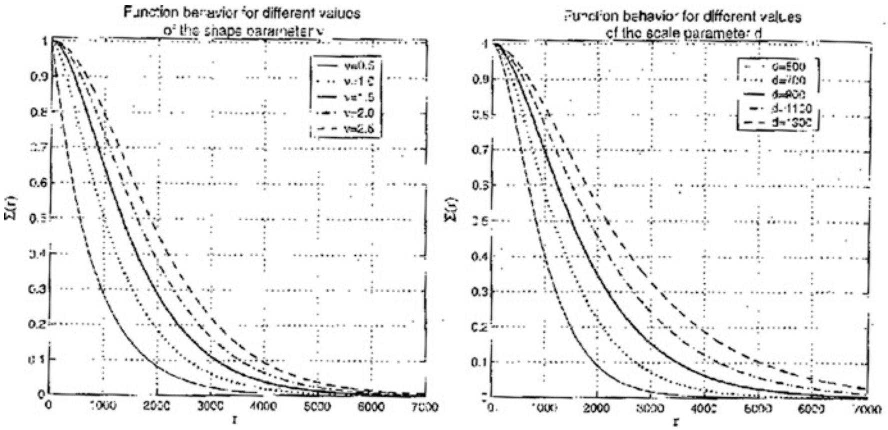


Fig. C34 The behavior of Tatarski's correlation function for different values of shape parameter $\nu(d = 900)$ and scale parameter $d(\nu = 3/2)$

Σ_{ij}, introduced here are explanatory functions, are the components of the correlation tensor $\Sigma_{ij}(r)$ in the new set of coordinates. The functions $\Sigma'_{ij}(r)$ clearly depend only on the length $r = \|r\|$ of the vector $\mathbf{r}$, since the direction of $\mathbf{r}$ is fixed in the new set of coordinates. In the space $\mathbb{R}^{\nvDash}$ exist a reflection which leaves the points $\mathbf{x}^*$ and $\mathbf{x}^{**} = (O')$ unmoved and it replaces the axis $O'x'_j$ by $-O'x'_j$, where $j \neq 1$ is a fixed number. However, it does not change the directions of all other coordinates axes $O'x'_l, l \in \{1, 2, 3\}$ and $l \neq j$. It follows that

$$\Sigma'_{ij}(r) = -\Sigma'_{ij}(r) = 0 textrm fori \neq j \tag{C447}$$

Hence, only the diagonal elements $\Sigma'_{ij}(r)$ of $\Sigma_{ij}(r)$ can differ from zero. Further more, if $i \neq j$ and $j \neq 1$, then the axis $O'x'_i$ by its rotation around the axis $O'x'_j$, can be transformed to the axis $O'x'_1$. Hence

$$\Sigma'_{22}(r) = \Sigma'_{33}(r) \tag{C448}$$

The tensor Σ'_{ij} and, consequently, Σ_{ij} are symmetric and their components $\Sigma'_{ij}(r)$ can take at most only tow non-equal non-zero values at the already introduces longitudinal correlation function $\Sigma'_{ii}(r) = \Sigma_l(r)$ and the lateral correlation function $\Sigma'_{22}(r) = \Sigma'_{33}(r) = \Sigma_m(r)$, which specify in a unique way the correlation tensor. In order to obtain the explicit form of $\Sigma_{ij}(r)$, as the function of $\Sigma_l(r)$ and $\Sigma_m(r)$, the unit vectors of the old coordinate axes Ox_1, Ox_2, Ox_3 along the axes of the new system $O'x'_1, O'x'_2, O'x'_3$ must be resolved and then $\Sigma_{ij}(r)$ can be represented as linear combination of the functions $\Sigma'_{kl}(r), (k \neq i, l \neq j$ and $k, l \in \{1, 2, 3\})$, which leads to (C442) and (C443).

Example C.1. homogenous and isotropic correlation functions

It was shown in the pervious section that for a homogenous and isotropic random fields defined on the Euclidean space $\mathbb{R}^{\ltimes}$, the correlation between $\mathbf{x}^*$ and $\mathbf{x}^{**}$ depends only on the Euclidean distance $\|\mathbf{x}^* - \mathbf{x}^{**}\|$. Therefore as shown in Table C32, we can distinguish a homogenous and isotropic correlation function on $\mathbb{R}^{\ltimes}$ with the real valued function $\Sigma(r)$ defined on $[0, \infty)$ and we denote by Φ_n the class of all continuous permissible functions $\Sigma(r) : [0, \infty) \rightarrow$ such that $\Sigma(0) = 1$ (we are working in terms od correlation not covariance) and the symmetric function $\Sigma(\|.\|)$ is a positive definite on $\mathbb{R}^{\ltimes}$. The characterization of classes Φ_n, also shown in Table C.9.10, is a well known result of J. F. Schoenberg (1938). The function $\Sigma(r) : [0, \infty) \rightarrow \mathbb{R}$ is an element of Φ_n if and only if it abmits a representation in the form of (C450), where W is a probability measure on $[0, \infty)$, $J_{(n-2)/2}$ is the Bessel function of the first kind of order $(n-2)/2$ and Γ stands for the Gamma function. Hence in for geodesy relevant spaces $\mathbb{R}^{\nvDash}$ and $\mathbb{R}^{\nVdash}$, (C450) reduces to the form presented in (C452).

Table C.9.10: Isotropic correlation and Schoenberg's characterization

"homogenous and isotropic correlation function"

$$\Sigma(\mathbf{x}^*, \mathbf{x}^{**}) = \Sigma(\|\mathbf{x}^*, \mathbf{x}^{**}\|), \quad \mathbf{x}^*, \mathbf{x}^{**} \in \mathbb{R}^{\ltimes} \tag{C449}$$

$$r := \|\mathbf{x}^*, \mathbf{x}^{**}\|$$

"the characterization of the Φ_n class"

$$\Sigma_n(r) = \int\limits_{[0,\infty)} \Omega_n(r\nu)dW(\nu) \tag{C450}$$

$$\Omega_n(r\nu) = \Gamma(n/2)(\tfrac{2}{r})^{(n-2)/2} J_{(n-2)/2}(r) \tag{C451}$$

"reduction of the Ω_n for $n = 2$ and $n = 3$"

$$\Omega_2(r) = J_0(r) \text{ and } \Omega_3(r) = r^{-1}\sin r \tag{C452}$$

End of Table C.9.10

Table C.9.11: Special classes of correlation functions

"Tatarski's class"

$$\Sigma_{[\nu]}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\tfrac{r}{d})^{\nu} K_\nu(\tfrac{r}{d}) \tag{C 453}$$

"Shkarofsky's class"

$$\Sigma_{[\nu,\delta]}(r) = \frac{(\frac{r^2}{d^2}+\delta^2)^{\nu/2} K_\nu((\frac{r^2}{d^2}+\delta^2)^{1/2})}{\delta^\nu K_\nu(\delta)} \tag{C454}$$

End of Table C.9.11

Example C.2. Tatarski's class of homogenous and isotropic correlation functions

Many analytical candidate models for Σ have been suggested in the literature (for example, *Buell, (1972); Haslett, (1989)*), but we refer to Tatarski (1962) as being the first who elaborated on such correlation functions which fulfill all the conditions presented in the previous *section*. The Tatarski's correlation function class is shown in Table C.9.11 and illustrated by Fig. C.34. In addition to *V. J. Tatarski's class*, a very general family of correlation function models due to J. P. *Shkarofsky* (1968) is introduced, that came as the generalization of Tatarski's correlation function family. These two classes, which have been proved to be the members of Φ_3 (Shkarofsky, (1968) and of Φ_∞ classes; Gneiting et. al. (1999)), can be applied to many geodetic problems, (e.g. Grafarend (1979); Meier (1982); Wimmer (1982); Grafarend (1985)).

In equation of Table C.9.11, K_ν stands for a modified Bessel function of order $\nu, d > 0$ is a scale parameter, and $\delta > 0$ and ν are shape parameters. In the case of $\delta = 0$ *J. P. Shkarofsky's* class reduces to *V. J. Tatarski's* class.

The special case of *Tatarski's class* appears if the shape parameter ν is sum of a non-negative integer k and $1/2$. Then the right-hand side of the equation can be written as a product of $\exp(-r/d)$ and a polynomial of degree k in r/d (eg. *T. Gneiting*, (1999)). In particular, in the case of $n = 3$ dimensional *Markov process* of the $p = 1$ order, shown in Fig. C34, the shape parameter is expressed as $\nu = (2p + 1)/(n - 1) = 3/2$ and results in the flowing simplification of (C454):

$$\Sigma_{[3/2]}(r) = (1 + \frac{r}{d}) \exp(-\frac{r}{d}) \quad \text{(C455)}$$

Example C.3. Three dimensional Taylor-Karman criterion matrix in the geometric and gravitational analysis of LEO satellite ephemeris

WE have so far analyzed the theoretical background and solution for design of the *homogenous and isotropic Taylor-Karman correlation tensor*. The question is, how this theoretical concept applies to the geometric and gravitational analysis of LEO satellite ephemeris.

The basic idea is, that the errors in the position vectors of LEO satellites constitute a vector-valued stochastic process. Following this concept, a satellite orbit of LEO-GPS tracked satellites is an homogenous and anisotropic field of error vectors and the error situation is described by the covariance function. As it is well known, the error situation os a newly determined position in a *three-dimensional Euclidean space* is the best, when the error ellipsoid is sphere (isotropy) with the minimal radius and if the error situation is uniform over the complete satellite orbit (homogeneity). This *"ideal"* situation can be explained by the three-dimensional *Taylor-Karman structured criterion matrix* of *W. Baarda - E. Grafarend* (potential) type. Then the correlations between the vectors of pseudo-observations of satellite ephemeris described by the longitudinal and lateral correlation functions.

The characteristic correlation functions can be estimated by matching the correlation tensor with a three-dimensional Markov process of 1^{st} order and with the introduction of some additional information about the underlying process. The correlation analysis is performed with the assumption that the vector valued three-dimensional random function is of "potential type" (E. Grafarend (1979)), i.e. the gradient of a random scalar function "signal $s(\mathbf{x})$". The structure of the random function $s(\mathbf{x})$ is outlined as an n-dimensional Markov process of the p^{th} order. Figure C35 illustrates the case $n = 3$ and $p = 1$:

One of the simples differential equation os such a process has the form given by

$$(\Delta^2 - \alpha^2)^p s(\mathbf{x}) = e(\mathbf{x}) \tag{C456}$$

where $e(\mathbf{x})$ is a white noise. If the Laplace operator can be applied to the homogenous random scalar function, then it transforms this function into a new homogenous random function, having the spectral density that corresponds to the spectral density of the correlation function of (C453) and (C454), see *P. Whittle* (1954), *V. Heine* (1955) and *P. Whittle* (1963). Hence the homogenous solution of C455 (if it exist) must have the spectral density that corresponds to the spectral density of the correlation function.

Table C.9.9 summarizes the representation of the homogenous and isotropic correlation functions of type

(i) Signal correlation function Σ,
(ii) Longitudinal correlation function Σ_l and
(iii) Lateral correlation function Σ_m

Example C.9.34: Results and conclusions

$\boxed{\text{Case study: Champ}}$

As the numerical test in this study, we processed two data sets: two three-dimensional $\{x(t_k),\ y(t_k), z(t_k)\}$ and $\{x(t_d), y(t_d), z(t_d)\}$ Cartesian ephemeris time-series of CHAMP satellite orbit for the test period from day 140 to 150 of 2001 (20 May to 30 May, both inclusive). We analyzed in total 27,360 triplets of satellite positions. Both time series are indexed with a 30 s sampling rate and referenced to a kinematic (index k) and a dynamic CHAMP orbit (index d). The dynamic orbit used as a reference, provides us with ephemeris differences between the two orbits. The estimation of ("real") auto and cross correlations between the

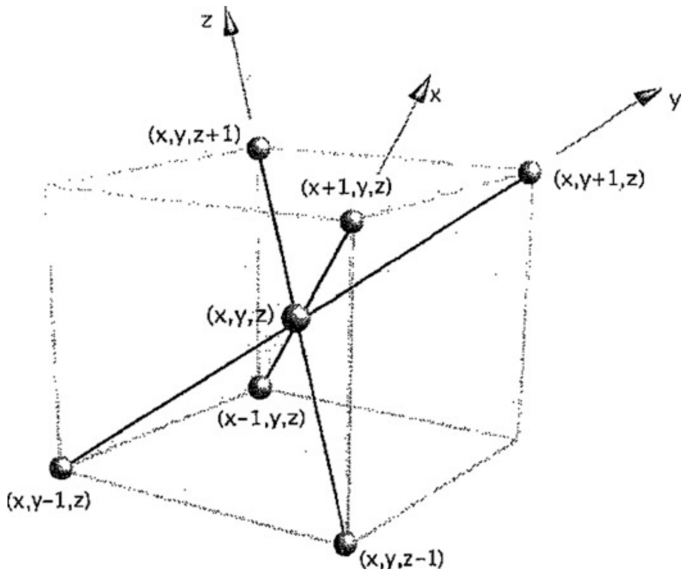


Fig. C35 The six point interaction in the grid; the three-dimensional Markov process of the 1^{st} order (auto regressive process); the gray rectangle represents the same process in two-dimensions

vectors of pseudo-observations as functions of time, can be performed as Priestley (1981).

According to Tables C.9.10 and C.9.11 the *Taylor-Karman structured* ("ideal") correlations are computed from the three-dimensional $\{x(t_k), y(t_k), z(t_k)\}$ time series. The adopted scale parameter is $d = (2/3)R_{char}$, where R_{char} is the *characteristic length of the process*. The characteristic length is defined by the arc length of 2400 seconds (80 points for the 30 second sampling rate). The both parameters are experimentally estimated. For further details on the scale parameter and characteristic length, please see Wimmer (1982).

The numerical results of the study are graphically presented in Fig. C37. The gray area represents the estimated ("real") correlation situation along the satellite arc as presupposed by Austen et al. (2001). The high auto and low cross-correlations between CHAMP satellite positions for approximately 20 min of an orbit arc are very evident. The *Taylor-Karman structured correlation* (black line), as theoretically assumed, gives an upper bound of the "real" correlation situation along the satellite orbit.

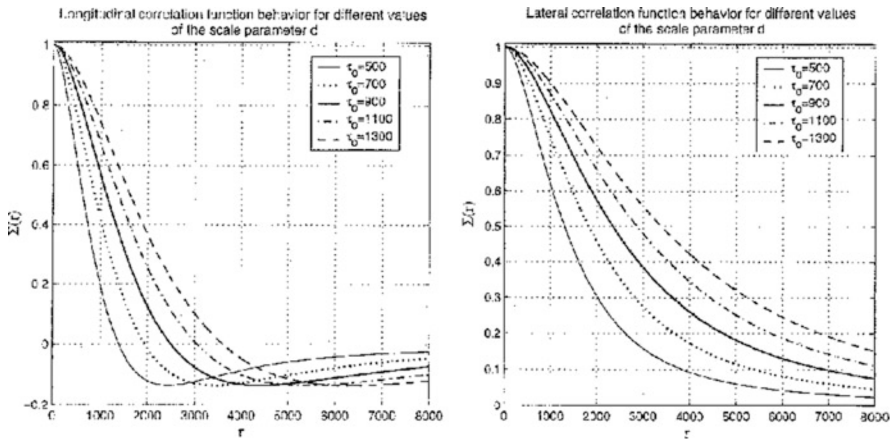


Fig. C36 The behavior of the longitudinal and lateral correlation functions, (C456) and (C 458), for different values of the scale parameter $d(\nu = 3/2)$

Table C.9.12 Longitudinal and lateral correlation function for a homogenous and isotropic vector-valued random field of potential type, 1^{st} order Markov process

"condition for a process of potential type"

$$\Sigma(x) \to \frac{2p}{r^2} \int_0^r x\Sigma(x)\, dx \to$$

$$\to \Sigma_l(r) = \frac{d}{dr}[r\Sigma_m(r)]$$

"input"

$$\Sigma(r) = \frac{2^{-1/2}}{\Gamma(3/2)}(\tfrac{r}{d})^{3/2} K_{3/2}(\tfrac{r}{d}) = (1 + \tfrac{r}{d}) \exp(-\tfrac{r}{d}) \tag{C457}$$

"output"

$$\Sigma_l(r) = -6(\tfrac{r}{d})^{-2} + \exp -(\tfrac{r}{d})[4 + 2(\tfrac{r}{d}) + 6(\tfrac{r}{d})^{-1} + 6(\tfrac{r}{d})^{-2}] \tag{C458}$$

$$\Sigma_m(r) = -6(\tfrac{r}{d})^{-2} + \exp -(\tfrac{r}{d})[2 + 6(\tfrac{r}{d})^{-1} + 6(\tfrac{r}{d})^{-2}] \tag{C459}$$

End of Table C 9.12

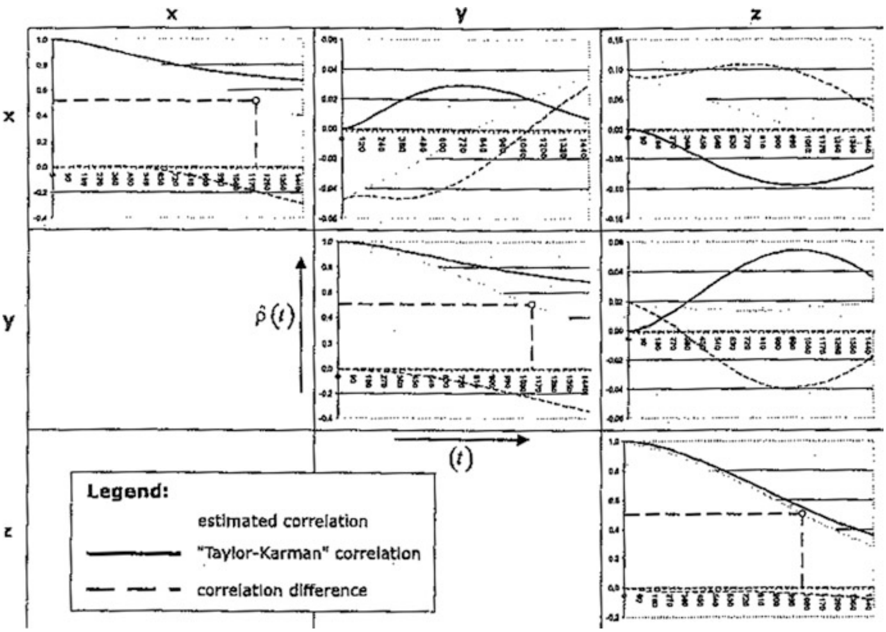


Fig. C37 Graphical representation of the numerical study results (20 minute correlation length, 40 points for the 30 second sampling rate)

Ex 4: Nonlinear Prediction

Example C.9.4: Nonlinear Prediction

For direct interpolation, extrapolation or filtering without trend analysis, linear prediction gives poor results, asking for a more general concept, namely *nonlinear prediction*. A set of *functional series* of Volterra type for nonlinear prediction of scalar-, vector-, in general, tensor-valued signals is set-up. Optimizing the *Helmert or Work Meister* point error, a set of linear integral equations, generalizing the Weiner-Hopf integral equations. Multi-point-correlation functions *(MPCF)* are characteristic for those equations. A criterion for linear prediction, namely the vanishing of all MPCFs *beside the two-point-correlation*, is formulated solutions of the discrete integral equations and the prediction error tensor of rank two are reviewed. A *structure analysis* of PCFs of *homogenous and isotropic vector fields* proves the number of characteristic functions to be $2(n = 2), 4(n = 3), 10(n = 4)$ and $26(n = 5)$.

The methods of statistical prediction were developed relating to stochastic processes by two famous scientists, by the Russian *A. N. Kolmorogov* (1941) and by the American *N. Wiener* (1941), nearly at the same time during the Second World War.

The world-wide resonance to these milestones began rather late: the work of *A. N. Kolmorogov* was due to the bad documentation of Eastern publications in Western countries nearly unknown, the great works of N. Wiener appeared as a classified report of section D_2 of the *American National Defence Research Committee*. At that time the difficult formula apparatus made the acceptance rather impossible. *Nowadays*, the difficulties are only history, the number of publications about linear filtering, interpolation and extrapolation, reaches ten thousands. For example, the *American Journal JRE, Transactions on Information Theory*, New York appears regularly every year with a *literature report* about the *State-of-the-Art*. Notable are the works of *R. Kalman* (1960) who succeeded to solve elegant by the celebrated perdition equation, the *Wiener-Hopf integral equation* in transforming to a differential equation which is much easier to solve. The methods became very well known as a method of data analysis applied to *Navigation of the Apollo Spacecraft*.

In practice we begin with a two-step procedure. *First*, we analyze the observed values with respect to *trend*: we fit a data set to a harmonic function, to a exponential function or a polynomial with respect to "position". *Second*, we analyze the data after the trend fit to a Wiener-Kolmogorov prediction assuming a *linear relation* between the predicted signal and the reduced unknown signal. The coefficients in this linear relation depend on the distance between the points and its direction. It is well known how to operate with the *discrete approximation* of the classical *Wiener-Hopf integral equation*.

The two step procedure accounts for the typical results that (a) the *direct application* of a *linear Kolmogorov-Wiener prediction* gives too inconsistent results, namely too large mean-square-errors of the predicted signal. The remarkable disadvantage of the two step procedure, in contrast is the immense computer work. Thus the central question is: *is the two-step-procedure necessary* when a concept for *nonlinear prediction* exists.

For all statistical prediction concepts the correlation function between various points is the key quantity. In practice, the computational problems are greatly reduced if we are able to find out whether the predicted signal is subject to a statistical homogenous and isotropic process. It is a key information if a correlation function depends only on the *distance* between the involved points or is a function of the *difference vector*, namely *isotropy and homogeneity*, our special topic in a special section.

Example C.9.41: The hierarchy for the equations in nonlinear prediction of scalar-valued signals

We review shortly the characteristical equations which are given in *nonlinear prediction*. Before hand we agree to the following notation: All Greek indices run $\{1, 2, \cdots, N-1, N\}$ counting the number of points P_α with a signal $s(P_\alpha)$. The signal may be a scalar signal, a vector signal or a tensor signal. A natural assumption is the "simple prediction model" that the predicted signal $s(P)$ at the point P is a linear function of the observed signal $s(P_\alpha)$ for $\alpha \in \{1, 2, \cdots, N-1, N\}$. The coefficients which describe the linear relation are dented by $a(PP_\alpha)$. The inconsistency between the predicted signal and the observed signal is denoted by the letter "ε'' leading the *prediction variance* $E\{\varepsilon(P)\varepsilon(P)\}$, namely assuming that the difference $s(P) - \sum_{\alpha=1}^{N} a(P, P_\alpha)s(P_\alpha)$ is *random*.

Nonlinear prediction is defined by a nonlinear relation between the predicted signal and the observed signal, *for instance* the sum of a linear and a quadratic relation described by our previous formula. The postulate os a minimum mean square error leads to the problem to estimate the coefficients a and b. Of course, we could continue with signal functions which depends of arbitrary power series. The open question is which is *signal relation of arbitrary powers*.

In order to give an answer to our key question, we have to change our notion from discrete to an integral concept. We arrive in this concept if we switch over to a continuous distribution of points P, namely from $\{P, P_\alpha\}$ to $\{\mathbf{r}, \mathbf{r}'\}$. In a one-dimensional problem we integrate over a *measured line*, in a two-dimensional problem we integrate over a *surface*. The integral set-up, in contrast, is reviewed by formula $(C461) - (C464)$ $O(s^3)$ expresses the order of neglected terms in the *functional series* $S[s(\mathbf{r}']$. We refer to a type of *Taylor series* in a *generalized Hilbert space* referred to *V. Volterra* (1959), called *Volterra functional series*. The coefficient functions given by $(C462)$ are called as *variational derivatives* in a *generalized Hilbert space*.

A special criticism originates from the *continuous version of the Volterra series*. Help comes from the application of functional analysis, namely the *Dirac function* as a *generalized function* illustrated by formulae (C464) and (C465). It is easy now the switch from the continuum to the discrete, namely for a regular grid. Nowadays the integral version is "simple" to end up with a discrete formulation, introduced already by *N. Wiener*.

Up to now we have introduced a set-up of our prediction formulae for scalar signals. In relating our work to practical problems we have to generalize to *vector-valued functions*.

Table C.9.13 Discrete and continuous Volterra series for scalar signals

"discrete 2^{nd} order Volterra series"

$$s(P) = \sum_{\alpha=1}^{N} a(P, P_\alpha)s(P_\alpha) + \sum_{\alpha=1}^{N}\sum_{\beta=1}^{N} b(P, P_\alpha, P_\beta)s(P_\alpha)s(P_\beta) + \varepsilon(P) \quad \text{(C460)}$$

"continuous 2^{nd} order Volterra series"

$$
\begin{aligned}
s(\mathbf{r}) = & \int d^n\mathbf{r}' a(\mathbf{r},\mathbf{r}') s(\mathbf{r}') \\
& + \int d^n\mathbf{r}' \int d^n\mathbf{r}'' a(\mathbf{r},\mathbf{r}',\mathbf{r}'') s(\mathbf{r}') s(\mathbf{r}'') \\
& + O(s^3) + \varepsilon(\mathbf{r}) \qquad \text{(C461)}
\end{aligned}
$$

"continuous 3^{rd} order Volterra series"

$$
\begin{aligned}
s(\mathbf{r}) = & \int d^n\mathbf{r}' a(\mathbf{r},\mathbf{r}') s(\mathbf{r}') \\
& + \int d^n\mathbf{r}' \int d^n\mathbf{r}'' a(\mathbf{r},\mathbf{r}',\mathbf{r}'') s(\mathbf{r}') s(\mathbf{r}'') \\
& + \int d^n\mathbf{r}' \int d^n\mathbf{r}'' \int d^n\mathbf{r}''' a(\mathbf{r},\mathbf{r}',\mathbf{r}'',\mathbf{r}''') s(\mathbf{r}') s(\mathbf{r}'') s(\mathbf{r}''') \\
& + O(s^4) + \varepsilon(\mathbf{r}) \qquad \text{(C462)}
\end{aligned}
$$

"variational derivatives"

$$
\begin{aligned}
a(\mathbf{r},\mathbf{r}') &= \frac{\delta s(\mathbf{r})}{\delta s(\mathbf{r}')} \qquad (C463) \\
a(\mathbf{r},\mathbf{r}',\mathbf{r}'') &= \frac{1}{2!} \frac{\delta^2 s(\mathbf{r})}{\delta s(\mathbf{r}')\delta s(\mathbf{r}'')} \\
a(\mathbf{r},\mathbf{r}',\mathbf{r}'',\mathbf{r}''') &= \frac{1}{3!} \frac{\delta^3 s(\mathbf{r})}{\delta s(\mathbf{r}')\delta s(\mathbf{r}'')\delta s(\mathbf{r}''')}
\end{aligned}
$$

"from continuous to discrete via Dirac functions: example"

$$
\int_{-\infty}^{+\infty} d^n\mathbf{r}' f(\mathbf{r}')\delta(\mathbf{r},\mathbf{r}') = f(\mathbf{r}) \qquad \text{(C464)}
$$

$$
\delta(\mathbf{r},\mathbf{r}') := \left[\begin{array}{ll} \to \infty & \text{for } \mathbf{r} = \mathbf{r}', \int_{-\infty}^{+\infty} d^n\mathbf{r}'\delta(\mathbf{r},\mathbf{r}') = 1 \\ 0 & otherwise. \end{array} \right.
$$

$$
\begin{aligned}
s(P) &= \int d^n\mathbf{r}' a(P,\mathbf{r}') s(\mathbf{r}') \qquad \text{(C465)} \\
&= \sum_{\alpha=1}^{N} K\alpha d^n\mathbf{r}'\delta(P,\mathbf{r}') = \sum_{\alpha=1}^{N} a(P,P_\alpha) s(P_\alpha)
\end{aligned}
$$

End of Table C.9.12

Example C.9.42: The hierarchy of the equations in nonlinear prediction of vector-valued signals

In order to take care of practice, we will present finally the concept of prediction of vector-valued signals set-up by *vector-valued Volterra series* of type (C641). The partial variational derivatives of the vector-valued functions, also called signal $S_i(\mathbf{r})$, are given by (C459) upto order $O(s^3$. We let the indices of Latin type run $1, 2, \cdots, n-1, n$ where n is the maximum dimension number. We apply the *Einstein summation convention*, namely the addition with respect to identical numbered indices.

In order to fill the deflections depend on the position $\mathbf{r}$ of the measurements. In order to determine the unknown coefficients of the nonlinear prediction "*ansatz*", we choose risk function to be optimized. For instance, we may choose the trace of the variance function $\mathfrak{T}_1 = tr\sigma_{ij}\mathbf{r} = min$, called *Helmert point error*, or the determinant of the variance function $\mathfrak{T}_2 = det\sigma_{ij}\mathbf{r} = min$, called *Werkmeister* *point error*, as *two basic invariants* called *Hilbert invariants*. Such an optimal design leads to a linear integral equation, a system of first kind, for the unknown tensors $a_{ij}(\mathbf{r}, \mathbf{r}')$, $a_{ijk}(\mathbf{r}, \mathbf{r}', \mathbf{r}'')$ etc. If we only refer to terms up to second order, namely to quadratic signal terms, we have eliminated cubic, biquadratic etc. term, the standard concept of *linear prediction*.

The first nonlinear term, the quadratic terms, lead to the concept of

two-point correlation functions:

$$\Phi_{ij}(\mathbf{r}, \mathbf{r}') := E\{s_i(\mathbf{r})s_j(\mathbf{r}')\}$$

three-point correlation functions:

$$\Phi_{ijk}(\mathbf{r}, \mathbf{r}', \mathbf{r}'') := E\{s_i(\mathbf{r})s_j(\mathbf{r}')s_k(\mathbf{r}'')\}$$

and

four-point correlation functions:

$$\Phi_{ijkl}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \mathbf{r}''') := E\{s_i(\mathbf{r})s_j(\mathbf{r}')s_k(\mathbf{r}'')s_l(\mathbf{r}''')\}$$

The term "*multipoint functions*" is to be understood that tensor valued functions between multipoints appear as correlation functions. With respect to *linear prediction* newly *three-point and four-point correlation functions appear*. Let us formulate the *discrete version* of the integral equations: here run all the bar indices $(1, 2, \cdots, Nn)$ up to the number Nn. Indeed, we approximated the integral (C470), (C471) etc. by (C472), (C473) etc. Indeed we switched from continuum to discrete by our summation convention by *neglecting the summation signs*. *For instance*, the two-point and three-point functions $a_{im}(\mathbf{r}, \mathbf{r_1})$, $a_{imn}(\mathbf{r}, \mathbf{r_1}, \mathbf{r_2})$ are written $a_{im}^{P\gamma}$, $b_{imn}^{P\gamma\delta}$ in the discrete versions.

In summary, our matrices have $Nn[(Nn)^2 + (Nn)]$ elements. They correspond to the correlation functions $\Phi_{\bar{i},\bar{j}}$, $\Phi_{\bar{i},\bar{j},\bar{k}}$ and $\Phi_{\bar{i},\bar{j},\bar{k},\bar{l}}$. Finally we restrict us to prediction terms of to *second order* when we write (C473) to (C476). Here we advantage of two conventions: *"Extensor" calculus* developed by Craig (1930–1939) abbreviates (a) as the *multiindex* the series $i_1 i_2, \cdots i_\infty$. Again we sum up indices in brackets. In the *new version* of *H. V. Craig's* multiindex we gain our characteristic equations (C476) - (C 479). The convention coefficients $a_i(i)$ have their origin in the *generalized Taylor series* with the condition that *higher order coefficients* are getting smaller with *increasing multiindex degree*. If we apply the *fundamental prediction equation* (C475) we have to sum up the coefficients $a_{\bar{i},\bar{i}}$, $a\bar{i}, \bar{i_1}, \bar{i_2}, a\bar{i}, \bar{i_1}, \bar{i_2}, \bar{i_3}$ etc. With the *generalized variance-covariance* matrix, function of the coordinates of the position vector $\mathbf{r}$. The Helmert point error reduces proportional to the number of prediction coefficients as long as the inequality (C481) holds. If we fix the coefficient functions $a_{ij}(\mathbf{r}, \mathbf{r}')$ for the distance zero to one, the following coefficients $a_{ijk}(\mathbf{r}, \mathbf{r}', \mathbf{r}'')$, $a_{ijkl}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \mathbf{r}''')$ are smaller than one.

Table C.9.44 Discrete and continuous Volterra series for vectors

continuous 2^{nd} order Volterra series

$$s_i(r) = \sum_{j=1}^{n} \int d^n\mathbf{r}' a_{ij}(\mathbf{r},\mathbf{r}') s_j(\mathbf{r}') + \sum_{j=1}^{n}\sum_{k=1}^{n} \int d^n\mathbf{r}' \int d^n\mathbf{r}'' a_{ijk}(\mathbf{r},\mathbf{r}',\mathbf{r}'') s_j\mathbf{r}' s_k\mathbf{r}'' + o(s^3) \tag{C466}$$

$$a_{ij}(\mathbf{r},\mathbf{r}') = \frac{\delta s_i(\mathbf{r})}{\delta s_j(\mathbf{r}')}, \quad a_{ijk}(\mathbf{r},\mathbf{r}',\mathbf{r}'') = \frac{\delta^2 s_i(\mathbf{r})}{\delta s_j(\mathbf{r}')\delta s_k(\mathbf{r}'')} \tag{C467}$$

Variance function as a tensor of 2^{nd} order

$$\sigma_{ij}(\mathbf{r}) = E\Big\{[s_i(\mathbf{r}) - \int d^n(\mathbf{r_1}) a_{ii_1}(\mathbf{r},\mathbf{r_1}) s_{i_1}(\mathbf{r_1}) - \int d^n(\mathbf{r_1}) \int d^n(\mathbf{r_2}) a_{ii_1i_2}(\mathbf{r},\mathbf{r_1},\mathbf{r_2}) s_{i_1}(\mathbf{r_1}) s_{i_2}(\mathbf{r_2}) - \cdots] [s_j(\mathbf{r}) - \int d^n(\mathbf{r_1}) a_{jj_1}(\mathbf{r},\mathbf{r_1}) s_{j_1}(\mathbf{r_1}) - \int d^n(\mathbf{r_1}) \int d^n(\mathbf{r_2}) a_{jj_1j_2}(\mathbf{r},\mathbf{r_1},\mathbf{r_2}) s_{j_1}(\mathbf{r_1}) s_{j_2}(\mathbf{r_2}) - \cdots]\Big\} \tag{C468}$$

$$\mathbf{J_1} = tr\sigma_{ij}(\mathbf{r}) = min, \quad \mathbf{J_2} = det\sigma_{ij}(\mathbf{r}) = min \tag{C469}$$

from continuous to discrete: multipoint correlation functions

$$\Phi_{ij}(\mathbf{r},\mathbf{r}') = \int d^n\mathbf{r_1} a_{im}(\mathbf{r},\mathbf{r_1}) \Phi_{jm}(\mathbf{r}',\mathbf{r_1}) + \int d^n\mathbf{r_1} \int d^n\mathbf{r_2} a_{imn}(\mathbf{r},\mathbf{r_1},\mathbf{r_2}) \Phi_{jmn}(\mathbf{r},\mathbf{r_1},\mathbf{r_2}) \tag{C470}$$

$$\Phi_{ijk}(\mathbf{r},\mathbf{r}',\mathbf{r}'') = \int d^n\mathbf{r_1} a_{im}(\mathbf{r},\mathbf{r_1}) \Phi_{jkm}(\mathbf{r}',\mathbf{r}'',\mathbf{r_1}) + \int d^n\mathbf{r_1} \int d^n\mathbf{r_2} a_{imn}(\mathbf{r},\mathbf{r_1},\mathbf{r_2}) \Phi_{jkmn}(\mathbf{r}',\mathbf{r}'',\mathbf{r_1},\mathbf{r_2}) \tag{C471}$$

$$\Phi_{jk}^{P\alpha} = a_{im}^{P\gamma}\Phi_{jm}^{\alpha\gamma} + b_{imn}^{P\gamma\delta}\Phi_{jmn}^{\alpha\gamma\delta}, \quad \Phi_{ijk}^{P\alpha\beta} = a_{im}^{P\gamma}\Phi_{jkm}^{\alpha\beta\gamma} + b_{imn}^{P\gamma\delta}\Phi_{jkmn}^{\alpha\beta\gamma\delta} \tag{C472}$$

$$\Phi_{\overline{j}\overline{k}} = a_{\overline{i}\overline{m}}\Phi_{\overline{j}\overline{m}} + b_{\overline{i}\overline{m}\overline{n}}\Phi_{\overline{j}\overline{m}\overline{n}}, \quad \Phi_{\overline{i}\overline{j}\overline{k}} = a_{\overline{i}\overline{m}}\Phi_{\overline{j}\overline{k}\overline{m}} + b_{\overline{i}\overline{m}\overline{n}}\Phi_{\overline{j}\overline{k}\overline{m}\overline{n}} \tag{C473}$$

$$\begin{gathered}
a_{\overline{i}\overline{1}} = c_{\overline{i}1}, \quad a_{\overline{i}\overline{2}} = c_{\overline{i}2}, \cdots, a_{\overline{i}\overline{m}} = c_{\overline{i}\overline{m}}, \cdots, a_{\overline{i}Nn} = c_{\overline{i}Nn} \\
b_{\overline{i}11} = c_{\overline{i}Nn+1}, \; b_{\overline{i}12} = c_{\overline{i}Nn+2}, \cdots, b_{\overline{i}1\overline{n}} = c_{\overline{i}Nn} + \overline{n}, \cdots, b_{\overline{i}1Nn} = c_{\overline{i}Nn} + Nn \\
b_{\overline{i}21} = c_{\overline{i}2Nn+1}, \; b_{\overline{i}22} = c_{\overline{i}2Nn+2}, \cdots, b_{\overline{i}2\overline{n}} = c_{\overline{i}2Nn} + \overline{n}, \cdots, b_{\overline{i}2Nn} = c_{\overline{i}2Nn} + Nn \\
b_{\overline{i}31} = c_{\overline{i}3Nn+1}, \; b_{\overline{i}32} = c_{\overline{i}3Nn+2}, \cdots, b_{\overline{i}3\overline{n}} = c_{\overline{i}3Nn} + \overline{n}, \cdots, b_{\overline{i}3Nn} = c_{\overline{i}3Nn} + Nn \\
\vdots \\
b_{\overline{i}\overline{m}1} = c_{\overline{i}\overline{m}Nn+1}, \; b_{\overline{i}\overline{m}2} = c_{\overline{i}\overline{m}Nn+2}, \cdots, b_{\overline{i}\overline{m}\overline{n}} = c_{\overline{i}\overline{m}Nn} + \overline{n}, \cdots,
\end{gathered}$$

$$b_{\bar{i}\overline{m}Nn} = c_{\bar{i}\overline{m}Nn} + Nn$$
$$b_{\bar{i}NnNn} = c_{\bar{i}(Nn)^2} \div Nn \tag{C474}$$

$$\Phi_{\mu'\nu'} = C_{\mu'\chi'}\Phi_{\chi'\nu'}, \quad C_{\mu'\lambda'} = \Phi_{\mu'\nu'}\Phi^{-1}_{\nu'\lambda'} \tag{C475}$$

$$\boxed{\Phi_{i(j)} - (a_{i(i)}, \Phi_{(j)(i)}) = 0} \tag{C476}$$

$$\Phi_{ij_1} - (a_{ii_1}, \Phi_{j_1i_1}) - (a_{ii_1i_2}, \Phi_{j_1i_1i_2}) - \cdots = 0 \tag{C477}$$

$$\Phi_{ij_1j_2} - (a_{ii_1}, \Phi_{j_1j_2i_1}) - (a_{ii_1i_2}, \Phi_{j_1j_2i_1i_2}) - \cdots = 0 \tag{C478}$$

$$(a_{ii_1}, \Phi_{j_1i_1}) = \int d^n\mathbf{r_1} a_{ii_1}(\mathbf{r}, \mathbf{r_1})\Phi_{j_1i_1})(\mathbf{r}', \mathbf{r_1}) \tag{C479}$$

$$(a_{ii_1i_2}, \Phi_{j_1j_2i_1i_2}) = \int d^n\mathbf{r_1} \int d^n\mathbf{r_2} a_{ii_1i_2}(\mathbf{r}, \mathbf{r_1}, \mathbf{r_2})\Phi_{j_1j_2i_1i_2})(\mathbf{r}', \mathbf{r}'', \mathbf{r_1}, \mathbf{r_2}) \tag{C480}$$

$$\sigma_{ij}(\mathbf{r}) = ((\delta_{i(i)} - a_{i(i)}), \Phi_{j(i)}) - ((\delta_{i(i)} - a_{i(i)}), a_{j(j)}\Phi_{(i)(j)}) \tag{C481}$$

$$2tr(a_{i(i)}, \Phi_{j(i)}) > tr(a_{i(i)}, a_{j(j)}, \Phi_{(i)(j)}) \tag{C482}$$

Table C.9.44: Volterra series

Example C.4. Structure analysis

Here we analyst the structure of multipoint functions of higher order with respect to statistical assumptions of *homogeneity and isotropy*. These assumptions made our prediction formula *much simpler*. If we were not assuming this statistical property, we had to compute for every point and every direction an independent higher order correlation function. For instance, if we had to calculate *two-point-correlation* functions in two dimensions every 1 Gon, we had to live with 1600 different correlation functions. If we would be bale to assume isotropy, the number of characteristic functions would reduce to *two functions* only! In the case of *three-point-functions* the reductions range from 3200 *non-isotropic functions to four* in case of isotropy. Of course, we will prove these results.

The structure of these multi-point-functions of statistically homogenous and isotropic vector-valued signals follow from the *form invariance* of the functions with respect to rotations and mirror functions of the coordinate system. Alternatively, we have to postulate invariance against transformations of the complete *orthogonality group* $O(n)$. *Form invariance* of a multi-point-correlation functions is to be

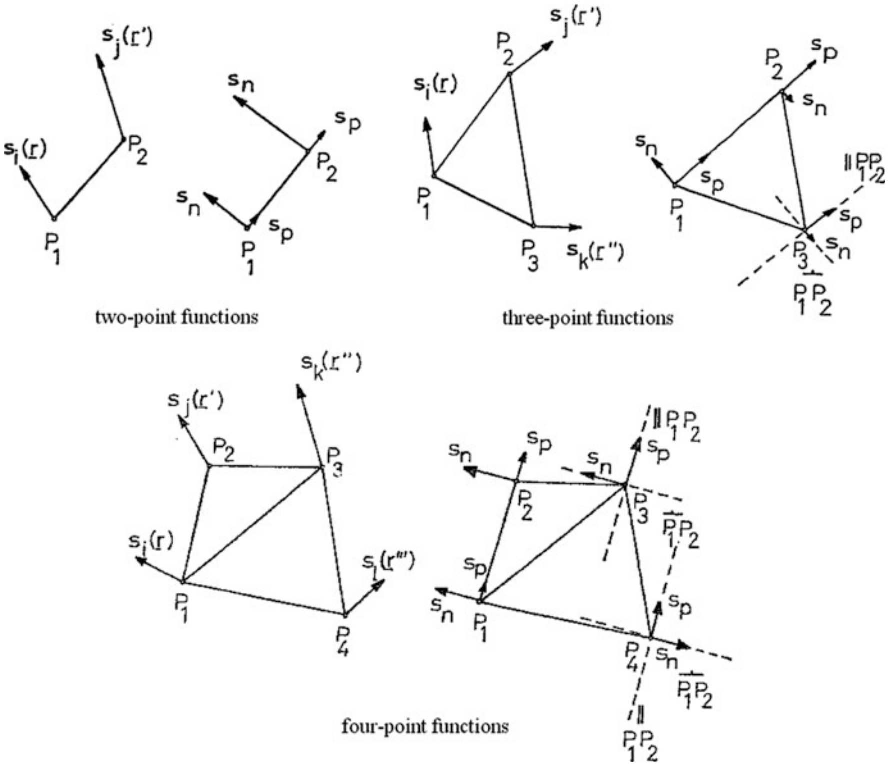


Fig. C38 Projections: *longitudinal and lateral* components for two-, three- and four point correlation functions, isotropy-homogeneity assumption

understood with respect to base functions which leads to scalar quantities. Our next section offers a better understanding of these theoretical results. As a short introduction to two-point-correlation os vector-valued signals we illustrate the theory of characteristic functions for two-point-functions, for three-point-functions and four-point-functions by Fig. C38.

In case of a vector valued functions the structure of the correlation function used for linear prediction is generated the coordinates $s_i(\mathbf{r})$ at the point $\mathbf{r}$ or P_1 and $s_j(\mathbf{r}')$ at the point $\mathbf{r}'$ or P_2 and the projected coordinate components along the connection line P_1P_2 (along track) and orthogonal to the line P_1P_2 (cross track) called "longitudinal" and "lateral" s_p and s_n ("parallel" p, "normal" n). These characteristic functions Φ_{nn} resp. Φ_{pp} given by the two-point correlation functions illustrated in Fig. C38.

Table C.15 Characteristic functions for two-point- and three-point correlation functions of vector-valued signals

"homogenous and isotropic two-point-function"

$$\Phi_{nn}(r) = E\{s_n(\mathbf{r})s_n(\mathbf{r}')\}, \quad \Phi_{pp}(r) = E\{s_p(\mathbf{r})s_p(\mathbf{r}')\}, \\ \Phi_{pn}(r) = \Phi_{np}(r) = 0 \tag{C483}$$

$$\Phi_j(\mathbf{r}, \mathbf{r}') = E\{s_i(\mathbf{r})s_j(\mathbf{r}')\} = \Phi_{nn}\delta_{ij} - [\Phi_{pp}(r) - \Phi_{nn}(r)]\frac{\Delta x_i \Delta x_j}{r^2} \tag{C484}$$

"homogenous and isotropic three-point-function"

$$\left.\begin{aligned} \Phi_{nnp}(r_1, r_2) &= E\{s_n(\mathbf{r}), s_n(\mathbf{r}'), s_p(\mathbf{r}'')\} \\ \Phi_{npn}(r_1, r_2) &= E\{s_n(\mathbf{r}), s_p(\mathbf{r}'), s_n(\mathbf{r}'')\} \\ \Phi_{pnn}(r_1, r_2) &= E\{s_p(\mathbf{r}), s_n(\mathbf{r}'), s_n(\mathbf{r}'')\} \\ \Phi_{ppp}(r_1, r_2) &= E\{s_p(\mathbf{r}), s_p(\mathbf{r}'), s_p(\mathbf{r}'')\} \end{aligned}\right\} \tag{C485}$$

$$\Phi_{ijk}(\mathbf{r}, \mathbf{r}', \mathbf{r}'') = \frac{x_{i'} - x_i}{|\mathbf{r} - \mathbf{r}'|}\delta_{jk}\Phi_{pnn}(r_1, r_2) + \frac{x_{j'} - x_j}{|\mathbf{r} - \mathbf{r}'|}\delta_{ik}\Phi_{npn}(r_1, r_2) + \\ = \frac{x_{k''} - x_k}{x_{1''} - x_1}\delta_{ij}\Phi_{nnp}(r_1, r_2) + \frac{(x_{i'} - x_i)(x_{j'} - x_j)(x_{k''} - x_k)}{|\mathbf{r} - \mathbf{r}'|^2(x_{1''} - x_1)} \cdot \\ \cdot \{\Phi_{ppp}(r_1, r_2) - \Phi_{pnn}(r_1, r_2) - \Phi_{npn}(r_1, r_2) - \Phi_{nnp}(r_1, r_2)\} \tag{C486}$$

End of Table: characteristic functions

In *Table C 16*, we present only results.
How is the detailed proof of our results?
This is here our central interest.

We begin with the definition of the popular scalar product and the norm of a *Hilbert space* by equations (c 484) and (c 488), extend by the variance-covariance function (C489) by *Craig calculus* generalizations of a *Hilbert space* are given by (C490) and (C491). In terms of *Craig calculus* we compute *Helmert's optimal design* of unknown coefficients by (C492), (C493), (C494) and (C495) and in contrast *Werkmeister's optimal design* by (C496), (C497) and (C498), again of the unknown coefficients.

Table C.16 Optimal variance function

Hilbert space: scalar product, norm

$$(f, g) = \int_\Omega f\omega\overline{g\omega}dP\omega \tag{C487}$$

$$|f_2| = \sqrt{(f,f)} = \sqrt{\int_\Omega |f\omega|^2 dP\omega} \tag{C488}$$

Variance function of vector-valued signals, Craig calculus

$$\sigma_{ij}(\mathbf{r}) = (\varepsilon_i(\mathbf{r}), \varepsilon_j(\mathbf{r})) = ((\delta_{i(i)} - a_{i(i)}), \Phi_{j(i)}) - ((\delta_{i(i)} - a_{i(i)}), a_{j(j)}\Phi_{i(j)}) \tag{C489}$$

Hilbert optimal design, Craig calculus

$$\left.\begin{array}{l} J_1 = \mathrm{tr}\,\sigma_{ij}(\mathbf{r}) = ((\delta_{i(i)} - a_{i(i)}), \Phi_{j(i)}) - ((\delta_{i(i)} - a_{i(i)}), a_{j(j)}\Phi_{i(j)}) = \min \\ \delta J_1[a_{I(i)}] = 0, \delta^2 J_1[a_{I(i)}] > 0 \end{array}\right\} \tag{C490}$$

$$\left.\begin{array}{l} (\frac{dJ_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon})_{\varepsilon=0} = 0, (\frac{d^2 J_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon^2})_{\varepsilon=0} > 0 \\ \\ J_1[a_{i(i)} + \varepsilon b_{i(i)}] = \mathrm{tr}\,((\delta_{i(i)} - (a_{i(i)} + \varepsilon b_{i(i)})), \Phi_{j(i)}) \\ -\mathrm{tr}\,((\delta_{i(i)} - (a_{i(i)} + \varepsilon b_{i(i)})), (a_{j(j)} + \varepsilon b_{j(j)})\Phi_{i(j)}) \\ \\ (\frac{dJ_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon})_{\varepsilon=0} = 0 : \mathrm{tr}\,(b_{i(i)}, (\Phi_{j(i)} - a_{j(j)}\Phi_{i(j)}) = 0 \end{array}\right\} \tag{C491}$$

$$(\frac{d^2 J_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon^2})_{\varepsilon=0} > 0 : \mathrm{tr}\,(b_{i(k)}, b_{j(l)}\Phi_{(k)(l)}) > 0 \tag{C492}$$

positive definiteness of the correlation function

$$\mathrm{tr}\, \int d^n\mathbf{r}' \int d^n\mathbf{r}'' b_{ik}(\mathbf{r}, \mathbf{r}') b_{ji}(\mathbf{r}, \mathbf{r}'') \Phi_{kl}(\mathbf{r}, \mathbf{r}'') > 0 \tag{C493}$$

Werkmeister optimal design, Craig calculus

$$\left.\begin{array}{l} J_2 = \det \sigma_{ij}(\mathbf{r}) = \{((\delta_{1(i)} - a_{1(i)}), \Phi_{1(i)}) - ((\delta_{1(i)} - a_{1(i)}), a_{1(j)}\Phi_{(i)(j)})\} \\ \quad \{((\delta_{2(i)} - a_{2(i)}), \Phi_{2(i)}) - ((\delta_{2(i)} - a_{2(i)}), a_{2(j)}\Phi_{(i)(j)})\} \\ \quad -\{((\delta_{1(i)} - a_{1(i)}), \Phi_{2(i)}) - ((\delta_{1(i)} - a_{1(i)}), a_{2(j)}\Phi_{(i)(j)})^2\} \end{array}\right\} \tag{C494}$$

$$\delta J_2[a_{i(i)}] = 0, \delta^2 J_2[a_{i(i)}] > 0$$

$$\left.\begin{array}{l} (\frac{dJ_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon})_{\varepsilon=0} = 0, (\frac{d^2 J_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon^2})_{\varepsilon=0} > 0 \\ \\ J_2[a_{i(i)} + \varepsilon b_{i(i)}] = A_0 + A_1\varepsilon + A_2\varepsilon^2 + A_3\varepsilon^3 + A_4\varepsilon^4 \\ \\ (\frac{d^2 J_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon^2})_{\varepsilon=0} = 0 : \quad A_1 = 0 \end{array}\right\} \tag{C495}$$

$$(\frac{d^2 J_1[a_{i(i)}+\varepsilon b_{i(i)}]}{d\varepsilon^2})_{\varepsilon=0} > 0: \quad A_2 > 0 \tag{C496}$$

End of Table: Craig optimal design

Finally, we study in more detail the concept of *structure functions*. We illustrate the form invariance of its *base invariants* which are invariant in terms of *rotational and mirror* symmetry. Let us call the base vector $e_i(\mathbf{r})$ relating to signal $s_i(\mathbf{r})$ at the placement vector $\mathbf{r}$, $e_j(\mathbf{r}')$ relating to signal $s_j(\mathbf{r}')$ at placement vector $\mathbf{r}'$ etc. The set-up of *base vectors* by (C449) and corresponding *structure functions* by (C580) are given. They lead us to the combination of base invariants Φ_1, Φ_2 following Taylor (1935), and Karman (1937), Φ_3, Φ_4, Φ_5 etc in Table (C501)–(C505). finally, we summarize up to *order three* the various isotropic and homogenous correlation functions in Table C.17, (C501)–(C505). They are based on the work of Karman and Howarth (1938), and Obuchow (1958). Let us present finally the number of characteristic functions for homogenous and isotropic correlation functions of higher order for vector-valued signals, collected in Table C.18.

Table C.17 Scalar products of base invariants

base invariants

$$\left.\begin{array}{l} x_i x_i, \quad x_{j'} x_{j'}, \quad x_{k''} x_{k''}, \quad x_{l'''} x_{l'''}, \text{ etc.} \\ e_i e_{i'}, \quad e_i e_{i''}, \quad e_i e_{i'''}, \text{ etc.} \\ e_{i'} e_{i''}, \quad e_{i'} e_{i'''}, \text{ etc.} \\ \text{etc.} \\ x_i e_i, \quad x_i e_{i'}, \quad x_i e_{i''}, \quad x_i e_{i'''}, \text{ etc.} \\ x_{i'} e_{i'}, \quad x_{i'} e_{i''}, \quad x_{i'} e_{i'''}, \text{ etc.} \\ x_{i''} e_{i''}, \quad x_{i''} e_{i'''}, \text{ etc.} \\ \text{etc.} \end{array}\right\} \tag{C497}$$

base functions

$$\left.\begin{array}{l} \Phi_1(e_i) = \Phi_i e_i \\ \Phi_2(e_i, e_{j'}) = e_i e_{j'} \Phi_{ij'}(\mathbf{r}, \mathbf{r}') \\ \Phi_3(e_i, e_{j'}, e_{k''}) = e_i e_{j'} e_{k''} \Phi_{ij'k''}(\mathbf{r}, \mathbf{r}', \mathbf{r}'') \\ \Phi_4(e_i, e_{j'}, e_{k''}, e_{l'''}) = e_i e_{j'} e_{k''} e_{l'''} \Phi_{ij'k''l'''}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \mathbf{r}''') \\ \ldots \end{array}\right\} \tag{C498}$$

combinations of base invariants

$$\Phi_1 = \Phi_1(e_i) = \Phi_1 e_i \tag{C499}$$

$$\Phi_2(e_i, e_{j'}) = e_i e_{j'} \Phi_{ij'}(\mathbf{r}, \mathbf{r}') = e_i e_{j'}\{\Delta x_i \Delta x_{j'} a_1(r) + \delta_{ij'} a_2(r)\} \tag{C500}$$

(Taylor (1935), Karman (1937))

$$\begin{aligned} \Phi_3(e_i, e_{j'}, e_{k''}) &= e_i e_{j'} e_{k''} \Phi_{ij'k''}(\mathbf{r}, \mathbf{r}', \mathbf{r}'') = \\ = e_i e_{j'} e_{k''}\{&\Delta x_i \Delta x_{j'} \Delta x_{k''} a_1(r_1, r_2) + \Delta x_i \delta_{j'k''} a_2(r_1, r_2) + \\ &+ \Delta x_{j'} \delta_{ik''} a_3(r_1, r_2) + \Delta x_{k''} \delta_{ij'} a_4(r_1, r_2)\} \end{aligned} \tag{C 501}$$

$$\left.\begin{aligned} &\Phi_4(e_i, e_{j'}, e_{k''}, e_{l'''}) = e_i e_{j'} e_{k''} e_{l'''} \Phi_{ij'k''l'''}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \mathbf{r}''') = \\ &= e_i e_{j'} e_{k''} e_{l'''}\{\Delta x_i \Delta x_{j'} \Delta x_{k''} \Delta x_{l'''} a_1(r_1, r_2, r_3) \\ &+ \Delta x_i \Delta x_{j'} \delta_{k''l'''} a_2(r_1, r_2, r_3) + \\ &+ \Delta x_{j'} \Delta x_{l'''} \delta_{ik''} a_3(r_1, r_2, r_3) + \Delta x_i \Delta x_{k''} \delta_{j'l'''} a_4(r_1, r_2, r_3) + \\ &+ \Delta x_i \Delta x_{l'''} \delta_{j'k''} a_5(r_1, r_2, r_3) + \Delta x_{j'} \Delta x_{k''} \delta_{il'''} a_6(r_1, r_2, r_3) + \\ &+ \Delta x_{k''} \Delta x_{l'''} \delta_{ij'} a_7(r_1, r_2, r_3) + \delta_{ij'} \delta_{k''l'''} a_8(r_1, r_2, r_3) + \\ &+ \delta_{ik''} \delta_{j'l'''} a_9(r_1, r_2, r_3) + \delta_{j'k''} \delta_{il'''} a_{10}(r_1, r_2, r_3)\} \end{aligned}\right\} \tag{C502}$$

$$\left.\begin{aligned} &\Phi_4(e_i, e_{j'}, e_{k''}, e_{l'''}, e_{m''''}) = e_i e_{j'} e_{k''} e_{l'''} e_{m''''} \Phi_{ij'k''l'''m''''}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \mathbf{r}''', \mathbf{r}'''') = \\ &= e_i e_{j'} e_{k''} e_{l'''} e_{m''''}\{\Delta x_i \Delta x_{j'} \Delta x_{k''} \Delta x_{l'''} \Delta x_{m''''} a_1(r_1, r_2, r_3, r_4) \\ &+ \Delta x_i \Delta x_{j'} \Delta x_{k''} \delta_{l'''m''''} a_2 + \\ &+ \Delta x_i \Delta x_{j'} \Delta x_{m''''} \delta_{k''l'''} a_3 + \Delta x_i \Delta x_{j'} \Delta x_{l''''} \delta_{k''m''''} a_4 \\ &+ \Delta x_i \Delta x_{l'''} \Delta x_{m''''} \delta_{j'k''} a_5 + \\ &+ \Delta x_i \Delta x_{k''} \Delta x_{m''''} \delta_{j'l'''} a_6 + \Delta x_i \Delta x_{k''} \Delta x_{l'''} \delta_{j'm''''} a_7 \\ &+ \Delta x_{k''} \Delta x_{l'''} \Delta x_{m''''} \delta_{ij'} a_8 + \\ &+ \Delta x_{j'} \Delta x_{l'''} \Delta x_{m''''} \delta_{ik''} a_9 + \Delta x_{j'} \Delta x_{k''} \Delta x_{m''''} \delta_{il'''} a_{10} \\ &+ \Delta x_{j'} \Delta x_{k''} \Delta x_{l'''} \delta_{im''''} a_{11} + \\ &+ \Delta x_i \delta_{j'k''} \delta_{l'''m''''} a_{12} + \Delta x_{j'} \delta_{ik''} \delta_{l'''m''''} a_{13} + \Delta x_{k''} \delta_{ij'} \delta_{l'''m''''} a_{14} + \\ &+ \Delta x_{l'''} \delta_{ij'} \delta_{k''m''''} a_{15} + \Delta x_{m''''} \delta_{ij'} \delta_{k''l'''} a_{16} + \Delta x_i \delta_{j'l'''} \delta_{k''m''''} a_{17} + \\ &+ \Delta x_{j'} \delta_{il'''} \delta_{k''m''''} a_{18} + \Delta x_{k''} \delta_{il'''} \delta_{j'm''''} a_{19} + \Delta x_{l'''} \delta_{ik''} \delta_{j'm''''} a_{20} + \\ &+ \Delta x_{m''''} \delta_{ik''} \delta_{j'l'''} a_{21} + \Delta x_i \delta_{k''l'''} \delta_{j'm''''} a_{22} + \Delta x_{j'} \delta_{im''''} \delta_{k''l'''} a_{23} + \\ &+ \Delta x_{k''} \delta_{im''''} \delta_{j'l'''} a_{24} + \Delta x_{l'''} \delta_{im''''} \delta_{j'k''} a_{25} + \Delta x_{m''''} \delta_{il'''} \delta_{j'k''} a_{26}\} \end{aligned}\right\} \tag{C503}$$

End of Table: Scalar products of basic invariants

Table C.18: Isotropic and homogenous correlation functions, two- and three- point functions, normal and longitudinal components

$$\left.\begin{aligned}\Phi_{11}(\mathbf{r},\mathbf{r}') &= a_1(r)r^2 + a_2(r) = \Phi_{pp}\\ \Phi_{12}(\mathbf{r},\mathbf{r}') &= \Phi_{12}(\mathbf{r},\mathbf{r}') = \Phi_{pn} = \Phi_{np} = 0\\ \Phi_{22}(\mathbf{r},\mathbf{r}') &= \Phi_{33}(\mathbf{r},\mathbf{r}') = \Phi_{nn}\end{aligned}\right\} \tag{C 504}$$

$$\left.\begin{aligned}\Phi_{111} &= r_1^2(x_1'' - x_1)a_1(r_1,r_2) + r_1[a_2(r_1,r_2) + a_3(r_1,r_2)]\\ &\quad +(x_1'' - x_1)a_4(r_1,r_2) = \Phi_{ppp}\\ \Phi_{112} &= r_1^2(x_3'' - x_2)a_1(r_1,r_2) + (x_2'' - x_2)a_4(r_1,r_2)\\ \Phi_{113} &= r_1^2(x_3'' - x_3)a_1(r_1,r_2) + (x_3'' - x_3)a_4(r_1,r_2)\\ \Phi_{121} &= 0,\ \ \Phi_{122} = r_1 a_2 = \Phi_{pnn},\ \ \Phi_{123} = 0\\ \Phi_{131} &= 0,\ \ \Phi_{132} = 0,\ \ \Phi_{133} = r_1 a_2 = \Phi_{pnn}\\ \Phi_{211} &= 0,\ \ \Phi_{212} = r_1 a_3 = \Phi_{npn},\ \ \Phi_{213} = 0\\ \Phi_{231} &= \Phi_{232} = \Phi_{233} = 0\\ \Phi_{311} &= \Phi_{312} = \Phi_{313} = r_1 a_3 = \Phi_{npn}\\ \Phi_{321} &= \Phi_{322} = \Phi_{323} = 0\\ \Phi_{331} &= (x_1'' - x_1)a_4 = \Phi_{nnp},\ \ \Phi_{332} = (x_2'' - x_2)a_4,\\ \Phi_{333} &= (x_3'' - x_3)a_4\end{aligned}\right\} \tag{C505}$$

$$a_1(r) = \frac{\Phi_{pp}-\Phi_{nn}}{r^2},\ \ a_2(r) = \Phi_{nn}$$

$$\left.\begin{aligned}a_1(r_2,r_2) &= \frac{\Phi_{ppp}-\Phi_{pnn}-\Phi_{npn}-\Phi_{nnp}}{r_1^2(x_1''-x_1)}\\ a_2(r_2,r_2) &= \frac{\Phi_{pnn}}{r_1}\\ a_3(r_2,r_2) &= \frac{\Phi_{npn}}{r_1}\\ a_4(r_2,r_2) &= \frac{\Phi_{nnp}}{(x_1''-x_1)}\end{aligned}\right\} \tag{C506}$$

$$\Phi_{ij}(\mathbf{r},\mathbf{r}') = \Phi_{nn}(r)\delta_{ij} + [\Phi_{pp}(r) - \Phi_{nn}(r)]\frac{\Delta x_i\,\Delta x_j}{r^2} \tag{C507}$$

$$\begin{aligned}\Phi_{ijk}(\mathbf{r},\mathbf{r}',\mathbf{r}'') &= \frac{\Delta x_i}{r_1}\delta_{jk}\Phi_{nnn} + \frac{\Delta x_j}{r_1}\delta_{ik}\Phi_{npn} + \frac{(x_k''-x_k)}{(x_1''-x_1)}\delta_{ij}\Phi_{nnp}\\ &\quad + \frac{\Delta x_i\,\Delta x_j\,(x_k''-x_k)}{r_1^2(x_1''-x_1)}(\Phi_{ppp} - \Phi_{pnn} - \Phi_{npn} - \Phi_{nnp})\end{aligned} \tag{C508}$$

$$\begin{aligned}\Phi_{ij'k'}(\mathbf{r},\mathbf{r}'=\mathbf{r}'') &= \Delta x_i\,\Delta x_{j'}\,\Delta x_{k'}a_1(r_1) + \Delta x_i\,\delta_{j'k'}a_2(r_1)\\ &\quad +(\Delta x_{j'}\delta_{ik'} + \Delta x_{k'}\delta_{ij'})a_3(r_1)\\ &= \frac{\Phi_{ppp}(r_1)-\Phi_{pnn}(r_1)-2\Phi_{nnp}(r_1)}{r_1^3}\Delta x_i\,\Delta x_{j'}\,\Delta x_{k'}\\ &\quad +\Phi_{pnn}(r_1)\frac{\Delta x_i}{r_1}\delta\, j'k' + \frac{(\Delta x_{j'}\delta_{ik'}+\Delta x_{k'}\delta_{ij'})}{r_1}\Phi_{nnp}(r_1)\end{aligned} \tag{C509}$$

End of Table: Isotropy and homogenity

Table C.19:Number of characteristic functions for multi-point functions based on vector-valued signals

multi-point correlation functions	*number of characteristic functions*
2	2
3	4
4	10
5	26

End of Table: characteristic functions

Ex 5: Nonlocal Time Series Analysis

Example: Nonlocal time series analysis

Classical time series analysis, for instance Earth tide measurements, is based on *local* harmonic and disharmonic effects. Here we present a *nonlocal concept* to analyze on a *global scale* measurements at different locations. The tool of the *generalized harmonic analysis* base on *correlation functions of higher order* between vector-valued signals at different points and at different times. Its *polyspectrum* is divided into *poly-coherence* and *poly-phase*. A 3d-filder of *Kolmogorov-Wiener type* is designed for an array analysis. The optional distance between stations and the optimal registration time- Nyquist frequency- is given to avoid aliasing.

Standard procedures of signal analysis treat a registered time series of *one station* only. It is our target here to *combine various observation stations*: we are interested in the *global picture* handled by *array analysis*. We like to refer to applications in seismology and *tidal analysis*, the great success of *tomography*. A 3d- or *multi-channel filter* makes possible to arrange an optimal signal-to-noise ratio in terms of an optimal design of the station distance and an optimal design of the registration time, namely statistical accuracy of numbers which decide upon a linear or nonlinear filter.

In the *first section* we introduce the classical *nonlocal analysis* for $N = 2$ stations for a vector-valued signal. We assume no correlation between the signal components. The *mono-spectrum*, namely the *Fourier-transform* of scalar-valued correlation functions is separated by *coherence-* and *phase-graphs*. The *second section* is a straight forward approach to registered vector-valued signals between $N = 2$ stations. Support has the origin in the analysis of *coherence surface* and of *phase surface* of Fourier-transformed *correlation-tensors*. *Poly-spectra* of signal *space-time functions* form the basis of $N-$ stations analysis in the *third section*. We separate real- and imaginary-parts in terms of a poly-coherence-graph and of a poly-phase-graph, both matrix forms. An extended array analysis is based on a *Kolmogorov-Wiener filter* in the *fourth section*. We give a simple analysis to predict an optimal registration and an optimal station distance. At the end we give references to (a) *array analysis*, (b) *mono-spectrum* and (c) *bi- and poly-spectra*.

Scalar signals, "classical analysis"

In terms of the "*classical time series analysis*" registrations at two different places are to be combined: Which are the frequency bands in which we can detect *coherences*, which is the spectral phase-dislocation between two registrations. We are to keep in mind the *Fundamental Lemma* of *N. Wiener* that a frequency assured by coherence has the property of *maximal probability*, characterized by signal time-functions at two stations. We sum up the following operations:

(i) Calculation of auto- and cross-correlation functions of scalar-valued signals,
(ii) Calculation of Fourier-transforms of auto- and cross-correlation functions, separation of mono-spectra in terms of active and passive (cos – versus sin – terms) components,
(iii) Plots of coherence graphs $H^2(\mathbf{k}_1, \omega_1)$ and phase graphs $\Psi^2(\mathbf{k}_1, \omega_1)$

Here we depart from a scalar space-time representation of a signal, namely by a signal $s(\mathbf{r}, t)$ at a placement $\mathbf{r}$ and at the time t *and* a signal $s(\mathbf{r}', t')$ at the placement $\mathbf{r}'$ and at the time t' by *correlation*, in detail by (C512 - C 519). $f[s(\mathbf{r}, t), s(\mathbf{r}', t')]$ denotes the probability density, to meet the signals s and s' at the placements $\mathbf{r}, \mathbf{r}'$ and the times t, t'. we intend to transform the probability integral into a space-time integral assuming *homogeneity and ergodicity* by (c 513) and (C514). The space-time complex mono-spectrum $\Phi(\mathbf{r}_i, t_1) \to \widetilde{\Phi}(\mathbf{k}_i, \omega_1)$ is defined by (C515), *direct and inverse*, (C516), $\widetilde{\Phi}(\mathbf{k}_i, \omega_1)$ decomposed into a *real-valued part* and an *imaginary part*. "$\widetilde{\ }$" denotes the *Fourier transform*, $i\ :\ -\sqrt{-1}$, n indicates the dimension. In analogy to wave propagation in electrodynamics, we define $Rc\{\widetilde{\Phi}\} =: \widetilde{\Gamma}(\mathbf{k}_i, \omega_1)$ active spectrum or co-spectrum, but $Jm\{\widetilde{\Phi}\} =: \widetilde{\Lambda}$ quadratic or "blind spectrum". Backward transformation $\Phi(\mathbf{r}_i, \tau_1)\cos(\omega_1\tau_1 - < \mathbf{k}_1|\mathbf{r}_1 >)$ and $\Phi(\mathbf{r}_i, \tau_1)\sin(\omega_1\tau_1 - < \mathbf{k}_1|\mathbf{r}_1 >)$ lead to $\{\widetilde{\Gamma}(\mathbf{k}_i, \omega_1), \widetilde{\Lambda}(\mathbf{k}_i, \omega_1)\}$ Let us finally $H^2 := \widetilde{\Gamma}^2(\mathbf{k}_i, \omega_1) + \widetilde{\Lambda}^2(\mathbf{k}_i, \omega_1)\}$ as the *space-time coherence graph*, and $\tan\Psi(\mathbf{k}_i, \omega_1)\} := \widetilde{\Lambda}(\mathbf{k}_i, \omega_1)/\widetilde{\Gamma}(\mathbf{k}_i, \omega_1), \Psi(\mathbf{k}_i, \omega_1)$ the *phase graph*, both very important in defining "coherent light". Figure C39 illustrates the *phase graph*, $\widetilde{\Gamma}^2 + \widetilde{\Lambda}^2$ the length of the complex vector in the polar coordinates. The auto correlation functions $\Gamma(\mathbf{r}, \mathbf{r})$ and $\Lambda(\mathbf{r}', \mathbf{r}')$ clearly lead us to "angle" and "length" in the complex plane Fig. C40 illustrates the fundamental coherence- and phase-functions.

Finally the contributions of lackman (1965); Granger and Hatanaka (1964); Harris (1967); Kertz (1969, 1971); Lee (1960); Pugachev (1965); Robinson (1967), Robinson and Treitel (1964).

Table C.20: Variance-covariance function of a scalar signal, direct and inverse Fourier transform, space-time representation

$$E[s(\mathbf{r},t)s(\mathbf{r}',t')] := \Phi(\mathbf{r},\mathbf{r}',t,t') \tag{C510}$$

$$:= \int_{-\infty}^{+\infty} ds(\mathbf{r},t) \int_{-\infty}^{+\infty} ds(\mathbf{r}',t')s(\mathbf{r},t)s(\mathbf{r}',t')f[s(\mathbf{r},t)s(\mathbf{r}',t)] \tag{C511}$$

$$\Phi(\mathbf{r}'-\mathbf{r},t'-t) = \lim_{T,x,y\to\infty} \frac{1}{8Txy} \int_{-T}^{+T} dt \int_{-x}^{+x} dx \int_{-y}^{+y} dy \tag{C512}$$

$$s(x,y,t)s[x+(x'+x),y+(y'+y),t+(t'+t)] \tag{C513}$$

$$\boxed{\mathbf{r}'-\mathbf{r} := \mathbf{r}_1, \quad t'-t := \tau_1} \tag{C514}$$

$$\widetilde{\Phi}(\mathbf{k}_1,\omega_1) = (2\pi)^{\frac{-n+1}{2}} \int_{-\infty}^{+\infty} d^n\mathbf{r}_1 \int_{-\infty}^{+\infty} d\tau_1 \Phi(\mathbf{r}_1,\tau_1) \exp +i(\omega_1\tau_1 - \mathbf{k}_1\mathbf{r}_1) \tag{C515}$$

$$\Phi(\mathbf{r}_1,\tau_1) = (2\pi)^{\frac{-n+1}{2}} \int_{-\infty}^{+\infty} d^n\mathbf{k}_1 \int_{-\infty}^{+\infty} d\omega_1 \widetilde{\Phi}(\mathbf{k}_1,\omega_1) \exp -i(\omega_1\tau_1 - \mathbf{k}_1\mathbf{r}_1) \tag{C516}$$

$$\widetilde{\Phi}(\mathbf{k}_1,\omega_1) = Re[\widetilde{\Phi}] + i \;\; Im[\widetilde{\Phi}] = \widetilde{\Gamma}(\mathbf{k}_1,\omega_1) + i \;\; \widetilde{\Lambda}(\mathbf{k}_1,\omega_1) \tag{C 517}$$

$$\widetilde{\Lambda}(\mathbf{k}_1,\omega_1) := (2\pi)^{\frac{-n+1}{2}} \int_{-\infty}^{+\infty} d^n\mathbf{r}_1 \int_{-\infty}^{+\infty} d\tau_1 \Phi(\mathbf{r}_1,\tau_1) \cos(\omega_1\tau_1 - \mathbf{k}_1\mathbf{r}_1) \tag{C518}$$

$$\widetilde{\Gamma}(\mathbf{k}_1,\omega_1) := (2\pi)^{\frac{-n+1}{2}} \int_{-\infty}^{+\infty} d^n\mathbf{r}_1 \int_{-\infty}^{+\infty} d\tau_1 \Phi(\mathbf{r}_1,\tau_1) \sin(\omega_1\tau_1 - \mathbf{k}_1\mathbf{r}_1) \tag{C519}$$

$$H^2 := \widetilde{\Gamma}^2(\mathbf{k}_1,\omega_1) + \widetilde{\Lambda}^2(\mathbf{k}_1,\omega_1) \tag{C520}$$

$$tg\Psi(\mathbf{k}_1,\omega_1) := \frac{\widetilde{\Lambda}^2(\mathbf{k}_1,\omega_1)}{\widetilde{\Gamma}^2(\mathbf{k}_1,\omega_1)} \tag{C521}$$

End of Table: variance-covariance functions

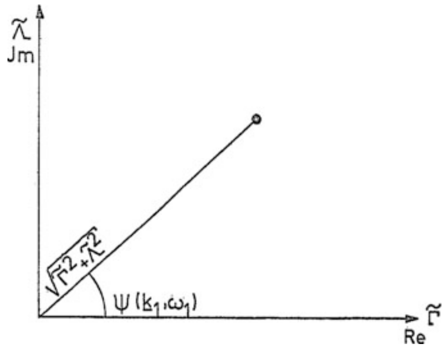


Fig. C39 Decomposition into real and imaginary component Gauss representation $\{\widetilde{\Lambda},\widetilde{\Gamma}\}$

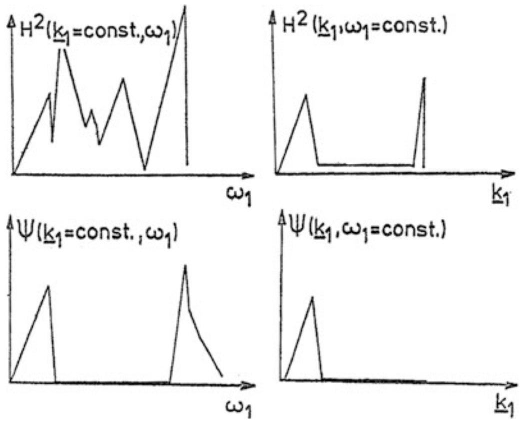


Fig. C40 Scalar coherence and phase graph

C.9.52 First generalization: *vector-valued classical analysis*

At first, we deal with separate single components of a general *three-component tidal recording*. We develop a joint analysis of all three components *observed at two points* on the surface of the *Earth*, for example. Its signal functions are correlated with each other, in the special case here the *vertical component* as well as the *horizontal component*. We receive a correlation tensor of second degree, namely in the form of the *coherence matrix* and the *phase matrix*. In detail, the analysis technique is based on the following operations:

(i) Computation of auto- and cross-correlations of the vector-valued signal,
(ii) Computation of the Fourier transform of the functions separation within the mono spectral matrix in "*active*" and "*blind*" spectra *cosine*- and *sine-Fourier transform*,
(iii) Part of the coherenec matrix $H_{ij}(\mathbf{k}_1,\omega_1)$ as well as the phase matrix $\Psi_{ij}(\mathbf{k}_1,\omega_1)$.

Theoretical Background

Let $s_i(\mathbf{x}, t) = \delta g_i(\mathbf{x}, t)$ be a vector-valued space-time-signal consisting of vertical and horizontal components subject to $i = 1$ or $\mathbf{x}, i = 2$ or y and $i = 3$ or z. The signal function $s_i(\mathbf{x}, t)$ at the placement $\mathbf{x}$ and at the time instant t *and* the signal function $s_j(\mathbf{x}', t')$ at the placement $\mathbf{x}'$ and the time instant t' are correlated by means of

$$E\{s_i(\mathbf{x}, t)s_j(\mathbf{x}', t')\} =: \Phi_{ij}(\mathbf{x}, \mathbf{x}', t, t') \text{ for } i, j \in \{1, 2, 3\} \tag{C522}$$

As an analogue of our previous formulae we define the *coherence matrix*, namely the *coherence tensor* H_{ij}, and the *phase matrix* Ψ_{ij}

$$H_{ij}(\mathbf{k}_1, \omega_1) := \widetilde{\Gamma}^2_{ij}(\mathbf{k}_1, \omega_1) + \widetilde{\Lambda}^2_{ij}(\mathbf{k}_1, \omega_1) \tag{C523}$$

$$tg \;\; \Psi_{ij}(\mathbf{k}_1, \omega_1) = \frac{\widetilde{\Lambda}^2_{ij}(\mathbf{k}_1, \omega_1)}{\widetilde{\Gamma}^2_{ij}(\mathbf{k}_1, \omega_1)} \tag{C524}$$

The *coherence matrix* as well as the *phase matrix* are represented by special surface for instance by coherence ellipses, coherence hyperbolae etc.

C.9.53 Second generalization: poly spectra

Up to now we analyzed only signal registrations at two different points by correlation, namely by coherence and phase. *Here*, we introduce the concept of *poly spectra* for the purpose of registration of Earth tides *at all points we have registrations*. In detail, the analysis technique is based on the following operations:

(i) Computation of multi-point correlation functions of a vector-valued tidal signal
(ii) Computation of the Fourier transforms called *polyspectra*. The concept is able to separate mono-, bi-, tri-, in short *polyspectra* in "active" and "blind" spectra, cosine- and sine-Fourier transform.
(iii) Part of *generalized coherence graphs*

$$H_{i_1 i_2 \cdots i_{N-1} i_N}(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_{N-1}, \omega_1, \omega_2, \cdots, \omega_{N-1}) \tag{C525}$$

and the *generalized phase graphs*

$$\Psi_{i_1 i_2 \cdots i_{N-1} i_N}(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_{N-1}, \omega_1, \omega_2, \cdots, \omega_{N-1}) \tag{C526}$$

at N tidal stations being functions of *wave number vectors* $\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_{N-1}$ and the *frequencies* $\omega_1, \omega_2, \cdots, \omega_{N-1}$.

Theoretical Background

Let $s_{i_1}(\mathbf{x}, t) - \delta g i(\mathbf{x}, t)$ be a tensor-valued space-time tidal signal at the first tidal station, $s_{i_2}(\mathbf{x}', t)$ be a tensor-valued tidal signal at the second tidal station. If we correlate all signal functions with each other, e receive a multi-point correlation function equal to the number of tidal stations.

$$E\{s_{i_1}(\mathbf{x},t)s_{i_2}(\mathbf{x}',t')\cdots s_{i_n}(\mathbf{x}_N,t_n)\} := \Phi_{i_1 i_2 \cdots i_{N-1} i_N} \tag{C527}$$

In case of space like homogeneity *and* the time like stationarity we are able to represent the various correlation functions by

$$\Phi_{i_1 i_2 \cdots i_{N-1} i_N}(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N-1}, \tau_1, \tau_2, \cdots, \tau_{N-1}) \tag{C528}$$

subject to

$$\mathbf{x}_1 := \mathbf{x}' - \mathbf{x}, \quad \mathbf{x}_2 := \mathbf{x}'' - \mathbf{x}, \quad \text{etc } .\tau_1 = t' - t, \quad \tau_2 = t'' - t, \quad \text{etc} \tag{C529}$$

and their polyspectra

$$\begin{aligned}
&\widetilde{\Phi}_{i_1 i_2 \cdots i_{N-1} i_N}(\mathbf{k}_1, \mathbf{k}_2, \cdots, \mathbf{k}_{N-1}, \omega_1, \omega_2, \cdots, \omega_{N-1}) = \\
&(2\pi)^{-\frac{n(N-1)+(N-1)}{2}} \int\limits_{-\infty}^{+\infty} d^{n(N-1)}\mathbf{r} \int\limits_{-\infty}^{+\infty} d^{n(N-1)}\tau \\
&\Phi_{i_1 i_2 \cdots i_{N-1} i_N}(r_1, r_2, \cdots, r_{N-1}, \tau_1, \tau_2, \cdots, \tau_{N-1}) * \\
&\exp + i(\omega_1\tau_1 + \omega_2\tau_2 + \cdots + \omega_{N-1}\tau_{N-1} - \mathbf{k}_1\mathbf{r}_1 - \mathbf{k}_2\mathbf{r}_2 - \cdots - \mathbf{k}_{N-1}\mathbf{r}_{N-1})
\end{aligned} \tag{C530}$$

n is the dimension number, N the number of tidal stations. For the case of $N = 3$ we use the term "*bispectrum*", for the case $N = 4$ we use the term "*trispectrum*" in general "*polyspectrum*". The common separation of $\Phi_{i_1 i_2 \cdots i_{N-1} i_N}$ in real- and imaginary part leads us to

$$\boxed{\text{the Polycoherence graph as well as the poly phase graph}}$$

$$H_{i_1 i_2 \cdots i_{N-1} i_N} = Rc^2[\widetilde{\Phi}_{i_1 i_2 \cdots i_{N-1} i_N}] + Jm^2[\widetilde{\Phi}_{i_1 i_2 \cdots i_{N-1} i_N}] \tag{C531}$$

$$tg\Psi_{i_1 i_2 \cdots i_{N-1} i_N} = \frac{Jm[\widetilde{\Phi}_{i_1 i_2 \cdots i_{N-1} i_N}]}{Rc[\widetilde{\Phi}_{i_1 i_2 \cdots i_{N-1} i_N}]} \tag{C532}$$

Numerical results are presented in Brillinger (1964), Grafarend (1972), Hasselmann et al. (1963), Rosenblatt (1965, 1966), Sinaj (1963 i, ii), Tukey (1959), and Van Nees (1965).

Table C.9.54: Generalized array analysis on the basis of a Kolmogorov-Wiener filter (3D-filter)

We do *not* attempt to review "*filter theory*", instead we present *array analysis* for the purpose to optimally *reduce noise* in *tidal measurements* in optimal station density. The method of analysis is based in the following operations:

(i) Computation of signal correlation functions from tidal registration in multi-stations and afterwards in noise correlation functions from an analysis of noise characteristic multi-dimensional stochastic process, for instance of type Markov, homogenous and isotropic
(ii) Computation of spectra from diverse correlation functions
(iii) Solving the linear matrix equation of type *Wiener-Hopf equations* of optimal filter functions.

The *optimal station density* is achieved from the postulate $< \Delta\mathbf{k}|\Delta\mathbf{x} >\leq 1$, the *optimal registration time* from the postulate $\Delta\omega\Delta t \leq 1$. $\Delta\mathbf{x}$ is the distance of the tidal stations in *position space*, $\Delta\mathbf{x}$ the distance in the *Fourier space*, for instance $\mathbf{k}_x = 2\pi/\lambda_x$ *wavelength aliasing* Δt the discretization parameter in the *time domain* and $\Delta\omega$ the corresponding *frequency distance* in the *Fourier space*, for instance $\Delta\omega = 2\pi/\Delta T$, $\Delta\omega$ the *frequency aliasing*.

Example

$$\Delta r \leq \frac{1}{\Delta k \cos \sphericalangle\mathbf{k},\mathbf{r}} = \frac{\Delta\lambda}{2\pi \cos \sphericalangle\mathbf{k},\mathbf{r}}$$

$$\Delta\omega \leq 1/\Delta t$$

$$\sphericalangle\mathbf{k},\mathbf{r} = 0, \quad \Delta\lambda = 10^3 \text{ Km}, \quad \Delta r \cong 140 \text{ Km}$$

$$\sphericalangle\mathbf{k},\mathbf{r} = 30^0, \quad \Delta\lambda = 10^3 \text{ Km}, \quad \Delta r \cong 330 \text{ Km}$$

$$\Delta t = 1 \text{ sec}, \quad \Delta\omega = 1 \text{ Hz}, \quad \Delta T \cong 6 \text{ sec}$$

$$\Delta t = 1 \text{ msec}, \quad \Delta\omega = 10^3 \text{ Hz}, \quad \Delta T \cong 6 \text{ msec}$$

Theoretical Background

Lets us assume that the vector-valued registration signal consists of *two parts*, the *pure signal* part and the *superimposed noise*:

$$\delta g_i(\mathbf{x},t) = s_i(\mathbf{x},t) + n_i(\mathbf{x},t) \tag{C533}$$

The filtered signal $\delta g_i'(\mathbf{x},t)$ has the linear structure of type

$$\delta g_i'(\mathbf{x},t) = s_i'(\mathbf{x},t) + n_i'(\mathbf{x},t) \tag{C534}$$

The *filter error* $\varepsilon_i(\mathbf{x},t) = s'_i(\mathbf{x},t) - s_i(\mathbf{x},t) + n'_i(\mathbf{x},t) - n_i(\mathbf{x},t)$ is designed by $I_1 tr\sigma_{ij}(\mathbf{x},t) = \min$ *or* $I_1 \det\sigma_{ij}(\mathbf{x},t) = \min$ subject to the *error tensor* $\sigma_{ij}(\mathbf{x},t) := E\{\varepsilon_i(\mathbf{x},t)\varepsilon_j(\mathbf{x},t)\} = \min$. The *filter* $f_{ij}(\mathbf{x},t)$ is defined by the *linear Volterra setup*

$$\boxed{n'_i(\mathbf{x},t) := \int\limits_{-\infty}^{+\infty} d^n\mathbf{x}_1 \int d\tau_1 f_{ij}(\mathbf{x}_1,\tau_1)n_j(\mathbf{x}-\mathbf{x}_1,t-\tau_1)} \tag{C535}$$

$J[\sigma_{ij}(f_{kl}(\mathbf{x},t))] = \min$ lead the "necessary" generalized Wiener-Hopf equations.

$$\Phi_{s_i s_j}(\mathbf{r},\mathbf{r}',t,t') = \int\limits_{-\infty}^{+\infty} d^n\mathbf{r}'' \int\limits_{-\infty}^{+\infty} dt'' f_{ik}(\mathbf{r},\mathbf{r}'')\Phi_{jk}(\mathbf{r},\mathbf{r}'') \tag{C536}$$

$$\widetilde{f}_{ij} = \widetilde{\Phi}_{s_i s_j}(\mathbf{k}_1,\omega_1)[\widetilde{\Phi}_{s_i s_k} + \widetilde{\Phi}_{n_i n_k}]^+ \tag{C537}$$

$[]^+$ identifies the *generalized inverse* of type *minimum norm least squares*.

$$\widetilde{F}_{[i\times i]} = \widetilde{\Phi}_{ss[i\times i]}[\widetilde{\Phi}_{ss[i\times i]} + \widetilde{\Phi}_{nn[i\times i]}] \tag{C538}$$

The result of 3D or linearized multi-channel leads to an easy solution. If we would care for a nonlinear filter subject to the multi-index $J := (j_1, j_2, \cdots, j_\alpha)$ and $K := (k_1, k_2, \cdots, k_\beta)$ we will receive the solution

$$\Phi_{ij} = (f_{ik}, \Phi_{jk}) \tag{C539}$$

with respect to the scalar product $(.,.)$ of the corresponding Hilbert space. Numericals examples you can find in C. B. Archambeau et. al. (1965), J. W. Britill and F. E. Whiteway (1965), H. W. Briscoe and P. L. Fleck (1965), J. P. Burg (1964), J. F. Claerbout (1964), P. Embree et. al. (1963), E. Grafarend (1972), A. N. Kolmogorov (1941), N. Wiener (1949, 1964, 1968).

Comments

In *experimental sciences*, namely geosciences like geophysics, geodesy, environmental sciences, namely, weather, air pollution, rainfall, oceanography, temperature affected sciences take action on a *spatially and temporally large scale to very large scales*. For this reason, they *never can be considered* stationary or homogenous-isotropic, but

non-stationary inhomogeneous or anisotropic

in the statistical sense Spock and Pilz (2008) if not Guttorp and Sampson (1994) as one of the attempts *to model non-stationary*, other examples are geodetic networks in two- and three-dimensional *Euclidean space* and on two-, three- and four-dimensional manifolds including the sphere, the ellipsoid or *General Relativity* as *an excellent example* for *inhomogeneity and anisotropy presented here. C. Calder* and *N. Cressie* (2007) review different directions in spectral and convolution based on *Gaussian non-stationary random modeling*. Speck and Pilz (2008) present a spectral decomposition without any restriction to the Gaussian approach. They advice the reader to apply a *special Generalized Linear Mixed Model* (GLMM), the topic of our book, namely derived from the spectral decomposition of *non-stationary random fields*.

Topics

(i) Stationary and isotropic random field

versus

(ii) Non-stationary, locally isotropic random field
(iii) Interpretation in the context of linear mixed models (LMMs)

$$Y = F\beta + Ab + \varepsilon_0$$

a. $F\beta$: design matrix of the first kind, result: regression function, trend behaviour F fixed design matrix
b. Ab: non-deterministic random fluctuations around the trend functions, A fixed design matrix.
c. ε_0: independent, homoscedastic error term centered at zero

(iv) Relation to non-stationary, locally isotropic random fields
(v) Generalized linear mixed models (GLMMs)
(vi) Extending generalized linear mixed models (DHGLMMs)
(vii) Limear and mixed models and Hermitic polynomials
(viii) Anisotropic stochastic process axisymmetric, Chandrasekhar (1950), Grafarend (1972)

Appendix D
Basics of Groebner Basis Algebra

D-1 Definitions

To enhance the understanding of the theory of Groebner bases presented in Chap. 15, the following definitions supplemented with examples are presented. Three commonly used monomial ordering systems are; *lexicographic ordering, graded lexicographic ordering* and *graded reverse lexicographic ordering*. First we define the monomial ordering before considering the three types.

Definition D.1. (Monomial ordering):

A monomial ordering on $k\,[x_1, \ldots, x_n]$ is any relation $>$ on $\mathbb{Z}^n_{\geq 0}$ or equivalently any relation on the set x^α, $\alpha \in \mathbb{Z}^n_{\geq 0}$ satisfying the following conditions:
(a) Is total (or linear) ordering on $\mathbb{Z}^n_{\geq 0}$
(b) If $\alpha > \beta$ and $\gamma \in \mathbb{Z}^n_{\geq 0}$, then $\alpha + \gamma > \beta + \gamma$
(c) $<$ is a well ordering on $\mathbb{Z}^n_{\geq 0}$.
This condition is satisfied if and only if every strictly decreasing sequence in $\mathbb{Z}^n_{\geq 0}$ eventually terminates.

Definition D.2. (Lexicographic ordering):

This is akin to the ordering of words used in dictionaries. If we define a polynomial in three variables as $P = k[x, y, z]$ and specify an ordering $x > y > z$, i.e., x comes before y and y comes before z, then any term with x will supersede that of y which in tern supersedes that of z. If the powers of the variables for respective monomials are given as $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n), \alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, then $\alpha >_{lex} \beta$ if in the vector difference $\alpha - \beta \in \mathbb{Z}^n$, the most left non-zero entry is positive. For the same variable (e.g., x) this subsequently means $x^\alpha >_{lex} x^\beta$.

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2,
© Springer-Verlag Berlin Heidelberg 2012

Example D.1.

$x > y^5z^9$ is an example of lexicographic ordering. As a second example, consider the polynomial $f = 2x^2y^8 - 3x^5yz^4 + xyz^3 - xy^4$, we have the *lexicographic order*; $f = -3x^5yz^4 + 2x^2y^8 - xy^4 + xyz^3 \,|x > y > z$.

Definition D.3. (Graded lexicographic ordering):

In this case, the total degree of the monomials is taken into account. First, one considers which monomial has the highest total degree before looking at the lexicographic ordering. This ordering looks at the left most (or largest) variable of a monomial and favours the largest power. Let $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, then $\alpha >_{grlex} \beta$ if $|\alpha| = \sum_{i=1}^{n} \alpha_i > |\beta| = \sum_{i=1}^{n} \beta_i$ or $|\alpha| = |\beta|$, and $\alpha >_{lex} \beta$, in $\alpha - \beta \in \mathbb{Z}^n$, the most left non zero entry is positive.

Example D.2.

$x^8y^3z^2 >_{grlex} x^6y^2z^3 \,|(8,3,2) >_{grlex} (6,2,3)$, since $|(8,3,2)| = 13 > |(6,2,3)| = 11$ and $\alpha - \beta = (2,1,-1)$. Since the left most term of the difference (2) is positive, the ordering is graded lexicographic. As a second example, consider the polynomial $f = 2x^2y^8 - 3x^5yz^4 + xyz^3 - xy^4$, we have the graded lexicographic order; $f = -3x^5yz^4 + 2x^2y^8 - xy^4 + xyz^3 \,|x > y > z$.

Definition D.4. (Graded reverse lexicographic ordering):

In this case, the total degree of the monomials is taken into account as in the case of graded lexicographic ordering. First, one considers which monomial has the highest total degree before looking at the lexicographic ordering. In contrast to the graded lexicographic ordering, one looks at the right most (or largest) variable of a monomial and favours the smallest power. Let $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, then $\alpha >_{grevlex} \beta$ if $|\alpha| = \sum_{i=1}^{n} \alpha_i > |\beta| = \sum_{i=1}^{n} \beta_i$ or $|\alpha| = |\beta|$, and $\alpha >_{grevlex} \beta$, and in $\alpha - \beta \in \mathbb{Z}^n$ the right most non zero entry is negative.

Example D.3.

$x^8y^3z^2 >_{grevlex} x^6y^2z^3 \,|(8,3,2) >_{grevlex} (6,2,3)$ since $|(8,3,2)| = 13 > |(6,2,3)| = 11$ and $\alpha - \beta = (2,1,-1)$. Since the right most term of the difference (-1) is negative, the ordering is *graded reverse lexicographic*. As a second example, consider the polynomial $f = 2x^2y^8 - 3x^5yz^4 + xyz^3 - xy^4$, we have the *graded reverse lexicographic order*: $f = 2x^2y^8 - 3x^5yz^4 - xy^4 + xyz^3\,|x > y > z$.

If we consider a non-zero polynomial $f = \sum_\alpha a_\alpha x^\alpha$ in $k\,[x_1, \ldots, x_n]$ and fix the monomial order, the following additional terms can be defined:

Definition D.5.

Multidegree of f : Multideg $(f) = \max(\alpha \in \mathbb{Z}^n_{\geq 0}\,|a_\alpha \neq 0)$ Leading Coefficient of f : LC $(f) = a_{\text{multideg}(f)} \in k$ Leading Monomial of f : LM $(f) = x^{\text{multideg}(f)}$ (with coefficient 1) Leading Term of f : LT $(f) = LC(f)$ LM (f).

Example D.4.

Consider the polynomial $f = 2x^2y^8 - 3x^5yz^4 + xyz^3 - xy^4$ with respect to lexicographic order $\{x > y > z\}$, we have
Multideg (f)=(5,1,4)
LC (f)= -3
LM (f)= x^5yz^4
LT (f)= $-3x^5yz^4$

The definitions of polynomial ordering above have been adopted from Cox et al. (1997), pp. 52–58.

D-2 Buchberger Algorithm

computes the Groebner basis using the computer algebraic systems ((CAS) of Mathematica and Maple. With Groebner basis, most nonlinear equations that are encountered in geodesy and geoinformatics can be solved. All that is required of the user is to write algorithms that can easily run in Mathematica or Maple using steps that are discussed in Sects. D-21 and D-22 below.

D-21 Mathematica Computation of Groebner Basis

Groebner basis can be computed using algebraic softwares of Mathematica (e.g., version 7 onwards). In Mathematica, Groebner basis command is executed by writing

$$In[1] := GroebnerBasis[\{polynomials\}, \{variables\}], \tag{D.1}$$

where In[1]:= is the Mathematica prompt which computes the Groebner basis for the Ideal generated by the polynomials with respect to the *monomial order* specified by *monomial order options*.

Example D.5. (Mathematica computation of Groebner basis):

In Example 15.3 on p. 536, the systems of polynomial equations were given as

$$\begin{bmatrix} f_1 = xy - 2y \\ f_2 = 2y^2 - x^2. \end{bmatrix}$$

Groebner basis of this system would be computed by

$$In[1] := GroebnerBasis[\{f_1, f_2\}, \{x, y\}], \tag{D.2}$$

leading to the same values as in (15.10) on p. 536. With this approach, however, one obtains too many elements of Groebner basis which may not be relevant to the task at hand. In a case where the solution of a specific variable is desired, one can avoid computing the undesired variables, and alleviate the need for back-substitution by simply computing the *reduced Groebner basis*. In this case (D.1) modifies to

$$In[1] := GroebnerBasis[\{polynomials\}, \{variables\}, \{elims\}], \tag{D.3}$$

where *elims* is for elimination order. Whereas the term reduced Groebner basis is widely used in many Groebner basis literature, we point out that within Mathematica software, the concept of a reduced basis has a different technical meaning. In this book, as in its predecessor, and to keep with the tradition, we maintain the use of the term reduced Groebner basis.

Example D.6. (Mathematica computation of reduced Groebner basis):

In Example D.2.1, one would compute the reduced Groebner basis using (D.3) as

$$In[1] := GroebnerBasis[\{f_1, f_2\}, \{x, y\}, \{y\}], \tag{D.4}$$

which will return only $-2x^2 + x^3$. Note that this form is correct only when the *retained* and *eliminated* variables are disjoint. If they are not, there is absolutely no guarantee as to which category a variable will be put in! As an example, consider the solution of three polynomials $x^2 + y^2 + z^2 - 1$, $xy - z + 2$, and $z^2 - 2x + 3y$. In the approach presented in (D.4), the solution would be

$$In[1] := GroebnerBasis[\{x^2+y^2+z^2-1, xy-z+2, z^2-2x+3y\}, \{x, y, z\}, \{x, y\}], \tag{D.5}$$

leading to

$$\{1024 - 832z - 215z^2 + 156z^3 - 25z^4 + 24z^5 + 13z^6 + z^8\}.$$

Lichtblau (Priv. Comm.) however suggests that the retained variable, in this case (z) and the eliminated variables (x, y) be separated, i.e.,

$$In[1] := GroebnerBasis[\{x^2 + y^2 + z^2 - 1, xy - z + 2, z^2 - 2x + 3y\}, \{z\}, \{x, y\}]. \tag{D.6}$$

The results of (D.6) are the same as those of (D.5), but with the advantage of a vastly better speed if one uses an ordering better suited for the task at hand, e.g., elimination of variables in this example. For the problems that are solved in our books, the condition of the retained and eliminated variables being disjoint is true, and hence the approach in (D.5) has been adopted. The univariate polynomial $-2x^2 + x^3$ is then solved for x using the roots command in Matlab (see e.g., Hanselman and Littlefield (1997), p. 146) by

$$roots(\,[1\ -2\ 0\ 0]\,). \tag{D.7}$$

The values of the row vector in (D.7) are the coefficients of the cubic polynomial $x^3 - 2x^2 + 0x + 0 = 0$ obtained from (D.4). The values of y from Example (15.3) on p. 536 can equally be computed from (D.4) by replacing y in the option part with x and thus removing the need for back substitution. We leave it for the reader to compute the values of y from Example (15.3) and also those of z in Example (15.3) using reduced Groebner basis (D.3) as an exercise. The reader should confirm that the solution of y leads to $y^3 - 2y$ with the roots $y = 0$ or $y = \pm 1.4142$. From experience, we recommend the use of reduced Groebner basis for applications in geodesy and geoinformatics. This will; fasten the computations, save on computer space, and alleviates the need for back-substitution.

D-22 Maple Computation of Groebner Basis

In Maple Version 10 the command is accessed by typing > *with (Grobner);* Basis (WL, T) is for Groebner, *gbasis* (WL, T) is for reduced Groebner basis. WL list or set of polynomials, T monomial order description, e.g. "plex" (variables) – pure lexicographic order and > is the Maple prompt and the semicolon ends the Maple command). Once the Groebner basis package has been loaded, the execution command then becomes > *gbasis (polynomials, variables)* which computes the Groebner basis for the *ideal* generated by the *polynomials* with respect to the *monomial ordering* specified by *variables* in the executable command. We should point out that at the time of writing this book, May 2012, version 16 of Maple has been released.

D.3 Gauss Combinatorial Formulation

CARL FRIEDRICH GAUSS

WERKE

NEUNTER BAND.

GG-4

Bücherei des Geodätischen Instituts
der Techn. Hochschule Stuttgart
Nr. 3774

HERAUSGEGEBEN
VON DER
KÖNIGLICHEN GESELLSCHAFT DER WISSENSCHAFTEN
ZU
GÖTTINGEN.
IN COMMISSION BEI B. G. TEUBNER IN LEIPZIG.
1903.

BESTIMMUNG
DES
BREITENUNTERSCHIEDES
ZWISCHEN DEN
STERNWARTEN VON GÖTTINGEN UND ALTONA
DURCH
BEOBACHTUNGEN AM RAMSDENSCHEN ZENITHSECTOR
VON
CARL FRIEDRICH GAUSS,
RITTER DES GUELPHEN- UND DANNEBROG-ORDENS; K. GROSSBR. HANNOVERSCHEM HOFRATH;
PROFESSOR DER ASTRONOMIE UND DIRECTOR DER STERNWARTE IN GÖTTINGEN;
MITGLIED DER AKADEMIEN UND SOCIETÄTEN VON BERLIN, COPENHAGEN, EDINBURG, GÖTTINGEN,
LONDON, MÜNCHEN, NEAPEL, PARIS, PETERSBURG, STOCKHOLM,
DER AMERIKANISCHEN, ITALIENISCHEN, KURLÄNDISCHEN, LONDONER ASTRONOMISCHEN U. A.

GÖTTINGEN,
BEI VANDENHOECK UND RUPRECHT.
1828.

IX. 1

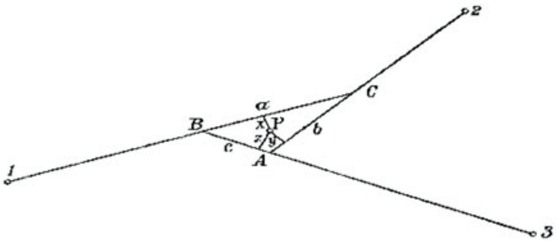
NACHLASS.

[1.]

Endresultat für den Ort eines Punktes in einer Ebene, der von drei bekannten aus angeschnitten ist.

Es bedeuten 10, 20, 30 die drei beobachteten Richtungen [nach P] und α, β, γ die entsprechenden Entfernungen.

Die drei einzelnen Resultate aus den Combinationen 2—3, 1—3, 1—2 seien A, B, C, zugleich die Winkel des durch jene gebildeten Dreiecks; die ihnen gegenüber stehenden Seiten a, b, c.



Perpendikel von dem gesuchten Orte auf a, b, c seien x, y, z. S doppelter Flächeninhalt des Dreiecks.

Es sind dann

$$\frac{x}{\alpha}, \quad \frac{y}{\beta}, \quad \frac{z}{\gamma}$$

die übrig bleibenden Fehler, also

$$\frac{xx}{\alpha\alpha} + \frac{yy}{\beta\beta} + \frac{zz}{\gamma\gamma} \quad \text{Minimum}$$

und

$$ax + by + cz = S.$$

Also werden x, y, z proportional den Grössen $\alpha\alpha a$, $\beta\beta b$, $\gamma\gamma c$;

$$x = \frac{\alpha\alpha a S}{\alpha\alpha aa + \beta\beta bb + \gamma\gamma cc},$$

etc.

[Bezeichnet (ABC) die Fläche des Dreiecks ABC, u. s. f., so ist

$$S = 2(ABC) = (\alpha\alpha aa + \beta\beta bb + \gamma\gamma cc)k$$
$$2(BPC) = \alpha\alpha aak$$
$$2(APC) = \beta\beta bbk$$
$$2(APB) = \gamma\gamma cck,$$

wo k die Correlate der Bedingungsgleichung ist. P ist der durch die Perpendikel x, y, z bestimmte Punkt.

Folglich wird, wenn A, B, C, P die complexen Grössen bedeuten, denen die Eckpunkte des Dreiecks ABC und der Punkt P entsprechen:

$$(\alpha\alpha aa + \beta\beta bb + \gamma\gamma cc)P = \alpha\alpha aaA + \beta\beta bbB + \gamma\gamma ccC.]$$

Es folgt hieraus, dass das Endresultat*)

$$\frac{\alpha\alpha aaA + \beta\beta bbB + \gamma\gamma ccC}{\alpha\alpha aa + \beta\beta bb + \gamma\gamma cc}$$

also ein Mittel aus den drei partiellen Resultaten A, B, C ist, indem man diesen die Gewichte

$$\alpha\alpha aa, \quad \beta\beta bb, \quad \gamma\gamma cc$$

beilegt, oder

$$\alpha\alpha \sin A^2, \quad \beta\beta \sin B^2, \quad \gamma\gamma \sin C^2.$$

Offenbar ist hier A zugleich der Winkel zwischen 20 und 30, u. s. f.

References

Aarts, E.H. and Korst, J. (1989): Simulated Annealing and Boltzman Machines, Wiley, New York

Abbe, E. (1906): Über die Gesetzmäßigkeit in der Verteilung der Fehler bei Beobachtungsreihen, Gesammelte Abhandlungen, Vol. II, Jena 1863 (1906)

Abdous, B. and R. Theodorescu (1998): Mean, median, mode IV, Statistica Neerlandica 52 (1998), 356-359

Abel, J.S. and Chaffee, J.W. (1991): Existence and uniqueness of GPS solutions, *IEEE Transactions on Aerospace and ElectronicSystems* 27 (1991) 952-956.

I. Abhandlung Monatsberichte der Akademie der Wissenschafte 1853, Werke 4 (1929) 1-11

Absil, P.A., Mahony, R., Sepulehre, R. and Van Dooren, P. (2002): A Grassmann-Rayleigh quotient iteration for computing invariant subspaces, SIAM Review 44 (2002), 57-73

Abramowitz, M. and J.A. Stegun (1965): Handbook of mathematical functions, Dover Publication, New York 1965

Adams, M. and V. Guillemin (1996): Measure theory and probability, 2nd edition, Birkhäuser-Verlag, Basel Boston Berlin 1996

Adelmalek, N.N. (1974): On the discrete linear L1 approximation and L1 solutions of overdetermined linear equations, J. Approximation Theory 11 (1974), 38-53

Adler, R. and Pelzer, H. (1994): Development of a Long Term Geodetic Monitoring System of Recent Crustal Activity Along the Dead Sea Rift. Final scientific report for the Gennan-Israel Foundation GIF

Aduol, F.W.O. (1989): Intergrierte geodätische Netzanalyse mit stochastischer Vorinformation über Schwerefeld und Referenzellipsoid, DGK, Reihe C, Heft Nr. 351

Afifi, A.A. and V. Clark (1996): Computer-aided multivariate analysis, Chapman and Hall, Boca Raton 1996

Agnew, D.C. (1992): The time-domain behaviour of power-law noises. Geophysical research letters, 19(4):333-336, 1992

Agostinelli, C. and M. Markatou (1998): A one-step robust estimator for regression based on the weighted likelihood reweighting scheme, Stat. & Prob. Letters 37 (1998), 341-350

Agréo, G. (1995): Maximum likelihood estimation for the exponential power function parameters, Comm. Statist. Simul. Comput. 24 (1995), 523-536

Aickin, M. and C. Ritenbaugh (1996): Analysis of multivariate reliability structures and the induced bias in linear model estimation, Statistics in Medicine 15 (1996), 1647-1661

Aitchinson, J. and I.R. Dunsmore (1975): Statistical prediction analysis, Cambridge University Press, Cambridge 1975

Aitken, A.C. (1935): On least squares and linear combinations of observations, Proc. Roy. Soc. Edinburgh 55 (1935), 42-48

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*, Springer Geophysics, DOI 10.1007/978-3-642-22241-2,

© Springer-Verlag Berlin Heidelberg 2012

Airy, G.B. (1861): On the algebraical and numerical theory of errors of observations and the combination of observations, Macmillan Publ., London 1861

Aki, K. and Richards P.G. (1980): Quantitative seismology, W.H. Freeman, San Francisco

Albert, A. (1969): Conditions for positive and nonnegative definiteness in terms of pseudo inverses, SIAM J. Appl. Math. 17 (1969), 434-440

Alberda, J.E. (1974): Planning and optimization of networks: some general considerations, Boll. di Geodesicl, e Scienze, Aff. 33 (1974) 209-240

Albertella, A. and F. Sacerdote (1995): Spectral analysis of block averaged data in geopotential global model determination, Journal of Geodesy 70 (1995), 166-175

Alcala, J.T., Cristobal, J.A. and W. Gonzalez-Manteiga (1999): Goodness-of-fit test for linear models based on local polynomials, Statistics & Probability Letters 42 (1999), 39-46

Aldrich, J. (1999): Determinacy in the linear model: Gauss to Bose and Koopmans, International Statistical Review 67 (1999), 211-219

Alefeld, G. and J. Herzberger (1974): Einfhrung in die Intervallrechung, Bibliographisches Inst. Mannheim 1974

Alefeld, G. and J. Herzberger (1983): Introduction to interval computation. Computer science and applied mathematics, Academic Press, New York London 1983

Ali, A.K.A., Lin, C.Y. and E.B. Burnside (1997): Detection of outliers in mixed model analysis, The Egyptian Statistical Journal 41 (1997), 182-194

Allende, S., Bouza, C. and I. Romero (1995): Fitting a linear regression model by combining least squares and least absolute value estimation, Questiio 19 (1995), 107-121

Allmer, F. Louis Krger (1857-1923), 25 pages, Technical University of Graz, Graz 2001 Alzaid, A.A. and L. Benkherouf (1995): First-order integer-valued autoregressive process with Euler marginal distributions, J. Statist. Res. 29 (1995), 89-92

Amiri-Simkooei, A.R. (2004): A new method for second order design of geodetic networks: aiming at high reliability, survey Review, 37 (2004), pp. 552-560

Amiri-Simkooei, A.R. (2007): Least-squares estimation of variance components, Theory and EPS applications, Ph.D. Thesis, Delft Univ. of Tech., delft, The Netherlands 2007

An, H.-Z., F.J. Hickernell and L.-X. Zhu (1997): A new class of consistent estimators for stochastic linear regressive models, J. Multivar. Anal. 63 (1997), 242-258

Andel, J. (1984): Statistische Analyse von Zeitreihen, Akademie Verlag, Stuttgart 1984

Anderson, P.L. and M.M. Meerscaert (1997): Periodic moving averages of random variables with regularly varying tails, Ann. Statist. 25 (1997), 771-785

Anderson, R.D., Henderson H.V., Pukelsheim F. and Searle S.R. (1984): Best estimation of variance components from balanced data, with arbitrary kurtosis, Mathematische Operationsforschung und Statistik, Series Statistics 15, pp. 163-176

Anderson, T.W. (1945): Non-central Wishart Distribution and Its Application to Statistics, doctoral dissertation, Princeton University

Anderson, T.W. (1946): The non-central Wishart distribution and certain problems of multivariate statistics, Ann. Math. Stat., 17, 409-431

Anderson, T.W. (1948): The asymptotic distributions of the roots of certain determinantal equations, J. Roy. Stat. Soc., B, 10, 132-139

Anderson, T.W. (1951): Classification by multivariate analysis, psychometrika, 16, 31-50

Anderson, T.W. (1973): Asymptotically efficient estimation of covariance matrices with linear structure, The Annals of Statistics 1 (1973), 135-141

Anderson, T.W. (1984): An introduction to multivariate statistical analysis, 2nd ed., J. Wiley, New York 1982

Anderson, T.W. (2003): An introduction to multivariate statistical analysis, J. Wiley, Stanford CA, 2003

Anderson, T.W. and M.A. Stephens (1972): Tests for randomness of directions against equatorial and bimodal alternatives, Biometrika 59 (1972), 613-621

Anderson, W.N. and R.J. Duffin (1969): Series and parallel addition of matrices, J. Math. Anal. Appl. 26 (1969), 576-594

Ando, T. (1979): Generalized Schur complements, Linear Algebra Appl. 27 (1979), 173-186

Andrews, D.F. (1971): Transformations of multivariate data, Biometrics 27 (1971), 825-840
Andrews, D.F. (1974): Technometrics 16 (1974), 523-531
Andrews, D.F., Bickel, P.J. and F.R. Hampel (1972): Robust estimates of location, Princeton University Press, Princeton 1972
Angelier, J., Tarantola, A., Vallete, B. and S. Manoussis (1982): Inversion of field data in fault tectonics to obtain the regional stress - I, Single phase fault populations: a new method of computing the stress tensor, Geophys. J. R. astr. Soc., 69, 607-621
Angulo, J.M. and Bueso, M.C. (2001): Random perturbation methods applied to multivariate Spatial Sampling design, Environmetrics, 12, 631-646.
Anh, V.V. and T. Chelliah (1999): Estimated generalized least squares for random coefficient regression models, Scandinavian J. Statist. 26 (1999), 31-46
Anido, C. and T. Valdèes (2000): Censored regression models with double exponential error distributions: an iterative estimation procedure based on medians for correcting bias, Revista Matemèatica Complutense 13 (2000), 137-159
Anscoube, F.J. (1960): Rejection of outliers, Technometrics, 2 (1960), pp. 123-147
Ansley, C.F. (1985): Quick proofs of some regression theorems via the QR Algorithm, The American Statistician 39 (1985), 55-59
Anton, H. (1994): Elementary linear algebra, J. Wiley, New York 1994
Arcadis (1997): Ingenieurgeologischer Situationsbericht für 1997, Bericht des Pumpspeicherwerks Hohenwarthe, VEAG, 1997
Archambeau, C.B. and Bradford, J.C. and Broome, P.W. and Dean, W.C. and Flinn, E.A. (1965): Data processing techniques for the detection and interpretation of teleseismic signals, Proc. IEEE 53 (1965), 1860-1865
Arellano-Valle, R. and Bolfarine, H. and Lachos, V (2005): Skew-normal linear mixed models. In Journal of Data Science vol 3, pp.415-438
Arnold, B.C. and R.M. Shavelle (1998): Joint confidence sets for the mean and variance of a normal distribution, J. Am. Statist. Ass. 52 (1998), 133-140
Arnold, B.F. and P. Stahlecker (1998): Prediction in linear regression combining crisp data and fuzzy prior information, Statistics & Decisions 16 (1998), 19-33
Arnold, B.F. and P. Stahlecker (1999): A note on the robustness of the generalized least squares estimator in linear regression, Allg. Statistisches Archiv 83 (1999), 224-229
Arnold, K.J. (1941): On spherical probability distributions, P.H. Thesis, M.I.T., Boston 1941
Arnold, S.F. (1981): The theory of linear models and multivariate analysis, J. Wiley, New York 1981
Arrowsmith, D.K. and C.M. Place (1995): Differential equations, maps and chaotic behavior, Champman and Hall, London 1995
Arun, K.S. (1992): A unitarily constrained total least squares problem in signal processing, SIAM J. Matrix Anal. Appl. 13 (1992), 729-745
Ashkenazi V. and Grist, S. (1982): Inter-comparison of 3-D geodetic network adjustment models, International sypmosium on geodetic networks and computations of the I.A.G. Munich, August 31 to September 5, 1981.
Atiyah, M.F., Hitchin, N.J. and Singer, J.M.: Self-duality in four-dimensional Riemannian geometry. Proc. Royal Soc. London A362 (1978) 425-461
Austen, G. and Grafarend, E.W. and Reubelt, T. (2001): Anal: Earth's gravitational field from semi continuous ephe low earth orbiting GPS-tracked satellite of type CRA or Goce, in: Vistas for Geodesy in the New Millenn Adam and Schwarz, International Association of Geo posia 125, 309-315, 2001
Aven, T. (1993): Reliability and risk analysis, Chapman and Hall, Boca Raton 1993
Awange, L.J. (1999): Partial procrustes solution of the three dimensional orientation problem from GPS/LPS observations, Quo vadis geodesia...? Festschrift to E.W. Grafarend on the occasion of his 60th birthday, Eds. F. Krumm and V.S. Schwarze, Report Nr. 1999.6-1
Awange, J.L. (2002): Gröbner bases, multi polynomial resultants and the Gauss-Jacobi combinatorial algorithms - adjustment of nonlinear GPS/LPS observations, Schriftenreihe der Institute des Studiengangs Geodäsie und Geoinformatik, Report 2002.1

Awange, J. and E.W. Grafarend (2002): Linearized Least Squares and nonlinear Gauss-Jacobi combinatorial algorithm applied to the 7-parameter datum transformation C7(3) problem, Zeitschrift für Vermessungswesen 127 (2002), 109-116

Awange, J.L. (2012): Environmental Monitoring using GNSS. Global Navigation Satellite Systems, Springer, Berlin-New York.

Azzalini, A. (1996): Statistical inference, Chapman and Hall, Boca Raton 1996

Azzam, A.M.H. (1996): Inference in linear models with non stochastic biased factors, Egyptian Statistical Journal, ISSR - Cairo University 40 (1996), 172-181

Azzam, A., Birkes, D. and J. Seely (1988): Admissibility in linear models with polyhedral covariance structure, Probability and Statistics, essays in honor of Franklin A. Graybill, J.N. Srivastava, Ed. Elsevier Science Publishers, B.V. (North-Holland), 1988

Baarda, W. (1967): Statistical Concepts in Geodesy. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 2, No.4, Delft.

Baarda, W. (1967): A generalization of the concept strength of the figure, Publications on Geodesy, New Series 2, Delft 1967

Baarda, W. (1968): A testing procedure for use in geodetic networks, Netherlands Geodetic Commission, New Series, Delft, Netherlands, 2 (5), 1968

Baarda, W. (1973): S-transformations and criterion matrices, Netherlands Geodetic Commission, Vol. 5, No. 1, Delft 1973

Baarda, W. (1977): Optimization of design and computations of control networks, F. Halmos and J. Somogyi (eds.), Akademiai Kiado, Budapest 1977, 419-436

Babai, L. (1986): On Lovasz' lattice reduction and the nearest lattice point problem, Combinatorica 6 (1986), 1-13

Babu, G.J. and E.D. Feigelson (1996): Astrostatistics, Chapman and Hall, Boca Raton 1996

Bai, J. (1994): Least squares estimation of a shift in linear processes, Journal of Time Series Analysis 15 (1994), 453-472

Bognarova, M., Kubacek, L. and Volaufova J. (1996): Comparison of MINQUE and LMVQUIE by simulation. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 35 (1996), 25-38.

Bachelier, L. (1900): Theorie de le speculation, Ann, Sci. EC. Norm. Super. 17 (1900) 21-26

Bähr, H.G. (1988): A quadratic approach to the non-linear treatment of non-redundant observations *Manuscripta Geodaetica*, 13 (1988) 191-197.

Bähr, H.G. (1991): Einfach überbestimmtes ebenes Einschneiden, differentialgeometrisch analysiert, *Zeitschrift für Vermessungswesen* 116 (1991) 545-552.

Bai, W.M., Vigny, C., Ricard, Y. and Froidevaus, C. (1992): On the origin of deviatoric stresses in the lithosphere, J. geophys. Res., B97, 11728-11737

Bai, Z.D., Chan, X.R., Krishnaiah, P.R. and L.C. Zhao (1987): Asymptotic properties of the EVLP estimation for superimposed exponential signals in noise, Technical Report 87-19, CMA, University of Pittsburgh, Pittsburgh 1987

Bai, Z.D. and Y. Wu (1997): General M-estimation, J. Multivar. Anal. 63 (1997), 119-135

Bajaj, C.L. (1989): Geometric Modelling with Algebraic Surfaces. In: The Mathematics of Surfaces III, Ed.: D.C. Handscomb, Clarendon Press, Oxford.

Bajaj, C., Garity, T. and Waren, J. (1988): On the applications of multi-equational resultants, Department of Computer Science, Purdue University, Technical Report CSD-TR-826, pp. 1-22, 1988.

Baksalary, J.K. and Kala, R. (1981): Linear transformations preserving best linear unbiased estimators in a general Gauss-Markoff model. Ann. Stat. 9 (1981), 913-916.

Baksalary, J.K. and A. Markiewicz (1988): Admissible linear estimators in the general Gauss-Markov model, J. Statist. Planning and Inference 19 (1988), 349-359

Baksalary, J.K. and Puntanen, S. (1989): Weighted-Least-squares estimation in the general Gauss-Markov model. In: Statistical Data Analysis and Inference, Pap. Int. Conf., Neuchatel/Switzerl. (Y. Dodge, ed.), North-Holland, Amsterdam, pp. 355-368.

Baksalary, J.K. and Liski, E.P. and G. Trenkler (1989): Mean square error matrix improvements and admissibility of linear estimators, J. Statist. Planning and Inference 23 (1989), 313-325

Baksalary, J.K. and F. Pukelsheim (1989): Some properties of matrix partial ordering, Linear Algebra and it Appl., 119 (1989), pp. 57-85

Baksalary, J.K. and F. Pukelsheim (1991): On the Löwner, Minus and Star partial orderings of nonnegative definite matrices and their squares, Linear Algebra and its Applications 151 (1991), 135-141

Baksalary, J.K. and Rao, C.R., Markiewicz, A. (1992): A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss-Markov model. J. Stat. Plann. Inference 31, 3 (1992), 335-351.

Balakrishnan, N. and Basu, A.P. (eds.) (1995): The exponential distribution, Gordon and Breach Publishers, Amsterdam 1995

Balakrishnan, N. and R.A. Sandhu (1996): Best linear unbiased and maximum likelihood estimation for exponential distributions under general progressive type-II censored samples, Sankhya 58 (1996), 1-9

Bamler, R. and P. Hartl (1998): Synthetic aperture radar interferometry, Inverse Problems 14 (1998), R1-R54

Banachiewicz, T. (1937): Zur Berechnung der Determinanten, wie auch der Inversen und zur darauf basierten Au ösung der Systeme linearen Gleichungen, Acta Astronom. Ser. C3 (1937), 41-67

Bancroft, S. (1985): An algebraic solution of the GPS equations, *IEEE Transaction on Aerospace and Electronic Systems* AES-21 (1985) 56-59.

Bandemer, H. (2005): Mathematik und Ungewissheit: Drei Essais zu Problemen der Anwendung, Edition avn Gutenberg platz, Leipzig 2005

Bandemer, H. et al. (1976): Optimale Versuchsplanning, Z. Auflage Akademic Verlag, Berlin 1976

Bandemer, H. et al. (1977): Theorie und Anwendung der optimalen Versuchsplanung I. Akademie-Verlag, Berlin

Bandemer, H. and W. Nather (1980): Theorie and Anwendung der optimalen Versuchsplanung II, akademic Verlag, Berlin 1980

Bandemer, H. and S. Gottwald (1993): Einfhrung in die Fuzzy-Methoden, Akademic-Verlag, Berlin 1993

Bahndorf, J. (1991): Zur Systematisierung der Seilnetzberechnung und zur Optimierung von Seilnetzen. Deutsche Geodätische Kommission, Reihe C, Nr. 373, Miinchen.

Bansal, N.K., Hamedani, G.G. and H. Zhang (1999): Non-linear regression with multidimensional indices, Statistics & Probability Letters 45 (1999), 175-186

Barankin, E.W. (1949): Locally best unbiased estimates, Ann. Math. Statist. 20 (1949), 477-501

Barber, B.C. (1993): The non-isotropic two-dimensional random walk, Waves in Random Media, 3 (1993) 243-256

Barham, R.H. and W. Drane (1972): An algorithm for least squares estimation of nonlinear parameters when some of the parameters are linear, Technometrics 14 (1972), 757-766

Barlett, M. (1933): On the theory of statistical regression, Proc. Roy. Soc .. Edinb., 53, 260-283.

Barlett, M. (1937): Property od sufficiency and statistical tests, Proc. Roy. Soc., A 160, 268-282

Barlett, M. (1947): Multivariate analysis, J. Roy. Stat. Soc., Supple., 9, 176-197.

Barlow, R.E., Clarotti, C.A. and F. Spizzichino (eds.) (1993): Reliability and decision making, Chapman and Hall, Boca Raton 1993

Barlow, R.E. and F. Proschan (1966): Tolerance and confidence limits for classes of distributions based on failure rate, Ann. Math. Statist 37 (1966), 1593-1601

Barnard, J., McCulloch, R. and X.-L. Meng (2000): Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, Statistica Sinica 10 (2000), 1281-1311

Barnard, M.M. (1935): The scalar variations of skull parameters in four series of Egyptian skulls, Ann. Eugen. 6 (1935), 352-371

Barndorff-Nielsen, O. (1978): Information and exponential families in statistical theory, J. Wiley, New York, 238 pp.

Barndorff-Nielsen, O.E., Cox, D.R. and C. Klüppelberg (2001): Complex stochastic systems, Chapman and Hall, Boca Raton, Florida 2001

Barnett, V. and Lewis, T. (1994): Outliers in Statistical Data. John Wiley, New York, 1994.

Barnett, V. (1999): Comparative statistical inference, 3rd ed., J. Wiley, New York 1999

Barone, J. and A. Novikoff (1978): A history of the axiomatic formulation of probability from Borel to Kolmogorov, Part I, Archive for History of Exact Sciences 18 (1978), 123-190

Barrlund, A. (1998): Efficient solution of constrained least squares problems with Kronecker product structure, SIAM J. Matrix Anal. Appl. 19 (1998), 154-160

Barrodale, I. and D.D. Oleski (1981): Exponential approximation using Prony's method, The Numerical Solution of Nonlinear Problems, eds. Baker, C.T.H. and C. Phillips, 258-269

Bartelme, N. and Meissl, P. (1975): Ein einfaches, rasches und numerisch stabiles Verfahren zur Bestimmung des kürzesten Abstandes eines Punktes von einem sphäroidischen Rotationsellipsoid, *Allgemeine Vermessungs-Nachrichten* 82 (1975) 436-439.

Bartlett, M.S. (1980): An introduction to stochastic processes with special reference to methods and applications, paperback edition, University press, Cambridge 1980

Bartlett, M.S. and D.G. Kendall (1946): The statistical analysis of variance-heterogeneity and the logarithmic transformation, Queen's College Cambridge, Magdalen College Oxford, Cambridge/Oxford 1946

Bartoszynski, R. and M. Niewiadomska-Bugaj (1996): Probability and statistical inference, J. Wiley, New York 1996

Barut, A.O. and R.B. Haugen (1972): Theory of the conformally invariant mass, Annals of Physics 71 (1972), 519-541

Basset JR., G. and R. Koenker (1978): Asymptotic theory of least absolute error 73 (1978), 618-622

Basawa, I.V. (2001): Inference in stochastic processes in: D.N. Shanbhag and C.R. Rao. Handbook of statistics, pp. 55-77, Vol 19, Elsevier Science 2001

Basch, A. (1913): Uber eine Anwendung der graphostatischen Methode auf den Ausgleich von Beobachtungsergebnissen. OZfV 11, 11-18, 42-46.

Baselga, S. (2007): Global Optimization Solution of Robust Estimation. J Surv. Eng. 2007, 133, 123-128.

Basu, S. and Mukhopadhyay, S. (2000): Binary response regression with normal scale mixture links. Dey, D.K., Ghosh S.K.-and Mallick, B.K. eds., Generalized linear models: A Bayesian perspective, Marcel Dekker, New York

Batchelor, G.K. (1946): Theory of axisymmetric turbulenz, Proc. Royal Soc. A186 (1946) 480

Batchelor, G.K. and A.A. Townsend (1950): The nature of turbulent motion at large wave numbers, Fluid Mech. (1952) 238-255

Batchelor, G.K. (1953): The theory of homogeneous turbulence, Cambridge University Press, Cambridge, 1953.

Batchelor, G.K. (1967): The theory of homogenous turbulence, Cambridge (1967)

Baselga, S. (2007): Critical Limitation in Use of't Test for Gross Error Detection. J Surv. Eng. 2007, 133,52-55.

Bateman, H. (1910): The transformation of the electrodynamical equations, Proc. London Math. Soc. 8 (1910), 223-264, 469-488

Bateman, H. (1910): The transformation of coordinates which can be used to transform one physical problem into another, Proc. London Math. Soc. 8 (1910), 469-488

Bates, D.M. and D.G. Watts (1980): Relative curvature measures of nonlinearity (with discussion), J. Roy. Statist. Soc. B42 (1980), 1-25

Bates, D.M. and M.J. Lindstorm (1986): Nonlinear least squares with conditionally linear parameters, Proceedings of the Statistical Computation Section, American Statistical Association, Washington 1986

Bates, D.M. and D.G. Watts (1988a): Nonlinear regression analysis and its applications, J. Wiley, New York 1988

Bates, D.M. and D.G. Watts (1988b): Applied nonlinear regression, J. Wiley, New York 1988

Bates, R.A., Riccomagno, E., Schwabe, R. and H.P. Wynn (1998): Lattices and dual lattices in optimal experimental design for Fourier models, Computational Statistics & Data Analysis 28 (1998), 283-296

Batschelet, E. (1965): Statistical methods for the analysis of problems in animal orientation and certain biological rhythms, Amer. Inst. Biol. Sciences, Washington

Batschelet, E. (1971): Recent statistical methods for orientation, (Animal Orientation Symposium 1970 on Wallops Island), Amer. Inst. Biol. Sciences, Washington, D.C., 1971

Batschelet, E. (1981): Circular statistics in biology, Academic Press, New York London 1981

Bauer, H. (1992): Maß- und Integrationstheorie, 2. Au age, Walter de Gruyter, Berlin New York 1992

Bauer, H. (1996): Probability theory, de Gruyter Verlag, Berlin New York 1996

Bayen, F. (1976): Conformal invariance in physics, in: Cahen, C. and M. Flato (eds.), Differential geometry and relativity, Reidel Publ., 171-195, Dordrecht 1976

Beale, E.M. (1960): Confidence regions in non-linear estimation, J. Roy. Statist. Soc. B22 (1960), 41-89

Beaton, A.E. and J.W. Tukey (1974): Technometrics 16 (1974), 147-185

Becker, T. and Weispfenning, V. (1993): Gröbner bases. A computational approach to commutative algebra. Graduate Text in Mathematics 141, Springer-Verlag, New York 1993.

Becker, T., Weispfennig, V. and H. Kredel (1998): Gröbner bases: a computational approach to commutative algebra, Corr. 2. print., Springer-Verlag, Heidelberg Berlin New York 1998

Beckermann, B. and E.B. Saff (1999): The sensitivity of least squares polynomial approximation, Int. Series of Numerical Mathematics, Vol. 131: Applications and computation of orthogonal polynomials (eds. W. Gautschi, G.H. Golub, G. Opfer), 1-19, Birkhäuser-Verlag, Basel Boston Berlin 1999

Beckers, J., Harnad, J., Perroud, M. and P. Winternitz (1978): Tensor fields invariant under subgroups of the conformal group of space-time, J. Math. Phys. 19 (1978), 2126-2153

Beder, C. and Forstner, W, (2006): Direct Solutions for Computing Cylinders from Minima/ Sets of 3D Points. European Conference on Computer Vision ECCV 2006. Graz, Austria 2006, S. 135-146.

Behnken, D.W. and N.R. Draper (1972): Residuals and their variance, Technometrics 11 (1972), 101-111

Behrens, W.A. (1929): Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen, Landwirtschaftliche Jahrbücher 68 (1929), 807-837

Belikov, M.V. (1991): Spherical harmonic analysis and synthesis with the use of column-wise recurrence relations, manuscripta geodaetica 16 (1991), 384-410

Belikov, M.V. and K.A. Taybatorov (1992): An efficient algorithm for computing the Earth's gravitational potential and its derivatives at satellite altitudes, manuscripta geodaetica 17 (1992), 104-116

Ben-Menahem, A. and Singh, S.J. (1981): Seismic waves and sources. Springer-Verlag New York 1981

Ben-Israel, A. and T. Greville (1974): Generalized inverses: Theory and applications, J. Wiley, New York 1974

Benbow, S.J. (1999): Solving generalized least-squares problems with LSQR, SIAM J. Matrix Anal. Appl. 21 (1999), 166-177

Benciolini. B. (1984): Continuous networks II. Erice Lecture Notes, this volume. Heidelberg.

Benda, N. and R. Schwabe (1998): Designing experiments for adaptively fitted models, in: MODA 5 - Advances in model-oriented data analysis and experimental design, Proceedings of the 5th International Workshop in Marseilles, eds. Atkinson, A.C., Pronzato, L. and H.P. Wynn, Physica-Verlag, Heidelberg 1998

Bennett, B. M. (1951): Note on a solution of the generalized Behrens-Fisher problem, Ann.Insr:Srat:XIath., 2, 87-90.

Bennett, R.J. (1979): Spatial time series, Pion Limited, London 1979

Benning, W. (1974): Der kürzeste Abstand eines in rechtwinkligen Koordinaten gegebenen Außenpunktes vom Ellipsoid, *Allgemeine Vermessungs-Nachrichten* 81 (1974) 429-433.

Benning, W. (1987): Iterative ellipsoidische Lotfußpunktberechung, *Allgemeine Vermessungs-Nachrichten* 94 (1987) 256-260.

Benning, W. and Ahrens, B. (1979): Konzept und Realisierung eines Systems zur automatischen Fehlerlokalisierung und automatischen Berechnung von Naherungskoordinaten. N achrichten aus dem 6ffentlichen Vermessungsdienst N ordrhein-Westfalen, Heft 2, 107-124.
Beran, R.J. (1968): Testing for uniformity on a compact homogeneous space, J. App. Prob. 5 (1968), 177-195
Beran, R.J. (1979): Exponential models for directional data, Ann. Statist. 7 (1979), 1162-1178
Beran, J. (1994): Statistical methods for long memory processes, Chapman and Hall, Boca Raton 1994
Berberan, A. (1992): Outlier detection and heterogeneous observations - a simulation case study, Australian Journal of Geodesy, Photogrammetry and Surveying 56 (1992), 49-61
Berberan, A. (1995): Multiple Outlier Detection-A Real Case Study. Surv. Rev. 1995, 33, 41-49.
Bergbauer, K. and Lechner, W. and Schuchert, M. and Walk, R. (1979): Anlage, Beobachtung und Ausgleichung eines dreidimensionalen Testnetzes. Diplomarbeit Hochschule der bundeswehr. Nunchen 1979.
Berger, M.P.F. and F.E.S. Tan (1998): Optimal designs for repeated measures experiments, Kwantitatieve Methoden 59 (1998), 45-67
Berman, A. and R.J. Plemmons (1979): Nonnegative matrices in the mathematical sciences, Academic Press, New York London 1979
Bertsekas, D.P. (1996): Incremental least squares methods and the extended Kalman filter, Siam J. Opt. 6 (1996), 807-822
Berry, J.C. (1994): Improving the James-Stein estimator using the Stein variance estimator, Statist. Probab. Lett. 20 (1994), 241-245
Bertuzzi, A., Gandolfi, A. and C. Sinisgalli (1998): Preference regions of ridge regression and OLS according to Pitman's criterion, Sankhya: The Indian Journal of Statistics 60 (1998), 437-447
Bessel, F.W. (1838): Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler, Astronomische Nachrichten 15 (1838), 368-404
Betensky, R.A. (1997): Local estimation of smooth curves for longitudinal data, Statistics in Medicine 16 (1997), 2429-2445
Betti, B. and Crespi, M. and Sanso, F. (1993): A geometric illustration of ambiguity resolution in GPS theory and a Bayesian approach. Manuscr Geod 18:317-330
Beylkin, G. and N. Saito (1993): Wavelets, their autocorrelation function and multidimensional representation of signals, in: Proceedings of SPIE - The international society of optical engineering, Vol. LB26, Int. Soc. for Optical Engineering, Bellingham 1993
Bezdek, J.C. (1973): Fuzzy Mathematics in Pattern Classification, Ph.D. Thesis, Cornell University, Ithaca, NY
Bhatia, R. (1996): Matrix analysis, Springer-Verlag, Heidelberg Berlin New York 1996
Bhattacharyya, D.P. and R.D. Narayan (1939): Moments of the D^2-statistic for population with unequal dispresions, Sankhya, 5,401-412.
Bhattacharya, R.N. and R. Ranga Rao (1976): Normal approximation and asymptotic expansions, J. Wiley, New York 1976
Bhimasankaram, P. and Sengupta, D. (1996): The linear zero functions approach to linear models. Sankhya, Ser. B 58, 3 (1996), 338-351.
Bhimasankaram, P. and SahaRay, R. (1997): On a partitioned linear model and some associated reduced models. Linear Algebra Appl. 264 (1997), 329-339.
Bhimasankaram, P. and Sengupta, D., and Ramanathan, S. (1995): Recursive inference in a general linear model. Sankhya, Ser. A 57, 2 (1995), 227-255.
Biacs, Z.F., Krakiwsky, E.J. and Lapucha, D. (1990): Reliability Analysis of Phase Observations in GPS Baseline Estimation. J Surv. Eng. 1990, 116, 204-224.
Biancale, R., Balmino, G., Lemoine, J.-M., Marty, J.-M., Moynot, B., Barlier, F., Exertier, P., Laurain, O., Gegout, P., Schwintzer, P., Reigber Ch., Bode A., Konig, R., Massmann, F.-H., Raimondo, J.C., Schmidt, R. and Zhu, S.Y., (2000): A new global Earth's gravity field model from satellite orbit perturbations: GRlMS-Sl. Geophys. Res. Lett., 27,3611-3614.

Bibby, J. (1974): Minimum mean square error estimation, ridge regression, and some unanswered questions, colloquia mathematica societatis Janos Bolyai, Progress in statistics, ed. J. Gani, K. Sarkadi, I. Vincze, Vol. I, Budapest 1972, North Holland Publication Comp., Amsterdam 1974

Bickel, P.J., Doksum, K. and J.L. Hodges (1982): A Festschrift for Erich L. Lehmann, Chapman and Hall, Boca Raton 1982

Bickel, P.J. and K.A. Doksum (1981): An analysis of transformations revisited, J. Am. Statist. Ass. 76 (1981), 296-311

Bierman, G.J. (1977): Factorization Methods for discrete sequential estimation, Academic Press, New York London 1997

Bill, R. (1985): Kriteriummatrizen ebener geodätischer Netze, Ph.D. Thesis, Akademic d. Wissenschafien, Deutsche Geodätische Kommission, München,

Bill, R. (1985b): Kriteriummatrizen ebener geodätischer Netze, Deutsche Geodätische Kommission, München, Reihe A, No. 102

Bill, R. and Kaltenbach H. (1986): Kriteriummatrizen ebener geodätischer Netze - Zur Schutzung von Korrelations funktionen uvittels Polynomen and Splines, Acl. Vermessungsnaeir AVN 96 (1986) 148-156

Bilodeau, M. and D. Brenner (1999): Theory of multivariate statistics, Springer-Verlag, Heidelberg Berlin New York 1999

Bingham, C. (1964): Distributions on the sphere and projective plane, Ph. D. Thesis, Yale University 1964

Bingham, C. (1974): An antipodally symmetric distribution on the sphere, Ann. Statist. 2 (1974), 1201-1225

Bingham, C., Chang, T. and D. Richards (1992): Approximating the matrix Fisher and Bingham distributions: Applications to spherical regression and Procrustes analysis, Journal of Multivariate Analysis 41 (1992), 314-337

Bini, D. and V. Pan (1994): Polynomial and matrix computations, Vol. 1: Fundamental Algorithms, Birkhäuser-Verlag, Basel Boston Berlin 1994

Bill, R. (1984): Eine Strategie zur Ausgleichung und Analyse von Verdichtungsnetzen, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Report C295, 91 pages, München 1984

Bill, R. (1985): Kriterium matrizen chener geodätischer Netze, Deutsche Geodätische Kommission, Series A, report 102, Munich 1985

Bill, R. and H. Kaltenbach (1985): Kriterium matrizen ebener geodesticher Netze, zur Schutzung von Korrelations functionen mittels Polynomen and Spline, All. Vermessungswesen 96 (1986), pp. 148-156

Birtill, J.W. and Whiteway, F.E. (1965): The application of phased arrays to the analysis of seismic body waves, Phil.Trans.Roy.Soc. London, A 258 (1965), 421-493

Bischof, C.H. and G. Quintana-Orti (1998): Computing rank-revealing QR factorizations of dense matrices, ACM Transactions on Mathematical Software 24 (1998), 226-253

Bischoff, W. (1992): On exact D-optimal designs for regression models with correlated observations, Ann. Inst. Statist. Math. 44 (1992), 229-238

Bischoff, W. and Heck, B., Howind, J. and Teusch, A. (2005): A procedure for testing the assumption of homoscedasticity in least squares residuals: a case study of GPS carrier-phase observations, Journal of Geodesy, 78:397-404, 2005.

Bischoff, W. and Heck, B. and Howind, J. and Teusch, A. (2006): A procedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances: a case study of GPS carrier-phase observations. J Geod 79(12):694-704

Bradley, J.V. (1968): Distribution-free statistical tests, Englewood Cliffs, Prentice-Hall 1968

Briscoe, H.W. and Fleck, P.L (1965): Data recording and processing for the experimental large aperture seismic array, Proc. IEEE 53 (1965), 1852-1859

Bjerhammar, A. (1951): Rectangular reciprocal matrices with special reference to calculation, Bull. Gèeodèesique 20 (1951), 188-210

Bjerhammar, A. (1951): Application of calculus of matrices to the method of least-squares with special reference to geodetic calculations, Trans. RIT, No. 49, Stockholm 1951
Bjerhammar, A. (1955): En ny matrisalgebra, SLT 211-288, Stockholm 1955
Bjerhammar, A. (1958): A generalized matrix algebra, Trans. RIT, No. 124, Stockholm 1958
Bjerhammar, A. (1969): Linear Prediction and Filtering, Royal Inst. Techn. Div. Geodesy, Stockholm 1969.
Bjerhammar, A. (1973): Theory of errors and generalized matrix inverses, Elsevier, Amsterdam 1973
Björck, A. (1967): Solving linear least squares problems by Gram-Schmidt orthogonalization, Nordisk Tidskr. Informationsbehandling (BIT) 7 (1967), 1-21
Björck, A. (1990): Least Squares Methods. In: Handbook of Numerical Analysis, Eds.: P.G. Ciarlet, J.L. Lions, North Holland, Amsterdam
Björck, A. (1996): Numerical methods for least squares problems, SIAM, Philadelphia 1996
Björck, A. and G.H. Golub (1973): Numerical methods for computing angles between linear subspaces, Mathematics of Computation 27 (1973), 579-594
Björkström, A. and R. Sundberg (1999): A generalized view on continuum regression, Scandinavian Journal of Statistics 26 (1999), 17-30
Blackman, R.B. (1965): Linear data - smoothing and prediction in theory and practice, Reading 1965
Blaha, G. (1987): Nonlinear Parametric Least-Squares Adjustment. AFGL Technical Report 87-0204, Air Force Geophysics Laboratory, Bedford, Mass.
Blaha, G. (1991): Non-Iterative Geometric Approach to Nonlinear Parametric Least-Squares Adjustment with or without Constraints. Report No. PL-TR-91-2136, Phillips Laboratory, Air Force Systems Command, Hanscom Air Force Base Massachusetts.
Blaha, G. and Besette, R.P. (1989): Nonlinear least squares method via an isomorphic geometrical setup, *Bulletin Geodesique* 63 (1989) 115-138.
Blaker, H. (1999): Shrinkage and orthogonal decomposition, Scandinavian Journal of Statistics 26 (1999), 1-15
Blaschke, W.; Leichtweiss, K. (1973): Elementare Differentialgeometrie. Springer-Verlag, Berlin Heidelberg.
Blewitt, G. (2000): Geodetic network optimization for geophysical parameters, Geophysical Research Letters 27 (2000), 2615-3618
Bloomfield, P. and W.L. Steiger (1983): Least absolute deviations - theory, applications and algorithms, Birkhäuser-Verlag, Basel Boston Berlin 1983
Bobrow, J.E. (1989): A direct minimization approach for obtaining the distance between convex polyhedra, Int. J. Robotics Research 8 (1989), 65-76
Boedecker, G. (1977): The development of an observation scheme for the variance-covariance matrix of the unknowns, Proceedings International Symposium on Optimization of Design and Computation of Control Networks, Sopron 1977.
Boggs, P.T., Byrd, R.H. and R.B. Schnabel (1987): A stable and efficient algorithm for nonlinear orthogonal distance regression, SIAM J. Sci. Stat. Comput. 8 (1987), 1052-1078
Boik, R.J. (1986): Testing the rank of a matrix with applications to the analysis of interaction in ANOVA. J. Am. Stat. Assoc. 81 (1986), 243-248.
Bollerslev, T. (1986): Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (1986), 307-327
Boogaart, K.G. and van den A. Brenning (2001): Why is universal kriging better than IRFkkriging: estimation of variograms in the presence of trend, Proceedings of 2001 Annual Conference of the International Association for Mathematical Geology, September 6-12, 2001, Cancun, Mexico
Boogaart, K.G. and van den H. Schaeben (2002a): Kriging of regionalized directions, axes, and orientations 1. Directions and axes, Mathematical Geology, to appear
Boogaart, K.G. and van den H. Schaeben (2002b): Krigging of normalised directions, axes, Month. Geology 34 (2002) 479

Boon, F. and Ambrosius, B. (1997): Results of real-time applications of the LAMBDA method in GPS based aircraft landings. In: Proceedings KIS97, pp 339-345
Boon, F., de Jonge, P.J. and Tiberius, C.C.J.M. (1997): Precise aircraft positioning by fast ambiguity resolution using improved troposphere modelling. In: Proceedings ION GPS-97. vol. 2, pp 1877-1884
Booth, J.G. and J.P. Hobert (1996): Standard errors of prediction in generalized linear mixed models, J. Am. Statist. Ass. 93 (1996), 262-272
Bopp, H.; Krauss, H. (1978): Strenge oder herk6mmliche bedingte Ausgleichung mit Unbekannten bei nichtlinearen Bedingungsgleichungen? Allgemeine Vermessungsnachrichten, 85. Jg., Heft 1, 27-3l.
Bordes L., Nikulin, M. and V. Voinov (1997): Unbiased estimation for a multivariate exponential whose components have a common shift, J. Multivar. Anal. 63 (1997), 199-221
Borg, I. and P. Groenen (1997): Modern multidimensional scaling, Springer Verlag, Berlin New York 1997
Borkowski, K.M. (1987): Transformation of geocentric to geodetic coordinates without approximation, *Astrophys. Space. Sci.* 139 (1987) 1-4.
Borkowski, K.M. (1989): Accurate algorithm to transform geocentric to geodetic coordinates, *Bull. Geod.* 63 (1989) 50-56.
Borovkov, A.A. (1998): Mathematical statistics, Gordon and Breach Science Publishers, Amsterdam 1998
Borre, K. (1977): Geodetic elasticity theory. Bulletin Geodesique, 63-71.
Borre, K. (1978): Error propagation in absolute geodetic networks - a continuous approach -. Studia geophysica et geodaetica, 213-223.
Borre, K. (1979a): Covariance functions for random error propagation in large geodetic networks. Scientific Bulletin of the Stanislaw I Staszic University of Mining and Metallurgy, Krakow, 23-33.
Borre, K. (1979b): On discrete networks versus continua. 2. Int. Symp. wei tgespannte Fl achentra.gwerke. Beri chtsheft 2. p. 79-85. Stuttgart.
Borre, K. (1980): Elasticity and stress-strain relations in geodetic networks, in: Proceedings of the International School of Advanced Geodesy, ed. E. Grafarend and A. Marussi, p. 335-374, Florence.
Borre, K. (2001): Plane networks and their applications, Birkhäuser-Verlag, Basel Boston Berlin 2001
Borre, K. and Meissl, P. (1974): Strength analysis of leveling type networks, an application of random walk theory, Geodaetisk Institute Meddelelse No. 50, Copenhagen.
Borre, K. and T. Krarup (1974): Foundation of a Theory of Elasticity, for Geodetic Networks. 7. nordiske geodaetm0de i K0benhavn.
Borre, K. and S.L. Lauritzen (1983): Some geometric aspects of adjustment, In. Festschrift to Torben Krarup, Eds. E. Kejlso et al., Geod. Inst., No. 58, pp. 70-90.
Bos, M.S., Fernandes, R.M., Williams, S.D.P. and L. Bastos (2008): Fast error analysis of continuous GPS observations, J. Geodesy 82 (2008) 157-166
Bose, R.C. (1944): The fundamental theorem of linear estimation, Proc. 31st Indian Scientific Congress (1944), 2-3
Bose, R.C. (1947): The design of experiments, Presidential Address to Statistics Section, 34th session of the Indian Science Congress Assoc. 1947
Bossler, J. (1973): A note on the meaning of generalized inverse solutions in geodesy, J. Geophys. Res. 78 (1973), 2616
Bossler, J., Grafarend, E. and R. Kelm (1973): Optimal design of geodetic nets II, J. Geophys. Res. 78 (1973), 5887-5897
Bossler, J., Grafarend, E. and R. Kelm (1973) : Optimal Design of Geodetic Nets II, J. Geophysical Research 78 (1973) 5837-5898
Boulware, D.G., Brown, L.S. and R.D. Peccei (1970): Deep inelastic electroproduction and conformal symmetry, Physical Review D2 (1970), 293-298

Bowring, B.R. (1976): Transformation from spatial to geographical coordinates, *Survey Review* 23 (1976) 323-327.
Bowring, B.R. (1985): The accuracy of geodetic latitude and height equations, *Survey Review* 28 (1985) 202-206.
Box, G.E.P. and D.R. Cox (1964): Ana analysis of transformations, J. Roy. Statist. Soc. B26 (1964), 211-252
Box, G.E.P. and G. Tiao (1973): Bayesian inference in statistical analysis, Addison-Wesley, Reading 1973
Box, G.E.P. and N.R. Draper (1987): Empirical model-building and response surfaces, J. Wiley, New York 1987
Box, M.J. (1971): Bias in nonlinear estimation, J. Roy. Statist. Soc. B33 (1971), 171-201
Branco, M.D. (2001): A general class of multivariate skew-elliptical distributions, J. Multivar. Anal. 79 (2001), 99-113
Brandt, S. (1992): Datenanalyse. Mit statistischen Methoden und Computerprogrammen, 3. Au., BI Wissenschaftsverlag, Mannheim 1992
Brandt, S. (1999): Data analysis: statistical and computational methods for scientists and engineers, 3rd ed., Springer-Verlag, Heidelberg Berlin New York 1999
Brandstätter, G. (1974): Notiz zur analytischen Lösung des ebenen Rückwärtsschnittes, *Österreichische Zeitschrift für Vermessungswesen* 61 (1974) 134-136.
Braess, D. (1986): Nonlinear approximation theory, Springer-Verlag, Heidelberg Berlin New York 1986
Branham, R.L. (1990): Scientific Data Analysis. An Introduction to Overdetermined Systems, Springer-Verlag, New York
Brauner, H. (1981): Differentialgeometrie. Vieweg, Wiesbaden.
Breckling, J. (1989): Analysis of directional time series: application to wind speed and direction, Springer-Verlag, Heidelberg Berlin New York 1989
Bremand, P. (1999): Markov chains, Gibbs fields, Monte Carlo simulation and queues, Spring Verlag, New York 1999
Breslow, N.E. and D.G. Clayton (1993): Approximate inference in generalized linear mixed models, J. Am. Statist. Ass. 88 (1993), 9-25
Brill, M. and E. Schock (1987): Iterative solution of ill-posed problems - a survey, in: Model optimization in exploration geophysics, ed. A. Vogel, Vieweg, Braunschweig 1987
Brillinger, E.D. (1964): An introduction to polyspectra, Economic Research Program Memorandum no 67, Princeton 1964
Bro, R. (1997): A fast non-negativity-constrained least squares algorithm, J. Chemometrics 11 (1997), 393-401
Bro, R. and S. de Jong (1997): A fast non-negativity-constrained least squares algorithm, Journal of Chemometrics 11 (1997), 393-401
Brock, J.E. (1968), Optimal matrices describing linear systems, AIAA J. 6 (1968), 1292-1296
Brockwell, P. and Davies, R. (2003). Time Series: Theorie and Methods. Springer, New York
Brokken, F.B. (1983): Orthogonal Procrustes rotation maximizing congruence, *Psychometrika* 48 (1983) 343-352.
Brovelli, M.A., Sanso, F. and G. Venuti (2003): A discussion on the Wiener-Kolmogorov prediction principle with easy-to-compute and robust variants, Journal of Geodesy 76 (2003), 673-683
Brown, B. and R. Mariano (1989): Measures of deterministic prediction bias in nonlinear models, Int. Econ. Rev. 30 (1989), 667-684
Brown, B.M., P. Hall, and G.A. Young (1997): On the effect of inliers on the spatial median, J. Multivar. Anal. 63 (1997), 88-104
Brown, H. and R. Prescott (1999): Applied mixed models in medicine, J. Wiley, New York 1999
Brown, K.G. (1976): Asymptotic behavior of Minque-type estimators of variance components, The Annals of Statistics 4 (1976), 746-754
Brown, P.J., Le, N.D. and J.V. Zidek (1994): Multivariate spatial interpolation and exposure to air pollutants. Can J Stat 2:489-509

Brown, W.S. (1971): On Euclid's Algorithm and the Computation of Polynomial Greatest Common Divisors. Journal of the ACM, 18(4), 478-504.

Brualdi, R.A. and H. Schneider (1983): Determinantal identities: Gauss, Schur, Cauchy, Sylvester, Kronecker, Jacobi, Binet, Laplace, Muir and Cayley, Linear Algebra Appl. 52/53 (1983), 765-791

Brunk, H.D. (1958): On the estimation of parameters restricted by inequalities, Ann. Math. Statist. 29 (1958), 437-454

Brunner, F.K. (1979): On the analysis of geodetic networks for the determination of the incremental strain tensor, *Survey Review* 25 (1979) 146-162.

Brunner, F.K. (1999): Zur Präzision von GPS Messungen, Geowiss. Mitteilungen, TU Wien, Heft 50:1-10, 1999.

Brunner, F.K., Hartinger, H. and L. Troyer (1999): GPS signal diffraction modelling: the stochastic sigma-model, Journal of Geodesy 73 (1999), 259-267

Brunner, F.K. (2009): Ranm-veit analyse von GPs messengun mittels Turbulenz theorie, First holloquium E. Grafarend, Stuttgart 2009

Bruno, A.D. (2000): Power geometry in algebraic and differential equations, Elsevier, Amsterdam Lausanne New York Oxford Shannon Singapore Tokyo 2000

Brus, D.J. and de Gruijter, J.J. (1997): Random sampling or geostatistical modeling? Chosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80:1-44

Brus, D.J. and Heuvelink, G.B.M. (2007): Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138: 86-95

Brzèezniak, Z. and T. Zastawniak (1959): Basic stochastic processes, Springer-Verlag, Heidelberg Berlin New York 1959

Buchberger, B. (1965a): Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen PolynomideaL Dissertation der Universitat Innsbruck.

Buchberger, B. (1965b): An algorithm for finding a basis for the residue class ring of a zero dimensional polynomial ideal (German), Ph.D. thesis, University of Innsbruck, Institute of Mathematics.

Buchberger, B. (1970): Ein algorithmisches Kriterium für die Lösbarkeit eines algebraischen Gleichungsystems, *Aequationes Mathematicae* 4 (1970) 374-383.

Buchberger, B. (1979): A criterion for detecting unnecessary reductions in the construction of Groebner bases. Proceedings of the 1979 European Symposium on Symbolic and Algebraic computation, Springer lecture notes in Computer Science 72, Springer-Verlag, pp. 3-21, Berlin-Heidelberg-New York 1979.

Buchberger, B. (1985): Groebner Bases: An Algorithmic Method in Polynomial Ideal Theory. In: Multidimensional Systems Theory, Ed.: N.K. Bose, D. Reidel Publishing Company, Dordrecht-Boston-Lancaster.

Buchberger, B. and Kutzler, B. (1986): Computer-Algebra fiir den Ingenieur. In: Mathematische Methoden in der Technik, Hrsg.: J. Lehn, H. Neunzert, H. Wacker, Band 4, Rechnerorientierte Verfahren, B.G. Teubner, Stuttgart

Buchberger, B. (1999): Personal communication.

Buchberger, B. and Winkler, F. (1998): Groebner bases and applications. London mathematical society lecture note series 251, Cambridge university press, 1998.

Buell, C.E. (1972): Correlation functions for wind and geopotential on isobaric surfaces, Journal of Applied Meteorology, 1972.

Bueso, M.C., Angulo, J.M., Qian, G. and Alonso, F.J. (1999): Spatial sampling design based on stochastic complexity. J. Multivar. Anal. 71: 94-110

Buhmann, M.D. (2001): Approximatipn and interpolation with radial functions, In: Multivariate Approximation and Applications, Cambridge University Press, Cambridge 2001, 25-43

Bunday, B.D., S.M.H. Bokhari, and K.H. Khan (1997): A new algorithm for the normal distribution function, Sociedad de Estadistica e Investigacion Operativa 6 (1997), 369-377

Bunke, H. and O. Bunke (1974): Identifiability and estimability, Math. Operationsforschg. Statist. 5 (1974), 223-233

Bunke, H. and O. Bunke (1986): Statistical inference in linear models, J. Wiley, New York 1986

Bunke, H. and O. Bunke (1989): Nonlinear regression, functional relations and robust methods, Vol 2, Wiley 1989

Bunke, O. (1977): Mixed models, empirical Bayes and Stein estimators, Math. Operationsforschg. Ser. Statistics 8 (1977), 55-68

Buonaccorsi, J., Demidenko, E. and T. Tosteson (2000): Estimation in longitudinal random effects models with measurement error, Statistica Sinica 10 (2000), 885-903

Burg, J.P (1964): Three-dimensional filtering with an array of seismometers, Geophysics 29 (1964), 693-713

Burns, F., Carlson, D., Haynsworth, E., and T. Markham (1974): Generalized inverse formulas using the Schur complement, SIAM J. Appl. Math. 26 (1974), 254-259

Businger, P. and G.H. Golub (1965): Linear least squares solutions by Householder transformations, Numer. Math., 7 (1965), 269-276

Butler, N.A. (1999): The efficiency of ordinary least squares in designed experiments subject to spatial or temporal variation, Statistics & Probability Letters 41 (1999), 73-81

Caboara, M. and E. Riccomagno (1998): An algebraic computational approach to the identifiability of Fourier models, J. Symbolic Computation 26 (1998), 245-260

Cadet, A. (1996): Polar coordinates in Rnp, application to the computation of the Wishart and Beta laws, Sankhya: The Indian Journal of Statistics 58 (1996), 101-114

Cai, J. (2004): Statistical inference of the eigenspace components of a symmetric random deformation tensor, Dissertation, Deutsche Geodätische Kommission (DGK) Reihe C, Heft Nr. 577, 138 Seiten, München, 2004

Cai, J. and Grafarend, E. and Schaffrin, B. (2004): The A-optimal regularization parameter in uniform Tykhonov-Phillips regularization - α-weighted BLE, IAG Symposia 127 "Hotine-Marussi Symposium on Mathematical Geodesy", Matera, Italy, 17-21 June 2002, edited by F. Sanso, Springer, 309-324

Cai, J. and Grafarend, E. and Schaffrin B. (2005): Statistical inference of the eigenspace components of a two-dimensional, symmetric rank two random tensor, J. of Geodesy, Vo1. 78, 426-437

Cai, J. and Grafarend, E. (2007): Statistical analysis of the eigenspace components of the twodimensicnial, symmetric rank-two strain rate tensor derived from the space geodetic measurements (ITRF92-ITRF2000 data sets) in central Mediterranean and Western Europe, Geophysical Journal International, Vol. 168, 449-472

Cai, J. and Koivula, H. and Poutanen, M. (2008): The statistical analysis of the eigenspace components of the strain rate tensor derived from FinnRef GPS measurements (1997-2004), IAG Symposia 132 "VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy", Wuhan, China, 29 May-2 June, 2006, eds. P. Xu, J. Liu and A. Dermanis, pp. 79-87, Springer, Berlin, Heidelberg

Cai, J. and Hu, C. and Grafarend, E. and Wang, J. (2008): The uniform Tykhonov-Phillips regularization (a weighted S-homBLE) and its application in GPS rapid static positioning, in Fennoscandia, IAG Symposia 132 "VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy", Wuhan, China, 29 May-2 June, 2006, eds. P. Xu, J. Liu and A. Dermanis, pp. 221-229, Springer, Berlin, Heidelberg

Cai, J. and Wang, J. and Wu, J. and Hu, C. and Grafarend, E. and Chen, J. (2008): Horizontal deformation rate analysis based on multi-epoch GPS measurements in Shanghai, Journal of Surveying Engineenng, Vol. 134(4), 132-137

Cai, J. and Grafarend, E. and Hu, C. (2009): The total optimal criterion in solving the mixed integer linear model with GNSS carrier phase observations, Journal of GPS Solutions, Vol. 13, 221-230, doi: 10.1007 Is 10291-008-0 115-y

Calder, C. and Cressie, N. (2007): Some topics in convolution-based modeling. In Proceedings of the 56th session of the International Statistics Institute, Lisbon.

Cambanis, S. and I. Fakhre-Zakeri (1996): Forward and reversed time prediction of autoregressive sequences, J. Appl. Prob. 33 (1996), 1053-1060

Campbell, H.G. (1977): An introduction to matrices, vectors and linear programming, 2nd ed., Printice Hall, Englewood Cliffs 1977
Candy, J.V. (1988): Signal processing, McGrow Hill, New York 1988
Canny, J.F. (1988): The complexity of robot motion planning, ACM doctoral dissertation award, MIT Press, 1988.
Canny, J.F., Kaltofen, E. and Yagati, L. (1989): Solving systems of nonlinear polynomial equations faster, Proceedings of the International Symposium on Symbolic and Algebraic Computations ISSAC, July 17-19, Portland, Oregon 1989. pp. 121-128.
Carlin, B.P. and T.A. Louis (1996): Bayes and empirical Bayes methods, Chapman and Hall, Boca Raton 1996
Carlitz, L. (1963): The inverse of the error function, Pacific J. Math. 13 (1963), 459-470
Carlson, D., Haynsworth, E. and T. Markham (1974): A generalization of the Schur complement by means of the Moore-Penrose inverse, SIAM J. Appl. Math. 26 (1974), 169-179
Carlson, D. (1986): What are Schur complements, anyway? Linear Algebra and its Applications 74 (1986), 257-275
Carmo Do, M. (1983): Diiferentialgeometrie von Kurven und Flachen. Vieweg-Studium, Braunschweig.
Carroll, J.D., Green, P.E. and A. Chaturvedi (1999): Mathematical tools for applied multivariate analysis, Academic Press, San Diego 1999
Carroll, J.D. and P.E. Green (1997): Mathematical tools for applied multivariate analysis, Academic Press, San Diego 1997
Carroll, R.J. and D. Ruppert (1982a): A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model, J. Am. Statist. Ass. 77 (1982), 878-882
Carroll, R.J. and D. Ruppert (1982b): Robust estimation in heteroscedastic linear models, Ann. Statist., 10 (1982), pp. 429-441
Carroll, R.J., Ruppert, D. and L. Stefanski (1995): Measurement error in nonlinear models, Chapman and Hall, Boca Raton 1995
Carruthers, P. (1971): Broken scale invariance in particle physics, Phys. Lett. Rep. 1 (1971), 1-30
Caspary, W. (1987): Concepts of Network and Deformation Analysis, Monograph 11, The University of New South Wales, School of Surveying, N.S.W., Australia, 1987
Caspary, W.; Borutta, H. (1986): Geometrische Deformationsanalysemit robusten Schätzverfahren, Allgemeine Vermessungsnachrichten 8-9/1986, S.315-326
Caspary, W. and Lohse, P. and Van Mierlo, J. (1991): SSG 4.120: Non-linear Adjustment. In: National Report of the Federal Republic of Germany on the Geodetic Activities in the Years 1987-1991, Eds.: E.W. Grafarend, K. Schnadelbach Deutsche Geodätische Kommission, Reihe B, Nr.294.
Caspary, W. and K. Wichmann (1994): Lineare Modelle. Algebraische Grundlagen und statistische Anwendungen, Oldenbourg Verlag, München/Wien 1994
Castillo, J. (1994): The singly truncated normal distribution, a non-steep exponential family, Ann. Inst. Statist. Math 46 (1994), 57-66
Castillo, J. and P. Puig (1997): Testing departures from gamma, Rayleigh and truncated normal distributions, Ann. Inst. Statist. Math. 49 (1997), 255-269
Cattani, E., Dickenstein, A. and Sturmfels, B. (1998): Residues and resultants, *J. Math. Sci. University of Tokyo* 5 (1998) 119-148.
Cayley, A. (1843): On the application os Quaternions to the theory of rotation, Phil. May. 33, (1848), pp. 196-200
Cayley, A. (1845): On Jacobi's elliptic functions, in reply to the Rev. B. Bronwix and on quaternions, Philos, May. 261 (1845), pp. 208-211
Cayley, A. (1855): Sept differents memoires d'analyse, No. 3, Remarque sur la notation des fonctions algebriques, Journal für die reine und angewandte Mathematik 50 (1855), 282-285
Cayley, A. (1857): A memoir on the theory of matrices, Phil. Trans. Roy. Soc. London, 148 (1857), pp. 17-37 (collected works, vol. 2, 475-496)
Cayley, A. (1858): A memoir on the theory of matrices, Phil. Transactions, Royal Society of London 148 (1858), 17-37

Cayley, A. (1863): On Jacobi's elliptic functions, in reply to the Rev. B. Bronwin, and on quaternions (appendix only), in: The collected mathematical Papers, Johnson reprint Co., p. 127, New York, 1963
Cayley, A. (1885): On the Quaternion equation $qQ - Qq' = 0$, Messenger, 14 (1885), pp. 108-112
Cenkov, N.N. (1972): Statistical decision rule and optimal inference, Nauka 1972
Champagne, F.H. (1978): The fine scale structure of the turbulent velocity field, J. Fluid. Mech. 86 (1978) 67-108
Chaffee, J. and Abel, J. (1994): On the exact solutions of the pseudorange equations, *IEEE Transactions on Aerospace and Electronic Systems* 30 (1994) 1021-1030.
Chan, K.-S. and H. Tong (2001): Chaos, a statistical perspective, Springer-Verlag, Heidelberg Berlin New York 2001
Chan, K.-S. and H. Tong (2002): A note on the equivalence of two approaches for specifying a Markov process, Bernoulli 8 (2002), 117-122
Chan, T.F. and P.C. Hansen (1991): Some applications of the rank revealing QR factorizations, Numer. Linear Algebra Appl., 1 (1991), 33-44
Chan, T.F. and P.C. Hansen (1992): Some applications of the rank revealing QR factorization, SIAM J. Sci. Statist. Comput., 13 (1992), 727-741
Chandrasekar, S. (1950): The theory of axisymmetric turbulence, Proc. Royal Soc. a2242 (1950) 557
Chandrasekar, S. and I.C.F. Ipsen (1995): Analysis of a QR algorithm for computing singular values, SIAM J. Matrix Anal. Appl. 16 (1995), 520-535
Chandrasekar, S., Gu, M. and A.H. Sayed (1998): A stable and efficient algorithm for the indefinite linear least-squares problem, SIAM J. Matrix Anal. Appl. 20 (1998), 354-362
Chandrasekar, S., Golub, G.H., Gu, M. and A.H. Sayed (1998): Parameter estimation in the presence of bounded data uncertainties, SIAM J. Matrix Anal. Appl. 19 (1998), 235-252
Chang, F.-C. and Y.-R. Yeh (1998): Exact A-optimal designs for quadratic regression, Statistica Sinica 8 (1998), 527-533
Chang, H. and Fu, A.Q. and Le, N.D. and Zidek J.V. (2005): Designing environmental monitoring networks for measuring extremes. http://www.samsi.info/TR/tr2005-04.pdf
Chang, T. (1986): Spherical regression, Annals of Statistics 14 (1986), 907-924
Chang, T. (1988): Estimating the relative rotations of two tectonic plates from boundary crossings, J. Am. Statist. Ass. 83 (1988), 1178-1183
Chang, T. (1993): Spherical regression and the statistics of tectonic plate reconstructions, International Statis. Rev. 51 (1993), 299-316
Chang, X.W., Yang, X., Zhou, T. (2005): MLAMBDA: a modified LAMBDA method for integer ambiguity determination, Proceedings ION GNSS2005, Long Beach
Chapman, D.G. and H. Robbins (1951): Minimum variance estimation without regularity assumptions, Ann. Math. Statist. 22 (1951), 581-586
Char, B.W. et al. (1990): MAPLE Reference Manual. Fifth Edition. Waterloo Maple Publishing
Chartres, B.A. (1963): A geometrical proof of a theorem due to Slepian, SIAM Review 5 (1963), 335-341
Chatfield, C. and A.J. Collins (1981): Introduction to multivariate analysis, Chapman and Hall, Boca Raton 1981
Chatterjee, S. and A.S. Hadi (1986): Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (1986), pp. 379-416
Chatterjee, S. and A.S. Hadi (1988): Sensitivity analysis in linear regression, J. Wiley, New York 1988
Chatterjee, S. and M. Mächler (1997): Robust regression: a weighted least-squares approach, Commun. Statist. Theor. Meth. 26 (1997), 1381-1394
Chaturvedi, A. and A.T.K. Wan (1998): Stein-rule estimation in a dynamic linear model, J. Appl. Stat. Science 7 (1998), 17-25
Chaturvedi, A. and A.T.K. Wan (1999): Estimation of regression coefficients subject to interval constraints in models with non-spherical errors, Indian Journal of Statistics 61 (1999), 433-442
Chaturvedi, A. and A.T.K. Wan (2001): Stein-rule restricted regression estimator in a linear regression model with nonspherical disturbances, Commun. Statist.-Theory Meth. 30 (2001), 55-68

Chauby, Y.P. (1980): Minimum norm quadratic estimators of variance components, Metrika 27 (1980), 255-262

Chen, Hsin-Chu. (1998): Generalized reexive matrices: Special properties and applications, Society for Industrial and Applied Mathematics, 009 (1998), 141-153

Chen, L: Math. Phys. 167 (1995) 443-469

Chen, X. (2001): On maxima of dual function of the CDT subproblem, J. Comput. Mathematics 19 (2001), 113-124

Chen Y.-S. (2001): Outliers detection and confidence interval modification in fuzzy regression Fuzzy Sets Syst. 119, 2 (2001), 259-272.

Chen, Z. and J. Mi (1996): Confidence interval for the mean of the exponential distribution, based on grouped data, IEEE Transactions on Reliability 45 (1996), 671-677

Chen, V.C.P., Tsui, K.L., Barton, R.R. and Meckesheimer, M. (2006): A review on design, modeling and applications of computer experiments. IIE Trans 38:273-291

Chen, Y.Q., Kavouuras, M. and Chrzanowski, A. (1987): A Strategy for Detection of Outlying Observations in Measurements of High Precision. The Canadian Surveyor. 41:529-540.

Cheng, C.L. and van J.W. Ness (1999): Statistical regression with measurement error, Arnold Publ., London 1999

Cheng, C.L. (1998): Polynomial regression with errors in variables, J. Roy. Statist. Soc. B60 (1998), 189-199

Cherioito, P. (2001): Mixed fractional Brownian motion, Bernoulli 7 (2001) 913-934

Chiang, C.-Y. (1998): Invariant parameters of measurement scales, British Journal of Mathematical and Statistical Psychology 51 (1998), 89-99

Chiang, C.L. (2003): Statistical methods of analysis, University of California, Berkeley, USA 2003

Chikuse, Y. (1999): Procrustes analysis on some special manifolds, Commun. Statist. Theory Meth. 28 (1999), 885-903

Chilées, J.P. and P. Delfiner (1999): Geostatistics - modelling spatial uncertainty, J. Wiley, New York 1999

Chiodi, M. (1986): Procedures for generating pseudo-random numbers from a normal distribution of order, Riv. Stat. Applic. 19 (1986), 7-26

Chmielewski, M.A. (1981): Elliptically symmetric distributions: a review and bibliography, International Statistical Review 49 (1981), 67-74

Chow, Y.S. and H. Teicher (1978): Probability theory, Springer-Verlag, Heidelberg Berlin New York 1978

Chow, T.L. (2000): Mathematical methods for physicists, Cambridge University Press, Cambridge 2000

Christensen, R. (1991): Linear models for multivariate time series and spatial data. Springer, Berlin

Christensen, R. (1996): Analysis of variance, design and regression, Chapman and Hall, Boca Raton 1996

Chrystal, G. (1964): Textbook of Algebra (Vol. 1), Chelsea, New York 1964.

Chrzanowski, A. and Chen, Y.Q. (1998): Deformation monitoring, analysis, and prediction - status report, Proc. FIG XIX into Congress, Helsinki, Vo1. 6, pp. 84-97

Chrzanowski, A. Chen, Y.Q. and Secord, M. (1983): On the strain analysis of tectonic movements using fault crossing geodetic surveys, Tectonophysics, 97, 297-315:

Crocetto, N. and Vettore, A. (2001): Best unbiased estimation of variance-covariance components: From condition adjustment to a generalized methodology. J. Inf. Optimization Sci. 22, 1 (2001), 113-122.

Chiles, J.P. and P. Delfiner (1999): Geostatistics: modeling spatial uncertainty, Wiley series, New York 1999

Chui, C.K. and G. Chen (1989): Linear Systems and optimal control, Springer-Verlag, Heidelberg Berlin New York 1989

Chui, C.K. and G. Chen (1991): Kalman filtering with real time applications, Springer-Verlag, Heidelberg Berlin New York 1991

Ciarlet, P.G. (1988): Mathematical elasticity, Vol. 1: three-dimensional elasticity, North Holland Publ. Comp., Amsterdam

Claerbout, J.F. (1964): Detection of p - waves from weak sources at great distances, Geophysics 29 (1964), 197-211

Clark, G.P.Y. (1980): Moments of the least squares estimators in a nonlinear regression model, JR. Statist. Soc. B42 (1980), 227-237

Clerc-Bèerod, A. and S. Morgenthaler (1997): A close look at the hat matrix, Student 2 (1997), 1-12

Cobb, L., Koppstein, P. and N.H. Chen (1983): Estimation and moment recursions relations for multimodal distributions of the exponential family, J. Am. Statist. Ass. 78 (1983), 124-130

Cobb, G.W. (1997): Introduction to design and analysis of experiments, Springer-Verlag, Heidelberg Berlin New York 1997

Cochran, W. (1972): Some effects of errors of measurement on linear regression, in: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, 527-539, UCP, Berkeley 1972

Cochran, W. (1972): Stichprobenverfahren, de Gruyter, Berlin 1972

Cohen, A. (1966): All admissible linear estimates of the mean vector, Ann. Math. Statist. 37 (1966), 458-463

Cohen, C. and A. Ben-Israel (1969): On the computation of canonical correlations, Cahiers Centre èEtudes Recherche Opèer 11 (1969), 121-132

Cohen J., Kesten M., and Newman C. (eds.) (1985): Random matrices and their applications, Amer. Math. Soc., Providence

Collett, D. (1992): Modelling binary data, Chapman and Hall, Boca Raton 1992

Collet, D. and T. Lewis (1981): Discriminating between the von Mises and wrapped normal distributions, Austr. J. Statist. 23 (1981), 73-79

Colton, D., Coyle, J. and P. Monk (2000): Recent developments in inverse acoustic scattering theory, SIAM Review 42 (2000), 369-414

Comte, F. and Rozenholc, Y. (2002): Adaptive estimation of volatility functions in (auto-) regressive models, Stoel cesses and their Applications 97, 111-145, 2002.

Cook, R.D., Tsai, C.L. and B.C. Wei (1986): Bias in nonlinear regression, Biometrika 73 (1986), 615-623

Cook, R.D. and S. Weisberg (1982): Residuals and influence in regression, Chapman and Hall, London 1982

Cook, R.D. and Goldberg, M.L. (1986): Curvatures for parameter subsets in nonlinear regression. Annals of Statistics, 14.

Corbeil, R.R. and S.R. Searle (1976): restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics, 18 (1976), pp. 31-38

Cottle, R.W. (1974): Manifestations of the Schur complement, Linear Algebra Appl. 8 (1974), 189-211

Coulman, C.E. and J. Vernin (1991): Significance of anisotropy and the outer scale of turbulence for optical and radio seeing, Applied Optics 30 (1991) 118-126

Cox, L.B. and J.S.Y. Wang (1993): Fractal surfaces: measurement and applications in the earth Sciences. Fractals 1 1 (1993), pp. 87-115

Cox, A.J. and N.J. Higham (1999): Row-wise backward stable elimination methods for the equality constrained least squares problem, SIAM J. Matrix Anal. Appl. 21 (1999), 313-326

Cox, D., Little, J. and O'Shea, D. (1997): Ideals, Varieties, and Algorithms, An introduction to computational algebraic geometry and commutative algebra, Springer-Verlag, New York 1997.

Cox, D., Little, J. and O'Shea, D. (1998): Using algebraic geometry. Graduate Text in Mathematics 185, Springer-Verlag, New York 1998.

Cox, D.A. (1998): Introduction to Gröbner bases, *Proceedings of Symposia in Applied Mathematics* 53 (1998) 1-24.

Cox, D.B. Brading, J.D.W. (1999): Integration of LAMBDA ambiguity resolution with Kalman filter for relative navigation of spacecraft. In: Proceedings ION NTM 99, pp 739-745

Cox, D.R. and D.V. Hinkley (1979): Theoretical statistics, Chapman and Hall, Boca Raton 1979

Cox, D.R. and V. Isham (1980): Point processes, Chapman and Hall, Boca Raton (1980)

Cox, D.R. and E.J. Snell (1989): Analysis of binary data, Chapman and Hall, Boca Raton 1989
Cox, D.R. and N. Wermuth (1996): Multivariate dependencies, Chapman and Hall, Boca Raton 1996
Cox, D.R. and N. Reid (2000): The theory of the design of experiments, Chapman & Hall, Boca Raton 2000
Cox, D.R. and P.J. Salomon (2003): Components of variance Chapman and Hall/CRC, Boca Raton London New York Washington DC
Cox, R., Litte, J. and D. O'shea (1997) Ideals, variates and algorithms, an introduction
Cox, T.F. and Cox, M.A. (1994): Multidimensional scaling, St. Edmundsbury Press, St. Edmunds, Suffolk 1994.
Cox, T.F. and M.A.A. Cox (2001): Multidimensional scaling, Chapman and Hall, Boca Raton, Florida 2001;
Chapman & Hall/CRC, Boca Raton London New York Washington D.C. 2003
Craig, H.V.: a) Bull. Amerie. Soc. 36, 558 (1930J, 37, 731 (1931), 39, 919 (1933); b) Americ. J. Math. 57, 457 (1935), 58, 833 (1936), 59, 764 (1937), 61, 791 (1939J.
Craig, A.T. (1943): Note on the independence of certain quadratic forms, The Annals of Mathematical Statistics 14 (1943), 195-197
Cressie, N. (1991): Statistics for spatial data. Wiley, New York
Croceto, N. (1993): Point projection of topographic surface onto the reference ellipsoid of revolution in geocentric Cartesian coordinates, *Survey Review* 32 (1993) 233-238.
Cross, P.A. and K. Thapa (1979): The optimal design of levelling networks. Surv. Rev. 25, 68-79.
Cross, P.A. (1985): The optimal design of levelling networks, Survey Review 25 (1979) 68-79
Cross, P.A. (1985): Numerical methods in network design, in: Grafarend, E.W. and F. Sanso (eds.), Optimization and design of geodetic networks, Springer-Verlag, Heidelberg Berlin New York 1985, 429-435
Crosilla, F. (1983a): A criterion matrix for a second order design of control networks, *Bull. Geod.* 57 (1983a) 226-239.
Crosilla, F. (1983b): Procrustean transformation as a tool for the construction of a criterion Matrix for control networks, *Manuscripta Geodetica* 8 (1983b) 343-370.
Crosilla, F. and A. Beinat, (2005): A forward scarth method for robust general procrustes analysis, Udine/Italy 2005
Crowder, M. (1987): On linear and quadratic estimating function, Biometrika 74 (1987), 591-597
Crowder, M.J. and D.J. Hand (1990): Analysis of repeated measures, Chapman and Hall, Boca Raton 1990
Crowder, M.J., Kimber, A., Sweeting, T., and R. Smith (1993): Statistical analysis of reliability data, Chapman and Hall, Boca Raton 1993
Crowder, J.M., Sweeting, T. and R. Smith (1994): Statistical analysis of reliability data, Chapman and Hall, Boca Raton 1994
Csörgö, M. and L. Horvath (1993): Weighted approximations in probability and statistics, J. Wiley, New York 1993
Csörgö, S. and L. Viharos (1997): Asymptotic normality of least-squares estimators of tail indices, Bernoulli 3 (1997), 351-370
Csörgö, S. and J. Mielniczuk (1999): Random-design regression under long-range dependent errors, Bernoulli 5 (1999), 209-224
Cummins, D. and A.C. Webster (1995): Iteratively reweighted partial least-squares: a performance analysis by Monte Carlo simulation, J. Chemometrics 9 (1995), 489-507
Cunningham, E. (1910): The principle of relativity in electrodynamics and an extension thereof, Proc. London Math. Soc. 8 (1910), 77-98
Czüber, E. (1891): Theorie der Beobachtungsfehler, Leipzig 1891
D'Agostino, R. and M.A. Stephens (1986): Goddness-of-fit techniques, Marcel Dekker, New York 1986
Daalen, D. van (1986): The Astrometry Satellite Hipparcos. Proc. I. Hotine-Marussi Symp. Math. Geodesy, Rome 1985, Milano 1986, 259-284.

Dach, R. (2000): Einfluß von Auflasteffekten auf Präzise GPS-Messungen, DGK, Reihe C, Heft Nr. 519.
Dai, L., Nagarajan, N., Hu, G. and Ling, K. (2005): Real-time attitude determination for micro satellites by LAMBDA method combined with Kalman filtering. In: AIAA proceedings 22nd ICSSC, Monterey
Daniel, J.W. (1967): The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer. Anal. 4 (1967), 10-26
Dantzig, G.B. (1940): On the nonexistence of tests of Student's hypothesis having power functions independent of 2, Ann. Math. Statist. 11 (1949), 186-192
Das, I. (1996): Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems, SIAM J. Optim. 3 (1998), 631ff.
Das, R. and B.K. Sinha (1987): Robust optimum invariant unbiased tests for variance components. In Proc. of the Second International Tampere Conference in Statistics. T. Pukkila and S. Puntanen (eds.), University of Tampere/Finland (1987), 317-342
Das Gupta, S., Mitra, S.K., Rao, P.S., Ghosh, J.K., Mukhopadhyay, A.C. and Y.R. Sarma (1994): Selected papers of C.R. Rao, Vol. 1, J. Wiley, New York 1994
Das Gupta, S., Mitra, S.K., Rao, P.S., Ghosh, J.K., Mukhopadhyay, A.C. and Y.R. Sarma (1994): Selected papers of C.R. Rao, Vol. 2, J. Wiley, New York 1994
Davenport, J.H., Siret, Y. and Tournier, E. (1988): Computer algebra. Systems and algorithms for algebraic computation, Academic Press Ltd., St. Edmundsbury, London 1988.
David, F.N. (1937): A note on unbiased limits for the correlation coefficient, Biometrika, 29, 157-160
David, F.N. (1938): Tables of the Ordinates and Probability Integral o/the Distribution of the Correlation Coefficient in Small Samples, London, Biometrika.
David, F.N. (1953): A note on the evaluation of the multivariate normal integral, Biometrika, 40, 458-459
David, F.N. (1954): Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples, Cambridge University Press, London 1954
David, F.N. and N.L. Johnson (1948): The probability integral transformation when parameters are estimated from the sample, Biometrika 35 (1948), 182-190
David, H.A. (1957): Some notes on the statistical papers of Friedrich Robert Helmert (1943-1917), Bull. Stat. Soc. NSW 19 (1957), 25-28
David, H.A. (1970): Order Statistics, J. Wiley, New York 1970
David, H.A., Hartley, H.O. and E.S. Pearson (1954): The distribution of the ratio in a single normal sample, of range to standard deviation, Biometrika 41 (1954), 482-293
Davidian, M. and A.R. Gallant (1993): The nonlinear mixed effects model with a smooth random effects density, Biometrika 80 (1993), 475-488
Davidian, M., and D.M. Giltinan (1995): Nonlinear models for repeated measurement data, Chapman and Hall, Boca Raton 1995
Davis, R.A. (1997): M-estimation for linear regression with infinite variance, Probability and Mathematical Statistics 17 (1997), 1-20
Davis, R.A. and W.T.M. Dunsmuir (1997): Least absolute deviation estimation for regression with ARMA errors, J. Theor. Prob. 10 (1997), 481-497
Davis, J.H. (2002): Foundations of deterministic and stochastic control, Birkhäuser-Verlag, Basel Boston Berlin 2002
Davison, A.C. and D.V. Hinkley (1997): Bootstrap methods and their application, Cambridge University Press, Cambridge 1997
Dawe, D.J. (1984): Matrix and finite element displacement analysis of structures, Clarendon Press, Oxford
Decreusefond, L. and A.S. Üstünel (1999): Stochastic analysis of the fractional Brownian motion, Potential Anal. 10 (1999), 177-214
Dedekind, R. (1901): Gauß in seiner Vorlesung über die Methode der kleinsten Quadrate, Berlin 1901

Defant, A. and K. Floret (1993): Tensor norms and operator ideals, North Holland, Amsterdam 1993
Deitmar, A. (2002): A first course in harmonic analysis, Springer-Verlag, Heidelberg Berlin New York 2002
Demidenko, E. (2000): Is this the least squares estimate?, Biometrika 87 (2000), 437-452
Denham, M.C. (1997): Prediction intervals in partial least-squares, J. Chemometrics 11 (1997), 39-52
Denham, W. and S. Pines (1966): Sequential estimation when measurement function nonlinearity is comparable to measurement error, AIAA J4 (1966), 1071-1076
Denis, J.-B. and A. Pazman (1999): Bias of LS estimators in nonlinear regression models with constraints. Part II: Bi additive models, Applications of Mathematics 44 (1999), 375-403
Dermanis, A. (1977): Geodetic linear estimation techniques and the norm choice problem, Manuscripta geodetica 2 (1977), 15-97
Dermanis, A. (1978): Adjustment of geodetic observations in the presence of signals, International School of Anvanced Geodesy, Erice, Sicily, May-June 1978, Bollettino di Geodesia e Scienze Affini 38 (1979), 513-539
Dermanis, A. (1980): Adjustment of geodetic observations in the presence of signals. In: Proceedings of the international school of advanced geodesy. Bollettino di Geodesia e Scienze Affini. Vol. 38, pp. 419-445
Dermanis, A. (1998): Generalized inverses of nonlinear mappings and the nonlinear geodetic datum problem, Journal of Geodesy 72 (1998), 71-100
Dermanis, A. and E. Grafarend (1981): Estimability analysis of geodetic, astrometric and geodynamical quantities in Very Long Baseline Interferometry, Geophys. J. R. Astronom. Soc. 64 (1981), 31-56
Dermanis, A. and F. Sanso (1995): Nonlinear estimation problems for nonlinear models, manuscripta geodaetica 20 (1995), 110-122
Dermanis, A. and R. Rummel (2000): Data analysis methods in geodesy, Lecture Notes in Earth Sciences 95, Springer-Verlag, Heidelberg Berlin New York 2000
Dette, H. (1993): A note on E-optimal designs for weighted polynomial regression, Ann. Stat. 21 (1993), 767-771
Dette, H. (1997): Designing experiments with respect to 'standardized' optimality criteria, J. R. Statist. Soc. B59 (1997), 97-110
Dette, H. (1997): E-optimal designs for regression models with quantitative factors - a reasonable choice?, The Canadian Journal of Statistics 25 (1997), 531-543
Dette, H. and W.J. Studden (1993): Geometry of E-optimality, Ann. Stat. 21 (1993), 416-433
Dette, H. and W.J. Studden (1997): The theory of canonical moments with applications in statistics, probability, and analysis, J. Wiley, New York 1997
Dette, H. and T.E. O'Brien (1999): Optimality criteria for regression models based on predicted variance, Biometrika 86 (1999), 93-106
Deutsch, F. (2001): Best approximation in inner product spaces, Springer-Verlag, Heidelberg Berlin New York 2001
Dewess, G. (1973): Zur Anwendung der Schätzmethode MINQUE auf Probleme der Prozeß-Bilanzierung, Math. Operationsforschg. Statistik 4 (1973), 299-313
Dey, D. and Ghosh, S. and Mallick, B. (2000): Generalized linear models: A Bayesian perspective. Marcel Dekker, New York
DiCiccio, T.J. and B. Efron (1996): Bootstrap confidence intervals, Statistical Science 11 (1996), 189-228
Diebolt, J. and J. Züber (1999): Goodness-of-fit tests for nonlinear heteroscedastic regression models, Statistics & Probability Letters 42 (1999), 53-60
Dieck, T. (1987): Transformation groups, W de Gruyter, Berlin/New York 1987
Diggle, P.J., Liang, K.Y. and S.L. Zeger (1994): Analysis of longitudinal data, Clarendon Press, Oxford 1994
Diggle, P., Lophaven S. (2006): Bayesian geostatistical design. Scand. J. Stat. 33:53-64

Ding, C.G. (1999): An efficient algorithm for computing quantiles of the noncentral chi-squared distribution, Computational Statistics & Data Analysis 29 (1999), 253-259

Dixon, A.L. (1908): The elimination of three quantics in two independent variables, *Proc. London Mathematical Society series 2* 6 (1908) 468-478.

Dixon, G.M. (1994): Division Algebras: Octonions, Quaternions, Complex Numbers and the Algebraic Design of Physics, Kluwer Aca. Publ., Dordrecht-Boston-London 1994

Dixon, W.J. (1951): Ratio involving extreme values, Ann. Math. Statistics 22 (1951), 68-78

Dobbie, M.J. and Henderson, B.L. and Stevens, D.L. Jr. (2008): Sparse sampling: spatial design for monitoring stream networks. Stat. Surv. 2: 113-153

Dobson, A.J. (1990): An introduction to generalized linear models, Chapman and Hall, Boca Raton 1990

Dobson, A.J. (2002): An introduction to generalized linear models, 2nd ed., Chapman/Hall/CRC, Boca Raton 2002

Dodge, Y. (1987): Statistical data analysis based on the L1-norm and related methods, Elsevier, Amsterdam 1987

Dodge, Y. (1997): LAD Regression for Detecting Outliers in Response and Explanatory Variables, J. Multivar. Anal. 61 (1997), 144-158

Dodge, Y. and D. Majumdar (1979): An algorithm for finding least square generalized inverses for classification models with arbitrary patterns, J. Statist. Comput. Simul. 9 (1979), 1-17

Dodge, Y. and J. Jureckovèa (1997): Adaptive choice of trimming proportion in trimmed least-squares estimation, Statistics & Probability Letters 33 (1997), 167-176

Donoho, D.L. and P.J. Hüber (1983): The notion of breakdown point, Festschrift für Erich L. Lehmann, eds. P.J. Bickel, K.A. Doksum and J.L. Hodges, Wadsworth, Belmont, Calif. 157-184, 1983

Dorea, C.C.Y. (1997): L1-convergence of a class of algorithms for global optimization, Student 2 (1997)

Dorrie, H. (1948): Kubische und biquadratische Gleichungen. Leibniz-Verlag, Miinchen

Downs, T.D. and A.L. Gould (1967): Some relationships between the normal and von Mises distributions, Biometrika 54 (1967), 684-687

Draper, N.R. and R. Craig van Nostrand (1979): Ridge regression and James-Stein estimation: review and comments, Technometrics 21 (1979), 451-466

Draper, N.R. and J.A. John (1981): Influential observations and outliers in regression, Technometrics 23 (1981), 21-26

Draper, N.R. and F. Pukelsheim (1996): An overview of design of experiments, Statistical Papers 37 (1996), 1-32

Draper, N.R. and F. Pukelsheim (1998): Mixture models based on homogenous polynomials, J. Stat., Planning and Inference, 71 (1998), pp. 303-311

Draper, N.R. and F. Pukelsheim (2002): Generalized Ridge Analysis Under Linear Restrictions, With Particular Applications to Mixture Experiments Problems, Technometrics, 44, 250-259.

Draper, N.R. and F. Pukelsheim (2003): Canonical reduction of second order fitted models subject to linear restrictions, Statistics and Probability Letters, 63 (2003), pp. 401-410

Drauschke, M., Forstner, W. and Brunn, A. (2009): Multidodging: Ein effizienter Algorithmus zur automatischen Verbesserung von digitalisierten Luftbildern. Seyfert, Eckhardt (Hg.): Publikationen der DGPF, Band 18: Zukunft mit Tradition. Jena 2009, S. 61-68.

Drauschke, M. and Roscher, R. and Labe, T. and Forstner, W. (2009): Improving Image Segmentation using Multiple View Analysis. Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluatin (CMRT09). 2009, S. 211-216.

Driscoll, M.F. (1999): An improved result relating quadratic forms and Chi-Square Distributions, The American Statistician 53 (1999), 273-275

Driscoll, M.F. and B. Krasnicka (1995): An accessible proof of Craig's theorem in the general case, The American Statistician 49 (1995), 59-62

Droge, B. (1998): Minimax regret analysis of orthogonal series regression estimation: selection versus shrinkage, Biometrika 85 (1998), 631-643

Dryden, I.L. (1998): General registration and shape analysis. STAT 96/03. Stochastic Geometry: Likelihood and Computation, by W.S. Kendell, O. Barndoff-Nielsen and M.N.N. van Lieshout, Chapman and Hall/CRC Press, 333-364, Boca Raton 1998.
Drygas, H. (1970): The coordinate-free approach to Gauss-Markov estimation. SpringerVerlag, Berlin-Heidelberg-New York, 1970.
Drygas, H. (1972): The estimation of residual variance in regression analysis, Math. Operations Forsch. Atatist., 3 (1972), pp. 373-388
Drygas, H. (1975a): Weak and strong consistency of the least squares estimators in regression models. Z. Wahrscheinlichkeitstheor. Verw. Geb. 34 (1975), 119-127.
Drygas, H. (1975b): Estimation and prediction for linear models in general spaces, Math. Operationsforsch. Statistik 6 (1975), 301-324
Drygas, H. (1977): Best quadratic unbiased estimation in variance-covariance component models, Math. Operationsforsch. Statistik, 8 (1977), pp. 211-231
Drygas, H. (1983): Sufficiency and completeness in the general Gauss-Markov model, Sankhya Ser. A45 (1983), 88-98
Duan, J.C. (1997): Augmented GARCH (p,q) process and its diffusion limit, J. Econometrics 79 (1997), 97-127
Duda, R.O., Hart, P.E. and Stork, D.G. (2000): Pattern classification, 2ed., Wiley, New York, 2000
Duncan, W.J. (1944): Some devices for the solution of large sets of simultaneous linear equations, London, Edinburgh and Dublin Philosophical Magazine and Journal of Science (7th series) 35 (1944), 660-670
Dunfour, J.M. (1986): Bias of s2 in linear regression with dependent errors, The American Statistician 40 (1986), 284-285
Dunnett, C.W. and M. Sobel (1954): A bivariate generalization of Student's t-distribution, with tables for certain special cases, Biometrika 41 (1954), 153-69
Dupuis, D.J., Hamilton, D.C. (2000): Regression residuals and test statistics: Assessing naive outlier deletion. Can. J. Stat. 28, 2 (2000), 259-275
Durand, D. and J.A. Greenwood (1957): Random unit vectors II : usefulness of Gram-Charlier and related series in approximating distributions, Ann. Math. Statist. 28 (1957), 978-986
Durbin, J. and G.S. Watson (1950): Testing for serial correlation in least squares regression, Biometrika 37 (1950), 409-428
Durbin, J. and G.S. Watson (1951): Testing for serial correlation in least squares regression II, Biometrika 38 (1951), 159-177
Dvorak, G. and Benveniste, Y. (1992): On transformation strains and uniform fields in multiphase elastic media, Proc. R. Soc. Lond., A437, 291-310
Ecker, E. (1977): Ausgleichung nach der Methode der kleinsten Quadrate, Öst. Z. Vermessungswesen 64 (1977), 41-53
Eckert, M (1935): Eine neue ächentrene (azimutale) Erdkarte, Petermann's Mitteilungen 81 (1935), 190-192
Eckhart, C. and G. Young (1939): A principal axis transformation for non-Hermitean matrices, Bull. Amer. Math. Soc. 45 (1939), 188-121
Eckl, M.C., Snay, R.A., Solder, T., Cline, M.W. and G.L. Mader (2001): Accuracy of GPS-derived positions as a function of interstation distance and observing-session duration, Journal of Geodesy 75 (2001), 633-640
Edelman, A. (1989): Eigenvalues and condition numbers of random matrices, PhD dissertation, Massachussetts Institute of Technology 1989
Edelman, A. (1998): The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20 (1998), 303-353
Edelman, A., Elmroth, E. and B. Kagström (1997): A geometric approach to perturbation theory of matrices and matrix pencils. Part I: Versal deformations, SIAM J. Matrix Anal. Appl. 18 (1997), 653-692
Edelman, A., Arias, T.A. and Smith, S.T. (1998): The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20 (1998), 303-353

Edgar, G.A. (1998): Integral, probability, and fractal measures, Springer-Verlag, Heidelberg Berlin New York 1998

Edgeworth, F.Y. (1883): The law of error, Philosophical Magazine 16 (1883), 300-309

Eeg, J. and T. Krarup (1973): Integrated geodesy, Danish Geodetic Institute, Report No. 7, Copenhagen 1973

Effros, E.G. (1997): Dimensions and C* algebras, Regional Conference Series in Mathematics 46, Rhode Island 1997

Efron, B. and R.J. Tibshirani (1994): An introduction to the bootstrap, Chapman and Hall, Boca Raton 1994

Eibassiouni, M.Y. and J. Seely (1980): Optimal tests for certain functions of the parameters in a covariance matrix with linear structure, Sankya 42 (1980), 64-77

Eichhorn, A. (2005): Ein Beitrag zur parametrischen Identifikation von dynamischen Strukturmodellen mit Methoden der adaptiven Kalman Filterung, Deutsche Geodätische Kommission, Bayerische Akademic der Wissenschaften, Reihe C 585, München 2005

Einstein, A. (1905): Ueber die von der molekularkinetische Theorie der Waerme geforderte Berechnung von in Fluessigkeiten suspendierten Teilchen, Ann. Phys. 17 (1905) 549-560

Ekblom, S. and S. Henriksson (1969): Lp-criteria for the estimation of location parameters, SIAM J. Appl. Math. 17 (1969), 1130-1141

Elden, L. (1977): Algorithms for the regularization of ill-conditioned least squares problems, 17 (1977), 134-145

Elhay, S., Golub, G.H. and J. Kautsky (1991): Updating and downdating of orthogonal polynomials with data fitting applications, SIAM J. Matrix Anal. Appl. 12 (1991), 327-353

Elian, S.N. (2000): Simple forms of the best linear unbiased predictor in the general linear regression model, American Statistician 54 (2000), 25-28

Ellis, R.L. and I. Gohberg (2003): Orthogonal systems and convolution operators, Birkhäuser-Verlag, Basel Boston Berlin 2003

Ellis, S.P. (1998): Instability of least squares, least absolute correlation and least median of squared linear regression, Stst. Sci., 13 (1998), pp. 337-350

Ellis, S.P. (2000): Singularity and outliers in linear regression with application to least squares, least absolute deviation, and least median of squares linear regression. Metron 58, 1-2 (2000), 121-129.

Ellenberg, J.H. (1973): The joint distribution of the standardized least squares residuals from a general linear regression, J. Am. Statist. Ass. 68 (1973), 941-943

Elpelt, B. (1989): On linear statistical models of commutative quadratic type, Commun. Statist.-Theory Method 18 (1989), 3407-3450

El-Basssiouni, M.Y. and Seely, J. (1980): Optimal tests for certain functions of the parameters in a covariance matrix with linear structure, Sankya A42 (1980), 64-77

Elfving, G. (1952): Optimum allocation in linear regression theory, Ann. Math. Stat. 23 (1952), 255-263

El-Sayed, S.M. (1996): The sampling distribution of ridge parameter estimator, Egyptian Statistical Journal, ISSR - Cairo University 40 (1996), 211-219

Embree, P.J., Burg, P. and Backus, M.M. (1963): Wide band velocity filtering - the pie slid process, Geophysics 28 (1963), 948-974

Engeln-Mullges, G. and Reutter, F. (1990): Formelsammlung zur numerischen Mathematik mit C-Programmen. Bibliographisches Institut Mannheim & F. A. Brockhaus AG

Engl, H.W., Hanke, M. and A. Neubauer (1996): Regularization of inverse problems, Kluwer Academic Publishers, Dordrecht 1996

Engl, H.W., A.K. Louis, and W. Rundell (1997): Inverse problems in geophysical applications, SIAM, Philadelphia 1997

Engler, K., Grafarend, E.W., Teunissen, P. and J. Zaiser (1982): Test computations of three-dimensional geodetic networks with observables in geometry and gravity space, Proceedings of the International Symposium on Geodetic Networks and Computations. Vol. VII, 119-141 Report B258/VII. Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, München 1982.

Eringen, A.C. (1962): Non linear theory of continuous media, McGraw Hill Book Comp., New York
Ernst, M.D. (1998): A multivariate generalized Laplace distribution, Computational Statistics 13 (1998), 227-232
Eshagh, M. and Sjöberg (2008): The modified best quadratic unbiased non-negative estimator (MBQUNE) of variance components. Stud. Geophysics Geod. 52 (2008) 305-320
Ethro, U. (1991): Statistical Test of Significance for Testing Outlying Observations. Survey Review. 31:62-70.
Euler, N. and W.H. Steeb (1992): Continuous symmetry, Lie algebras and differential equations, B.I. Wissenschaftsverlag, Mannheim 1992
Even Tzur, G. (2004): Variance Factor Estimation for Two-Step Analysis of Deformation Networks. Journal of surveying engineering. Journal of Surveying Engineering, 130(3): 113-118.
Even Tzur, G. (2006): Datuln Definition and its Influence on the Reliability of Geodetic Networks. ZN, 131: 87-95.
Even Tzur, G. (2009): Two-Steps Analysis of Movement of the Kfar-Hanassi Network. FIG Working week 2009, Eilat, Israel.
Everitt, B.S. (1987): Introduction to optimization methods and their application in statistics, Chapman and Hall, London 1987
Fahrmeir, L. and G. Tutz (2001): Multivariate statistical modelling based on generalized linear models, Springer-Verlag, Heidelberg Berlin New York 2001
Fakeev, A.G. (1981): A class of iterative processes for solving degenerate systems of linear algebraic equations, USSR. Comp. Maths. Math. Phys. 21 (1981), 15-22
Falk, M., Hüsler, J. and R.D. Reiss (1994): Law of small numbers, extremes and rare events, Birkhäuser-Verlag, Basel Boston Berlin 1994
Fan, J. and I. Gijbels (1996): Local polynomial modelling and its applications, Chapman and Hall, Boca Raton 1996
Fang, K.-T. and Y. Wang (1993): Number-theoretic methods in statistics, Chapman and Hall, Boca Raton 1993
Fang, K.-T. and Y.-T. Zhang (1990): Generalized multivariate analysis, Science Press Beijing, Springer Verlag, Bejing, Berlin New York 1990
Fang, K.-T., Kotz, S. and K.W. Ng (1990): Symmetric multivariate and related distributions, Chapman and Hall, London 1990
Farahmand, K. (1996): Random polynomials with complex coefficients, Statistics & Probability Letters 27 (1996), 347-355
Farahmand, K. (1999): On random algebraic polynomials, Proceedings of the American Math. Soc. 127 (1999), 3339-3344
Faraway, J. (2006). Extending the linear model with R. Chapman and Hall, CRE.
Farebrother, R.W. (1988): Linear least squares computations, Dekker, New York 1988
Farebrother, R.W. (1999): Fitting linear relationships, Springer-Verlag, Heidelberg Berlin New York 1999
Farrel, R.H. (1964): Estimators of a location parameter in the absolutely continuous case, Ann. Math. Statist. 35 (1964), 949-998
Fasséo, A. (1997): On a rank test for autoregressive conditional heteroscedasticity, Student 2 (1997), 85-94
Faulkenberry, G.D. (1973): A method of obtaining prediction intervals, J. Am. Statist. Ass. 68 (1973), 433-435
Fausett, D.W. and C.T. Fulton (1994): Large least squares problems involving Kronecker products, SIAM J. Matrix Anal. Appl. 15 (1994), 219-227
Fazekas, I. and Kukush, A. S. Zwanzig (2004): Correction of nonlinear orthogonal regression estimator. Ukr. Math. J. 56, 1101-1118
Fedorov, V.V. (1972): Theory of Optimal Experiments. Academic Press
Fedorov, V.V. and Hackl, P. (1994): Optimal experimental design: spatial sampling. Calcutta Stat Assoc Bull 44:173-194

Federov, V.V. and Müller, W.G. (1988): Two approaches in optimization of observing networks, in Optimal Design and Analysis of Experiments, Y. Dodge, V.V. Federov & H.P. Wynn, eds, North-Holland, Amsterdam, pp. 239-256

Fedorov, V.V. and Flagnagan, D. (1997): Optimal monitoring on Merer‘s expansion of covariance kernel, J. Comb. Inf. Syst. Sci. 23(1997) 237-250

Fedorov, V.V. and P. Hackl (1997): Model-oriented design of experiments, Springer-Verlag, Heidelberg Berlin New York 1997

Fedorov, V.V., Montepiedra, G. and C.J. Nachtsheim (1999): Design of experiments for locally weighted regression, J. Statist. Planning and Inference 81 (1999), 363-382

Fedorov, V.V. and Müller, W.C. (2007): Optimum designhn of correlated fields via covariance kernel expansion, in: Lopez Fidalgo, Rodriguez, J. and Forney, B. (eds) Physica Verlag Heidelberg

Feinstein, A.R. (1996): Multivariate analysis, Yale University Press, New Haven 1996

Feldmann, R.V. (1962): Matrix theory I: Arthur Cayley - founder of matrix theory, Mathematics teacher, 57, (1962), pp. 482-484

Felus, Y.A. and R.C. Burtch (2005): A new procrustes algorithm for three dimensional datum conversion, Ferris State University 2005

Fengler, M., Freeden, W. and V. Michel (2003): The Kaiserslautern multiscale geopotential model SWITCH-03 from orbit pertubations of the satellite CHAMP and its comparison to the models EGM96, UCPH2002-02-0.5, EIGEN-1s, and EIGEN-2, Geophysical Journal International (submitted) 2003

Fenyo, S. and H.W. Stolle (1984): Theorie und Praxis der linearen Integralgleichungen 4. Berlin

Feuerverger, A. and P. Hall (1998): On statistical inference based on record values, Extremes 1:2 (1998), 169-190

Fiebig, D.G., Bartels, R. and W.Krämer (1996): The Frisch-Waugh theorem and generalized least squares, Econometric Reviews 15 (1996), 431-443

Fierro, R.D. (1996): Pertubation analysis for twp-sided (or complete) orthogonal decompositions, SIAM J. Matrix Anal. Appl. 17 (1996), 383-400

Fierro, R.D. and J.R. Bunch (1995): Bounding the subspaces from rank revealing two-sided orthogonal decompositions, SIAM J. Matrix Anal. Appl. 16 (1995), 743-759

Fierro, R.D. and P.C. Hansen (1995): Accuracy of TSVD solutions computed from rank-revealing decompositions, Numer. Math. 70 (1995), 453-471

Fierro, R.D. and P.C. Hansen (1997): Low-rank revealing UTV decompositions, Numer. Algorithms 15 (1997), 37-55

Fill, J.A. and D.E. Fishkind (1999): The Moore-Penrose generalized inverse for sums of matrices, SIAM J. Matrix Anal. Appl. 21 (1999), 629-635

Finsterwalder, S. and Scheufele, W. (1937): Das Rückwartseinschneiden im Raum, Sebastian Finsterwalder zum 75 Geburtstage, pp. 86-100, Verlag Hebert Wichmann, Berlin 1937.

Fischer, F.A. (1969): Einfuhrung in die Statische Übertragungstheorie, Bib. Institute, Maurheim, Zurigh 1969

Fischler, M.A. and Bolles, R.C. (1981): Random Sample Consensus: A Paradigm for Medell Fitting with Application to Image Analysis and Automated Cartography, *Communications of the ACM* 24 (1981) 381-395

Fiserova, E. (2004): Estimation in universal models with restrictions. Discuss. Math., Probab. Stat. 24, 2 (2004), 233-253.

Fiserova, E. (2006): Testing hypotheses in universal models. Discuss. Math., Probab. Stat. 26 (2006), 127-149

Fiserova, E. and Kubacek, L. (2003) Sensitivity analysis in singular mixed linear models with constraints. Kybernetika 39, 3 (2003), 317-332

Fiserova, E. and Kubacek, L. (2004): Statistical problems of measurement in triangle. Folia Fac. Sci. Nat. Univ. Masarykianae Brunensis, Math. 15 (2004), 77-94

Fiserova, E. and Kubacek, L. (2003): Insensitivity regions and outliers in mixed models with constraints. Austrian Journal of Statistic 35, 2-3 (2003), 245-252

Fisher, N.I. (1993): Statistical Analysis of Circular Data, Cambridge University Press, Cambridge 1993

Fisher, N.I. and Hall, P. (1989): Bootstrap confidence regions for directional data, J. Am. Statist. Ass. 84 (1989), 996-1002

Fisher, N.J. (1985): Spherical medians, J. Roy. Statist. Soc. B47 (1985), 342-348

Fisher, N.J. and A.J. Lee (1983): A correlation coefficient for circular data, Biometrika 70 (1983), 327-332

Fisher, N.J. and A.J. Lee (1986): Correlation coefficients for random variables on a sphere or hypersphere, Biometrika 73 (1986), 159-164

Fisher, R.A. (1915): Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika 10 (1915), 507-521

Fisher, R.A. (1935): The fiducial argument in statistical inference, Annals of Eugenics 6 (1935), 391-398

Fisher, R.A. (1939): The sampling distribution of some statistics obtained from nonlinear equations, Ann. Eugen. 9 (1939), 238-249

Fisher, R.A. (1953): Dispersion on a sphere, Pro. Roy. Soc. Lond. A217 (1953), 295-305

Fisher, R.A. (1956): Statistical Methods and Scientific Inference, Edinburg: Oliver and Boyd, 1956

Fisher, R.A. and F. Yates (1942): Statistical tables for biological, agricultural and medical research, 2nd edition, Oliver and Boyd, Edinburgh 1942

Fisz, M. (1971): Wahrscheinlichkeitsrechnung und mathematische Statistik, Berlin (Ost) 1971

Fitzgerald, W.J., Smith, R.L., Walden, A.T. and P.C. Young (2001): Non-linear and nonstationary signal processing, Cambridge University Press, Cambridge 2001

Fitzgibbon, A., Pilu, M. and Fisher, R.B. (1999): Direct least squares fitting of ellipses, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999) 476-480.

Fletcher, R. and C.M. Reeves (1964): Function minimization by conjugate gradients, Comput. J. 7 (1964) 149-154

Fletcher, R. (1987): Practical Methods of Optimization. Second Edition, John Wiley & Sons, Chicester.

Fletling, R. (2007): Fuzzy clusterung zur Analyse von Überwachungsmessungen, in: F.K. Brunner (ed.) Ingenienr vermessung, pp. 311-316, Wichmann Verlag, Heidelberg 2007

Flury, B. (1997): A first course in multivariate statistics, Springer-Verlag, Heidelberg Berlin New York 1997

Focke, J. and G. Dewess (1972): Über die Schätzmethode MINQUE von C.R. Rao und ihre Verallgemeinerung, Math. Operationsforschg. Statistik 3 (1972), 129-143

Förstner, W. (1976): Statistical Test Methods for Blunder Detection in Planimetric Block Triangulation. Presented paper to Comm. III, ISP Helsinki 1976

Förstner, W. (1978a): PrOfung auf grobe Bildkoordinatenfehler bei der relativen Orientierung? BuL, 1978, 205-212

Förstner, W. (1978b): Die Suche nach groben Fehlern in photogrammetrischen Lageblocken, Dissertation der Universitat Stuttgart .

Förstner, W. (1979a): Ein Verfahren zur Schatzung von Varianz- und Kovarianzkomponenten, Ally. Vermess. Nachrichten 86 (1979), pp. 446-453

Förstner, W. (1979b): On Internal and External Reliability of Photogrammetric Coordinates. Proceedings of the Annual Convention of the ASP-ACSM, Washington 1979

Förstner, W. (1979c): Konvergenzbeschleunigung bei der a posteriori Varianzschätzung, Zeitschrift für Vermessungswesen 104 (1979), 149-156

Förstner, W. (1979d): Ein Verfahren zur Schätzung von Varianz- und Kovarianz-Komponenten, Allgemeine Vermessungsnachrichten 86 (1979), 446-453

Förstner, W. (1980a): The Theoretical Reliability of Photogrammetric Coordinates. Archives of the ISPRS. XIV Congress of the Int. Society for Photogrammetry Comm. 111, Hamburg, 1980

Förstner, W. (1980b): Zur PrOfung zusatzlicher Parameter in Ausgleichungen. ZN, 1980, 105, 510-519

Förstner, W. (1983): Reliability and Discernability of Extended Gauss-Markov Models. Ackermann, F. (ed.) Deutsche Geodätische Kommission, A98, 1983, 79-103

Förstner, W. (1985): Determination of the Additive Noise Variance in Observed Autoregressive Processes using Variance Component Estimation Technique. Statistics and Decision, Supplement Issue No. 2. München 1985, S. 263-274.
Förstner, W. (1987): Reliability Analysis of Parameter Estimation in Linear Models with Applications to Mensuration Problems in Computer Vision. CVGIP - Computer Vision, Graphics, and Image Processing. 1987, S. 273-310.
Förstner, W. (1990): Modelle intelligenter Bildsensoren und ihre Qualität. Grol1 Kopf, R.E. (ed.) Springer, 1990, 254, 1-21
Förstner, W. (1993a): Image Matching. In: Haralick, R.M. I Shapiro, L.G. (Hg.): Computer and -RObot Vision. Addison-Wesley, S. 289-379
Förstner, W. (1993c): Uncertain Spatial Relationships and their Use for Object location in Digital/mages. Forstner, W.; Haralick, R.M. Radig, B. (ed.)lnstitut für Photogrammetrie, Universitat Bonn, 1992 (Buchbeitrag)
Förstner, W. (1994a): Diagnostics and Performance Evaluation in Computer Vision. Performance versus Methodology in Computer Vision, NSF/ARPA Workshop. Seattle 1994, S. 11-25.
Förstner, W. (1994b): Generic Estimation Procedures for Orientation with Minimum and Redundant Information Technical Report. Institute for Photogrammetry, University Bonn.
Förstner, W. (1995a): A 2D-data Model for Representing Buildings. Technical Report, Institut für Photogrammetrie, Universitat Bonn
Förstner, W. (1995b): The Role of Robustness in Computer Vision. Proc. Workshop "Milestones in CCfmi5uier Vision". Vorau 1995.
Förstner, W. (1997a):Evaluation of Restoration Filters. Technical Report, Institut für Photogrammetrie, Universitat Bonn
Förstner, W. (1997b): Parameter free information Preserving filter and Noise Variance Equalization, Technical Report, Institut für Photogrammetrie, Universitat Bonn
Förstner, W. (1998): On the Theoretical Accuracy of Multi Image Matching, Restoration and Triangulation. Festschrift zum 65. Geburtstag von Prof. Dr.-Ing. mult. G. Konecny. Institut für Photogrammetrie, Universitat Hannover 1998
Förstner, W. (1999a): On Estimating Rotations. Heipke, C. / Mayer, H. (Hg.): Festschrift für Prof. Dr.-Ing. Heinrich Ebner zum 60. Geburtstag. Lehrstuhl für Photogrammetrie und Fernerkundung, TU München 1999.
Förstner, W. (1999b): Uncertain Neighborhood Relations of Point Sets and Fuzzy Delaunay Triangulation. Proceedings of DAGM Symposium Mustererkennung. Bonn, Germany 1999.
Förstner, W. (1999c): Choosing Constraints for Points and Lines in Image Triplets 1999. Technical Report, Institut für Photogrammetrie, University Bonn.
Förstner, W. (1999d): Interior Orientation from Three Vanishing Points, Technical Report, Institut für Photogrammetrie, Universitat Bonn
Förstner, W. (1999e): Tensor-based Distance Transform. Technical Report, Institut für Photogrammetrie, Universitat Bonn.
Förstner, W. (2000): Optimally Reconstructing the Geometry of Image Triplets. Vernon, David (Hg.): Appeared in: Computer Vision - ECCV 2000. 2000, S. 669-684.
Förstner, W. (2001a): Algebraic Projective Geometry and Direct Optimal Estimation of Geometric Entities. Appeared at the Annual OeAGM 2001 Workshop.
Förstner, W. (2001b): Direct optimal estimation of geometric entities using algebraic projective geometry. Festschrift anlasslich des 60. Geburtstages von Prof. Dr.-Ing. Bernhard Wrobel. Technische Universitat Darmstadt 2001, S. 69-87.
Förstner, W. (2002): Computer Vision and Photogrammetry – Mutual Questions: Geometry, Statistics and Cognition. Bildteknik/lmage Science, Swedish Society for Photogrammetry and Remote Sensing. 2002, S. 151-164.
Förstner, W. (2003): Notions of Scale in Geosciences. Neugebauer, Horst J. I Simmer, Clemens (Hg.): Dynamics of Multi-Scale Earth Systems. 2003, S. 17-39.
Förstner, W. (2005): Uncertainty and Projective Geometry. Bayro Corrochano, Eduardo (Hg.): Handbook of Geometric Computing. 2005, S. 493-535.

Förstner, W. (2009): Computer Vision and Remote Sensing - Lessons Learned. Fritsch, Dieter (Hg.): Photogrammetric Week 2009. Heidelberg 2009, S. 241-249.

Förstner, W. and Schroth, R. (1981): On the estimation of covariance matrices for photogrammetric image coordinates. VI: Adjustment procedures of geodetic networks, 43-70.

Förstner, W. and Wrobel, B. (1986): Digitale Photogrammetrie, Springer (Buch)

Förstner, W. and Gulch, E. (1987): A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. Proceedings of the ISPRS Intercommission Workshop, 1987. Jg. 1987.

Förstner, W. and Vosselmann, G. (1988): The Precision of a Digital Camera. ISPRS 16th Congress. Kyoto 1988, S. 148-157.

Förstner, W. and Braun, C. (1993): Direkte Schatzung von Richtungen und Rotationen. Institut für Photogrammetrie, Bonn

Förstner, W. and Shao, J. (1994): Gabor Wavelets for Texture Edge Extraction. ISPRS Commission III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision. Munich, Germany

Förstner, W. and Brugelmann, R. (1994): ENOVA - Estimation of Signal Dependent Noise Variance. 3rd ECCV, Stockholm

Förstner, W. and Moonen, B. (1999): A Metric for Covariance Matrices. Krumm, F. Schwarze, V.S. (Hg.): Quo vadis geodesia ...?, Festschrift for Erik W. Grafarend on the occasion of his 60th birthday. TR Dept. of Geodesy and Geoinformatics, Stuttgart University 1999.

Förstner, W. and B. Moonen (2003): A metric for covariance matrices, in: E. Grafarend, F. Krumm and V. Schwarze: Geodesy - the Challenge of the 3rd Millenium, 299-309, Springer-Verlag, Heidelberg Berlin New York 2003

Förstner, W. and Dickscheid, T. and Schindler, F. (2009): Detecting Interpretable and Accurate Scale-Invariant Keypoint. 12th IEEE International Conference on Computer Vision (ICCV'09). Kyoto, Japan 2009, S. 2256-2263.

Forsgren, A. and W. Murray (1997): Newton methods for large-scale linear inequality-constrained minimization, Siam J. Optim. to appear

Forsythe, A.R. (1960): Calculus of Variations. New York

Forsythe, A.B. (1972): Robust estimation of straight line regression coefficients by minimizing p-th power deviations, Technometrics 14 (1972), 159-166

Foster, L.V. (2003): Solving rank-deficient and ill-posed problems using UTV and QR factorizations, SIAM J. Matrix Anal. Appl. 25 (2003), 582-600

Fotiou, A. and D. Rossikopoulos (1993): Adjustment, variance component estimation and testing with the affine and similarity transformations, Zeitschrift für Vermessungswesen 118 (1993), 494-503

Fotiou, A.(1998): A pair of closed expressions to transform geocentric to geodetic coordinates, *Zeitschrift für Vermessungswesen* 123 (1998) 133-135.

Fotopoulos, G. (2003): An Analysis on the Optimal Combination of Geoid, Orthometric and Ellipsoidal Height Data. Report No. 20185, Department of Geomatics Engineering, University of Calgary, Calgary, Canada.

Fotopoulos, G., (2005): Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetrical geoid height data. 1. Geodesy, 79, 111-123.

Forstner, W., (1979): Ein Verfahren zur Achatzung von varianz und kovarianzkomponenten. Allgemeine Vermessungs-Nachrichten, 86, 446-453.

Foucart, T. (1999): Stability of the inverse correlation matrix. Partial ridge regression, J. Statist. Planning and Inference 77 (1999), 141-154

Fox, A.J. (1972): Outliers in time series. J. Royal Stat. Soc., Series B, 43, (1972), 350-363

Fox, M. and H. Rubin (1964): Admissibility of quantile estimates of a single location parameter, Ann. Math. Statist. 35 (1964), 1019-1031

Franses, P.H. (1998): Time series models for business and economic forecasting, Cambridge University Press, Cambridge 1998

Fraser, D.A.S. (1963): On sufficiency and the exponential family, J. Roy. Statist. Soc. 25 (1963), 115-123

Fraser, D.A.S. (1968): The structure of inference, J. Wiley, New York 1968

Fraser, D.A.S. and I. Guttman (1963): Tolerance regions, Ann. Math. Statist. 27 (1957), 162-179

Freeman, R.A. and P.V. Kokotovic (1996): Robust nonlinear control design, Birkhäuser-Verlag, Basel Boston Berlin 1996

Freund, P.G.O. (1974): Local scale invariance and gravitation, Annals of Physics 84 (1974), 440-454

Frey, M. and J.C. Kern (1997): The Pitman Closeness of a Class of Scaled Estimators, The American Statistician, May 1997, Vol. 51 (1997), 151-154

Frisch, U. (1995): Turbulence, the legacy of A.N. Kolmogorov, Cambridge University Press, Cambridge 1995

Fristedt, B. and L. Gray (1995): A modern approach to probability theory, Birkhäuser-Verlag, Basel Boston Berlin 1997

Frobenius, F.G. (1893): Gedächtnisrede auf Leopold Kronecker (1893), Ferdinand Georg Frobenius, Gesammelte Abhandlungen, ed. J.S. Serre, Band III, 705-724, Springer-Verlag, Heidelberg Berlin New York 1968

Frobenius, G. (1908): Über Matrizen aus positiven Elementen, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften von Berlin, 471-476, Berlin 1908

Fröhlich, H. and Hansen, H.H. (1976): Zur Lotfußpunktrechnung bei rotationsellipsoidischer Bezugsfläche, *Allgemeine Vermessungs-Nachrichten* 83 (1976) 175-179.

Fuentes, M. (2002): lnterpolation of non stationary air pollution processes: a spatial spectral approach. In statistical Modeling, vol. 2, pp. 281-298.

Fuentes, M. and Smith, R. (2001): A new class of non stationary models. In Tech. report at North Carolina State University, Institute of Statistics

Fuentes, M., Chaudhuri, A. and Holland, D.M. (2007): Bayesian entropy for spatial sampling design of environmental data. J. Environ. Ecol. Stat. 14:323-340

Fujikoshi, Y. (1980): Asymptotic expansions for the distributions of sample roots under non-normality, Biometrika 67 (1980), 45-51

Fukushima, T. (1999): Fast transform from geocentric to geodetic coordinates, *Journal of Geodesy* 73 (1999) 603-610.

Fuller, W.A. (1987): Measurement error models, Wiley, New York 1987

Fulton, T., Rohrlich, F. and L. Witten (1962): Conformal invariance in physics, Reviews of Modern Physics 34 (1962), 442-457

Furno, M. (1997): A robust heteroskedasticity consistent covariance matrix estimator, Statistics 30 (1997), 201-219

Galil, Z. (1985): Computing d-optimum weighing designs: Where statistics, combinatorics, and computation meet, in: Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II, eds. L.M. LeCam and R.A. Olshen, Wadsworth 1985

Gallant, A.R. (1987): Nonlinear statistical models, J. Wiley, New York 1987

Gallavotti, G. (1999): Statistical mechanics: A short treatise, Springer-Verlag, Heidelberg Berlin New York 1999

Gander, W. (1981): Least squares with a quadratic constraint, Numer. Math. 36 (1981), 291-307

Gander, W., Golub, G.H. and Strebel, R. (1994): Least-Squares fitting of circles and ellipses, *BIT* No. 43 (1994) 558-578.

Gandin, L.S. (1963): Objective analysis of meteorological fields. Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad

Gao, X.S. and Chou, S.C. (1990): Implicitization of rational parametric equations. Technical report, Department of Computer Science TR-90-34, University of Texas, Austin 1990.

Gao, S. and T.M.F. Smith (1995): On the nonexistence of a global nonengative minimum bias invariant quadratic estimator of variance components, Statistics and Probability letters 25 (1995), 117-120

Gao, S. and T.M.F. Smith (1998): A constrained MINQU estimator of correlated response variance from unbalanced dara in complex surveys, Statistica Sinica 8 (1998), 1175-1188

Gao, Y., Lahaye, F., Heroux, P., Liao, X., Beck, N. and M. Olynik (2001): Modeling and estimation of C1-P1 bias in GPS receivers, Journal of Geodesy 74 (2001), 621-626

Garcèìa-Escudero, L.A., Gordaliza, A. and C. Matrèan (1997): k-medians and trimmed k-medians, Student 2 (1997), 139-148

Garcia-Ligero, M.J., Hermoso, A. and J. Linares (1998): Least squared estimation for distributed parameter systems with uncertain observations: Part 1: Linear prediction and filtering, Applied Stochastic Models and Data Analysis 14 (1998), 11-18

Garderen K.J. van, (1999): Exact geometry of autoregressive models, Journal of Time Series Analysis 20 (1999), 1-21

Gaspari, G. and Cohn, S.E. (1999): Construction of correlation I in two and three dimensions, Q. 1. R. Meteorol. Soc., 757, 1999.

Gastwirth, J.L. and H. Rubin (1975): The behaviour of robust estimators on dependent data, Ann. Stat., 3 (1975), pp. 1070-1100

Gatti, M. (2004): An empirical method of estimation of the variance-covariance matrix in GPS network design, Survey Review 37, (2004), pp. 531-541

Gauss, C.F. (1809): Theoria Motus, Corporum Coelesium, Lib. 2, Sec. III, Perthes u. Besser Publ., 205-224, Hamburg 1809

Gauss, C.F. (1816): Bestimmung der Genauigkeit der Beobachtungen, Z. Astronomi 1 (1816), 185-197

Gauss, C.F. (1923): Anwendung der Wahrscheinlichkeitsrechnung auf eine Aufgabe der praktischen Geometrie, Gauss Works IV: 1923

Gautschi, W. (1982): On generating orthogonal polynomials, SIAM Journal on Scientific and Statistical Computing 3 (1982), 289-317

Gautschi, W. (1985): Orthogonal polynomials - constructive theory and applications, J. Comput. Appl. Math. 12/13 (1985), 61-76

Gautschi, W. (1997): Numerical analysis - an introduction, Birkhäuser-Verlag, Basel Boston Berlin 1997

Gauss, C.F. (1809): Theoria motus corporum coelestium. Gattingen

Geisser, S. (1975): The predictive sample reuse method with applications. J. Math. Statist. 28: 385-394

Gelb, A. (1974): Applied optimal estimation, Ed., MIT Press, Cambridge/Mass and London

Gelfand, A.E. and D.K. Dey (1988): Improved estimation of the disturbance variance in a linear regression model, J. Econometrics 39 (1988), 387-395

Gelfand, I.M., Kapranov, M.M. and Zelevinsky, A.V. (1990): Generalized Euler Integrals and A-Hypergeometry Functions, *Advances in Mathematics* 84 (1990) 255-271.

Gelfand, I.M., Kapranov, M.M. and Zelevinsky, A.V. (1994): Discriminants, resultants and multidimensional determinants, Birkhäuser, Boston 1994.

Gelderen, M. van and R. Rummel (2001): The solution of the general geodetic boundary value problem by least squares, Journal of Geodesy 75 (2001), 1-11

Gelman, A., Carlin, J.B., Stern, H.S. and D.B. Rubin (1995): Bayesian data analysis, Chapman and Hall, London 1995

Genton, M.G. (1998): Asymptotic variance of M-estimators for dependent Gaussian random variables, Statistics and Probability Lett. 38 (1998), 255-261

Genton, M.G. and Y. Ma (1999): Robustness properties of dispersion estimators, Statistics & Probability Letters 44 (1999), 343-350

Genugten, B. and Van Der, B. (1997): Testing in the restricted linear model using canonical partitions. Linear Algebra Appl. 264 (1997), 349-353.

Gephart, J. and Forsyth D. (1984): An improved method for determining the regional stress tensor - using earthquake focal mechanism data: application to the San Fernando earthquake sequence, J. geophys. Res., B89, 9305-9320

Gere, G.M. and W. Weaver (1965): Matrix algebra for engineers, Van Nostrand Co., New York 1956

Ghosh, M., Mukhopadhyay, N. and P.K. Sen (1997): Sequential estimation, J. Wiley, New York 1997

Ghosh, S. (1996): Wishart distribution via induction, The American Statistician 50 (1996), 243-246
Ghosh, S. (1999): Multivariate analysis, design of experiments, and survey sampling, Marcel Dekker, Basel 1999
Ghosh, S., Beran, J. and J. Innes (1997): Nonparametric conditional quantile estimation in the presence of long memory, Student 2 (1997), 109-117
Ghosh, S. (ed.) (1999): Multivariate analysis, design of experiments, and survey sampling, Marcel Dekker, New York 1999
Giacolone, M. (1997): Lp-norm estimation for nonlinear regression models, Student 2 (1997), 119-130
Giering, O. (1984): Bemerkungen zum ebenen Trilaterationsproblem. Deutsche Geodätische Kommission, Reihe A, Nr. 101.
Giering, O. (1986): Analytische Behandlung des raumlichen Trilaterationsproblems 4, 6, 0, o. Deutsche Geodätische Kommission, Reihe A, Nr. 104
Gil, A. and J. Segura (1998): A code to evaluate prolate and oblate spheroidal harmonics, Computer Physics Communications 108 (1998), 267-278
Gilbert, E.G. and C.P. Foo (1990): Computing the distance between general convex objects in three-dimensional space, JEEE Transactions on Robotics and Automation 6 (1990), 53-61
Gilchrist, W. (1976): Statistical forecasting, Wiley, London 1976
Gill, P.E., Murray, W. and Wright, M.H. (1981): Practical Optimization. Academic Press, London.
Gill, P.E., Murray, W. and Saunders, M.A. (2002): Snopt: An SQP algorithm for large scale constrained optimization, Siam J. Optim. 12 (2002), 979-1006
Gillard, D., Wyss, M. and J. Nakata (1992): A seismotectonic model for Western Hawaii based on stress tensor inversion from fault plane solutions, J. Geophys., Res., B97, 6629-6641
Giri, N. (1977): Multivariate statistical inference, Academic Press, New York London 1977
Giri, N. (1993): Introduction to probability and statistics, 2nd edition, Marcel Dekker, New York 1993
Giri, N. (1996a): Multivariate statistical analysis, Marcel Dekker, New York 1996
Giri, N. (1996b): Group invariance in statistical inference, World Scientific, Singapore 1996
Girko, V.L. (1979): Distribution of eigenvalues and eigenvectors of hermitian stochastic matrices, Ukrain. Math. J., 31, 421-424
Girko, V.L. (1985): Spectral theory of random matrices, Russian Math. Surveys, 40, 77-120
Girko, V.L. (1988): Spectral theory of random matrices, Nauka, Moscow 1988
Girko, V.L. (1990): Theory of random determinants, Kluwer Academic Publishers, Dordrecht 1990
Girko, V.L. and A.K. Gupta (1996): Multivariate elliptically contoured linear models and some aspects of the theory of random matrices, in: Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, A.K. and V.L. Girko, 327-386, VSP, Utrecht 1996
Glatzer, E. (1999): über Versuchsplanungsalgorithmen bei korrelierten Beobachtungen, Master's thesis, Wirtschaftsuniversität Wien
Gleick, J. (1987): Chaos, viking, New York 1987
Gleser, L.J. and I. Olkin (1972): Estimation for a regression model with an unknown covariance matrix, in: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, Vol. 1, 541-568, University of California Press, Berkeley 1972
Glimm, J. (1960): On a certain class of operator algebras, Trans. American Mathematical Society 95 (1960), 318-340
Gnedenko, B.V. and A.N. Kolmogorov (1968): Limit distributions for sums of independent random variables, Addison-Wesley Publ., Reading, Mass. 1968
Gnedin, A.V. (1993): On multivariate extremal processes, J. Multivar. Anal. 46 (1993), 207-213
Gnedin, A.V. (1994): On a best choice problem with dependent criteria, J. Applied Probability 31 (1994), 221-234
Gneiting, T. (1999): Correlation functions for atmospheric data analysis, Q. J. R. Meteorol. Soc. 125 (1999), 2449-2464

Gneiting, T. and Savati, Z. (1999): The characterization problem for isotropic covariance functions, Analy. Royal Met. Society 115 (1999) 2449-2450

Gnot, S. and Michalski, A. (1994): Tests based on admissible estimators in two variance components models, Statistics 25 (1994), 213-223

Gnot, S. and Trenkler, G. (1996): Nonnegative quadratic estimation of the mean squared errors of minimax estimators in the linear regression model, Acta Applicandae Mathematicae 43 (1996), 71-80

Godambe, V.P. (1991): Estimating Functions, Oxford University Press 1991

Godambe, V.P. (1995): A unified theory of sampling from finite populations, J. Roy. Statist. Soc. B17 (1955), 268-278

Göbel, M. (1998): A constructive description of SAGBI bases for polynomial invariants of permutation groups, J. Symbolic Computation 26 (1998), 261-272

Goldberger, A.S. (1962): Best linear unbiased prediction in the generalized linear regression model, J. Am. Statist. Ass. 57 (1962), 369-375

Goldberger, A.S. (1964): Economic theory, Wiley, New York, 1964

Goldberg, M.L., Bates, D.M. and Watts, D.G. (1983): Simplified methods for assessing nonlinearity. American Statistical Association/Business and Economic Statistics Section, 67-74.

Goldie, C.M. and S. Resnick (1989): Records in a partially ordered set, Annals Probability 17 (1989), 678-689

Goldie, C.M. and S. Resnick (1995): Many multivariate records, Stochastic Processes Appl. 59 (1995), 185-216

Goldie, C.M. and S. Resnick (1996): Ordered independent scattering, Commun. Statist. Stochastic Models 12 (1996), 523-528

Goldman, A.J. and Zelen, M. (2003): Weak generalized inverses and minimum variance linear unbiased estimation. J. Res. Natl. Bur. Stand., Sect. B 68 (1964), 151-172.

Goldstine, H. (1977): A history of numerical analysis from the 16th through the 19th century, Springer-Verlag, Heidelberg Berlin New York 1977

Golshstein, E.G. and N.V. Tretyakov (1996): Modified Lagrangian and monotone maps in optimization, J. Wiley, New York 1996

Golub, G.H. (1968): Least squares, singular values and matrix approximations, Aplikace Matematiky 13 (1968), 44-51

Golub, G.H. (1973): Some modified matrix eigenvalue problems, SIAM Review 15 (1973), 318-334

Golub, G.H. and C.F. van Loan (1983): Matrix Computations. North Oxford Academic, Oxford.

Golub, G.H. and C.F. van Loan (1996): Matrix computations, 3rd edition, John Hopkins University Press, Baltimore 1996

Golub, G.H. and W. Kahan (1965): Calculating the singular values and pseudo-inverse of a matrix, SIAM J Numer. Anal. 2 (1965), 205-224

Golub, G.H. and U. von Matt (1991): Quadratically constrained least squares and quadratic problems, Numer. Math. 59 (1991), 561-580

Golub, G.H., Hansen, P.C. and O'Leary, D.P. (1999): Tikhonov regularization and total least squares, SIAM J. Matrix Anal. Appl. 21 (1999), 185-194

Gèomez, E., Gèomez-Villegas, M.A. and Marèìn, J.M. (1998): A multivariate generalization of the power exponential family of distributions, Commun. Statist. - Theory Meth. 27 (1998), 589-600

Gonin, R. and A.H. Money (1987): A review of computational methods for solving the nonlinear L 1 norm estimation problem, in: Statistical data analysis based on the L1 norm and related methods, Ed. Y. Dodge, North Holland 1987

Gonin, R. and A.H. Money (1989): Nonlinear Lp-norm estimation, Marcel Dekker, New York 1989

Goodall, C. (1991): Procrustes methods in the statistical analysis of shape, J. Roy. Stat. Soc. B53 (1991), 285-339

Goodman, J.W. (1985): Statistical optics, J. Wiley, New York 1985

Gordon, A.D. (1997): L1-norm and L2-norm methodology in cluster analysis, Student 2 (1997), 181-193

Gordon, A.D. (1999): Classification, 2nd edition, Chapman and Hall, Yew York 1999

Gordon, L. and M. Hudson (1977): A characterization of the Von Mises Distribution, Ann. Statist. 5 (1977), 813-814
Gordon, R. and Stein, S. (1992): Global tectonics and space geodesy, Science, 256, 333-342
Gordonova, V.I. (1973): The validation of algorithms for choosing the regularization parameter, Zh. vychisl. Mat. mat. Fiz. 13 (1973), 1328-1332
Gorman, O.T.W. (2001): Adaptive estimation using weighted least squares, Aust. N. Z. J. Stat. 43 (2001), 287-297
Gosset, W.S. (1908): The probable error of a mean, Biometrika, 6, 1-25.
Gotthardt, E. (1940): Zur Unbestimmtheit des räumlichen Rückwärtseinschnittes, Mitteilungen der Ges. f. Photogrammetry e.V., Jänner 1940, Heft 5.
Gotthardt, E. (1971): Grundsutzliches zur Fehlertheorie and Ausgleichung von Polygonzugen and Polygonnetzen, Wichmann Verlag, Karlsruhe 1971
Gotthardt, E. (1974): Ein neuer gefährlicher Ort zum räumlichen Rückwärtseinschneiden, Bildm. u. Luftbildw., 1974.
Gotthardt E. (1978): Einführung in die Ausgleichungsrechnung, 2nd ed., Wichmann, Karlsruhe 1978
Gould, A.L. (1969): A regression technique for angular varietes, Biometrica 25 (1969), 683-700
Gower, J.C. (1974): The median center, Applied Statistics 2, (1974), pp. 466-470
Grabisch, M. (1998): Fuzzy Integral as a Flexible and Interpretable Tool of Aggregation, In (Ed. B. Bouchon-Meunier, Aggregation and Fusion of Imperfect Information, pp. 51-72, Physica Verlag, Heidelberg, 1998
Grafarend, E.W.: a) DGK C 153, München 1970; b) AVN 77, 17 (1970); c) Vermessungstechnik 19, 66 (1971); d) ZfV 96, 41 (1971); e) ZAMM. imDruck.
Grafarend, E.W. (1967a): Bergbaubedingte Deformation und ihr Deformationstensor, Bergbauwissenschaften 14 (1967), 125-132
Grafarend, E.W. (1967b): Allgemeiner Fehlertensor bei a priori und a posteriori-Korrelationen, Zeitschrift für Vermessungswesen 92 (1967), 157-165
Grafarend, E.W. (1969): Helmertsche Fußpunktkurve oder Mohrscher Kreis? Allgemeine Vermessungsnachrichten 76 (1969), 239-240
Grafarend, E.W. (1970a): Praeidiktion und Punktamasse, AVN (1970), S. 17
Grafarend, E.W. (1970b): Verallgemeinerte Methode der kleinsten Quadrate für zyklische Variable, Zeitschrift für Vermessungswesen 4 (1970), 117-121
Grafarend, E.W. (1970c): Die Genauigkeit eines Punktes im mehrdimensionalen Euklidischen Raum, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften C 153, München 1970
Grafarend, E.W. (1970d): Fehlertheoretische Unschärferelation, Festschrift Professor Dr.-Ing. Helmut Wolf, 60. Geburtstag, Bonn 1970
Grafarend, E.W. (1971a): Mittlere Punktfehler und Vorwärtseinschneiden, Zeitschrift für Vermessungswesen 96 (1971), 41-54
Grafarend, E.W. (1971b): Isotropietests von Lotabweichungen Westdeutschlands, Z. Geophysik 37 (1971), 719-733
Grafarend, E.W. (1972a): Nichtlineare Prädiktion, Zeitschrift für Vermessungswesen 97 (1972), 245-255
Grafarend, E.W. (1972b): Isotropietests von Lotabweichungsverteilungen Westdeutschlands II, Z. Geophysik 38 (1972), 243-255
Grafarend, E.W. (1972c): Genauigkeitsmaße geodätischer Netze, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften A73, München 1972
Grafarend, E.W. (1973a): Nichtlokale Gezeitenanalyse, Mitt. Institut für Theoretische Geodäsie No. 13, Bonn 1973
Grafarend, E.W. (1973b): Optimales Design geodätische Netze 1 (zus. P. Harland), Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften A74, München 1973

Grafarend, E.W. (1973c): Eine Lotabweichungskarte Westdeutschlands nach einem geodätisch konsistenten Kolmogorov-Wiener Modell , Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Report A 82, München 1975
Grafarend, E.W. (1974): Optimization of geodetic networks, Bollettino di Geodesia e Scienze Affini 33 (1974), 351-406
Grafarend, E.W. (1975a): Three dimensional Geodesy 1. The holonomity problem, *Zeitschrift für Vermessungswesen* 100 (1975) 269-280.
Grafarend, E.W. (1975b): Second order design of geodetic nets, Zeitschrift für Vermessungswesen 100 (1975), 158-168
Grafarend, E.W. (1976): Geodetic applications of stochastic processes, Physics of the Earth and Planetory Interiors 12 (1976), 151-179
Grafarend, E.W. (1977a): Geodasie - Gaufische oder Cartansche Flachengeometrie. Allgemeine Vermessungs-Nachrichten, 84. Jg., Heft 4, 139-150.
Grafarend, E.W. (1977b): Stress-strain relations in geodetic networks Report 16, Geodetic Institute, University of Uppsala.
Grafarend, E.W. (1978): Operational geodesy, in: Approximation Methods in Geodesy, eds. H. Moritz and H. Sünkel, 235-284, H. Wichmann Verlag, Karlsruhe 1978
Grafarend, E.W. (1981): Die Beobachtungsgleichungen der dreidimensionalen Geodäsie im Geometrie- und Schwereraum. Ein Beitrag zur operationellen Geodäsie, *Zeitschrift für Vermessungswesen* 106 (1981) 411-429.
Grafarend, E.W. (1982): Six lectures on geodesy and global geodynamics, In: H. Moritz and H. Snkel (Ed.), Lecture Notes, Mitt. Geod .. lnst. Techn. UnL Graz, 41, 531-685
Grafarend, E.W. (1983): Stochastic models for point manifolds, in: Mathematical models of geodetic/photogrammetric point determination with regard to outliers and systematic errors, ed. F.E. Ackermann, Report A98, 29-52, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, München 1983
Grafarend, E.W. (1984): Variance-covariance component estimation of Helmert type in the Gauss-Helmert model, Zeitschrift für Vermessungswesen 109 (1984), 34-44
Grafarend, E.W. (1985a): Variance-covariance component estimation, theoretical results and geodetic applications, Statistics and Decision, Supplement Issue No. 2 (1985), 407-447
Grafarend, E.W. (1985b): Criterion matrices of heterogeneously observed three-dimensional networks, manuscripta geodaetica 10 (1985), 3-22
Grafarend, E.W. (1985c): Criterion matrices for deforming networks, in: Optimization and Design of Geodetic Networks, E. Grafarend and F. Sanso (eds.), 363-428, Springer-Verlag, Heidelberg Berlin New York 1985
Grafarend, E.W. (1986a): Generating classes of equivalent linear models by nuisance parameter elimination - applications to GPS observations, manuscripta geodaetica 11 (1986) 262-271
Grafarend, E.W. (1986b): Three dimensional deformation analysis: global vector spherical harmonic and local element representation, Tectonophysics, 130, 337-359
Grafarend, E.W. (1988): Azimuth transport and the problem of orientation within geodetic traverses and geodetic networks, *Vermessung, Photogrammetrie, Kulturtechnik* 86 (1988) 132-150.
Grafarend, E.W. (1989a): Four lectures on special and general relativity, Lecture Notes in Earth Sciences, F. Sanso and R. Rummel (eds.), Theory of Satellite Geodesy and Gravity Field Determination, Nr. 25, 115-151, Springer-Verlag Berlin New York 1989
Grafarend, E.W. (1989b): Photogrammetrische Positionierung, Festschrift Prof. Dr.-Ing. Dr. H.C. Friedrich Ackermann zum 60. Geburtstag, Institut für Photogrammetrie, Universität Stuttgart, Report 14, 45-55, Suttgart
Grafarend, E.W. (1990): Dreidimensionaler Vorwartsschnitt. Zeitschrift fiir Vermessungswesen, 10, 414-419.
Grafarend, E.W. (1991a): Relativistic effects in geodesy, Report Special Study Group 4.119, International Association of Geodesy, Contribution to Geodetic Theory and Methodology ed. F. Sanso, 163-175, Politecnico di Milano, Milano/Italy 1991

Grafarend, E.W. (1991b): The Frontiers of Statistical Scientific Theory and Industrial Applications (Volume II of the Proceedings of ICOSCO-I), American Sciences Press, 405-427, New York 1991

Grafarend, E.W. (1991c): Application of Geodesy to Engineering. IAG-Symposium No. 108, Eds. K. Linkwitz, V. Eisele, H. J. Mönicke, Springer-Verlag, Berlin-Heidelberg-New York 1991.

E. Grafarend, T. Krarup, R. Syffus: Journal of Geodesy 70 (1996) 276-286

Grafarend, E.W. (1998): Helmut Wolf - das wissenschaftliche Werk/the scientific work , Heft A115, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, C.H. Beck'sche Verlagsbuchhandlung, 97 Seiten, München 1998

Grafarend, E.W. (2000): Mixed integer-real valued adjustment (IRA) problems, GPS Solutions 4 (2000), 31-45

Grafarend, E.W. and Kunz, J. (1965): Der Rückwärtseinschnitt mit dem Vermessungskreisel, *Bergbauwissenschaften* 12 (1965) 285-297.

Grafarend, E. and E. Livieratos (1978): Rand defect analysis of satellite geodetic networks I : geometric and semi-dynamic mode, manuscipta geodaetica 3 (1978) 107-134

Grafarend, E. and K. Heinz (1978): Rank defect analysis of satellite geodetic networks II : dynamic mode, manuscipta geodaetica 3 (1978) 135-156

Grafarend, E.W. and P. Harland (1973) : Optimales Design geodaetischer Netze I, Deutsche Geodaetische Kommission, Bayerische Akademie der Wissenschaften, Series A, C.H. Beck Verlag, Munich 1973

Grafarend, E.W. and B. Schaffrin (1974): Unbiased free net adjustment, Surv. Rev. XXII, 171 (1974), 200-218

Grafarend, E.W., Heister, H., Kelm, R., Kropff, H. and B. Schaffrin (1979): Optimierung geodätischer Messoperationen , 499 pages, H. Wichmann Verlag, Karlsruhe 1979

Grafarend, E.W. and G. Offermanns (1975): Eine Lotabweichungskarte Westdeutschlands nach einem geodätisch konsistenten Kolmogorov-Wiener Modell, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften A82, München 1975

Grafarend, E.W. and B. Schaffrin (1976): Equivalence of estimable quantities and invariants in geodetic networks, Z. Vermessungswesen 191 (1976), 485-491

Grafarend, E.W., Schmitt, G. and B. Schaffrin (1976): Über die Optimierung lokaler geodätischer Netze (Optimal design of local geodetic networks), 7th course, High precision Surveying Engineering (7. Int. Kurs für Ingenieurvermessung hoher Präzision) 29 Sept - 8 Oct 1976, Darmstadt 1976

Grafarend, E.W. and Richter, B. (1977): Generalized Laplace condition, *Bull. Geod.* 51 (1977) 287-293.

Grafarend, E.W. and B. Richter (1978): Threedimensional geodesy II - the datum problem -, Zeitschrift für Vermessungswesen 103 (1978), 44-59

Grafarend, E.W. and A. d'Hone (1978): Gewichtsschätzung in geodätischen Netzen, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften A88, München 1978

Grafarend, E.W. and B. Schaffrin (1979): Kriterion-Matrizen I - zweidimensional homogene und isotope geodätische Netze - ZfV 104 (1979), 133-149

Grafarend, E.W., J.J. Mueller, H.B. Papo, and B. Richter (1979): Concepts for reference frames in geodesy and geodynamics: the reference directions, Bull. Gèeodèesique 53 (1979), 195-213

Grafarend, E.W., Heister, H., Kelm, R., Knopff, H. and Schaffrin, B. (1979): Optimierung Geodätischer Messoperationen, Herbat Wichmann Verlag, Karlsruhe 1979.

Grafarend, E.W. and R.H. Rapp (1980): Advances in geodesy. Selected papers from Rev. Geophys. Space Phys., Richmond, Virg., Am Geophys Union,Washington

Grafarend, E.W. and A. Kleusberg (1980): Expectation and variance component estimation of multivariate gyrotheodolite observations, I. Allgemeine Vermessungs-Nachrichten 87 (1980), 129-137

Grafarend, E.W., Kleusberg, A. and B. Schaffrin (1980): An introduction to the variance-covariance component estimation of Helmert type, Zeitschrift für Vermessungswesen 105 (1980), 161-180

Grafarend, E.W. and B. Schaffrin (1982): Kriterion Matrizen II: Zweidimensionale homogene und isotrope geodätische Netze, Teil II a: Relative cartesische Koordinaten, Zeitschrift für Vermessungswesen 107 (1982), 183-194, Teil IIb: Absolute cartesische Koordinaten, Zeitschrift für Vermessungswesen 107 (1982), 485-493

Grafarend, E.W., Knickemeyer, E.H. and B. Schaffrin (1982): Geodätische Datumtransformationen, Zeitschrift für Vermessungswesen 107 (1982), 15-25

Grafarend, E.W., Krumm, F. and B. Schaffrin (1985): Criterion matrices of heterogeneously observed threedimensional networks, manuscripta geodaetica 10 (1985), 3-22

Grafarend, E.W. and F. Sanso (1985): Optimization and design of geodetic networks, Springer-Verlag, Heidelberg Berlin New York 1985

Grafarend, E.W. and V. Müller (1985): The critical configuration of satellite networks, especially of Laser and Doppler type, for planar configurations of terrestrial points, manuscripta geodaetica 10 (1985), 131-152

Grafarend, E.W. and F. Krumm (1985): Continuous networks I, in: Optimization and Design of Geodetic Networks. E. Grafarend and F. Sanso (eds.), 301-341, Springer-Verlag, Heidelberg Berlin New York 1985

Grafarend, E.W., Krumm, F. and B. Schaffrin (1985): Criterion matrices of heterogenously observed three dimensional networks, manuscripta geodetica 10 (1985) 3-22

Grafarend, E.W., Krumm, F. and B. Schaffrin (1986): Kriterion-Matrizen III: Zweidimensionale homogene und isotrope geodätische Netze, Zeitschrift für Vermessungswesen 111 (1986), 197-207

Grafarend, E.W., Krumm, F. and B. Schaffrin (1986): Kriterion-Matrizen III: Zweidimensional homogene und isotrope geodätische Netze, Zeitschrift für Vermessungswesen 111 (1986), 197-207

Grafarend, E.W. and B. Schaffrin (1988): Von der statistischen zur dynamischen Auffasung geodätischer Netze, Zeitschrift für Vermessungswesen 113 (1988), 79-103

Grafarend, E.W. and B. Schaffrin (1989a): The Planar Trisection Problem. Festschrift to Torben Krarup, Eds.: E. Kejlso, K. Poder, C.C. Tscherning, Geodaetisk Institute, No. 58, Kobenhavn, 149-172.

Grafarend, E.W. and B. Schaffrin (1989b): The geometry of nonlinear adjustment - the planar trisection problem, Festschrift to Torben Krarup eds. E. Kejlso, K. Poder and C.C. Tscherning, Geodaetisk Institut, Meddelelse No. 58, 149-172, Kobenhavn 1989

Grafarend, E.W. and A. Mader (1989): A graph-theoretical algorithm for detecting configuration defects in triangular geodetic networks, Bull. Gèeodèesique 63 (1989), 387-394

Grafarend, E.W., Lohse, P. and Schaffrin, B. (1989): Dreidimensionaler Rückwärtsschnitt, *Zeitschrift für Vermessungswesen* 114 (1989) 61-67, 127-137, 172-175, 225-234, 278-287.

Grafarend, E.W. and B. Schaffrin (1991): The planar trisection problem and the impact of curvature on non-linear least-squares estimation, Comput. Stat. Data Anal. 12 (1991), 187-199

Grafarend, E.W. and Lohse, P. (1991): The minimal distance mapping of the topographic surface onto the (reference) ellipsoid of revolution, *Manuscripta Geodaetica* 16 (1991) 92-110.

Grafarend, E.W. and B.Schaffrin (1993): Ausgleichungsrechnung in linearen Modellen, Brockhaus, Mannheim 1993

Grafarend, E.W. and Xu, P. (1994): Observability analysis of integrated INS/GPS system, Bollettino di Geodesia e Scienze Affini 103 (1994), 266-284

Grafarend, E.W., Krumm, F. and F. Okeke (1995): Curvilinear geodetic datum transformations, Zeitschrift für Vermessungswesen 120 (1995), 334-350

Grafarend, E.W. and P. Xu (1995): A multi-objective second-order optimal design for deforming networks, Geophys. Journal Int. 120 (1995), 577-589

Grafarend, E.W. and Keller, W. (1995): Setup of observational functionals in gravity space as well as in geometry space, *Manuscripta Geodetica* 20 (1995) 301-325.

Grafarend, E.W., Krumm, F. and Okeke, F. (1995): Curvilinear geodetic datum transformation, *Zeitschrift für Vermessungswesen* 120 (1995) 334-350.

Grafarend, E.W., Syffus, R. and You, R.J. (1995): Projective heights in geometry and gravity space, *Allgemeine Vermessungs-Nachrichten* 102 (1995) 382-402.

Grafarend, E.W., Krarup, T. and R. Syffus (1996): An algorithm for the inverse of a multivariate homogeneous polynomial of degree n, Journal of Geodesy 70 (1996), 276-286

Grafarend, E.W. and G. Kampmann (1996): C10(3): The ten parameter conformal group as a datum transformation in three dimensional Euclidean space, Zeitschrift für Vermessungswesen 121 (1996), 68-77

Grafarend, E.W. and Shan, J. (1996): Closed-form solution of the nonlinear pseudo-ranging equations (GPS), Artificial Satellites, Planetary Geodesy No. 28, Special issue on the 30th anniversary of the Department of Planetary Geodesy, Vol. 31, No. 3, pp. 133-147, Warszawa 1996.

Grafarend, E.W. and Shan, J. (1997a): Closed-form solution of P4P or the three-dimensional resection problem in terms of Möbius barycentric coordinates, *Journal of Geodesy* 71 (1997) 217-231.

Grafarend, E.W., and Shan, J. (1997b): Closed form solution of the twin P4P or the combined three dimensional resection-intersection problem in terms of Möbius barycentric coordinates, *Journal of Geodesy* 71 (1997) 232-239.

Grafarend, E.W. and Shan J. (1997c): Estimable quantities in projective networks, Zeitschrift für Vermessungswesen, Part I, 122 (1997), 218-226, Part II, 122 (1997), 323-333

Grafarend, E.W. and Okeke, F. (1998): Transformation of conformal coordinates of type Mercator from global datum (WGS 84) to local datum (regional, national), *Marine Geodesy* 21 (1998) 169-180.

Grafarend, E.W. and Syfuss, R. (1998): Transformation of conformal coordinates of type Gauss-Krüger or UTM from local datum (regional, national, European) to global datum (WGS 84) part 1: The transformation equations, *Allgemeine Vermessungs-Nachrichten* 105 (1998) 134-141.

Grafarend, E.W. and Ardalan, A. (1999): World geodetic datum 2000, *Journal of Geodesy* 73 (1999) 611-623.

Grafarend, E.W. and Awange, J.L. (2000): Determination of vertical deflections by GPS/LPS measurements, *Zeitschrift für Vermessungswesen* 125 (2000) 279-288.

Grafarend, E.W. and J. Shan (2002): GPS Solutions: closed forms, critical and special configurations of P4P, GPS Solutions 5 (2002), 29-42

Grafarend, E.W. and J. Awange (2002): Algebraic solution of GPS pseudo-ranging equations, GPS Solutions 5 (2002), 20-32

Grafarend, E.W. and J. Awange (2002): Nonlinear adjustment of GPS observations of type pseudo-ranges, GPS Solutions 5 (2002), 80-93

Graham, A. (1981): Kronecker products and matrix calculus, J. Wiley, New York 1981

Gram, J.P. (1883): Über die Entwicklung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate, J. Reine Angew. Math. 94 (1883), 41-73

Granger, C.W.J. and Hatanka, M. (1964): Spectral analysis of economic time series, Princeton 1964

Granger, C.W.J. and P. Newbold (1986): Forecasting economic time series, 2nd ed., Academic Press, New York London 1986

Granger, C.W.J. and T. Teräsvirta (1993): Modelling nonlinear economic relations, Oxford University Press, New York 1993

J.T. Graves: Transactions of the Irish Academy 21 (1848) 338

Graybill, F.A. (1954): On quadratic estimates of variance components, The Annals of Mathematical Statistics 25 (1954), 267-372

Graybill, F.A. (1961): An intorduction to linear statistical models, Mc Graw-Hill, New York, 1961

Graybill, F.A. (1983): Matrices with applications in statistics, 2nd ed., Wadsworth, Beltmont 1983

Graybill, F.A. and R.A. Hultquist (1961): Theorems concerning Eisenhart's model II, The Annals of Mathematical Statistics 32 (1961), 261-269

Green, B. (1952): The orthogonal approximation of an oblique structure in factor analysis, Psychometrika 17 (1952), 429-440

Green, P.J. and B.W. Silverman (1993): Nonparametric regression and generalized linear models, Chapman and Hall, Boca Raton 1993

Green, P.E., Frosch, R.A. and Romney, C.F. (1965): Principles of an experimental large aperture seismic array, Proc. IEEE 53 (1965) 1821-1833

Green, P.E., Kelly, E.J. and Levin, M.J. (1966): A comparison of seismic array processing methods, Geophys. J. R. astr. Soc. 11 (1966) 67-84

Greenbaum, A. (1997): Iterative methods for solving linear systems, SIAM, Philadelphia 1997

Greenwood, J.A. and D. Durand (1955): The distribution of length and components of the sum of n random unit vectors, Ann. Math. Statist. 26 (1955), 233-246

Greenwood, P.E. and G. Hooghiemstra (1991): On the domain of an operator between supremum and sum, Probability Theory Related Fields 89 (1991), 201-210

Gregersen, S. (1992): Crustal stress regime in Fennoscandia from focal mechanisms, J. geophys. Res., B97, 11821-11827

Grenander, U. (1981): Abstract inference, J. Wiley, New York 1981

Greub, W.H. (1967): Multilinear algebra, Springer Verlag, Berlin 1967

Grewal, M.S., Weill, L.R. and Andrews, A.P. (2001): Global Positioning Systems, Inertial Navigation and Integration, John Wiley & Sons, New York 2001.

Griffith, D.F. and D.J. Higham (1997): Learning LaTeX, SIAM, Philadelphia 1997

Grigoriu, M. (2002): Stochastic calculus, Birkhanser, Boston-Basel-Baein 2002

Grimstad, A. and T.Mannseth (2000): Nonlinearity, sale and sensitivity for parameter estimation problems, SIAM J. Sci. Comput. 21 (2000), 2096-2113

Grodecki, J. (1999): Generalized maximum-likelihood estimation of variance components with inverted gamma prior, Journal of Geodesy 73 (1999), 367-374

Gröbner, W. (1968): Algebraische Geometrie I. Bibliographisches Institut, Mannheim.

Gröbner, W. (1970): Algebraische Geometrie II. Bibliographisches Institut, Mannheim

Gröbner, W. and N. Hofreiter (eds., 1973): Integraltafel, zweiter Teil, bestimmte Integrale. 5. 'verbesserte Auflage, Wien.

Gröchenig, K. (2001): Foundations of time-frequency analysis, Birkäuser Verlag, Boston Basel Berlin 2001

Groenigen, van J.W. and Siderius, W. and Stein, A. (1999): Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87:239-259

Groß, J. (1996): On a class of estimators in the general Gauss-Markov model, Commun. Statist. - Theory Meth. 25 (1996), 381-388

Groß, J. (1996): Estimation using the linear regression model with incomplete ellipsoidal restrictions, Acta Applicandae Mathematicae 43 (1996), 81-85

Groß, J. (1998): Statistical estimation by a linear combination of two given statistics, Statistics and Probability Lett. 39 (1998), 379-384

Groß, J. (2003): Linear regression. Springer, Berlin, 2003.

Groß, J. and G. Trenkler (1997): When do linear transforms of ordinary least squares and Gauss-Markov estimator coincide?, Sankhya 59 (1997), 175-178

Groß, J., Trenkler, G. and E.P. Liski (1998): Necessary and sufficient conditions for superiority of misspecified restricted least squares regression estimator, J. Statist. Planning and Inference 71 (1998), 109-116

Groß, J., Puntanen, S. (2000): Estimation under a general partitioned linear model. Linear Algebra Appl. 321, 1-3 (2000), 131-144.

Groß, J., Trenkler, G. and H.J. Werner (2001): The equality of linear transforms of the ordinary least squares estimator and the best linear unbiased estimator, Sankhya: The Indian Journal of Statistics 63 (2001), 118-127

Großmann, W. (1973): Grundzüge der Ausgleichungsrechnung, Springer-Verlag, Heidelberg Berlin New York 1973

Großman, N. (1980): Four Lectures on Geometry for-Geodesists. In: Proceedings of the International School of Advanced Geodesy, Erice (Italy), May 18.-June 2.1978, Eds.: E. W. Grafarend, A. Marussi, published by the University FAF, Munich.

Groten, E. (1964): Symposium "Extension of the Gravity Anomalies to Unsurveyed Areas", Columbus (Ohio) 1964

Groten, E. (1964): lnst. f. Astr. Phys. Geod., Publ. No. 39, Miinchen 1967.

Grubbs, F.E. (1973): Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments, Technometrics 15 (1973), 53-66

Grubbs, F.E. and G. Beck (1972): Extension of sample sizes and precentage points for significance tests of outlying observations, Technometrics 14 (1972), 847-854

Gruen, A. and Kahmen, H. (Eds.) (1989): Optical 3-D Measurement Techniques. Wichmann, Karlsruhe.

Gründig, L. (1975): Die Berechnung vorgespannter Seil- und Hangenetze unter der Beriicksichtigung ihrer topologischen und physikalischen Eigenschaften und der Ausgleichungsrechnung. Sonderforschungsbereich 64, Universitat Stuttgart, Mitteilungen 34/1975.

Gründig, L. (1976): Ausgleichung grofier geodätischer Netze, Bestimmung von Naherungskoordinaten, Ausgleichungstechniken und Fehlersuche. VII. Internationaler Kurs fiir Ingenieurvermessungen, Darmstadt, 41-51.

Grunert, J.A. (1841): Das Pothenotsche Problem in erweiterter Gestalt; nebst Bemerkungen über seine Anwendungen in der Geodäsie, Grunerts Archiv für Mathematik und Phsyik 1 pp. 238-241, 1841.

Guckenheimer, J., Myers, M. and Sturmfels, B. (1997): Computing Hopf birfucations, *SIAM J. Numerical Analysis* 34 (1997) 1-21.

Guenther, W.C. (1964): Another derivation of the non-central Chi-Square distribution, J. Am. Statist. Ass. 59 (1964), 957-960

Gui, Q. and J. Zhang (1998): Robust biased estimation and its applications in geodetic adjustments, Journal of Geodesy 72 (1998), 430-435

Gui, Q.M. and J.S. Liu (2000): Biased estimation in the Gauss-Markov model, Allgemeine Vermessungsnachrichten 107 (2000), 104-108

Gulliksen, H. and S.S. Wilks (1950), Regression tests for several samples Psychometrica, 15, 91-114

Gullikson, M. (1995a): The partial Procrustes problem - A first look, Department of Computing Science, Umea University, Report UMINF-95.11, Sweden 1995.

Gullikson, M. (1995b): Algorithms for the partial Procrustes problem, Department of Industrial Technology, Mid Sweden University s-891 18, Report 1995:27, Ornskoldsvik, Sweden 1995.

Gullikson, M. and Söderkvist, I. (1995): Surface fitting and parameter estimation with nonlinear least squares, *Zeitschrift für Vermessungswesen* 25 (1989) 611-636.

Gullikson, M., Soederkvist, I. and P.A. Wedin (1997): Algorithms for constrained and weighted nonlinear least-squares, Siam J. Optim. 7 (1997), 208-224

Gullikson, M. and P.A. Wedin (2000): The use and properties of Tikhonov filter matrices, SIAM J. Matrix Anal. Appl. 22 (2000), 276-281

Gumbel, E.J., Greenwood, J.A. and D. Durand (1953): The circular normal distribution: theory and tables, J. Am. Statist. Ass. 48 (1953), 131-152

Gundlich, B. and Koch, K.R. (2002): Confidence regions for GPS baselines by Bayesian statistics. J Geod 76:55-62

Guolin, L. (2000): Nonlinear curvature measures of strength and nonlinear diagnosis, Allgemeine Vermessungsnachrichten 107 (2000), 109-111

Gupta, A.K. and D.G. Kabe (1997): Linear restrictions and two step multivariate least squares with aplications, Statistics & Probability Letters 32 (1997), 413-416

Gupta, A.K. and D.K. Nagar (1998): Quadratic forms in disguised matrix T-variate, Statistics 30 (1998), 357-374

Gupta, S.S. (1963): Probability integrals of multivariate normal and multivariate t1, Annals of Mathematical Statistics 34 (1963), 792-828

Gurtin, M.E. (1981): An introduction to continuum mechanics, Academic Press, New York

Guttman, I. (1970): Statistical tolerance regions, Griffin, London 1970

Guttman, L. (1946): Enlargement methods for computing the inverse matrix, Ann. Math. Statist. 17 (1946), 336-343

Guttorp, P. and Sampson, P. (1994). Methods for estimating heterogeneous spatial covariance. In Patil, G. and Rao, C. eds., Handbook of Statistics, vol. 12, pp. 661-689.

Guu, S.M., Lur, Y.Y. and C.T. Pang (2001): On infinite products of fuzzy matrices, SIAM J. Matrix Anal. Appl. 22 (2001), 1190-1203

Haantjes, J. (1937): Conformal representations of an n-dimensional Euclidean space with a non-definite fundamental form on itself, in: Nederl. Akademie van Wetenschappen, Proc. Section of Sciences, Vol. 40, 700-705, Noord-Hollandsche Uitgeversmaatschappij, Amsterdam 1937

Haantjes, J. (1940): Die Gleichberechtigung gleichförmig beschleunigter Beobachter für die elektromagnetischen Erscheinungen, in: Nederl. Akademie van Wetenschappen, Proc. Section of Sciences, Vol. 43, 1288-1299, Noord-Hollandsche Uitgeversmaatschappij, Amsterdam 1940

Habermann, S.J. (1996): Advanced statistics, volmue I: description of populations, Springer-Verlag, Heidelberg Berlin New York 1996

Hadamard, J. (1893): Leson sur la propagation des ondes et les quations de l'hydrodynamique, Paris 1893, reprint Chelsea Publ., New York 1949

Hadamard, J. (1899): Theorem sur les series entieres, Acta Math. 22 (1899), 1-28

Hadamard, J. Lecons sur la propagation des ondes et les équations de l'hydrodynamique, Paris 1893, reprint Chelsea Publ., New York 1949

Härdle, W., Liang, H. and J. Gao (2000): Partially linear models, Physica-Verlag, Heidelberg 2000

Hager, W.W. (1989): Updating the inverse of a matrix, SIAM Rev. 31 (1989), 221-239

Hager, W.W. (2000): Iterative methods for nearly singular linear systems, SIAM J. Sci. Comput. 22 (2000), 747-766

Hahn, M. and R. Bill (1984): Ein Vergleich der L1- und L2-Norm am Beispiel. Helmerttransformation, AVN 11-12, pp. 440, 1984

Hahn, W. and P. Weibel (1996): Evolutionäre Symmetrietheorie, Wiss. Verlagsgesellschaft, Stuttgart 1996

Hajos G. (1970): Einfiihrung in die Geometrie. BSB B. G. Teubner, Leipzig.

Hald, A. (1998): A history of mathematical statistics from 1750 to 1930, J. Wiley, New York 1998

Hald, A. (2000): The early history of the cumulants and the Gram-Charlier series, International Statistical Review 68 (2000), 137-153

Halmos, P.R. (1946): The theory of unbiased estimation, Ann. Math. Statist. 17 (1946), 34-43

Hamilton, W.C. (1964): Statistics in Physical Science. The Ronald Press Company, New-York

Hammer, E. (1896): Zur graphischen Ausgleichung beim trigonometrischen Einschneiden von Punkten, *Optimization methods and softwares* 5 (1995) 247-269.

Hammersley, J.M. (1950): On estimating restricted parameters, J. R. Statist. Soc. B12 (1950), 192

Hampel, F.R. (1971) A general qualitative definition of robustness, Ann. Math. Phys. Statist. 42 (1971) DOI 10.1214/aoms/1177693054

Hampel, F.R. (1973): Robust estimation: a condensed partial survey, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 27 (1973), 87-104

Hampel, F.R. (1974): The influence curve and its role in robust estimation, J. Ann. Stat. Assoc., 69 (1974), pp. 383-393

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and W.A. Stahel (1986): Robust statistics, J. Wiley, New York 1986

Han, S. (1995): Ambiguity resolution techniques using integer least squares estimation for rapid static or kinematic positioning. In: Symposium Satellite Navigation Technology: 1995 and beyond, Brisbane, Australia, 10 p

Han, S. C., Kwon, J. H. and Jekeli, C. (2001): Accurate absolute GPS positioning through satellite clock error estimation, *Journal of Geodesy* 75 (2001) 33-43.

Han, Y.Y. (1992): Geometric analysis of finite strain tensors in tectonodeformation mechanics, Se. China, B35, 102-116

Hanagal, D.D. (1996): UMPU tests for testing symmetry and stress-passing in some bivariate exponential models, Statistics 28 (1996), 227-239

Hand, D.J. and M.J. Crowder (1996): Practical longitudinal data analysis, Chapman and Hall, Boca Raton 1996

Hand, D.J., Daly, F., McConway, K., Lunn, D. and E. Ostrowski (1993): Handbook of small data sets, Chapman and Hall, Boca Raton 1993

Hand, D.J. and C.C. Taylor (1987): Multivariate analysis of variance and repeated measures, Chapman and Hall, Boca Raton 1987

Handl, A. (2010): Multivariate Analysemethoden. Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS, Springer-Verlag, Heidelberg Berlin New York

Hanke, M. (1991): Accelerated Landweber iterations for the solution of ill-posed equations, Numer. Math. 60 (1991), 341-375

Hanke, M. and P.C. Hansen (1993): Regularization methods for large-scale problems, Surveys Math. Indust. 3 (1993), 253-315

Hanselman, D. and Littlefield, B. (1997): The student edition of Matlab, Prentice-Hall, New Jersey 1997.

Hansen, E. (1992): Global Optimization Using Interval Analysis, Marcel Dekker, New York

Hansen, P.C. (1987): The truncated SVD as a method for regularization, BIT 27 (1987), 534-553

Hansen, P.C. (1990): The discrete Picard condition for discrete ill-posed problems, BIT 30 (1990), 658-672

Hansen, P.C. (1990): Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank, SIAM J. Sci. Statist. Comput. 11 (1990), 503-518

Hansen, P.C. (1994): Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems, Numer. Algorithms 6 (1994), 1-35

Hansen, P.C. (1995): Test matrices for regularization methods, SIAM J. Sci. Comput. 16 (1995), 506-512

Hansen, P.C. (1998): Rank-deficient and discrete ill-posed problems, SIAM, Philadelphia 1998

Hanssen, R.F. (2001): Radar infriometry, Kluwer Academic Pule, Dordecht-Boston-London 2001

Hanssen, R.F., Teunissen, P.J.G. and Joosten, P. (2001): Phase ambiguity resolution for stacked radar interferometric data. In: Proc KIS2001, international symposium on kinematic systems in geodesy, Geomatics and Navigation, Banff, pp 317-320

Haralick, R.M., Lee, C., Ottenberg, K., and Nölle, M. (1991): Analysis and Solution of the Three Point Perspective Pose Estimation Problem, Proc. IEEE Org. on Computer Vision and Pattern Recognition, pp. 592-598, 1991.

Haralick, R.M., Lee, C., Ottenberg, K., and Nölle, M. (1994): Review and Analysis of Solution of the Three Point Perspective Pose Estimation Problem, *International Journal of Computer Vision 13* 3 (1994) 331-356.

Hardtwig, E. (1968): Fehler- und Ausgleichsrechung, Bibliographisches Institut, Mannheim 1968

Harley, B.I. (1954): A note on the probability integral of the correlation coefficient, Biometrika, 41, 278-280

Harley, B.I. (1956): Some properties of an angular transformation for the correlation coefficient, Biometrika 43 (1956), 219-223

Harris, B. (1967): Spectral analysis of time series, New York 1967

Harter, H.L. (1964): Criteria for best substitute interval estimators with an application to the normal distribution, J. Am. Statist. Ass. 59 (1964), 1133-1140

Harter, H.L. (1974/75): The method of least squares and some alternatives (five parts) International Statistics Review 42 (1974), 147-174, 235-264, 43 (1975), 1-44, 125-190, 269-278

Harter, H.L. (1977): The non-uniqueness of absolute values regression, Commun. Statist. Simul. Comput. 6 (1977), 829-838

Hartley, H.O. and J.N.K. Rao (1967): Maximum likelihood estimation for the mixed analysis of variance model, Biometrika 54 (1967), 93-108

Hartinger, H. and Brunner, F.K. (1999): Variances of GPS Phase Observations: The SIGMA-Model. GPS Solution 2/4: 35-43, 1999

Hartman, P. and G.S. Watson (1974): Normal distribution functions on spheres and the modified Bessel function, Ann.Prob. 2 (1974), 593-607

Hartmann, C., Van Keer Berghen, P., Smeyersverbeke, J. and D.L. Massart (1997): Robust orthogonal regression for the outlier detection when comparing two series of measurement results, Analytica Chimica Acta 344 (1997), 17-28

Hartung, J. (1978): Zur Verwendung von Vorinformation in der Regressionsanalyse, Tech. report, Inst. für Angew. Statistik, Universität Bonn, Germany 1978

Hartung, J. (1981): Non-negative minimum biased invariant estimation in variance componet models, Annals of Statistics 9(1981), 278-292

Hartung, J. (1982): Statistik, pp. 611-613, Oldenburg Verlag, Miinchen
Hartung, J. (1999): Ordnungserhaltende positive Varianzschätzer bei gepaarten Messungen ohne Wiederholungen, Allg. Statistisches Archiv 83 (1999), 230-247
Hartung, J. and B. Voet (1986): Best invariant unbiased estimators for the mean squared error of variance component estimators, J. Am. Statist. Ass. 81 (1986), 689-691
Hartung, J. and Elpelt, B. (1989): Multivariate Statistik, Oldenbourg Verlag, München 1989
Hartung, J., Elpelt, B. and K.H. Klösener (1995): Statistik, Oldenbourg Verlag, München 1995
Hartung, J. and K.H. Jöckel (1982): Zuverlässigkeits- und Wirtschaftlichkeitsüberlegungen bei Straßenverkehrssignalanlagen, Qualität und Zuverlässigkeit 27 (1982), 65-68
Hartung, J. and D. Kalin (1980): Zur Zuverlässigkeit von Straßenverkehrssignalanlagen, Qualität und Zuverlässigkeit 25 (1980), 305-308
Hartung, J. and H.J. Werner (1980): Zur Verwendung der restringierten Moore-Penrose-Inversen beim Testen von linearen Hypothesen, Z. Angew. Math. Mechanik 60 (1980), T344-T346
Harvey, A.C. (1993): Time series models, 2nd ed., Harvester Wheatsheaf, New York 1993
Harville, D.A. (1976): Extension of the Gauss-Markov theorem to include the estimation of random effects, Annals of Statistics 4 (1976), 384-395
Harville, D.A. (1977): Maximum likelihood approaches to variance component estimation and to related problems, J. Am. Statist. Ass. 72 (1977), 320-339
Harville, D.A. (1997): Matrix algebra from a statistician's perspective, Springer-Verlag, Heidelberg Berlin New York 1997
Harville, D.A. (2001): Matrix algebra: exercises and solutions, Springer-Verlag, Heidelberg Berlin New York 2001
Haselett, J. (1989): Space time modeling in meterology- a review, Bul. Int. Statist. 53 (1989) 229-246
Hasselman, K., Munk, W. and MacDonald, G. (1963): Bispectrum of ocean waves in: M. Rosenblatt: Timeseries analysis, New York 1963
Hassibi, A. and S. Boyd (1998): Integer parameter estimation in linear models with applications to GPS, JEEE Trans. on Signal Processing 46 (1998), 2938-2952
Haslett, J. (1989): Space time modeling in meteorology - a review, Bull. lnt. Stat. lnst., 53, 229-246, 1989
Haslett, J. and K. Hayes (1998): Residuals for the linear model with general covariance structure, J. Roy. Statist. Soc. B60 (1998), 201-215
Hastie, T.J. and R.J. Tibshirani (1990): Generalized additive models, Chapman and Hall, Boca Raton 1990
Hauser, M.A., Pötscher, B.M. and E. Reschenhofer (1999): Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures, Empirical Economics 24 (1999), 243-269
Haussdorff, F. (1901): Beiträge zur Wahrscheinlichkeitsrechnung, Königlich Sächsische Gesellschaft der Wissenschaften zu Leipzig, berichte Math. Phys. Chasse 53 (1901), 152-178
Hawkins, D.M. (1993): The accuracy of elemental set approximation for regression, J. Am. Statist. Ass. 88 (1993), 580-589
Hayakawa, T. and Puri, M.L. (1985): Asymptotic distributions of likelihood ratio criterion for testing latent roots and latent vectors of a covariance matrix under an elliptical population, Biometrica, 72, 331-338
Hayes, K. and J. Haslett (1999): Simplifying general least squares, American Statistician 53 (1999), 376-381
O. Hazlett: On the theory of associative division algebras, Trans. American Math. Soc. 18 (1917) 167-176
He, K. (1995): The robustness of bootstrap estimator of variance, J. Ital. Statist. Soc. 2 (1995), 183-193
He, X. (1991): A local breakdown property of robust tests in linear regression, J. Multivar. Anal. 38, 294-305, 1991
He, X., Simpson, D.G. and Portnoy, S.L. (1990): Breakdown robustness of tests, J. Am. Statist. Ass. 85, 446-452, 1990

Healy, D.M. (1998): Spherical Deconvolution, J. Multivar. Anal. 67 (1998), 1-22
Heck, B. (1981): Der Ein uß einzelner Beobachtungen auf das Ergebnis einer Ausgleichung und die Suche nach Ausreißern in den Beobachtungen, Allgemeine Vermessungsnachrichten 88 (1981), 17-34
Heck, B.(1987): Rechenverfahren und Auswertemodelle der Landesvermessung, Wichmann Verlag, Karlsruhe 1987.
Heideman, M.T., Johnson, D.H. and C.S. Burrus (1984): Gauss and the history of the fast Fourier transform, JEEE ASSP Magazine 1 (1984), 14-21
Heiligers, B. (1994): E-optimal designs in weighted polynomial regression, Ann. Stat. 22 (1994), 917-929
Hein, G. (1986): Integrated geodesy. In: Suenkel H (ed), Mathematical and numerical techniques in physical geodesy. Lecture Notes in Earth Sciences vol. 7. Springer, Heidelberg, pp. 505-548
Hein, G. and Landau, H. (1983): A contribution to 3D - operational geodesy, Part 3: Opera - a multi-purpose program for operational adjustment of geodetic observations of terrestrial type. Deutsche Geodätische Kommission, 8ayerische Akademie der Wissenschaften, Report B 264, München
Heine, V. (1955): Models for two-dimensional stationary stochastic processes, Biometrika 42 (1955), 170-178
Heinrich, L. (1985): Nonuniform estimates, moderate and large derivations in the central limit theorem for m-dependent random variable, Math. Nachr. 121 (1985), 107-121
Heitz, S. (1967): Habilitationsschrift, Bonn 1967
Hekimoglu, S. (1998): Application of equiredundancy design to M-estimation, Journal of Surveying Engineering 124 (1998), 103-124
Hekimoglu, S. and M. Berber (2003): Effectiveness of robust methods in heterogeneous linear models, Journal of Geodesy 76 (2003), 706-713
Helmert, F.R. (1875): Über die Berechnung des wahrscheinlichen Fehlers aus einer endlichen Anzahl wahrer Beobachtungsfehler, Z. Math. U. Physik 20 (1875), 300-303
Helmert, F.R. (1876): Diskussion der Beobachtungsfehler in Koppes Vermessung für die Gotthardtunnelachse, Z. Vermessungswesen 5 (1876), 129-155
Helmert, F.R. (1876): Die Genauigkeit der Formel von Peters zur Berechnung der wahrscheinlichen Beobachtungsfehler direkter Beobachtungen gleicher Genauigkeit, Astronomische Nachrichten 88 (1876), 113-120
Helmert, F.R. (1876): Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Fehlers direkter Beobachtungen gleicher Genauigkeit, Astron. Nachrichten 88 (1976), 113-132
Helmert, F.R. (1876a): Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler, Z. Math. U. Phys. 21 (1876), 192-218
Helmert, F.R. (1907): Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, mit Anwendungen auf die Geodäsie, die Physik und die Theorie der Messinstrumente, B.G. Teubner, Leipzig Berlin 1907
Helmert, F.R. (1924): Die Ausgleichungsrechnung nach der. Methode der Kleinsten Quadrate. 3. Auflage. Teubner. Leipzig. 1924
Henderson, H.V. (1981): The vec-permutation matrix, the vec operator and Kronecker products: a review, Linear and Multilinear Algebra 9 (1981), 271-288
Henderson, H.V. and S.R. Searle (1981): Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics
Henderson, H.V. and S.R. Searle (1981): On deriving the inverse of a sum of matrices, SIAM Review 23 (1981), 53-60
Henderson, H.V., Pukelsheim, F. and S.R. Searle (1983): On the history of the Kronecker product, Linear and Multilinear Algebra 14 (1983), 113-120
Hendriks, H. and Z. Landsman (1998): Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions, J. Multivar. Anal. 67 (1998), 227-243
Hengst, M. (1967): Einfuehrung in die Mathematische Statistik und ihre Anwendung, Bibliographisches Institut, Mannheim 1967

Henrici, P. (1962): Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices, Numer. Math. 4 (1962), 24-40
Hensel, K. (1968): Leopold Kronecker's Werke, Chelsea Publ. Comp., New York 1968
Herzberg, A.M. and A.V. Tsukanov (1999): A note on the choice of the best selection criterion for the optimal regression model, Utilitas Mathematica 55 (1999), 243-254
Hesse, K. (2003): Domain decomposition methods in multiscale geopotential determination from SST and SGG, Berichte aus der Mathematik, Shaker Verlag, Aachen 2003
Hetherington, T.J. (1981): Analysis of directional data by exponential models, Ph.D. Thesis, University of California, Berkeley 1981
Heuel, S. and Forstner, W. (2001): Matching, Reconstructing and Grouping 30 Lines From Multiple Views Using Uncertain Projective Geometry. CVPR 2001, Posters Session 4.2001, S. 721.
Heyde, C.C. (1997): Quasi-likelihood and its application. A general approach to optimal parameter estimation, Springer-Verlag, Heidelberg Berlin New York 1997
Hickernell, F.J. (1999): Goodness-of-fit statistics, discrepancies and robust designs, Statistics and Probability Letters 44 (1999), 73-78
Hida, T. and Si, S. (2004): An innovation approach to random fields: application of white noice theory, World Scientific 2004
Higham, N.J. and F. Tisseur (2000): A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra, SIAM J. Matrix Anal. Appl. 21 (2000), 1185-1201
Hinze, J.O. (1975): Turbulence, Mc Graw Hill, New York 1975
Heikkinen, M. (1982): Geschlossene Formeln zur Berechnung räumlicher geodätischer Koordinaten aus rechtwinkligen Koordinaten, *Zeitschrift für Vermessungswesen* 107 (1982) 207-211.
Hein, G. (1982a): A contribution to 3D-operational Geodesy (Principles and observation equations of terrestrial type), DGK Reihe B, Heft Nr. 258/VII pp. 31-64.
Hein, G. (1982b): A contribution to 3D-operational Geodesy part 2, DGK Reihe B, Heft Nr. 258/VII pp. 65-85.
Heiskanen, W. A. and Moritz, H. (1976): Physical Geodesy, Freeman and Company, London 1976.
Hemmleb, G. (1986): Vor hundert Fahren wurde Friedvich Robert Helmert Direkto des Pvenssischen Geodetischen Institutes, Vermessungstechnik 34, (1986), pp. 178
Hide, R. (2010): A path od discovery in geophysical fluid dynamics, Royal Astron. Soc., Astronomy and Geophysics, 51 (2010) 4.16-4.23
Hinde, J. (1998): Overdispersion: models and estimation, Comput. Stat. & Data Anal. 27 (1998), 151-170
Hinkley, D. (1979): Predictive likelihood, Ann. Statist. 7 (1979), 718-728
Hinkley, D., Reid, N. and E.J. Snell (1990): Statistical theory and modelling, Chapman and Hall, Boca Raton 1990
Hiroven, R.A.: a) Publ. Inst. Geod. Photogr. Cart., Ohio State University Report No.4, 1956; b) Ann. Acad. Scient. 1962.
Hirvonen, R.A. (1971): Adjustment by least squares in geodesy and photogrammetry, Frederick Ungar Publ. Co., New York
Hirvonen, R.A. and Moritz, H. (1963): Practical computation of gravity at high altitudes, Institute of Geodesy, Photogrammetry and Cartography, Ohio State University, Report No. 27, Ohio 1963.
Hjorth, J.S.U. (1993): Computer intensive statistical methods, Chapman and Hall, Boca Raton 1993
Ho, L.L. (1997): Regression models for bivariate counts, Brazilian J. Probability and Statistics 11 (1997), 175-197
Hoaglin, D.C. and R.E. Welsh (1978): The Hat Matrix in regression and ANOVA, The American Statistician 32 (1978), 17-22
Hodge, W.V.D. (1941): Theory and applications of harmonic integrals, Cambridge University Press, Cambridge 1941
Hodge, W.V.D. and D. Pedoe (1968): Methods of algebraic geometry, I, Cambridge University Press, Cambridge 1968

Hoel, P.G. (1965): Minimax distance designs in two dimensional regression, Ann. Math. Statist. 36 (1965), 1097-1106

Hoel, P.G., S.C. Port and C.J. Stone (1972): Introduction to stochastic processes, Houghton Mifflin Publ., Boston 1972

Hoerl, A.E. and Kennard, R.W. (1970): Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1970), 55-67

Hoerl, A.E. and R.W. Kennard (2000): Ridge regression: biased estimation for nonorthogonal problems, Technometrics 42 (2000), 80-86

Höpke, W. (1980): Fehlerlehre und Ausgleichungsrechnung, De Gruyter, Berlin 1980

Hoffmann, K. (1992): Improved estimation of distribution parameters: Stein-type estimators, Teubner-Texte zur Mathematik, Stuttgart/Leipzig 1992

Hofmann, B. (1986): Regularization for applied inverse and ill-posed problems, Teubner Texte zur Mathematik 85, Leipzig 1986

Hofman-Wellenhof, B., Lichtenegger, H. and Collins, J. (1994): Global Positioning System: Theory and practice, Springer-Verlag, Wien 1994.

Hofmann-Wellenhof, B., Moritz, H. (2005): Physical Geodesy, Springer Wien, New York

Hogg, R.V. (1972): Adaptive robust procedures: a partial review and some suggestions for future applications and theory, J. Am. Statist. Ass. 43 (1972), 1041-1067

Hogg, R.V. and R.H. Randles (1975): Adaptive distribution free regression methods and their applications, Technometrics 17 (1975), 399-407

Hong, C.S. and H.J. Choi (1997): On L1 regression coefficients, Commun. Statist. Simul. Comp. 26 (1997), 531-537

Hora, R.B. and R.J. Buehler (1965): Fiducial theory and invariant estimation, Ann. Math. Statist. 37 (1965), 643-656

Horaud, R., Conio, B. and Leboulleux, O. (1989): An Analytical Solution for the Perspective 4-Point Problem, *Computer Vision, Graphics and Image Processing* 47 (1989) 33-44.

Horn, R.A. (1989): The Hadamard product, in Matrix Theory and Applications, C.R. Johnson, ed., Proc. Sympos. Appl. Math. 40 (1989), 87-169

Horn, R.A. and C.R. Johnson (1990): Matrix analysis, Cambridge University Press, Cambridge 1990

Horn, R.A. and C.R. Johnson (1991): Topics on Matrix analysis, Cambridge University Press, Cambridge 1991

Horn, S.D. and R.A. Horn (1975): Comparison of estimators of heteroscedastic variances in linear models, Am. Stat. Assoc., 70, 872-879.

Hornoch, A.T. (1950): über die Zurückführung der Methode der kleinsten Quadrate auf das Prinzip des arithmetischen Mittels, Österr. Zeitschrift für Vermessungswesen 38 (1950), 13-18

Hoschek, J. and Lasser, D. (1992): Grundlagen der geometrischen Datenverarbeitung. B.G. Teubner, Stuttgart.

Hosking, J.R.M. and J.R. Wallis (1997): Regional frequency analysis. An approach based on L-moments, Cambridge University Press 1997

Hosoda, Y. (1999): Truncated least-squares least norm solutions by applying the QR decomposition twice, trans. Inform. Process. Soc. Japan 40 (1999), 1051-1055

Hotelling, H. (1931): The generalization of Student's ratio, Ann. Math. Stat., 2, 360-378.

Hotelling, H. (1933): analysis of a complex of statistical variables into principal components, J. Educ. Psych., 24, 417-441, 498-520.

Hotelling, H. (1951): A generalized T test and measure of multivariate dispersion, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley and Los Angeles, 23-41.

Hotelling, H. (1953): New light on the correlation coefficient and its transform, J. Roy. Statist. Soc. B15 (1953), 225-232

Householder, A.S. (1958): Unitary Triangulation of a Nonsymmetric Matrix. J. Assoc. Compo Mach., Vol. 5, 339.

Howind, J., Böhringer, M., Mayer, M., Kutterer, H., Lindner, K.v. and Heck, B (2000): Korrelationsstudien bei GPS-Phasenbeobachtungen. In: Dietrich, R. (ed): Deutsche Beiträge

zu GPS-Kampagnen des Scientific Committee on Antartic Research (SCAR) 1995-1998, Deutsche Geodätische Kommission B310: 201-205, Munich, 2000.
Hoyle, M.H. (1973), Transformations - an introduction and a bibliography, Int. Statist. Review 41 (1973), 203-223
Hsu, J.C. (1996): Multiple comparisons, Chapman and Hall, Boca Raton 1996
Hsu, P.L. (1938a): Notes on Hotelling's, generalized T, Ann. Math. Stat., 9, 231-243
Hsu, P.L. (1938b): On the best unbiassed quadratic estimate of the variance, Statist. Res. Mem., Univ. London 2 (1938), 91-104.
Hsu, P.L. (1940a): On generalized analysis of variance (1), Biometrika, 31, 221-237
Hsu, P.L. (1940s): An algebraic derivation of the distribution of rectangular coordinates, Proc. Edinburgh Math. Soc. 6 (1940), 185-189
Hsu, R. (1999): An alternative expression for the variance factors in using Iterated Almost Unbiased Estimation, Journal of Geodesy 73 (1999), 173-179
Hsu, Y.S., Metry, M.H. and Y.L. Tong (1999): Hypotheses testing for means of dependent and heterogeneous normal random variables, J. Statist. Planning and Inference 78 (1999), 89-99
Huang, J.S. (1999): Third-order expansion of mean squared error of medians, Statistics & Probability Letters 42 (1999), 185-192
Hüber, P.J. (1964): Robust estimation of a location parameter, Annals Mathematical Statistics 35 (1964), 73-101
Hüber, P.J. (1972): Robust statistics: a review, Annals Mathematical Statistics 43 (1972), 1041-1067
Hüber, P.J. (1981): Robust Statistics, J. Wiley, New York 1981
Hüber, P.J. (1984): Finite sample breakdown of M- and P- estimators, Ann. Stat., 12 (1984), pp. 119-126
Hüber, P.J. (1985): Projection persuit, Annals of Statistics 13, (1984), pp. 435-475
Huckeman, S., Hotz, T. and A. Munk (2007): Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions, Statistica Sinicia 2007
Huckeman, S. (2009): On the meaning of the mean slope, Digital Signal Processing 16 (2009) 468-478
Huckeman, S. (2010): On the meaning of mean slope, arXiv. (1002.07950.) (stat. ME), preprint 2010
Huet, S., A. Bouvier, M.A. Gruet and E. Jolivet (1996): Statistical tools for nonlinear regression, Springer-Verlag, Heidelberg Berlin New York 1996
Huffel, van S. and H. Zha (1991a): The restricted total least squares problem: Formulation, algorithm, and properties, SIAM J. Matrix Anal. Appl. 12 (1991), 292-309
Huffel, van S. and H. Zha (1991b): The total least squares problem, SIAM J. Matrix Anal. Appl. 12 (1991), 377-407
Huffel, van S. and H. Zha (1993): The total least square problem, in: C.R. Rao (ed.), Handbook of Statistics, Vol. 2, pp. 337-408, North Holland, Amsterdam-London-New York-Tokyo 1993
Hugentobler, U. and Dach, R. and Fridez, P. (2004): Bernese GPS Software Version 5.0. University of Bern, 2004.
Humak K M. S. (1977): Statistische Methoden der Modellbildung Band I. Statistische Inferenz für lineare Parameter. Akademie-Verlag, Berlin, 1977.
Hunter, D.B. (1995): The evaluation of Legendre functions of the second kind, Numerical Algorithms 10 (1995), 41-49
Huwang, L. and Y.H.S. Huang (2000): On errors-in-variables in polynomical regression-Berkson case, Statistica Sinica 10 (2000), 923-936
Hwang, C. (1993): Fast algorithm for the formation of normal equations in a least-squares spherical harmonic analysis by FFT, manuscripta geodaetica 18 (1993), 46-52
Ibragimov, F.A. and R.Z. Kasminskii (1981): Statistical estimation, asymptotic theory, Springer-Verlag, Heidelberg Berlin New York 1981
ICD (2000): Interface Control Document, Navstar GPS Space Segment/Navigation User Interfaces, ICDGPS-200C

Ihorst, G. and G. Trenkler (1996): A general investigation of mean square error matrix superiority in linear regression, Statistica 56 (1996), 15-23

Imhof, L. (2000): Exact designs minimizing the integrated variance in quadratic regression, Statistics 34 (2000), 103-115

Imhof, J.P. (1961): Computing the distribution of quadratic forms in normal variables, Biometrika 48 (1961), 419-426

Imon, A.H.M.R. (2002): On deletion residuals. Calcutta Stat. Assoc. Bull. 53, 209-210 (2002), 65-79.

Inda, de M.A. et al. (1999): Parallel fast Legendre transform, proceedings of the ECMWF Workshop Towards TeraComputing - the Use of Parallel Processors in Meteorology", Worls Scientific Publishing Co 1999

Ireland, K. and Rosen, M. (1990): A classical introduction to modern number theory, Springer-Verlag, New York 1990.

Irle, A. (1990): Sequentialanalyse: Optimale sequentielle Tests, Teubner Skripten zur Mathematischen Stochastik. Stuttgart 1990

Irwin, J.O. (1937): On the frequency distribution of the means of samples from a population having any law of frequency with finite moments with special reference to Pearson's Type II, Biometrika 19 (1937), 225-239

Isham, V. (1993): Statistical aspects of chaos, in: Networks and Chaos, Statistical and Probabilistic Aspects (ed. D.E. Barndorff-Nielsen et al.), 124-200, Chapman and Hall, London 1993

Ishibuchi, H., Nozaki, K. and Tanaka H. (1992): Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and Systems 52 (1992), 21-32

Ishibuchi, H., Nozaki, K., Yamamoto, N. and Tanaka, H. (1995): Selecting fuzzy if-then rules for classification problems using genetic algorithms, IEEE Transactions on Fuzzy Systems 3 (1995), 260-270

Ishibuchi, H. and Murata, T. (1997): Minimizing the fuzzy rule base and maximizing its performance by a multi-objective genetic algorithm, in: Sixth FUZZ-IEEE Conference, Barcelona 1997, 259-264

Ishimaru, A. (1978): Wave Propagation and Scattering in Random Media, Vol.2. Academic Press, New York, 1978.

Ivanov, A.V., S. Zwanzig, (1981): An asymptotic expansion for the distribution of the least squares estimator of a vector parameter in nonlinear regression. Soviet Math. Dokl. Vol. 23 (1981), No.1, 118-121.

Ivanov, A.V., S. Zwanzig, (1996): Saddlepoint approximation in nonlinear regression. Theory of Stochastic Processes, 2, (18),1996,156-162.

Ivanov, A.V., S. Zwanzig, (1983): An asymptotic expansion of the distribution of the least squares estimator in the nonlinear regression model. Statistics, VoL 14 (1983), No. 1, 7-27.

Izenman, A.J. (1975): Reduced-rank regression for the multivariate linear model, J. Multivar. Anal. 5 (1975), 248-264

Jacob, N. (1996): Pseudo-differential operators and Markov processes, Akademie Verlag, Berlin 1996

Jacobi, C.G.J. (1841): Deformatione et proprietatibus determinatum, Crelle's J. reine angewandte Mathematik, Bd.22

Jacod, J., Protter, P. (2000): Probability essentials, Springer-Verlag, Heidelberg Berlin New York 2000

Jaeckel, L.A. (1972): Estimating regression coefficients by minimizing the dispersion of the residuals, Annals Mathematical Statistics 43 (1972), 1449-1458

Jajuga, K. (1995): On the principal components of time series, Statistics in Transition 2 (1995), 201-205

James, A.T. (1954): Normal multivariate analysis and the orthogonal group, Ann. Math. Statist. 25 (1954), 40-75

Jammalamadaka S. R., Sengupta D. (1999): Changes in the general linear model: A unified approach. Linear Algebra Appl. 289, 1-3 (1999), 225-242.

Jammalamadaka, S.R., SenGupta, A. (2001): Topics in circular statistics, World Scientific, Singapore 2001
Janacek, G. (2001): Practical time series, Arnold, London 2001
Jara, A., Quintana, E. and E. Sammertin (2008): Linear effects mixed models with skew-elliptical distributions: A Bayesian approach. In Computational Statist. Data Analysis 52 (2008) 5033-5045
Jeffery, H. (1961): Cartesian tensors, Cambridge University Press
Jekeli, C. (2001): Inertial Navigation Systems with Geodetic Applications. Walter deGruyter, Berlin-New York, 352p
Jennison, C. and B.W. Turnbull (1997): Distribution theory of group sequential t, x2 and F-Tests for general linear models, Sequential Analysis 16 (1997), 295-317
Jennrich, R.I. (1969): Asymptotic properties of nonlinear least squares estimation, Ann. Math. Statist. 40 (1969), 633-643
Jensen, J.L. (1981): On the hyperboloid distribution, Scand. J. Statist. 8 (1981), 193-206
Jiang, J. (1997): A derivation of BLUP - Best linear unbiased predictor, Statistics & Probability Letters 32 (1997), 321-324
Jiang, J. (1999): On unbiasedness of the empirical BLUE and BLUP, Statistics & Probability 41 (1999), 19-24
Jiang, J., Jia, H. and H. Chen (2001): Maximum posterior estimation of random effects in generalized linear mixed models, Statistica Sinica 11 (2001), 97-120
Joe, H. (1997): Multivariate models and dependence concepts, Chapman and Hall, Boca Raton 1997
Jörgensen, B. (1984): The delta algorithm and GLIM, Int. Statist. Review 52 (1984), 283-300
Jörgensen, B. (1997): The theory of dispersion models, Chapman and Hall, Boca Raton 1997
Jörgensen, B., Lundbye-Christensen, S., Song, P.X.-K. and L. Sun (1996b): State space models for multivariate longitudinal data of mixed types, Canad. J. Statist. 24 (1996b), 385-402
Jorgensen, P.C., Kubik, K., Frederiksen, P. and W. Weng (1985): Ah, robust estimation! Australian Journal of Geodesy, Photogrammetry and Surveying 42 (1985), 19-32
John, S. (1962): A tolerance region for multivariate normal distributions, Sankya A24 (1962), 363-368
Johnson, N.L. and S. Kotz (1972): Distributions in statistics: continuous multivariate distributions, J. Wiley, New York 1972
Johnson, N.L., Kotz, S. and X. Wu (1991): Inspection errors for attributes in quality control, Chapman and Hall, Boca Raton 1991
Johnson, N.L. and S. Kotz and Balakrishnan, N. (1994): Continuous Univariate Distributions, vol 1. 2nd ed, John Wiley & Sons, New York, 784p
Johnson, N.L. and S. Kotz and Balakrishnan, N. (1995): Continuous Univariate Distributions, vol 2. 2nd ed, John Wiley & Sons, New York, 752p
Jonge de, P.J., Tiberius, C.C.J.M. (1996a): The LAMBDA method for integer ambiguity estimation: implementation aspects. Publications of the Delft Computing Centre, LGR-Series No. 12
Jonge de, P.J., Tiberius, C.C.J.M. (1996b): Integer estimation with the LAMBDA method. In: Beutler, G et al. (eds) Proceedings IAG symposium No. 115, GPS trends in terrestrial, airborne and spaceborne applications. Springer, Heidelberg, pp 280-284
Jonge de, P.J., Tiberius, C.C.J.M., Teunissen, P.J.G. (1996): Computational aspects of the LAMBDA method for GPS ambiguity resolution. In: Proceedings ION GPS-96, pp 935-944
Jordan, W., Eggert, O. and Kneissl, M. (1956): Handbuch der Vermessungskunde, Vol. III, Hohenmessung. Metzlersche Verlagsbuchhandlung, 10th edition, Stuttgart.
Jorgenson, B. (1999): Multivariate dispersion models, J. multivariate Analysis, 74, (1999), pp. 267-281
Joshi, V.M. (1966): Admissibility of confidence intervals, Ann. Math. Statist. 37 (1966), 629-638
Judge, G.G. and M.E. Bock (1978): The statistical implications of pre-test and Stein-rule estimators in econometrics, Amsterdam 1978

Judge, G.G., W.E. Griffiths, R.C. Hill, and T.C. Lee (1980): The Theory and Practice of Econometrics, New York: John Wiley & Sons, 793 pages
Judge, G.G. and T.A. Yancey (1981): Sampling properties of an inequality restricted estimator, Economics Lett. 7 (1981), 327-333
Judge, G.G. and T.A. Yancey (1986): Improved methods of inference in econometrics, Amsterdam 1986
Jukièc, D. and R. Scitovski (1997): Existence of optimal solution for exponential model by least squares, J. Comput. Appl. Math. 78 (1997), 317-328
Jupp, P.E. and K.V. Mardia (1980): A general correlation coefficient for directional data and related regression problems, Biometrika 67 (1980), 163-173
Jupp, P.E. and K.V. Mardia (1989): A unified view of the theory of directional statistics, 1975-1988, International Statist. Rev. 57 (1989), 261-294
Jureckova, J. (1995): Affine- and scale-equivariant M-estimators in linear model, Probability and Mathematical Statistics 15 (1995), 397-407
Jureckova, J.; Sen, P.K. (1996): In Robust Statistical Procedures Asymptotics and Interrelation, John Wiley & Sons: New York, USA, 1996.
Jurisch, R. and G. Kampmann (1997): Eine Verallgemeinerung des arithmetischen Mittels für einen Freiheitsgrad bei der Ausgleichung nach vermittelnden Beobachtungen, Zeitschrift für Vermessungswesen 122 (1997), 509-521
Jurisch, R. and G. Kampmann (1998): Vermittelnde Ausgleichungsrechnung mit balancierten Beobachtungen - erste Schritte zu einem neuen Ansatz, Zeitschrift für Vermessungswesen 123 (1998), 87-92
Jurisch, R. and G. Kampmann (2001): Plücker-Koordinaten - ein neues Hilfsmittel zur Geometrie-Analyse und Ausreissersuche, Vermessung, Photogrammetrie und Kulturtechnik 3 (2001), 146-150
Jurisch, R. and G. Kampmann (2002): Teilredundanzen und ihre natürlichen Verallgemeinerungen, Zeitschrift für Vermessungswesen 127 (2002), 117-123
Jurisch, R., Kampmann, G. and B. Krause (1997): über eine Eigenschaft der Methode der kleinsten Quadrate unter Verwendung von balancierten Beobachtungen, Zeitschrift für Vermessungswesen 122 (1997), 159-166
Jurisch, R., Kampmann, G. and J. Linke (1999a): über die Analyse von Beobachtungen in der Ausgleichungsrechnung - Äußere und innere Restriktionen, In "Qua Vadis Geodesia", Festschrift für Erik W. Grafarend, Schriftenreihe der Institute des Studienganges Geodäsie und Geoinformatik 1999, 207-230
Jurisch, R., Kampmann, G. and J. Linke (1999b): über die Analyse von Beobachtungen in der Ausgleichungsrechnung - Teil I, Zeitschrift für Vermessungswesen 124 (1999), 350-357
Jurisch, R., Kampmann, G. and J. Linke: über die Analyse von Beobachtungen in der Ausgleichungsrechnung - Teil II, Zeitschrift für Vermessungswesen 124 (1999), 350-357
Kagan, A.M., Linnik, J.V. and C.R. Rao (1965): Characterization problems of the normal law based on a property of the sample average, Sankya A27 (1965), 405-406
Kagan, A. and L.A. Shepp (1998): Why the variance?, Statist. Prob. Letters 38 (1998), 329-333
Kagan, A. and Z. Landsman (1999): Relation between the covariance and Fisher information matrices, Statistics & Probability Letters 42 (1999), 7-13
Kagan, Y.Y. (1992): Correlations of earthquake focal mechanisms, Geophys. J. Int., 110, 305-320
Kagan, Y.Y. (1990): Random stress and earthquake statistics: spatial dependence, Geophys. J. Int., 102, 573-583
Kagan, Y.Y. (1991): Likelihood analysis of earthquake catalogues, Geophys. J. Int., 106, 135-148
Kagan, Y.Y. and L. Knopoff (1985):Lorder statistical moment of the seismic moment tensor, Geophys. J. R. astr. Soc., 81,429-444
Kahmen, H. and Faig, W. (1988): Surveying, Walter de Gruyter, Berlin 1988.
Kahn, M., Mackisack, M.S., Osborne, M.R. and G.K. Smyth (1992): On the consistency of Prony's method and related algorithms, Jour. Comp. and Graph. Statist. 1 (1992), 329-349

Kahng, M.W. (1995): Testing outliers in nonlinear regression, Journal of the Korean Stat. Soc. 24 (1995), 419-437

Kakkuri, J. and Chen, R. (1992): On horizontal crustal strain in Finland, Bull. Geod., 66, 12-20

Kakwani, N.C. (1968): Note on the unbiasedness of mixed regression estimation, Econometrica 36 (1968), pp. 610-611

Kalkbrener, M. (1990a): Primitive Polynomial Remainder Sequences. Technical Report RISCLinz Series no. 90-01.0, Research Institute for Symbolic Computation, Univ. of Linz.

Kalkbrener, M. (1990b): Solving Systems of Bivariate Algebraic Equations by using Primitive Polynomial Remainder Sequences. Technical Report RISC-Linz Series no. 90-21.0, Research Institute for Symbolic Computation, Univ. of Linz.

Kalkbrener, M. (1990c): Schreiben von M. Kalkbrener, Research Institute for Symbolic Computation, Johannes Kepler Universitat Linz an P. Lohse, Geodätisches Institut der Universitat Stuttgart vom 30. Mai 1990. nicht veroffentlicht.

Kalkbrener, M. (1991): Three Contributions to Elimination Theory. Dissertation, Institut fiir Mathematik, Technisch-Naturwissenschaftliche Fakultat, Johannes Kepler Universit.at Linz.

Kallianpur, G. (1963): von Mises functionals and maximum likelihood estimation, Sankya A25 (1963), 149-158

Kallianpur, G. and Y.-T. Kim (1996): A curious example from statistical differential geometry, Theory Probab. Appl. 43 (1996), 42-62

Kallenberg, O. (1986): Random measures, Academic Press, London 1984

Kallenberg, O. (1997): Foundations of modern probability, Springer-Verlag, Heidelberg Berlin New York 1997

Kaleva, O. (1994): Interpolation of fuzzy data, Fuzzy Sets Sys, 61, 63-70.

Kalman, R.: A New Approach to Linear Filtering and Prediction Problems, Trans. ASME Ser. D, J. Basic Eng. 82, 35 (1960J.

Kaminsky, K.S. and P.I. Nelson (1975): Best linear unbiased prediction of order statistics in location and scale families, J. Am. Statist. Ass. 70 (1975), 145-150

Kaminsky, K.S. and L.S. Rhodin (1985): Maximum likelihood prediction, Ann. Inst. Statist. Math. A37 (1985), 507-517

Kampmann, G. (1988): Zur kombinativen Norm-Schätzung mit Hilfe der L1-, der L2- und der Boskovic-Laplace-Methode mit den Mittlen der linearen Programmierung, Ph.D. Thesis, Bonn University, Bonn 1988

Kampmann, G. (1992): Zur numerischen überführung verschiedener linearer Modelle der Ausgleichungsrechnung, Zeitschrift für Vermessungswesen 117 (1992), 278-287

Kampmann, G. (1994): Robuste Deformationsanalyse mittels balancierter Ausgleichung, Allgemeine Vermessungs-Nachrichten 1 (1994), 8-17

Kampmann, G. (1997): Eine Beschreibung der Geometrie von Beobachtungen in der Ausgleichungsrechnung, Zeitschrift für Vermessungswesen 122 (1997), 369-377

Kampmann, G. and B. Krause (1996): Balanced observations with a straight line fit, Bolletino di Geodesia e Scienze Affini 2 (1996), 134-141

Kampmann, G. and B. Krause (1997): A breakdown point analysis for the straight line fit based on balanced observations, Bolletino di Geodesia e Scienze Affini 3 (1997), 294-303

Kampmann, G. and B. Renner (1999): über Modellüberführungen bei der linearen Ausgleichungsrechnung, Allgemeine Vermessungs-Nachrichten 2 (1999), 42-52

Kampmann, G. and B. Renner (2000): Numerische Beispiele zur Bearbeitung latenter Bedingungen und zur Interpretation von Mehrfachbeobachtungen in der Ausgleichungsrechnung, Zeitschrift für Vermessungswesen 125 (2000), 190-197

Kampmann, G. and B. Krause (2004): Zur statistischen Begründung des Regressionsmodells der balanzierten Ausgleichungsrechnung, Z. Vermessungswesen 129 (2004), 176-183

Kannan, N. and D. Kundu (1994): On modified EVLP and ML methods for estimating superimposed exponential signals, Signal Processing 39 (1994), 223-233

Kantz, H. and Scheiber, T. (1997): Nonlinear rime series analysis, Cambridge University Press, Cambridge 1997

Kanwal, R.P. (1971): Linear integral equations, theory and techniques. Academic Press, New York, London 1971

Kaplan, E.D. and Hegarty, C.J. (eds) (2006): Understanding GPS: Principles and Applications. 2nd ed, Artech

Karatzas, I. and S.E. Shreve (1991): Brownian motion and stochastic calculus, Springer-Verlag, Heidelberg Berlin New York 1991

Karcz, I., Forrai, J. and Steinberg, G. (1992): Geodetic Network for Study of Crustal Moven1ents Across the Jordan-Deas Sea Rift. Journal of Geodynamics. 16:123-133.

Kariya, T. (1989): Equivariant estimation in a model with an ancillary statistic, Ann. Statist 17 (1989), 920-928

Karlin, S. and W.J. Studden (1966): Tchebychev systems, Interscience, New York (1966)

Karlin, S. and W.J. Studden (1966): Optimal experimental designs, Ann. Math. Statist. 57 (1966), 783-815

Karman, T. (1937): Proc. Nat. Acad. Sci. Volume 23, Washington 1937

Karman, T. and L. Howarth: Proc. Roy. Soc. A 164 (1938).

Karman, T. (1937): On the statistical theory of turbulence, Proc. Nat. Acad. Sci. 23 (1937) DS-1U5.

Karman, T. and L. Howrath (1938): On the statistical theory of isotropic turbulence, Proc. Royal Soc. London A 164 (1938) 192-215

Karr, A.F. (1986): Pointe process and their statistical interference, Marcel Dekker, New York, 1986

Kaula, W.M.: a) J. Geoph. Res. 64, 2401 (1959); b, Rev. Geoph. 507 (1963).

Kasala, S. and T. Mathew (1997): Exact confidence regions and tests in some linear functional relationships, Statistics & Probability Letters 32 (1997), 325-328

Kasietczuk, B. (2000): Geodetic network adjustment by the maximum likelihood method with application of local variance, asymmetry and excess coefficients, Anno LIX - Bollettino di Geodesia e Scienze Affini 3 (2000), 221-235

Kass, R.E. and P.W. Vos (1997): Geometrical foundations of asymptotic inference, J. Wiley, New York 1997

Kastrup, H.A. (1962): Zur physikalischen Deutung und darstellungstheoretischen Analyse der konformen Transformationen von Raum und Zeit, Annalen der Physik 9 (1962), 388-428

Kastrup, H.A. (1966): Gauge properties of the Minkowski space, Physical Review 150 (1966), 1183-1193

Kay, S.M. (1988): Sinusoidal parameter estimation, Prentice Hall, Englewood Cliffs, N.J. 1988

Keller, J.B. (1975): Closest unitary, orthogonal and Hermitian operators to a given operator, Math. Mag. 46 (1975), 192-197

Kelly, R.J. and T. Mathew (1993): Improved estimators of variance components having smaller probability of negativity, J. Roy. Stat. Soc. B55 (1993), 897-911

Kelm R, (1978): Ist die Variantzshatzung nach Helmert MINQUE? Allgemeine VermessungsNachrichten, 85, 49-54.

Kemperman, J.H.B. (1956): Generalized tolerance limits, Ann. Math. Statist. 27 (1956), 180-186

Kendall, D.G. (1974): Pole seeking Brownian motion and bird navigation, J. Roy. Statist. Soc. B36 (1974), 365-417

Kendall, D.G. (1984): Shape manifolds, Procrustean metrics, and complex projective space, Bulletin of the London Mathematical Society 16 (1984), 81-121

Kendall, M.G. (1960): The evergreen correlation coefficient, 274-277, in: Essays on Honor of Harold Hotelling, ed. I. Olkin, Stanford University Press, Stanford 1960

Kendall, M.G. and Stuart, A. (1958): The advanced theory of statistics, Vol. 1, distribution theory, Charles Griffin and Company Ltd, London

Kendall, M.G. and Stuart, A. (1977): The advanced theory of statistics. Vol. 1: Distribution theory. Charles Griffin & Company Lt., London-High Wycombe, 1977 (4th ed.)

Kendall, M.G. and Stuart, A. (1979): The advanced theory of statistics. Vol. 2: Inference and relationship. Charles Griffin & Company Lt., London-High Wycombe, 1979 (4th ed.).

Kenney, C.S., Laub, A.J. and M.S. Reese (1998): Statistical condition estimation for linear systems, SIAM J. Scientific Computing 19 (1998), 566-584

Kent, J. (1976): Distributions, processes and statistics on spheres, Ph. D. Thesis, University of Cambridge
Kent, J.T. (1983): Information gain and a general measure of correlation, Biometrika 70 (1983), 163-173
Kent, J.T. (1997): Consistency of Procrustes estimators, J. R. Statist. Soc. B59 (1997), 281-290
Kerr, R.M. (1985): Higher order derivative correlations and alignment of small scale structures in isotropic numerical turbulence, J. Fluid Mech. 153 (1985) 31-58
Kertz, W. (1969): Einfuehrung in die Geophysik, BI - Taschenbuch, Mannheim 1969/71
Keshin, M. (2004): Directional statistics of satellite-satellite single-difference widelane phase biases, Artificial satellites, 39 (2004), pp. 305-324
Khan, R.A. (1998): Fixed-width confidence sequences for the normal mean and the binomial probability, Sequential Analysis 17 (1998), 205-217
Khan, R.A. (2000): A note on Hammersley's estimator of an integer mean, J. Statist. Planning and Inference 88 (2000), 37-45
Khatri, C.G. and C.R. Rao (1968): Solutions to some fundamental equations and their applications to characterization of probability distributions, Sankya A30 (1968), 167-180
Khatri, C.G. and S.K. Mitra (1976): Hermitian and nonnegative definite solutions of linear matrix equations, SIAM J. Appl. Math. 31 (1976), 597-585
Khuri, A.I., Mathew, T. and B.K. Sinha (1998): Statistical tests for mixed linear models, J. Wiley, New York 1998
Khuri, A.I. (1999): A necessary condition for a quadratic form to have a chi-squared distribution: an accessible proof, Int. J. Math. Educ. Sci. Technol. 30 (1999), 335-339
Kiamehr R., (2006): Precise Gravimetric Geoid Model for Iran Based o'n GRACE and SRTM Data and the Least-Squares Modification of Stokes' Formula with Some Geodynamic Interpretations. Ph.D. Thesjs, Division of Geodesy, Royal Institute of Technology, Stockholm, Sweden
Kiamehr, R. and Eshagh, M. (2008): Estimating variance components of ellipsoidal, orthometric and geoidal heights through the GPS/levelling network in Iran, J. Earth and Space Physics 34 (2008) 1-13
Kianifard F., Swallow W. H. (1996): A review of the development and application of recursive residuals in linear models. J. Am. Stat. Assoc. 91, 433 (1996), 391-400.
Kida, S. (1989): Statistics of velocity gradients in turbulence at moderate Reynolts number, Fluid dynamics research 4 (1989) 347-370
Kidd, M. and N.F. Laubscher (1995): Robust confidence intervals for scale and its application to the Rayleigh distribution, South African Statist. J. 29 (1995), 199-217
Kiefer, J. (1974): General equivalence theory for optimal designs (approximate theory), Ann. Stat. 2 (1974), 849-879
Kiefer, J. (1959): Optimum experimental design. J Royal Stat Soc Ser B 21:272-319
Kiefer, J.C. and J. Wolfowitz (1959): Optimum design in regression problem, Ann. Math. Statist. 30 (1959), 271-294
Killian, K. (1990): Der gefährliche Ort des überbestimmten räumlichen Rückwärtseinschneidens, *Öst.Zeitschrift für Vermessungs wesen und Photogrammetry* 78 (1990) 1-12.
King, J.T. and D. Chillingworth (1979): Approximation of generalized inverses by iterated regularization, Numer. Funct. Anal. Optim. 1 (1979), 499-513
King, M.L. (1980): Robust tests for spherical symmetry and their application to least squares regression, Ann. Statist. 8 (1980), 1265-1271
Kirkwood, B.H., Royer, J.Y., Chang, T.C. and R.G. Gordon (1999): Statistical tools for estimating and combining finite rotations and their uncertainties, Geophys. J. Int. 137 (1999), 408-428
Kirsch, A. (1996): An introduction to the mathematical theory of inverse problems, Springer-Verlag, Heidelberg Berlin New York 1996
Kitagawa, G. and W. Gersch (1996): Smoothness priors analysis of time series, Springer-Verlag, Heidelberg Berlin New York 1996
Klebanov, L.B. (1976): A general definition of unbiasedness, Theory of Probability and Appl. 21 (1976), 571-585

Klees, R., Ditmar, P. and P. Broersen (2003): How to handle colored observation noise in large least-squares problems, Journal of Geodesy 76 (2003), 629-640

Kleffe, J. (1976): A note on MINQUE for normal models, Math. Operationsforschg. Statist. 7 (1976), 707-714

Kleffe, J. (1977): Invariant methods for estimating variance components in mixed linear models, Math. Operations forsch. Statistic. Series Ststist., 8 (1977), pp. 233-250

Kleffe, J. and R. Pincus (1974): Bayes and the best quadratic unbiased estimators for parameters of the covariance matrix in a normal linear model, Math. Operationsforschg. Statistik 5 (1974), 43-76

Kleffe, J. and R. Pincus (1974): Bayes and the best quadratic unbiased estimators for variance components and heteroscedastic variances in linear models, Math. Operationsforschg. Statistik 5 (1974), 147-159

Kleffe, J. and J.N.K. Rao (1986): The existence of asymptotically unbiased nonnegative quadratic estimates of variance components in ANOVA models, J. Am. Statist. Ass. 81 (1986), 692-698

Kleijnen, J.P.C. and Beers, van W.C.M. (2004): Application driven sequential design for simulation experiments: Kriging metamodeling. J Oper Res Soc 55:876-893

Kleijnen, J.P.C. (2004): An overview of the design and analysis of simulation experiments for sensitivity analysis. Eur J Oper Res 164:287-300

Klein, U. (1997): Analyse und Vergleich unterschiedlicher Modelle der dreidimensionale, DGK, Reihe C, Heft Nr. 479.

Kleusberg, A. and E.W. Grafarend (1981): Expectation and variance component estimation of multivariate gyrotheodolite observation II , Allgemeine Vermessungsnachrichten 88 (1981), 104-108

Kleusberg, A. (1994): Die direkte Lösung des räumlichen Hyperbelschnitts, *Zeitschrift für Vermessungswesen* 119 (1994) 188-192.

Kleusberg, A. (1999): Analytical GPS navigation solution, Quo vadis geodesia...? Festschrift for E.W. Grafarend on the occasion of his 60th birthday, Eds. F. Krumm and V.S. Schwarze, Report Nr. 1999.6-1.

Klingbeil, E. (1977): Variationsrechnung. Bibliographisches Institut. B.I. Wissenschaftsverlag, Mannheim.

Klingenberg, W. (1973): Eine Vorlesung iiber Differentialgeometrie. Springer-Verlag, Berlin Heidelberg.

Klingenberg, W. (1982): Riemannian Geometry. Walter de Gruyter, Berlin.

Klonecki, W. and S. Zontek (1996): Improved estimators for simultaneous estimation of variance components, Statistics & Probability Letters 29 (1996), 33-43

Kmenta, J. (1971): Elements of econometrics, Macmillan, New York, 1971

Knautz, H. (1996): Linear plus quadratic (LPQ) quasiminimax estimation in the linear regression model, Acta Applicandae Mathematicae 43 (1996), 97-111

Knautz, H. (1999): Nonlinear unbiased estimation in the linear regression model with nonnormal disturbances, J. Statist. Planning and Inference 81 (1999), 293-309

Knickmeyer, E.H. (1984): Eine approximative Lösung der allgemeinen linearen Geodätischen Randwertaufgabe durch Reihenentwicklungen nach Kugelfunktionen, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, München 1984

Knuth D. E. (1973): The Art of Computer Programming. Vol. I, Fundamental Algorithms, 2nd Edn., Addison-Wesley

Koch, G.G. (1968): Some futher remarks on A general approach to the estimation of variance components, Technometrics 10 (1968), 551-558

Koch, K.R. (1991): Moment tensor inversion of local earthquake data - I. investigation of the method and its numerical stability with model calculations, Geophys. J. Int., 106, 305-319

Koch, K.R. (1969): Solution of the Geodetic Boundary Value Problem for a Reference Ellipsoid. Journal of Geophysical Research, 74, 3796-3803; 1969

Koch, K.R. (1970a): Lunar Shape and Gravity Field. Photogrammetric Engineering, 36, 375-380, 1970

Koch, K.R. (1970b): Darstellung'des Erdschwerefeldes in der Satellitengeodasie durchdas Potential einer einfachen Schicht Zeitschrift für Vermessungswesen, 95, 173-179, 1970

Koch, K.R. (1970c): Reformulation of the Geodetic Boundary Value Problem in View of the Results of Geometric Satellite Geodesy. Advances in Dynamic Gravimetry, edited by W.T. Kattner, 111-114, Instrument Society of America, Pittsburgh 1970

Koch, K.R. (1970d): Über die Bestimmung des Gravitationsfeldes eines Himmelskorpers mit Hilfe der Saterliten photogrammetrie. Bildmessung und Luftbildwesen, 38, 331-338, Karlsruhe 1970

Koch, K.R. (1970e): Surface Density Values for the Earthfrom Satellite and Gravity Observations. Geophysical Journal of the Royal Astronomical. Society, 21, 1-12 1970

Koch, K.R. (1971a): Die geodätische Randwertaufgabe bei bekannter ErdoberfHiche. Zeitschrift fiiI Vermessungswesen, 96, 218-224, 1971

Koch, K.R. (1971b): Simple Layer Potential of a Level. Ellipsoid. Bollettino di Geofisica teorica ed applicata, 13, 264-270, 1971

Koch, K.R. (1971c): Geophysikalische Interpretation von mittleren D'ichtewerten der Erdoberflache aus Satelitenbeobachtungen und Schwerem,essungen. Das Unternehmen Erdmantel, herausgegeben von W; Kertz u.a., 244-246, Deutsche Forschungsgemeinschaft, Bonn 1971

Koch, K.R. (1972a): Geophysical Interpretation of Density Anomalies of the Earth Computed from Satellite Observations and Gravity Measurements. Zeitschrift für Geophysik, 38, 75-84, 1972

Koch, K.R. (1972b): Method of Integral Equations for the Geodetic Boundary Value Problem. Mitteilungen aus dem Institut für Theoretische Geodasie der Universitat Bonn, Nr, 4, 38-49, Bonn 1972 .

Koch, K.R. (1973): Kontinentalverschiebung und Erdschwerefeld. Zeitschrift für Vermessungswesen, 98, 8-12, 1973

Koch, K.R. (1974): Earth's Gravity Field and Station Coordinates From Doppler Data, Satellite Triangulation, and Gravity Anomalies. NOAA Technical Report NOS 62, U.S. Department of Commerce, Rockville, Md 1974

Koch, K.R. (1975): Rekursive numerische Filter. Zeitschrift fiir Vermessungswesen,. 100, 281-292, 1975

Koch, K.R. (1976): Schatzung. einer Kovarianzmatrix und Test ihrer Identitat mit einer gegebenen Matrix. Allgemeine Vermessungs-'-Nachrichten, 83, 328-333, 1976

Koch, K.R. (1978a): Hypothesentests bei singulren Ausgleichungsproblemen, ZfV, Vol. 103, pp. 1-10

Koch, K.R. (1978b): Schatzung von Varianzkomponenten. Allgemeine Vermessungs-Nachrichten, 85, 264-269, 1978

Koch, K.R. (1979a): Parameter Estimation in Mixed Models. Correlation Analysis and Systematical Errors with Special Reference, to Geodetic Problems, Report No.6, 1-10, Geodetic Institute, Uppsala University, Uppsala 1979

Koch, K.R. (1979b): Parameter Estimation in the Gauss-Helmert Model. Bollettino di Geodesia e Scienze Affini, 38, 553-563, 1979

Koch, K.R. (1980): Parameterschatzung und Hypothesentests in linearep. Modellen. Dummler, Bonn, 1980, pp. 308

Koch, K.R. (1981): Varianz- und Kovarianzkomponentenschatzung für Streckenmessungen auf Eichlinien Allgemeine Vermessungs-Nachrichten, 88, 135-132; 1981

Koch, K.R. (1982): S-transformations and projections for obtaining estimable parameters, in: Blotwijk, M.J. et al. (eds.): 40 Years of Thought, Anniversary volume for Prof. Baarda's 65th Birthday Vol. 1, Technische Hogeschool Delft, Delft (1982), 136-144

Koch, K.R. (1983): Estimation of Variances and Covariances in the Multivariate and in the Incomplete Multivariate Model. Deutsche Geodätische Kommission, Reihe A, Nr. 98, 53-59, München 1983

Koch, K.R. (1985): Ein Statistisches Auswerteverfahren für Deformationsmessungen, AVN, Vol. 92, pp. 97-108

Koch, K.R. (1986): Maximum Likelihood Estimate of Variance Components; Ideas by A.J. Pope. Bulletin Geodesique 60 329-338 1986

Koch, K.R. (1987a): Zur Auswertung von Streckenmessungen auf Eichlinien mi'ttels Varianzkomponentenschatzung. Allgemeine Vennessungs-Nachrichten, 94, 63-71, 1987

Koch, K.R. (1987b): Bayesian Inference for Variance Components. Manuscripta Geodaetica, 12, 309-313 1987

Koch, K.R. (1987c): Parameterschaetzung und Hypothesentests in linearen Modellen, 2nd ed., Duemmler, Bonn 1987

Koch, K.R. (1988a): Parameter estimation and hypothesis testing in linear models, Springer-Verlag, Heidelberg Berlin New York 1988

Koch, K.R. (1988b): Konfidenzinte:r:valle der Bayes-Statistik für die Varianzen von Streckenmessungen 'auf Eichlinien. Vermessung Photogrammetrie, Kultur'technik, 86, 337-340, 1988

Koch, K.R. (1988c): Bayesian Statistics for Variance Components with Informative and Noninformative Priors. Manuscripta GeodaetIca, 13, 370-373; 1988

Koch, K.R. (1990): Bayesian Inference with Geodetic Applications. Springer-Verlag, Berlin Heidelberg NewYork 1990, pp. 208

Koch, K.R. (1999): Parameter estimation and hypothesis testing in linear models, 2nd ed., Springer-Verlag, Heidelberg Berlin New York 1999

Koch, K.R. (2008): Determining uncertainties of correlated measurements by Monte Carlo simulations applied to laserscanning, Journal of Applied Geodesy, 2, 139-147 (2008).

Koch, K.R. and J. Kusche (2002): Regularization of geopotential determination from satellite data by variance components, Journal of Geodesy 76 (2002), 259-268

Koch, K.R. and A.J. Pope, (1969): Least Squares Adjustment with Zero Variances. Zeitschrift für Vermessungswesen, 94, 390-393, 1969 (zusammen mit A.J. Pope)

Koch, K.R. and B.U. Witte, (1971): Earth's' Gravity Field Represented by a Simple-Layer Potential from Doppler Tracking of Satellites, Journal of Geophysical Research, 76, 8471-8479, 1971

Koch, K.R. and A.J. Pope, (1972): Punktbestimmung durch die Vereinigung von geometrischer und dynamischer Satelllitengeodasie. Zeitschrift für Vermessungswesen, 97, 1-6, 1972

Koch, K.R. and A.J. Pope, (1972): Uniqueness and Existence for the Geodetic Boundary Value Problem Using the Known Surface of the Earth. Bulletin Geodesique, 106, 467-476, 1972

Koch, K.R. and H. Frohlich (1974): Integrationsfehler in den Variationsgleichungen für das' Modell der einfachen Schicht in der Satellitengeodasie. Mitteilungen aU8 dem Institut für Theoretische Geodasie der Universitat Bonn, Nr. 25; Bonn 1974

Koch, K.R. and W. Bosch, (1980): The geopotential from Gravity Measurements, Levelling Data and Satellite Results. Bulletin Geodesique, 54, 73-79, 1980

Koch, K.R. and M. Schmidt, (1994): Deterministiche and Stochastische signals, Dummler, Bonn 1994, pp. 358

Koch, K.R. and Z. Ou, (1994): Analytical expressions for Bayes estimates of variance components. Manuscripta Geodetica 19 284-293, 1994

König, D. and V. Schmidt (1992): Zufällige Punktprozesse, Teubner Skripten zur Mathematischen Stochastik, Stuttgart 1992

Koenker, R. and G. Basset (1978): Regression quantiles, Econometrica 46 (1978), 33-50

Kollo, T. and H. Neudecker (1993): Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices, J. Multivar. Anal. 47 (91993), 283-300

Kollo, T. and D. von Rosen (1996): Formal density expansions via multivariate mixtures, in: Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, A.K. and V.L.Girko, 129-138, VSP, Utrecht 1996

Kolmogorov, A.N. (1941): Interpolation' und Extrapolation von stationaeren zufaelligen Folgen, Bull. Acad. Sci. (USSR), Ser. Math. 5. (1941), 3-14

Kolmogorov, A.N. (1950): Foundations of the Theory of Probability, New York, Chelsea Pub. co.

Kolmogorov, A.N. (1962): A refinement of previous hypothesis concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number, Fluid mechanics 13 (1962) 82-85

Konoshi, S. and Rao, C.R. (1992): Principal component analysis for multivariate familial data, Biometrica, 79, 631-641

Koopman, R. (1982): Parameterschätzung ber a priori information, Vandenhoeck and Ruprecht, Göttingen 1982

Koopmans, T.C. and O. Reiersol (1950): The identification of structural characteristics, Ann. Math. Statistics 21 (1950), 165-181

Korc, F. and Forstner, W. (2008): Interpreting Terrestrial/mages of Urban Scenes Using Discriminative Random Fields. 21 st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS). Beijing, China 2008, S. 291-296 Part B3a.

Kosko, B. (1992): Networks and fuzzy systems, Prentice-Hall, Englewood Cliffs, N.J. 1992

Kosmol, P. (1989): Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben, B. G. Teubner, Stuttgart.

Kotecky, R. and J. Niederle (1975): Conformally covariant field equations: First order equations with non-vanishing mass, Czech. J. Phys. B25 (1975), 123-149

Kotsakis, C. (2005): A type of biased estimators for linear models with uniformly biased data, Journal of Geodesy, vol. 79, no. 6/7, pp. 341-350

Kott, P.S. (1998): A model-based evaluation of several well-known variance estimators for the combined ratio estimator, Statistica Sinica 8 (1998), 1165-1173

Koukouvinos, C. and J. Seberry (1996): New weighing matrices, Sankhya: The Indian Journal of Statistics B, 58 (1996), 221-230

Kourouklis S., Paige C. C. (1996): A constrained least squares approach to the general GaussMarkov linear model. J. Am. Stat. Assoc. 76, 375 (1981), 620-625.

Kowalewski, G. (1995): Robust estimators in regression, Statistics in Transition 2 (1995), 123-135

Kraichnan R.H. (1974): On Kolmogorov's inertial-range theories, J . Fluid Mech. (1974), vol. 62, part 2, pp. 305-330

Kraemer, J. and B. Eisfeller (2009): A-GNSS, a different approach, Inside GNSS, Sep/Oct issue, (2009), 52-61

Krämer, W., Bartels, R. and D.G. Fiebig (1996): Another twist on the equality of OLS and GLS, Statistical Papers 37 (1996), 277-281

Krantz, S.G. and H.R. Parks (2002): The implicit function theorem - history, theory and applications, Birkhäuser-Verlag, Basel Boston Berlin 2002

Krarup, T. (1969): A Contribution to the mathematical foundation of physical geodesy, Publ. Danish Geod. Inst. 44, Copenhagen

Krarup, T. (1979): S-transformations or how to live without the generalized inverse-almost, Danish Geodetic Institute, Copenhagen 1979

Krarup, T. (1980): Integrated geodesy. In: Proceedings of the international school of advanced geodesy. Bollettino di Geodesia e Scienze Affini 38:480-496

Krarup, T. (1982): Nonlinear adjustment and curvature, In. Forty years of thought, Delft, pp. 145-159, 1982.

Krarup, T., Juhl, J. and K. Kubik (1980): Götterdämmerung over least squares adjustment, in: Proc. 14th Congress of the International Society of Photogrammetry, Vol. B3, Hamburg 1980, 369-378

Krause, L.O. (1987): A direct solution of GPS-Type Navigation equations, *IEEE Transactions on Aerospace and Electronic Systems* 23 (1987) 225-232.

Krauss, K.: a) ZfV 95,387 (1970); b) ZfV 96, 233 (1971).

Krengel, U. (1985): Ergodic theorems, de Gruyter, Berlin New York 1985

Kres, H. (1983): Statistical tables for multivariate analysis, Springer-Verlag, Heidelberg Berlin New York 1985

Krige, D. (1951): A statistical approach to some basic mine valuation problems on the Witwatersrand. J Chem Metall Mining Soc South Afr 52:119-139

Krishna, S. and Manocha, D. (1995): Numerical algorithms for evaluating one-dimensional algebraic sets, Proceedings of the International Symposium on Symbolic and Algebraic Computation ISSAC, July 10-12, pp. 59-67, Montreal, Canada, 1995.

Kronecker, L. (1903): Vorlesungen über die Theorie der Determinanten, Erster Band, Bearbeitet und fortgeführt von K.Hensch, B.G.Teubner, Leipzig 1903

Krumbein, W.C. (1939): Preferred orientation of pebbles in sedimentary deposits, J. Geol. 47 (1939), 673-706

Krumm, F. (1982): Criterion matrices for estimable quantities. Proceedings Survey Control Networks, ed. Borre, K., Welsch, W.M., Vol. 7, pp. 245-257, Schriftenreih Wiss. Studiengang Vermessungswesen, Hochschule der Bundeswehr München.

Krumm, F., Grafarend, E. and B. Schaffrin (1986): Continuous networks, Fourier analysis and criterion matrices, manuscripta geodaetica 11 (1986), 57-78

Krumm, F. (1987): Geodätische Netze im Kontinuum: Inversionsfreie Ausgleichung und Konstruktion von Kriterionmatrizen, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Ph. D. Thesis, Report C334, München 1987

Krumm, F. and F. Okeke (1998): Graph, graph spectra, and partitioning algorithms in a geodetic network structural analysis and adjustment, Bolletino di Geodesia e Science Affini 57 (1998), 1-24

Kruskal, W. (1946): Helmert's distribution, American Math. Monthly 53 (1946), 435-438

Kruskal, W. (1968): When are Gauß-Markov and least squares estimators identical? A coordinate-free approach, Ann. Statistics 39 (1968), 70-75

Kryanev, A.V. (1974): An iterative method for solving incorrectly posed problem, USSR. Comp. Math. Math. Phys. 14 (1974), 24-33

Krzanowski, W.J. and F.H.C. Marriott (1994): Multivariate analysis: part I - distribution, ordination and inference, Arnold Publ., London 1994

Krzanowski, W.J. and F.H.C. Marriott (1995): Multivariate analysis: part II - classification, covariance structures and repeated measurements, Arnold Publ., London 1995

Kuang, S.L. (1991): Optimization and design of deformations monitoring schemes, PhD dissertation, Department of Surveying Engineering, University of New Brunswick, Tech. Rep. 91, Fredericton 1991

Kuang, S. (1996): Geodetic network analysis and optimal design, Ann Arbor Press, Chelsea, Michigan 1996

Kubacek, L. (1988): Foundations of the Estimation Theory. Elsevier, Amsterdam-Oxford, New York-Tokyo, 1988.

Kubacek, L. (1995): Linear statistical models with constraints revisited. Math. Slovaca 45, 3 (1995), 287-307.

Kubacek, L. (1995): On a linearization of regression models. Appl. Math., Praha, 40, 1 (1995), 61-78.

Kubacek, L. (1996a): Linear model with inaccurate variance components, Applications of Mathematics 41 (1996), 433-445

Kubacek, L. (1996b): Nonlinear error propagation law, Applications of Mathematics 41 (1996), 329-345

Kubacek, L. (1996c): Models with a low nonlinearity. Tatra Mt. Math. Publ. 7 (1996), 149-155.

Kubacek, L. (1997): Notice on the Chipman generalization of the matrix inverse. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 36 (1997), 95-98.

Kubacek, L. (2002): On an accuracy of change points. Math. Slovaca 52,4 (2002), 469-484.

Kubacek, L. (2005): Underparametrization in a regression model with constraints II. Math. Slovaca 66 (2005), 579-596.

Kubacek, L. and Kubackova, L. (1978): The present approach to the study of the least-squares method. Studia geoph. et geod. 22 (1978), 140-147.

Kubacek, L., Kubackova, L. and J. Kukaca (1987): Probability and statistics in geodesy and geophysics, Elsevier, Amsterdam 1987

Kubacek, L., Kubackova, L. and Volaufova, J. (1995): Statistical Models with Linear Structures. Veda, Bratislava, 1995.

Kubacek, L. and Kubackova, L. and J. Kukaca (1995): Statistical models with linear structures, slovak Academy of Sciences, Bratislava 1995

Kubacek, L. and Kubackova, L. (1997): One of the calibration problems. Acta Univ. Palacki. 010-muc., Fac. Rerum Nat., Math. 36 (1997), 117-130.

Kubacek, L. and Kubackova, L. (1998): Testing statistical hypotheses in deformation measurement; one generalization of the Scheffe theorem. Acta Univ. Palacki. Olomu., Fac. Rerum Nat., Math. 37 (1998), 81-88.

Kubacek, L. and Kubackova, L. (2000): Nonsensitiveness r:egions in universal models. Math. 810- vaca 50, 2 (2000), 219-240.

Kubacek, L., Kubackova L., SevCik J. (2002): Linear conform transformation: errors in both coordinate systems. Appl. Math., Praha 47, 4 (2002), 361-380.

Kubacek, L., Fiserova, E. (2003): Problems of sensitiveness and linearization in a determination of isobestic points. Math. Slovaca 53, 4 (2003), 407-426.

Kubackova, L. (1996): Joint confidence and threshold ellipsoids in regression models. Tatra Mt. Math. Publ. 7 (1996), 157-160.

Kubik, K.K. (1967): Iterative Methoden zur Lösunge des nichtlinearen Ausgleichungsproblemes, *Zeitschrift für Vermessungswesen* 91 (1967) 145-159.

Kubik, K. (1970): The estimation of the weights of measured quantities within the method of least squares. Bulletin Geodesique, 44, 21-40.

Kubik, K. (1982): Kleinste Quadrate und andere Ausgleichsverfahren, Vermessung Photogrammetrie Kulturtechnik 80 (1982), 369-371

Kubik, K., and Y. Wang (1991): Comparison of different principles for outlier detection, Australian Journal of Geodesy, Photogrammetry and Surveying 54 (1991), 67-80

Kuechler, U. and M. Soerensen (1997): Exponential families of stochastic processes, Springer-Verlag, Heidelberg Berlin New York 1997

Kukush, A. and S. Zwanzig (2002): On consistent estimatiors in nonlinear functional errors-in-variables models. p. 145-154. In Total Least squares and Erros-in-Variables Modeling ed. Sabine Huffel, Philippe Lemmerling, Kluwer

Kullback, S. (1934): An application of characteristic functions to the distribution problem of statistics, Annals Math. Statistics 4 (1934), 263-305

Kullback, S. (1935): On samples from a multivariate normal population, Ann. Math. Stat. 6, 202-213

Kullback, S.(1952): An application of information theory to multivariate analysis, Ann. Math. Stat., 23, 88-102

Kullback, S. (1956): An application of information theory to multivariate analysis II, Ann. Math. Stat., 27, 122-146

Kumaresan, R. (1982): Estimating the parameters of exponentially damped or undamped sinusoidal signals in noise, PhD thesis, The University of Rhode Island, Rhode Island 1982

Kunderova, P. (2000): Linear models with nuisance parameters and deformation measurement. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 39 (2000), 95-105.

Kunderova, P. (2001): Locally best and uniformly best estimators in linear models with nuisance parameters. Tatra Mt. Math. Publ. 22 (2001), 27-36.

Kunderova, P. (2001): Regular linear model with the nuisance parameters with constraints of the type I. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 40 (2001), 151-159.

Kunderova, P. (2002): Regular linear model with nuisance parameters with constraints of the type II. Folia Fac. Sci. Nat. Univ. Masarykianae Brunensis, Math. 11 (2002), 151-162.

Kunderova, P. (2003): Eliminating transformations for nuisance parameters in linear model. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 42 (2003), 59-68.

Kunderova, P. (2005): One singular multivariate linear model with nuisance parameters. Acta Univ. Palacki. Olomuc., Fac. Rerum Nat., Math. 44 (2005), 57-69.

Kundu, D. (1993a): Estimating the parameters of undamped exponential signals, Technometrics 35 (1993), 215-218

Kundu, D. (1993b): Asymptotic theory of least squares estimator of a particular nonlinear regression model, Statistics and Probability Letters 18 (1993), 13-17

Kundu, D. (1994a): Estimating the parameters of complex valued exponential signals, Computational Statistics and Data Analysis 18 (1994), 525-534

Kundu, D. (1994b): A modified Prony algorithm for sum of damped or undamped exponential signals, Sankhya A56 (1994), 525-544

Kundu, D. (1997): Asymptotic theory of the least squares estimators of sinusoidal signal, Statistics 30 (1997), 221-238
Kundu, D. and A. Mitra (1998): Fitting a sum of exponentials to equispaced data, Sankhya: The Indian Journal of Statistics 60 (1998), 448-463
Kundu, D. and R.D. Gupta (1998): Asymptotic properties of the least squares estimators of a two dimensional model, Metrika 48 (1998), 83-97
Kunst, R.M. (1997): Fourth order moments of augmented ARCH processes, Commun. Statist. Theor. Meth. 26 (1997), 1425-1441
Kurz, S. (1996): Positionierung mittels Rückwartsschnitt in drei Dimensionen. Studienarbeit, Geodätisches Institut, University of Stuttgart, Stuttgart 1996.
Kuo, W., Prasad, V.R., Tillman, F.A. and C.-L. Hwang (2001): Optimal reliability design, Cambridge University Press, Cambridge 2001
Kuo, Y. (1976): An extended Kronecker product of matrices, J. Math. Anal. Appl. 56 (1976), 346-350
Kupper, L.L. (1972): Fourier series and spherical harmonic regression, J. Roy. Statist. Soc. C21 (1972), 121-130
Kupper, L.L. (1973): Minimax designs of Fourier series and spherical harmonic regressions: a characterization of rotatable arrangements, J. Roy. Statist. Soc. B35 (1973), 493-500
Kurata, H. (1998): A generalization of Rao's covariance structure with applications to several linear models, J. Multivar. Anal. 67 (1998), 297-305
Kurz, L. and M.H. Benteftifa (1997): Analysis of variance in statistical image processing, Cambridge University Press, Cambridge 1997
Kusche, J. (2001): Implementation of multigrid solvers for satellite gravity anomaly recovery, Journal of Geodesy 74 (2001), 773-782
Kusche, J. (2002): Inverse Probleme bei der Gravitationsfeldbestimmung mittels SST- und SGG-Satellitenmissionen , Deutsche Geodätische Kommission, Report C548, München 2002
Kusche, J. (2003): A Monte-Carlo technique for weight estimation in satellite geodesy, Journal of Geodesy 76 (2003), 641-652
Kushner, H. (1967): Dynamical equations for optimal nonlinear filtering, J. Diff. Eq. 3 (1967), 179-190
Kushner, H. (1967): Approximations to optimal nonlinear filters, IEEE Trans. Auto. Contr. AC-12 (1967), 546-556
Kutoyants, Y.A. (1984): Parameter estimation for stochastic processes, Heldermann, Berlin 1984
Kutterer, H. (1994): Intervallmathematische Behandlung endlicher Unschärfen linearer Ausgleichsmodelle, PhD Thesis, Deutsche Geodätische Kommission DGK C423, München 1994
Kutterer, H. (1999): On the sensitivity of the results of least-squares adjustments concerning the stochastic model, Journal of Geodesy 73 (1999), 350-361
Kutterer, H. (2002): Zum Umgang mit Ungewissheit in der Geodäsie - Bausteine für eine neue Fehlertheorie - , Deutsche Geodätische Kommission, Report C553, München 2002
Kutterer, H. and S.Schön (1999): Statistische Analyse quadratischer Formen - der Determinantenansatz, Allg. Vermessungsnachrichten 10 (1999), 322-330
Lagrange, J. L. (1877): Lecons eIementaires sur les mathematiques donnees a l'Ecole Normale en 1795. In: Oeuvres de Lagrange, Ed.: M. J. A. Serret, Tome 7, Section IV, Paris.
Laha, R.G. (1956): On the stochastic independence of two second-degree polynomial statistics in normally distributed variates, The Annals of Mathematical Statistics 27 (1956), 790-796
Laha R.G. and E. Lukacs (1960): On a problem connected with quadratic regression, Biometrika 47 (1960), 335-343
Lai, T.L. and C.Z. Wie (1982): Least squares estimates in stochastic regression model with applications to stochastic regression in linear dynamic systems, Anals of Statistics 10 (1982), 154-166
Lai, T.L. and C.P. Lee (1997): Information and prediction criteria for model selection in stochastic regression and ARMA models, Statistical Sinica 7 (1997) 285-309
Laird, N.M. and J.H. Ware (1982): Random-effects models for longitudinal data, Biometrics 38 (1982), 963-974

Lame, M. G. (1818): Examen des differentes methodes employees pour resoudre les problemes de geometrie. Paris

LaMotte, L.R. (1973a): Quadratic estimation of variance components, Biometrics 29 (1973), 311-330

LaMotte, L.R. (1973b): On non-negative quadratic unbiased estimation of variance components, J. Am. Statist. Ass. 68(343): 728-730.

LaMotte, L.R. (1976): Invariant quadratic estimators in the random, one-way ANOVA model, Biometrics 32 (1976), 793-804

LaMotte, L.R. (1996): Some procedures for testing linear hypotheses and computing RLS estimates in multiple regression. Tatra Mt. Math. Publ. 7 (1996), 113-126.

LaMotte, L.R., McWhorter, A. and R.A. Prasad (1988): Confidence intervals and tests on the variance ratio in random models with two variance components, Com. Stat. - Theory Meth. 17 (1988), 1135-1164

Lamperti, J. (1966): Probility: a survey of the mathematical theory, WA Benjamin Inc., New York 1966

Lancaster, H.O. (1965): The helmert matrices, American Math. Monthly 72 (1965), 4-11

Lancaster, H.O. (1966): Forerunners of the Pearson Chi2 , Australian Journal of Statistics 8 (1966), 117-126

Langevin, P. (1905): Magnetisme et theorie des electrons, Ann. De Chim. et de Phys. 5 (1905), 70-127

Lanzinger, H. and U. Stadtmüller (2000): Weighted sums for i.i.d. random variables with relatively thin tails, Bernoulli 6 (2000), 45-61

Lapaine, M. (1990): "A new direct solution of the transformation problem of Cartesian into ellipsoidal coordinates". Determination of the geoid: Present and future. Eds. R.H. Rapp and F. Sanso, pp. 395-404. Springer-Verlag, New York 1990.

Laplace, P.S. (1793): Sur quelques points du Systeme du monde, Memoire de I'Academie royale des Sciences de Paris, in Oeuvres Completes, 11 (1793), pp. 477-558

Laplace, P.S. (1811): Mémoiresur les intégrales définies et leur application aux probabilites et specialement a la recherche du milieu qu'il faut choisir entre les resultats des observations, Euvres completes XII, Paris, 1811, 357-412.

Lardy, L.J. (1975): A series representation for the generalized inverse of a closed linear operator, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. 58 (1975), 152-157

Lauer, S.: Dissertation, Bonn 1971

Lauritzen, S. (1973): The probabilistic background of some statistical methods in Physical Geodesy, Danish Geodetic Institute, Maddelelse 48, Copenhagen 1973

Lauritzen, S. (1974): Sufficiency, prediction and extreme models, Scand. J. Statist. 1 (1974), 128-134

Läuter, H. (1970): Optimale Vorhersage und Schätzung in regulären und singulären Regressionsmodellen, Math. Operationsforschg. Statistik 1 (1970), 229-243

Läuter, H. (1971): Vorhersage bei stochastischen Prozessen mit linearem Regressionsanteil, Math. Operationsforschg. Statistik 2 (1971), 69-85, 147-166

Lawless, J.F. (1982): Statistical models and lifetime data, Wiley, New York 1982

Lawley, D.N. (1938): A generalization of Fisher's z-test. Biometrika, 30, 180-187

Lawson, C.L. and R.J. Hanson (1995): Solving least squares problems, SIAM, Philadelphia 1995 (reprinting with corrections and a new appendix of a 1974 Prentice Hall text)

Lawson, C.L. and R.J. Hanson (1974): Solving least squares problems, Prentice-Hal, Englewod Cliffs, NJ 1974

Laycock, P.J. (1975): Optimal design: regression models for directions, Biometrika 62 (1975), 305-311

LeCam, L. (1960): Locally asymptotically normal families of distributions, University of California Publication 3, Los Angeles 1960, 37-98

LeCam, L. (1970): On the assumptions used to prove asymptotic normality of maximum likelihood estimators, Ann. Math. Statistics 41 (1970), 802-828

LeCam, L. (1986): Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Chapman and Hall, Boca Raton 1986
Leichtweiss, K. (1962): -ober eine Art von Kriimmungsinvarianten beliebiger Untermannigfaltigkeiten des n-dimensionalen euklidischen Raums. Hbg. Math. Abh., Bd. XXVI, 156-190.
Leis, R. (1967): Vorlesung"en Über partielle Differentialgleichungen zweiter Ordnung. Mannheim
Lee, Y. W. (1960): Statistical theory of communication, New York 1960
Lee, Y. W., in: Selected Papers of N. Wiener, MIT-Press 1964
Lee, J.C. and S. Geisser (1996): On the Prediction of Growth Curves, in: Lee, C., Johnson, W.O. and A. Zellner (eds.): Modelling and Prediction Honoring Seymour Geisser, Springer-Verlag, Heidelberg Berlin New York 1996, 77-103
Lee, J.C., Johnson, W.O. and A. Zellner (1996): Modeling and prediction, Springer-Verlag, Heidelberg Berlin New York 1996
Lee, M. (1996): Methods of moments and semiparametric econometrics for limited dependent variable models, Springer-Verlag, Heidelberg Berlin New York 1996
Lee, P. (1992): Bayesian statistics, J. Wiley, New York 1992
Lee, S.L. (1995): A practical upper bound for departure from normality, Siam J. Matrix Anal. Appl. 16 (1995), 462-468
Lee, S.L. (1996): Best available bounds for departure from normality, Siam J. Matrix Anal. Appl. 17 (1996), 984-991
Lee, S.L. and J.-Q. Shi (1998): Analysis of covariance structures with independent and non-identically distributed observations, Statistica Sinica 8 (1998), 543-557
Lee, Y. and J.A. Nelder (1996): Hierarchical generalized linear models, J. Roy. Statist. Soc. B58 (1996), 619-678
Lee, Y., Neider, J. and Pawitan, Y. (2006): Generalized linear models with random effects, Chapman & Hall/CRC, New York
Lehmann, E.L. (1959): Testing statistical hypotheses, J. Wiley, New York 1959
Lehmann, E.L. and H. Scheffèe (1950): Completeness, similar regions and unbiased estimation, Part I, Sankya 10 (1950), 305-340
Lehmann, E.L. and G.Casella (1998): Theory of point estimation, Springer-Verlag, Heidelberg Berlin New York 1998
Leick, A. (1995): GPS satellite surveying, 2nd Edition, John Wiley & Sons, New York 1995.
Leick, A. (2004): GPS Satellite Surveying, 3rd ed, John Wiley & Sons, New Jersey, 435p
Lemmerling H., Van Huffel S. (2002): Structured total least squares. Analysis, algorithms and applications. In: Total least squares and errors-in-variables modelling. Analysis, algorithms and applications (S. Van Huffel et al., ed), Kluwer Academic Publishers, Dordrecht, 2002, pp. 79-91.
Lenstra, H.W. (1983): Integer programming with a fixed number of variables, Math. Operations Res. 8 (1983), 538-548
Lenth, R.V. (1981): Robust measures of location for directional data, Technometrics 23 (1981), 77-81
Lenstra, A.K., Lenstra, H.W. and L. Lovacz (1982): Factoring polynomials with rational coefficients, Math. Ann. 261 (1982), 515-534
Lentner, M.N. (1969): Generalized least-squares estimation of a subvector of parameters in randomized fractional factorial experiments, Ann. Math. Statist. 40 (1969), 1344-1352
Lenzmann, L. (2003): Strenge Auswertung des nichtlnearen Gauß-Helmert-Modells, AVN 2 (2004), 68-73
Lesaffre, E. and G. Verbeke (1998): Local influence in linear mixed models, Biometrics 54 (1998), 570-582
Lesanska E. (2001): Insensitivity regions for testing hypotheses in mixed models with constraint. Tatra Mt. Math. Publ. 22 (2001), 209-222.
Lesanska E. (2002): Effect of inaccurate variance components in mixed models with constraints. Folia Fac. Sci. Nat. Univ. Masarykianae Brunensis, Math. 11 (2002), 163-172.
Lesanska E. (2002): Insensitivity regions and their properties. J. Electr. Eng. 53 (2002), 68-71.
Lesanska E. (2002): Nonsensitiveness regions for threshold ellipsoids. Appl. Math., Praha 47, 5 (2002), 411-426.

Lesanska E. (2002): Optimization of the size of nonsensitive regions. Appl. Math., Praha 47, 1 (2002), 9-23.

Leslie, M. (1990): Turbulence in fluids, 2nd ed., Kluwer publ., Dordicht 1990

Letac, G. and M. Mora (1990): Natural real exponential families with cubic variance functions, Ann. Statist. 18 (1990), 1-37

Lether, F.G. and P.R. Wentson (1995): Minimax approximations to the zeros of Pn (x) and Gauss-Legendre quadrature, J. Comput. Appl. Math. 59 (1995), 245-252

Levenberg, K. (1944): A method for the solution of certain non-linear problems in least-squares, Quaterly Appl. Math. 2 (1944) 164-168

Levin, J. (1976): A parametric algorithm for drawing pictures of solid objects composed of quadratic surfaces, Communications of the ACM 19 (1976), 555-563

Lewis, R.M. and V. Torczon (2000): Pattern search methods for linearly constrained minimization, SIAM J. Optim. 10 (2000), 917-941

Lewis, T.O. and T.G. Newman (1968): Pseudoinverses of positive semidefinite matrices, SIAM J. Appl. Math. 16 (1968), 701-703

Li, B.L. and C. Loehle (1995): Wavelet analysis of multiscale permeabilities in the subsurface, Geophys. Res. Lett. 22 (1995), 3123-3126

Li, C.K. and R. Mathias (2000): Extremal characterizations of the Schur complement and resulting inequalities, SIAM Review 42 (2000), 233-246

Liang, K. and K. Ryu (1996): Selecting the form of combining regressions based on recursive prediction criteria, in: Lee, C., Johnson, W.O. and A. Zellner (eds.): Modelling and prediction honoring Seymour Geisser, Springer-Verlag, Heidelberg Berlin New York 1996, 122-135

Liang, K.Y. and S.L. Zeger (1986): Longitudinal data analysis using generalized linear models, Biometrika 73 (1986), 13-22

Liang, K.Y. and S.L. Zeger (1995): Inference based on estimating functions in the presence of nuisance parameters, Statist. Sci. 10 (1995), 158-199

Lichtenegger, H. (1995): Eine direkte Lösung des räumlichen Bogenschnitts, *Österreichische Zeitschrift für Vermessung und Geoinformation* 83 (1995) 224-226.

Lilliefors, H.W. (1967): On the Kolmogorov-Smirnov test for normality with mean and variance unknown, J. Am. Statist. Ass.. 62 (1967), 399-402

Lim, Y.B., Sacks, J. and Studden, W.J. (2002): Design and analysis of computer experiments when the output is highly correlated over the input space. Canad J Stat 30:109-126

Lin, K.C. and Wang, J. (1995): Transformation from geocentric to geodetic coordinates using Newton's iteration, *Bulletin Geodesique* 69 (1995) 300-303.

Lin, X. and N.E. Breslow (1996): Bias correction in generalized linear mixed models with multiple components of dispersion, J. Am. Statist. Ass. 91 (1996), 1007-1016

Lin, X.H. (1997): Variance component testing in generalised linear models with random effects, Biometrika 84 (1997), 309-326

Lin, T. and Lee, J. (2007):.Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. In Statistics in Medicine, published online

Lindley, D.V. and A.F.M. Smith (1972): Bayes estimates for the linear model, J. Roy. Stat. Soc. 34 (1972), 1-41

Lindsey, J.K. (1997): Applying generalized linear models, Springer-Verlag, Heidelberg Berlin New York 1997

Lindsey, J.K. (1999): Models for repeated measurements, 2nd edition, Oxford University Press, Oxford 1999

Linke, J., Jurisch, R., Kampmann, G. and H. Runne (2000): Numerisches Beispiel zur Strukturanalyse von geodätischen und mechanischen Netzen mit latenten Restriktionen, Allgemeine Vermessungsnachrichten 107 (2000), 364-368

Linkwitz, K. (1961): Fehlertheorie und Ausgleichung von Streckennetzen nach der Theorie elastischer Systeme. Deutsche Geodätische Kommission, Reihe C, Nr. 46.

Linkwitz, K. (1989): Bemerkungen zu Linearisierungen in der Ausgleichungsrechnung. In: Prof. Dr.Ing. Dr. h.c. Friedrich Ackermann zum 60. Geburtstag, Schriftenreihe des Institutes fiir Photogrammetrie, Heft 14.

Linkwitz, K. and Schek, H.J. (1971): Einige Bemerkungen zur Berechnung von vorgespannten Seilnetz-Konstruktionen. Ingenieur-Archiv, 40.

Linnainmaa, S., Harwood, D. and Davis, L.S. (1988): Pose Determination of a Three-Dimensional Object Using Triangle Pairs, *IEEE transaction on pattern analysis and Machine intelligence* 105 (1988) 634-647.

Linnik, J.V. and I.V. Ostrovskii (1977): Decomposition of random variables and vectors, Transl. Math. Monographs Vol. 48, American Mathematical Society, Providence 1977

Lippitsch, A. (2007): A deformation analysis method for the metrological ATLAS cavern network at CERN, Ph.D. thesis, Graz University of Technology, Graz/Austria/2007

Liptser, R.S. and A.N. Shiryayev (1977): Statistics of random processes, Vol. 1, Springer-Verlag, Heidelberg Berlin New York 1977

Liski, E.P. and S. Puntanen (1989): A further note on a theorem on the difference of the generalized inverses of two nonnegative definite matrices, Commun. Statist.-Theory Meth. 18 (1989), 1747-1751

Liski, E.P. and T. Nummi (1996): Prediction in repeated-measures models with engineering applications, Technometrics 38 (1996), 25-36

Liski, E.P., Luoma, A., Mandal, N.K. and B.K. Sinha (1998): Pitman nearness, distance criterion and optimal regression designs, Calcutta Statistical Ass. Bull. 48 (1998), 191-192

Liski, E.P., Mandal, N.K., Shah, K.R. and B.K. Sinha (2002): Topics in optimal design, Springer-Verlag, Heidelberg Berlin New York 2002

Lisle, R. (1992): New method of estimating regional stress orientations: application to focal mechanism of data of recent British earthquakes, Geophys. J. Int., 110,276.,282

Liu, J. (2000): MSEM dominance of estimators in two seemingly unrelated regressions, J. Statist. Planning and Inference 88 (2000), 255-266

Liu, S. (2000): Efficiency comparisons between the OLSE and the BLUE in a singular linear model, J. Statist. Planning and Inference 84 (2000), 191-200

Liu, X.-W. and Y.-X. Yuan (2000): A robust algorithm for optimization with general equality and inequality constraints, SIAM J. Sci. Comput. 22 (2000), 517-534

Ljung, L. (1979): Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, IEEE Trans. Auto. Contr. AC-24 (1979), 36-50

Lloyd, E.H. (1952): Least squares estimation of location and scale parameters using order statistics, Biometrika 39 (1952), 88-95

Lohse, P. (1990): Dreidimensionaler Rückwärtsschnit. Ein Algorithmus zur Streckenberechnung ohne Hauptachsentransformation, *Zeitschrift für Vermessungswesen* 115 (1990) 162-167.

Lohse, P. (1994): Ausgleichungsrechnung in nichtlinearen Modellen, Deutsche Geod. Kommission C 429, München 1994

Lohse, P., Grafarend, E.W. and Schaffriin, B. (1989a): Dreidimensionaler Riickwartsschnitt. Teil IV, V, Zeitschrift fiir Vermessungswesen, 5, 225-234; 6, 278-287.

Lohse, P., Grafarend, E.W. and Schaffriin, B. (1989b): Three-Dimensional Point Determination by Means of Combined Resection and Intersection. Paper presented at the Conference on 3-D Measurement Techniques, Vienna, Austria, September 18-20, 1989.

Longford, N.T. (1993): Random coefficient models, Clarendon Press, Oxford 1993

Longford, N. (1995): Random coefficient models, Oxford University Press, 1995

Longley, J.W. and R.D. Longley (1997): Accuracy of Gram-Schmidt orthogonalization and Householder transformation for the solution of linear least squares problems, Numerical Linear Algebra with Applications 4 (1997), 295-303

Lord, R.D. (1948): A problem with random vectors, Phil. Mag. 39 (1948), 66-71

Lord, R.D. (1995): The use of the Hankel transform in statistics, I. General theory and examples, Biometrika 41 (1954), 44-55

Lorentz, G.G. (1966): Metric entropy and approximation, Bull. Amer. Math. Soc. 72 (1966), 903-937

Loskowski, P. (1991): Is Newton's iteration faster than simple iteration for transformation between geocentric and geodetic coordinates? *Bulletin Geodesique* 65 (1991) 14-17.

Lu, S., Molz, F.J. and H.H. Liu (2003): An efficient three dimensional anisotropic, fractional Brownian motion and truncated fractional LEVY motion simulation algorithm based on successive random additions, Computers and Geoscience 29 (2003) 15-25

Ludlow, J. and W. Enders (2000): Estimating non-linear ARMA models using Fourier coefficients, International Journal of Forecasting 16 (2000), 333-347

Lumley, J.L. (1970): Stochastic tools in turbulence, Academic Press, New York 1970

Lumley, J.L. (1978): Computational modelling of turbulence, Advances in applied mechanics 18 (1978) 123-176

Lumley, J.L. (1983): Turbulence modelling, J. App. Mech. Trans. ASME 50 (1983) 1097-1103

Lund, U. (1999): Least circular distance regression for directional data, Journal of Applied Statistics 26 (1999), 723-733

Lütkepol, H. (1996): Handbook of matrices, J. Wiley, New York 1996

Lyubeznik, G. (1995): Minimal resultant system, *Journal of Algebra* 177 (1995) 612-616.

Ma, X.Q. and Kusznir, N.J. (1992): 3-D subsurface displacement and strain fields for faults and fault arrays in a layered elastic half-space, Geophys. J. Int., 111, 542-558

Mac Duffee, C.C. (1946): The theory of matrices (1933), reprint Chelsea Publ., New York 1946

Macaulay, F. (1902): On some formulae in elimination, *Proceeding in London Mathematical Society* pp. 3-27, 1902.

Macaulay, F. (1916): The algebraic theory of modular systems, Cambridge Tracts in Mathematics 19, Cambridge University Press, Cambridge 1916.

Macaulay, F. (1921): Note on the resultant of a number of polynomials of the same degree, *Proceeding in London Mathematical Society* 21 (1921) 14-21.

Macinnes, C.S. (1999): The solution to a structured matrix approximation problem using Grassmann coordinates, SIAM J. Matrix Analysis Appl. 21 (1999), 446-453

Madansky, A. (1959): The fitting of straight lines when both variables are subject to error, J. Am. Statist. Ass. 54 (1959), 173-205

Madansky, A. (1962): More on length of confidence intervals, J. Am. Statist. Ass. 57 (1962), 586-589

Mader, G.L. (2003): GPS Antenna Calibration at the National Geodetic Survey. http://www.ngs.noaa.gov/ANTCAL/images/summary.pdf

Maejima, M. (1978): Some Lp versions for the central limit theorem, Ann. Probability 6 (1978), 341-344

Maekkinen, J. (2002): A bound for the Euclidean norm of the difference between the best linear unbiased estimator and a linear unbiased estimator, Journal of Geodesy 76 (2002), 317-322

Maess, G. (1988): Vorlesungen über numerische Mathematik II. Analysis, Birkhäuser-Verlag, Basel Boston Berlin 1988

Magder, L.S. and S.L. Zeger (1996): A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians, J. Am. Statist. Ass. 91 (1996), 1141-1151

Magee, L. (1998): Improving survey-weighted least squares regression, J. Roy. Statist. Soc. B60 (1998), 115-126

Magness, T.A. and J.B. McGuire (1962): Comparison of least-squares and minimum variance estimates of regression parameters, Ann. Math. Statist. 33 (1962), 462-470

Magnus, J.R. and H. Neudecker (1988): Matrix differential calculus with applications in statistics and econometrics, J. Wiley, New York 1988

Mahalanobis, P. C. (1930): On tests and measures of group divergence, J. and Proc. Asiat. Soc. Beng., 26, 341-5S8.

Mahalanobis, P. C. (1936): On the generalized distance in statistics, Proc. Nat. Inst. Sci. India, 2, 49-55.

Mahalanobis, A. and M. Farooq (1971): A second-order method for state estimation of nonlinear dynamical systems, Int. J. Contr. 14 (1971), 631-639

Mahalanobis, P.C., Bose, R.C. and S.N. Roy (1937): Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions, Sankhya 3 (1937), 1-40

Mallet, A. (1986): A maximum likelihood estimation method for random coefficient regression models, Biometrika 73 (1986), 645-656

Malliavin, P. (1997): Stochastic analysis, Springer-Verlag, Heidelberg Berlin New York 1997
Mallick, B., Denison, D. and Smith, A. (2000): Semiparamefric generalized linear models. Dey, D., Ghosh, S:ai1d MaJITck B.K. eds., Generalized linear models: A Bayesian perspective, Marcel Dekker, New York.
Mallows, C.L. (1959): Latent roots and vectors of random matrices, Tech. Report 35, Statistical Tech. Research Group, Section of Math. Stat., Dept. of Math., Princeton University, Princeton, 1959
Mallows, C.L. (1961): Latent vectors of random symmetric matrices, Biometrika 48 (1961), 133-149
Malyutov, M.B. and R.S. Protassov (1999): Functional approach to the asymptotic normality of the non-linear least squares estimator, Statistics & Probability Letters 44 (1999), 409-416
Mamontov, Y. and M.Willander (2001): High-dimensional nonlinear diffusion stochastic processes, World Scientific, Singapore 2001
Mandel, J. (1994): The analysis of two-way layouts, Chapman and Hall, Boca Raton 1994
Mangoubi, R.S. (1998): Robust estimation and failure detection, Springer-Verlag, Heidelberg Berlin New York 1998
Manly, B.F. (1976): Exponential data transformation, The Statistican 25 (1976) 37-42
Manocha, D. (1992): Algebraic and Numeric techniques for modeling and Robotics, Ph.D. thesis, Computer Science Division, Department of Electrical Engineering and Computer Science, University of California, Berkeley 1992.
Manocha, D. (1993): Efficient algorithms for multipolynomial resultant, *the Computer Journal* 36 (1993) 485-496.
Manocha, D. (1994a): Algorithms for computing selected solutions of polynomial equations, Extended abstract appearing in the proceedings of the ACM ISSAC 94.
Manocha, D. (1994b): Computing selected solutions of polynomial equations, Proceedings of the International Sypmosium on Symbolic and Algebraic Computations ISSAC, July 20-22, pp. 1-8, Oxford 1994.
Manocha, D. (1994c): Solving systems of polynomial equations, IEEE Computer Graphics and application, pp. 46-55, March 1994.
Manocha, D. (1998): Numerical methods for solving polynomial equations, *Proceedings of Symposia in Applied Mathematics* 53 (1998) 41-66.
Manocha, D. and Canny, J. (1991): Efficient techniques for multipolynomial resultant algorithms, Proceedings of the International Symposium on Symbolic Computations, July 15-17, 1991, pp. 86-95, Bonn 1991.
Manocha, D. and Canny, J. (1992): Multipolynomial resultant and linear algebra, Proceedings of the International Symposium on Symbolic and Algebraic Computations ISSAC, July 27-29, 1992, pp. 158-167, Berkeley 1992.
Manocha, D. and Canny, J. (1993): Multipolynomial resultant algorithms, *Journal of Symbolic Computations* 15 (1993) 99-122.
Mardia, K.V. (1962): Multivariate Pareto distributions, Ann. Math. Statistics 33 (1962), 1008-1015
Mardia, K.V. (1970a): Families of bivariate distributions, Griffin, London 1970
Mardia, K.V. (1970b): Measures of multivariate skewness and kurtosis with applications, Biometrika 57 (1970), 519-530
Mardia, K.V. (1972): Statistics of directional data, Academic Press, New York London 1972
Mardia, K.V. (1975): Characterization of directional distributions, Statistical Distributions, Scientific Work 3 (1975), G.P. Patil et al. (eds.), 365-385
Mardia, K.V. (1975): Statistics of directional data, J. Roy. Statist. Soc. B37 (1975), 349-393
Mardia, K.V. (1976): Linear-circular correlation coefficients and rhythmometry, Biometrika 63 (1976), 403-405
Mardia, K.V. (1988): Directional data analysis: an overview, J. Applied Statistics 2 (1988), 115-122
Mardia, K.V. and M.L. Puri (1978): A robust spherical correlation coefficient against scale, Biometrika 65 (1978), 391-396

Mardia, K.V., Kent, J. and J.M. Bibby (1979): Multivariate analysis, Academic Press, New York London 1979

Mardia, K.V., Southworth, H.R. and C.C. Taylor (1999): On bias in maximum likelihood estimators, J. Statist. Planning and Inference 76 (1999), 31-39

Mardia, K.V. and P.E. Jupp (1999): Directional statistics, J. Wiley, New York 1999

Marinkovic, P, Grafarend, E. and T. Reubelt (2003): Space gravity spectroscopy: the benefits of Taylor-Karman structured criterion matrices, Advances in Geosciences 1 (2003), 113-120

Maritz, J. S. (1953): Estimation of the correlation coefficient in the case of a bivariate normal population when one of the variables is dichotomized, Psychometrika, 18, 97-110.

Mariwalla, K.H. (1971): Coordinate transformations that form groups in the large, in: De Sitter and Conformal Groups and their Applications, A.O. Barut and W.E. Brittin (Hrsg.), Vol. 13, 177-191, Colorado Associated University Press, Boulder 1971

Markatou, M. and X. He (1994): Bounded influence and high breakdown point testing procedures in linear models, J. Am. Statist. Ass. 89, 543-49, 1994

Markiewicz, A. (1996): Characterization of general ridge estimators, Statistics & Probability Letters 27 (1996), 145-148

Markov, A.A. (1912): Wahrscheinlichkeitsrechnung, 2nd edition, Teubner, Leipzig 1912

Marošsevièc, T. and D. Jukièc (1997): Least orthogonal absolute deviations problem for exponential function, Student 2 (1997), 131-138

Marquardt, D.W. (1963): An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Indust. Appl. Math. 11 (1963), 431-441

Marquardt, D.W. (1970): Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, Technometrics 12 (1970), 591-612

Marriott, J. and P. Newbold (1998): Bayesian comparison of ARIMA and stationary ARMA models, Journal of Statistical Review 66 (1998), 323-336

Marsaglia, G. and G.P.H. Styan (1972): When does rank (A+B) = rank A + rank B?, Canad. Math. Bull. 15 (1972), 451-452

Marsaglia G. and G.P.H. Styan (1974a): Equalities and inequalities for ranks of matrices. Linear Multilinear Algebra 2 (1974), 269-292.

Marsaglia G. and G.P.H. Styan (1974b): Rank conditions for generalized inverses of partitioned matrices. Sankhya, Ser. A 36 (1974), 437-442.

Marshall, J. (2002): L1-norm pre-analysis measures for geodetic networks, Journal of Geodesy 76 (2002), 334-344

Martinec, Z. (2002): Lecture Notes 2002. Scalar surface spherical harmonics, GeoForschungsZentrum Potsdam 2002

Maruyama, Y. (1998): Minimax estimators of a normal variance, Metrika 48 (1998), 209-214

Masry, E. (1997): Additive nonlinear ARX time series and projection estimates, Econometric Theory 13 (1997), 214-252

Mastronardi, N., Lemmerling, P. and S. van Huffel (2000): Fast structured total least squares algorithm for solving the basic deconvolution problem, Siam J. Matrix Anal. Appl. 22 (2000), 533-553

Masuyama, M. (1939): Correlation between tensor quantities, Proc. Phys.-Math. Soc. Japan, (3): 21, 638-647

Masuyama, M. (1939b): Tensor characteristic of vector set,-Proc. Phys.-Math. Soc. Japan (3): 21-648

Mathai, A.M. (1997): Jacobians of matrix transformations and functions of matrix arguments, World Scientific, Singapore 1997

Mathai, A.M. and Provost, S.B. (1992): Quadratic Forms in Random Variables. Marcel Dekker, New York, 367p

Mathar, R. (1997): Multidimensionale Skalierung, B. G. Teubner Verlag, Stuttgart 1997.

Matheron, G. (1970): The theory of regionalized variables and its applications. Fascicule 5, Les Cahiers du Centre de Morphologie Mathematique. Ecole des Mines de Paris, Fontainebleau, 211 p

Matheron, G. (1993): The intrinsic random functions and their applications, Adv. App. Prob. 5 (1973) 439-468

Mathes, A. (1998): GPS und GLONASS als Teil eines hybrid Meßsystems in der geodäsie am Beispiel des Systems HIGGINS, Dissertationen, DGK, Reihe C, Nr. 500.

Mathew, T. (1989): Optimum invariant tests in mixed linear models with two variance components, Statistical Data Analysis and Inference, Y. Dodge (ed.), North-Holland 1989, 381-388

Mathew, T. (1997): Wishart and Chi-Square Distributions Associated with Matrix Quadratic Forms, J. Multivar. Anal. 61 (1997), 129-143

Mathew, T. and B.K. Sinha (1988): Optimum tests in unbalanced two-way models without interaction, Ann. Statist. 16 (1988), 1727-1740

Mathew, T. and S. Kasala (1994): An exact confidence region in multivariate calibration, The Annals of Statistics 22 (1994), 94-105

Mathew, T. and W. Zha (1996): Conservative confidence regions in multivariate calibration, The Annals of Statistics 24 (1996), 707-725

Mathew, T. and K. Nordström (1997): An inequality for a measure of deviation in linear models, The American Statistician 51 (1997), 344-349

Mathew, T. and W. Zha (1997): Multiple use confidence regions in multivariate calibration, J. Am. Statist. Ass. 92 (1997), 1141-1150

Mathew, T. and W. Zha (1998): Some single use confidence regions in a multivariate calibration problem, Applied Statist. Science III (1998), 351-363

Mathew, T., Sharma, M.K. and K. Nordström (1998): Tolerance regions and multiple-use confidence regions in multivariate calibration, The Annals of Statistics 26 (1998), 1989-2013

Mathias, R. and G.W. Stewart (1993): A block QR algorithm and the singular value decomposition, Linear Algebra Appl. 182 (1993), 91-100

Matz, G., Hlawatsch, F. and Kozek, W. (1997): Generalized evolutionary spectral analysis and the Weyl-spectrum of non stationary random processes. In IEEE Trans. Signal. Proc., vol. 45, no. 6, pp. 1520-1534.

Mauly, B.F.J. (1976): Exponential data transformations, Statistician 27 (1976), 37-42

Mautz, R. (2001): Zur Lösung nichtlinearer Ausgleichungsprobleme bei der Bestimmung von Frequenzen in Zeitreihen, DGK, Reihe C, Nr. 532.

Mautz, R. (2002): Solving nonlinear adjustment problems by global optimization, Boll. di Geodesia e Scienze Affini 61 (2002), 123-134

Maxwell, S.E. (2003): Designig experiments and analyzing data. A model comparison perspective, Lawrence Erlbaum Associates, Publishers, London New Jersey 2003

Maybeck, P. (1979): Stochastic models, estimation, and control, Vol. 1, Academic Press, New York London 1979

Maybeck, P. (1982): Stochastic models, estimation and control, Vol. 1 and 2, Academic Press, Washington 1982

Mayer, D.H. (1975): Vector and tensor fields on conformal space, J. Math. Physics 16 (1975), 884-893

McComb, W.D. (1990): the physics of fluid turbulence, Clarendon Press, 572 pages, Oxford 1990

McCullagh, P. (1983): Quasi-likelihood functions, The Annals of Statistics 11 (1983), 59-67

McCullagh, P. (1987): Tensor methods in statistics, Chapman and Hall, London 1987

McCullagh, P. and Nelder, J.A. (1989): Generalized linear models, Chapman and Hall, London 1989

McCullagh, P. and Nelder, L. (2006): Generalized linear models. Chapman and Hall, CRC, New York

McCulloch, C.E. (1997): Maximum likelihood algorithms for generalized linear mixed models, J. Am. Statist. Ass. 92 (1997), 162-170

McCulloch, C.E. and S.R. Searle (2001): Generalized, linear, and mixed models, J. Wiley, New York 2001

McElroy, F.W. (1967): A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased, J. Am. Statist. Ass. 62 (1967), 1302-1304

McGaughey, D.R. and G.J.M Aitken (2000): Statistical analysis of successive random additions for generating fractional Brownian motion, Physics A 277 (2000) 25-34
McGilchrist, C.A. (1994): Estimation in generalized mixed models, J. Roy. Statist. Soc. B56 (1994), 61-69
McGilchrist, C.A. and C.W. Aisbett (1991): Restricted BLUP for mixed linear models, Biometrical Journal 33 (1991), 131-141
McGilchrist, C.A. and K.K.W. Yau (1995): The derivation of BLUP, ML, REML estimation methods for generalised linear mixed models, Commun. Statist.-Theor. Meth. 24 (1995), 2963-2980
McMorris, F.R. (1997): The median function on structured metric spaces, Student 2 (1997), 195-201
Meditch, J.S. (1969): Stochastic optimal linear estimation and control, McGraw Hill, New York 1969
Meetz, K. and W.L. Engl (1980): Electromagnetic field, Springer
Mehta, M.L. (1967): Random/matrices and the statistical theory of energy levels, Academic Press, New York
Mehta, M.L. (1991): Random matrices, Academic Press, New York London 1991
Meidow, J., Beder, C. and Forstner, W. (2009): Reasoning with uncertain points, straight lines, and straight line segments in 20. ISPRS Journal of Photogrammetry and Remote Sensing, 64. Jg. 2009, Heft: 2, S. 125-139.
Meidow, J., Förstner, W. and C. Beder (2009): Optimal parameter estimation with homogeneous entities and arbitrary constraints, pappers recognition, proc. 31st DAGM Sym. , Jena, Germany, pages 292-301, Springer Verlag, Berlin-Heidelberg 2010
Meier, S. (1967): Die Terrestriche Refraktion in Kongsfjord (West Spitzbergen), Geod. Geophys. Vertiff., Reihe 3, pp. 32-51, Berlin 1967
Meier, S. (1970): Beitrdge zur Refraktion in Kohen Breiten, Geod. Geophys. Veroff., Reihe, Berlin 1970
Meier, S. (1975): Zur Geninigkert der Trigonometrischen Hohen messung über Eis bei stabiler Temperatur and Refrektions schichtung, Vermessungstechnik 23 (1975) 382-385
Meier, S. (1976): Über die Abhungigkeit der Refraktions schwankung von der Zeileite, vermessungstechnik 24 (1976) 378-382
Meier, S. (1977): Zur Refraktions schwankung in der turbulenten unterschicht der Atmosphere Vermassungstechnik 25 (1977) 332-335
Meier, S. (1978): Stochastische Refreaktions modelle, Geod. Geophys. Reihe3, 40 (1978) 63-73
Meier, S. (1978): Refraktions schwankung and Turbilenz, Geod. Geophy. Reihe3, 41, (1978) 26-27
Meier, S. (1981): Planar geodetic covariance functions , Rev. of Geophysics and Space Physics 19 (1981) 673-686
Meier, S. (1982): Stochastiche Refraktions modell für Zielstrahlen in dreidimensionalen Euklischen Raum Vermessungstechnik 30 (1982) 55-57
Meier, S. (1982): Stochastiche Refraktions modell für Zielstrahlen in eiven Vertikalchene der Freien Atmosphere Vermessungstechnik 30 (1982) 55-57
Meier, S. (1982): Varianz and Erhaltungs neiyung der lokalen Lichtstrahlkrummung in Bodennahe, Vermessungstechnik 30 (1982) 420-423
Meier, S. and W. Keller (1990): Geostatistik, Einfuhrung in der Theorie der Zufulls prozesse, Springer Verley, Wein New York 1990
Meinhold, P. und E. Wagner (1979): Partielle Differentialgleichungen. Frankfurt/Main.
Meissl, P. (1965): über die Innere Genauigkeit dreidimensionaler Punkthaufen, Z. Vermessungswesen 90 (1965), 198-118
Meissl, P. (1969): Zusammenfassung und Ausbau der inneren Fehlertheorie eines Punkthaufens, Deutsche Geod. Kommission A61, München 1994, 8-21
Meissl, P. (1970): Ober die Fehlerfortpflanzung in gewissen regelmaBigen flachig ausgebreiteten Nivellementnetzen. Zeitschrift für Vermessungswesen 2i, 103-109.
Meissl, P. (1976): Hilbert spaces and their application to geodetic least-squares problems, Bolletino di Geodesia e Scienze Affini 35 (1976), 49-80

Meissl, P. (1982): Least squares adjustment. A modern approach, Mitteilungen der geodätischen Institut der Technischen Universität Craz, Folge 43.
Melas, V.B. (2006): Functional approach to optimal experimental design. Springer, Berlin
Melbourne, W. (1985): The case of ranging in GPS-based geodetic systems, Proc.1st Int.Symp. on Precise Positioning with GPS, Rockville, Maryland (1985), 373-386
Menz, J. (2000): Forschungsergebnisse zur Geomodellierung und deren Bedeutung, Manuskript 2000
Menz, J. and N. Kolesnikov (2000): Bestimmung der Parameter der Kovarianzfunktionen aus den Differenzen zweier Vorhersageverfahren, Manuskript, 2000
Merriman, M. (1877): On the history of the method of least squares, The Analyst 4 (1877)
Merriman, M. (1884): A textbook on the method of least squares, J. Wiley, New York 1884
Merrit, E. L. (1949): Explicit Three-point resection in space, *Phot. Eng.* 15 (1949) 649-665.
Meyer, C.D. (1973): Generalized inverses and ranks of block matrices, SIAM J. Appl. Math. 25 (1973), 597-602
Mhaskar, H.N., Narcowich, F.J and J.D. Ward (2001): Representing and analyzing scattered data on spheres, In: Multivariate Approximations and Applications, Cambridge University Press, Cambridge 2001, 44-72
Mierlo van, J. (1980): Free network adjustment and S-transformations, Deutsche Geod. Kommission B252, München 1980, 41-54
Mierlo van, J. (1982): Difficulties in defining the quality of geodetic networks. Proceedings Survey Control Networks, ed. Borre, K., Welsch, W.M., Vol. 7, pp. 259-274. Schriftenreihe Wiss. Studiengang Vermessungswesen, Hochschule der Bundeswehr, München.
Migon, H.S. and D. Gammermann (1999): Statistical inference, Arnold London 1999
Milbert, D. (1979): Optimization of Horizontal Control Networks by nonlinear Programming. NOAA, Technical Report NOS 79 NGS 12, Rockville
Millar, A.M. and Hamilton, D.C. (1999): Modern outlier detection methods and their effect on subsequent inference. J. Stat. Comput. Simulation 64, 2 (1999), 125-150.
Miller, R.G. (1966): Simultaneous statistical inference, Mc Graw-Hill Book Comp., New York 1966
Mills, T.C. (1991): Time series techniques for economists, Cambridge University Press, Cambridge 1991
Minzhong, J. and C. Xiru (1999): Strong consistency of least squares estimate in multiple regression when the error variance is infinite, Statistica Sinica 9 (1999), 289-296
Minkler, G. and J. Minkler (1993): Theory and application of Kalman filtering, Magellan Book Company 1993
Minoux, M. (1986): Mathematical Programming. John Wiley & Sons, Chicester.
Mirsky, C. (1960): Symmetric gauge functions and unitarily invariant norms, Quarterly Journal of Mathematics, 11 (1960), pp. 50-59
Mises, von R. (1918): Ueber die Ganzzahligkeit der Atomgewichte und verwandte Fragen, Phys.Z. 19 (1918) 490-500
Mises von, R. (1945): On the classification of observation data into distinct groups, Annals of Mathematical Statistics, 16, 68-73.
Misra, P. and Enge, P. (2001): Global Positioning System-Signals, Measurement, and Performance. Ganga-Jamuna Press, Lincoln, Massachusetts, 2001
Misra, R.K. (1996): A multivariate procedure for comparing mean vectors for populations with unequal regression coefficient and residual covariance matrices, Biometrical Journal 38 (1996), 415-424
Mitchell, T.J. and Morris, M.D. (1992): Bayesian design and analysis of computer experiments: two examples. Stat Sinica 2:359-379
Mitra, S.K. (1971): Another look at Rao's MINQUE of variance components, Bull. Inst. Inernat. Stat. Assoc., 44 (1971), pp. 279-283
Mitra, S.K. (1982): Simultaneous diagonalization of rectangular matrices, Linear Algebra Appl. 47 (1982), 139-150

Mitra, S.K. and Rao, C.R. (1968): Some results in estimation and tests of linear h-ypotheses under the Gauss-Markov model. Sankhya, Ser. A, 30 (1968), 281-290.

Mittermayer, E. (1972): A generalization of least squares adjustment of free networks, *Bull. Geod.* No. 104 (1972) 139-155.

Moenikes, R., Wendel, J. and Trommer, G.F. (2005): A modified LAMBDA method for ambiguity resolution in the presence of position domain constraints. In: Proceedings ION GNSS2005. Long Beach

Mohan, S.R. and S.K. Neogy (1996): Algorithms for the generalized linear complementarity problem with a vertical block Z-matrix, Siam J. Optimization 6 (1996), 994-1006

Moire, C. and J.A. Dawson (1992): Distribution, Chapman and Hall, Boca Raton 1996

Molenaar, M. (1981): A further inquiry into the theory of S - transformations and criterion matrices. Neth. Geod. Comm., Publ. on Geodesy. New Series, Vol. 6, Nr. 1, Delft

Möller, H. M. (1998): Gröbner bases and numerical analysis, In Groebner bases and applications. Eds. B. Buchberger and F. Winkler, London mathematical society lecture note series 251. Cambridge university press, pp. 159-177, 1998.

Molz, F.J., Liu, H.H. and Szulga, J. (1987): Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties and extensions, Water Resour. Res., 33(10), 2273-2286z

Money, A.H. et al. (1982): The linear regression model: Lp-norm estimation and the choice of p, Commun. Statist. Simul. Comput. 11 (1982), 89-109

Monin, A.S. and A.M. Yaglom (1981): Statistical fluid mechanics: mechanics of turbulence, Vol. 2, The Mit Press, Cambridge 1981

Monin, A.S. and A.M. Yaglom (1971): Statistical fluid mechanics, Vol I, 769 pages, MIT Press, Cambridge/Mass 1971

Monin, A.S. and A.M. Yaglom (1975): Statistical fluid mechanics, Vol II, 874 pages, pages, MIT Press, Cambridge/Mass 1975

Montfort van, K. (1989): Estimating in structural models with non-normal distributed variables: some alternative approaches, 'Reprodienst, Subfaculteit Psychologie', Leiden 1989

Montgomery, D.C. (1996): Introduction to statistical quality control, 3rd edition, J. Wiley, New York 1996

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2001): Introduction to linear regression analysis. John Wiley & Sons, Chichester, 2001 (3rd ed).

Mood, A.M., Graybill, F.A. and D.C. Boes (1974): Introduction to the theory of statistics, 3rd ed., McGraw-Hill, New York 1974

Moon, M.S. and R.F. Gunst (1994): Estimation of the polynomial errors-in-variables model with decreasing error variances, J. Korean Statist. Soc. 23 (1994), 115-134

Moon, M.S. and R.F. Gunst (1995): Polynomial measurement error modeling, Comput. Statist. Data Anal. 19 (1995), 1-21

Moore, E.H. (1900): A fundamental remark concerning determinantal notations with the evaluation of an important determinant of special form, Ann. Math. 2 (1900), 177-188

Moore, E.H. (1920): On the reciprocal of the general algebraic matrix, Bull. Amer. Math. Soc. 26 (1920), 394-395

Moore, E.H. (1935): General Analysis, Memoirs American Philosophical Society, pp. 147-209, New York

Moore, E.H. and M.Z. Nashed (1974): A general approximation theory of generalized inverses of linear operators in Banoch spaces, SIAM J. App. Math. 27 (1974)

Moors, J.J.A., and van Houwelingen, V.C. (1987): Estimation of linear models with inequality restrictions, Tilburg University Research report FEW 291, The Netherlands

Morgan, B.J.T. (1992): Analysis of quantal response data, Chapman and Hall, Boca Raton 1992

Morgan, A.P. (1992): Polynomial continuation and its relationship to the symbolic reduction of polynomial systems, In Symbolic and Numerical Computations for Artificial Intelligence, pp. 23-45, 1992.

Morgenthaler, S. (1992): Least-absolute-deviations fits for generalized linear models, Biometrika 79 (1992), 747-754
Morgera, S. (1992): The role of abstract algebra in structured estimation theory, IEEE Trans. Inform. Theory 38 (1992), 1053-1065
Moritz, H. (1972): Advanced least - squares estimation. The Ohio State University, Department of Geodetic Science, Report No. 175, Columbus..
Moritz, H. (1973): Least-squares collocation. DGK, A 59, Muenchen
Moritz, H. (1976): Covariance functions in least-squares collocation, Rep. Ohio State Uni. 240
Moritz, H. (1978): The operational approach to physical geodesy. The Ohio State University, Department of Geodetic Science, Report 277, Columbus
Moritz, H. (1980): Advanced physical geodesy. Herbert Wichmann Verlag Karlsruhe
Moritz, H. and Suenkel, H. (eds) (1978): Approximation methods in geodesy. Sammlung Wichmann Neue Folge, Band 10, Herbert Wichmann Verlag
Morris, C.N. (1982): Natural exponential families with quadratic variance functions, Ann. Statist. 10 (1982), 65-80
Morris, M.D. and Mitchell, T.J. and Ylvisaker, D. (1993): Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. Technometrics 35:243-255
Morrison, D.F. (1967): Multivariate statistical methods, Mc Graw-Hill Book Comp., New York 1967
Morrison, T.P. (1997): The art of computerized measurement, Oxford University Press, Oxford 1997
Morsing, T. and C. Ekman (1998): Comments on construction of confidence intervals in connection with partial least-squares, J. Chemometrics 12 (1998), 295-299
Moser, B.K. and J.K. Sawyer (1998): Algorithms for sums of squares and covariance matrices using Kronecker Products, The American Statistician 52 (1998), 54-57
Mount, V.S. and Suppe J. (1992): Present-day stress orientations adjacent to active strike-slip faults: California and Sumatra, J. geophys. Res., B97, 11995-12013
Moutard, T. (1894): Notes sur la propagation des ondes et les èequations de l'hydroudynamique, Paris 1893, reprint Chelsea Publ., New York 1949
Mudholkar, G.S. (1997): On the efficiencies of some common quick estimators, Commun. Statist.-Theory Meth. 26 (1997), 1623-1647
Mukhopadhyay, P. and R. Schwabe (1998): On the performance of the ordinary least squares method under an error component model, Metrika 47 (1998), 215-226
Mueller, C.H. (1997): Robust planning and analysis of experiments, Springer-Verlag, Heidelberg Berlin New York 1997
Mueller, C.H. (1998): Breakdown points of estimators for aspects of linear models, in: MODA 5 - Advances in model-oriented data analysis and experimental design, Proceedings of the 5th International Workshop in Marseilles, eds. Atkinson, A.C., Pronzato, L. and H.P. Wynn, Physica-Verlag, Heidelberg 1998
Mueller, H. (1985): Second-order design of combined linear-angular geodetic networks, Bull. Geodèesique 59 (1985), 316-331
Muller, B., Zoback, M.L., Fuchs, K., Mastin, L., Geregersen, S., Pavoni, N., Stephansson, O. and Ljunggren, C. (1992): Regional patterns of tectonic stress in Europe, J. geophys. Res., B97, 11783-11803
Muller, D.E. (1956): A method for solving algebraic equations using an automatic computer. Mathematical Tables and other Aids to Computation, Vol. X, No. 56, 208-215.
Müller, F.J. (1925): Direkte (Exakte) Lösungen des einfachen Rückwarscnittseinschneidens im Raum. 1 Teil, *Zeitschrift für Vermessungswesen* 37 (1925) 249-255, 265-272, 349-353, 365-370, 569-580.
Müller, H. and M. Illner (1984): Gewichtsoptimierung geodätischer Netze. Zur Anpassung von Kriteriummatrizen bei der Gewichtsoptimierung, Allgemeine Vermessungsnachrichten (1984), 253-269

Müller, H. and G. Schmitt (1985): SODES2 - Ein Programm-System zur Gewichtsoptimierung zweidimensionaler geodätischer Netze. Deutsche Geodätische Kommission, München, Reihe B, 276 (1985)
Müller, W.G. (2005): A comparison of spatial design methods for correlated observations. Environmetrics 16:495-505
Müller, W.G. and Pazman, A. (1998): Design measures and approximate information matrices for experiments without replications. J Stat Plan Inf 71:349-362
Müller, W.G. and Pazman, A. (1999): An algorithm for the computation of optimum designs under a given covariance structure. J Comput Stat 14:197-211
Müller, W.G. and Pazman, A. (2003): Measures for designs in experiments with correlated errors. Biometrika 90:423-434
Müller, P., Sanso, B. and De Iorio, M. (2004): Optimal Bayesian design by inhomogeneous Markov Chain simulation. J Am Stat Assoc 99: 788-798
Mueller, J. (1987): Sufficiency and completeness in the linear model, J. Multivar. Anal. 21 (1987), 312-323
Mueller, J., Rao, C.R. and B.K. Sinha (1984): Inference on parameters in a linear model: a review of recent results, in: Experimental design, statistical models and genetic statistics, K. Hinkelmann (ed.), Chap. 16, Marcel Dekker, New York 1984
Mueller, W. (1995): An example of optimal design for correlated observations, Österreichische Zeitschrift für Statistik 24 (1995), 9-15
Mueller, W. (1998): Spatial data collection, contributions to statistics, Physica Verlag, Heidelberg 1998
Mueller, W. (1998): Collecting spatial data - optimum design of experiments for random fields, Physica-Verlag, Heidelberg 1998
Mueller, W. (2001): Collecting spatial data - optimum design of experiments for random fields, 2nd ed., Physica-Verlag, Heidelberg 2001
Mueller, W. and A. Pèazman (1998): Design measures and extended information matrices for experiments without replications, J. Statist. Planning and Inference 71 (1998), 349-362
Mueller, W. and A.Pèazman (1999): An algorithm for the computation of optimum designs under a given covariance structure, Computational Statistics 14 (1999), 197-211
Mueller-Gronbach, T. (1996): Optimal designs for approximating the path of a stochastic process, J. Statist. Planning and Inference 49 (1996), 371-385
Muir, T. (1911): The theory of determinants in the historical order of development, volumes 1-4, Dover, New York 1911, reprinted 1960
Muirhead, R.J. (1982): Aspects of multivariate statistical theory, J. Wiley, New York 1982
Mukherjee, K. (1996): Robust estimation in nonlinear regression via minimum distance method, Mathematical Methods of Statistics 5 (1996), 99-112
Mukhopadhyay, N. (2000): Probability and statistical inference, Dekker, New York 2000
Muller, C. (1966): Spherical harmonics - Lecture notes in mathematics 17 (1966), Springer-Verlag, Heidelberg Berlin New York, 45 pp.
Muller, D. and Wei, W.W.S. (1997): Iterative least squares estimation and identification of the tranfer function model, J. Time Series Analysis 18 (1997), 579-592
Mundt, W. (1969): Statistische Analyse geophysikalischer Potentialfelder hinsichtlich Aufbau und Struktur der tieferen, Erdkruste, Deutsche Akademic der Wissen schaften, Berlin 1969
Munoz-Pichardo, J.M., Munoz-Garcèìa, J., Fernèandez-Ponce, J.M. and F. Lèopez-Blèaquez (1998): Local influence on the general linear model, Sankhya: The Indian Journal of Statistics 60 (1998), 269-292
Musyoka, S. M. (1999): Ein Modell für ein vierdimensionales integriertes regionales geodätisches Netz, Ph. D. thesis, Department of Geodesy, University of Karlsruhe (Internet Publication), 1999.
Murray, F.J. and J. von Neumann (1936): On rings of operator I, Ann. Math. 37 (1936), 116-229
Murray, J.K. and J.W. Rice (1993): Differential geometry and statistics, Chapman and Hall, Boca Raton 1993
Myers, J.L. (1979): Fundamentals of experimental designs, Allyn and Bacon, Boston 1979

Naes, T. and H. Martens (1985): Comparison of prediction methods for multicollinear data, Commun. Statist. Simulat. Computa. 14 (1985), 545-576
Nagaev, S.V. (1979): Large deviations of sums of independent random variables, Ann. Probability 7 (1979), 745-789
Nagar, A.L. and N.C. Kakwani (1964): The bias and moment matrix of a mixed regression estimator, Econometrica, 32 (1964), pp. 174-182
Nagar, A.L. and N.C. Kakwani (1969): Note on the use of prior information in statistical estimation of econometrics relation, Sankhya series A 27, (1969), pp. 105-112
Nagar, D.K. and A.K. Gupta (1996): On a test statistic useful in Manova with structured covariance matrices, J. Appl. Stat. Science 4 (1996), 185-202
Nagaraja, H.N. (1982): Record values and extreme value distributions, J. Appl. Prob. 19 (1982), 233-239
Nagarsenker, B.N. (1977): On the exact non-null distributions of the LR criterion in a general MANOVA model, Sankya A39 (1977), 251-263
Nakamura, N. and S.Konishi (1998): Estimation of a number of components for multivariate normal mixture models based on information criteria, Japanese J. Appl. Statist. 27 (1998), 165-180
Nakamura, T. (1990): Corrected score function for errors-invariables models: methodology and application to generalized linear models, Biometrika 77 (1990), 127-137
Namba, A. (2001): MSE performance of the 2SHI estimator in a regression model with multivariate t error terms, Statistical Papers 42(2001), 82-96
Nashed, M.Z. (1970): Steepest descent for singular linear operator equations, SIAM J. Number. Anal. 7 (1970), pp. 358-362
Nashed, M.Z. (1971): Generalized inverses, normal solvability, and iteration for singular operator equations pp. 311-359 in Nonlinear Functional Analysis and Applications (L. B. Rail, editor), Academic Press, New York, 1971
Nashed, M.Z. (1974): Generalized Inverses and Applications, Academic Press, New York 1974
Nashed, M.Z. (1976): Generalized inverses and applications, Academic Press, New York London 1976
Näther, W. (1985): Exact designs for regression models with correlated errors, Statistics 16 (1985), 479-484
Näther, W. and Wälder, K. (2003): Experimental design and statistical inference for cluster point processeswith applications to the fruit dispersion of Anemochorous forest trees, Biometrical J 45, pp. 1006-1022, 2003
Näther, W. and Wälder, K. (2007): Applying fuzzy measures for considering interaction effects in root dispersal models in Fuzzy Sets and Systems 158; S. 572-582, 2007
Nelder, J.A. and R.W.M. Wedderburn (1972): Generalized linear models, J. Roy. Stat. Soc., Series A 35, (1972), pp. 370-384
Nelson, C.R. (1973): Applied time series analysis for managerial forecasting, Holden Day, San Francisco, 1973
Nelson, R. (1995): Probability, stochastic processes, and queueing theory, Springer-Verlag, Heidelberg Berlin New York 1995
Nesterov, Y.E. and A.S. Nemirovskii (1992): Interior point polynomial methods in convex programming, Springer-Verlag, Heidelberg Berlin New York 1992
Nesterov, Y.E. and A.S. Nemirovskii (1994): Interior point polynomial algorithms in convex programming, SIAM Vol. 13, Pennsylvania 1994
Neudecker, H. (1968): The Kronecker matrix product and some of its applications in econometrics, Statistica Neerlandica 22 (1968), 69-82
Neudecker, H. (1969): Some theorems on matrix differentiation with special reference to Kronecker matrix products, J. Am. Statist. Ass. 64 (1969), 953-963
Neudecker, H. (1978): Bounds for the bias of the least squares estimator of 2 in the case of a first-order autoregressive process (positive autocorrelation), Econometrica 45 (1977), 1257-1262
Neudecker, H. (1978): Bounds for the bias of the LS estimator of 2 in the case of a first-order (positive) autoregressive process when the regression contains a constant term, Econometrica 46 (1978), 1223-1226

Neumaier, A. (1998): Solving ill-conditioned and singular systems: a tutorial on regularization, SIAM Rev. 40 (1998), 636-666

Neumann, J. (2009): Zur modellierung cines erweitertur unischerheits hanshaltes in Parameter Schetzung and hyphothesen test, Bayer, Akad. Wissen Schaften, Deutsche Geodeitische Kommission Reiche C634, München 2009

Neumann, J. and H. Kutterer (2007): Adaptative Kalman Filtering mit imprazisen Daten, in: F. K. Brunner (ed.) Ingenienrvermessung, pages 420-424, Wickmann Verlag, Heidelberg 2007

Neuts, M.F. (1995): Algorithmic probability, Chapman and Hall, Boca Raton 195

Neutsch, W. (1995): Koordinaten: Theorie und Anwendungen, Spektrum Akademischer Verlag, Heidelberg 1995

Neway, W.K. and J.K. Powell (1987): Asymmetric least squares estimation and testing, Econometrica 55 (1987), 819-847

Newman, D. (1939): The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation, Biometrika 31 (1939), 20-30

Neykov, N.M. and C.H. Mueller (2003): Breakdown point and computation of trimmed likelihood estimators in generalized linear models, in: R. Dutter, P. Filzmoser, U. Gather, P.J. Rousseeuw (eds.), Developments in Robust Statistics, 277-286, Physica Verlag, Heidelberg 2003

Neyman, J. (1934): On the two different aspects of the representative method, J. Roy. Statist. Soc. 97 (1934), 558-606

Neyman, J. (1937): Outline of the theory of statistical estimation based on the classical theory of probability, Phil. Trans. Roy. Soc. London 236 (1937), 333-380

Neytchev P. N. (1995): SWEEP operator for least-squares subject to linear constraints. Comput. Stat. Data Anal. 20, 6 (1995), 599-609.

Nicholson, W.K. (1999): Introduction to abstract algebra, 2nd ed., J. Wiley, New York 1999

Niemeier, W. (1985): Deformationsanalyse, in: Geodätische Netze in Landes- und Ingenieurvermessung II, Kontaktstudium, ed. H. Pelzer, Wittwer, Stuttgart 1985

Nobre, S. and M. Teixeiras (2000): Der Geodät Wilhelm Jordan und C.F. Gauss, Gauss-Gesellschaft e.V. Goettingen, Mitt. 38, 49-54 Goettingen 2000

Nyquist, H. (1988): Least orthogonal absolute deviations, Comput. Statist. Data Anal. 6 (1988), 361-367

Nyquist, H., Rice, S.O. and J. Riordan (l959): The distribution of random determinants, Quart. Appl. Math., 12, 97-104.

O'Neill, M., Sinclair, L.G. and F.J. Smith (1969): Polynomial curve fitting when abscissa and ordinate are both subject of error, Comput. J. 12 (1969), 52-56

O'Neill, M.E. and K. Mathews (2000): A weighted least squares approach to Levene's test of homogeneity of variance, Australia & New Zealand J. Statist. 42 (2000), 81-100

Nicholson, W. K. (1999): Introduction to abstract algebra, Second Edition, John Wiley & Sons, New York-Chichester-weinheim-Brisbane-Singapore 1999.

Nkuite G. (1998): Adjustment with singular variance-covariance matrix for models of geometric deformation analysis. Ph.D. thesis, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften Reihe C: Dissertationen, 501 Mlinchen: Verlag der Bayerischen Akademie der Wissenschaften, 1998 (in German).

Nordstrom K. (1985): On a decomposition of the singular Gauss-Markov model. In: Linear Statistical Inference, Proc. Int. Conf., Poznan, Poland, 1984, Lect. Notes Stat. (Calh1ski T., Klonecki W., eds.) Springer-Verlag, Berlin, 1985, pp. 231-245.

Nordstrom, K. and Fellman, J. (1990): Characterizations and dispersion matrix robustness of efficiently estimable parametric functionals in linear models with nuisance parameters. Linear Algebra Appl. 127 (1990), 341-36l.

Nurhonen, M. and Puntanen, S. (1992): A property of partitioned generalized regression. Commun. Stat., Theory Methods 21, 6 (1992), 1579-1583.

Nussbaum, M. and S. Zwanzig(1988): A minimax result in a model with infmitely many nuisance parameters, Transaction of the Tenth Prague Conference on Information Theory, Statistical Decisions, Random Processes, Prague 1988, 215-222.

Oberlack, M. (1997): Non-Isotropic Dissipation in Non-Homogeneous Turbulence, J. Fluid Mech., vol. 350, pp. 351-374, (1997)
Obuchow, A.M. (1958): Statistische Beschribung stetiger Felder, in: H. Goering (ed.) Sammelband zur ststischen Theorie der Turbolenz, pp. 1-42, Akademic Verlag 1958.
Obuchow, A.M. (1962): Some specific features of atmospheric turbulence, Fluid mechanics 13 (1962) 77-81
Obuchow, A.M. (1962): Some specific features of atmospheric turbulence, J. Geophys. Res. 67 (1962) 3011-3014
Odijk, D. (2002): Fast precise GPS positioning in the presence of ionospheric delays. Publications on Geodesy, vol. 52, Netherlands Geodetic Commission, Delft
Offlinger, R. (1998): Least-squares and minimum distance estimation in the three-parameter Weibull and Frèechet models with applications to river drain data, in: Kahle, et al. (eds.), Advances in stochastic models for reliability, quality and safety, 81-97, Birkhäuser-Verlag, Basel Boston Berlin 1998
Ogawa, J. (1950): On the independence of quadratic forms in a non-central normal system, Osaka Mathematical Journal 2 (1950), 151-159
Ohtani, K. (1988): Optimal levels of significance of a pre-test in estimating the disturbance variance after the pre-test for a linear hypothesis on coefficients in a linear regression, Econom. Lett. 28 (1988), 151-156
Ohtani, K. (1996): Further improving the Stein-rule estimator using the Stein variance estimator in a misspecified linear regression model, Statist. Probab. Lett. 29 (1996), 191-199
Ohtani, K. (1998): On the sampling performance of an improved Stein inequality restricted estimator, Austral. & New Zealand J. Statist. 40 (1998), 181-187
Ohtani, K. (1998): The exact risk of a weighted average estimator of the OLS and Stein-rule estimators in regression under balanced loss, Statistics & Decisions 16 (1998), 35-45
Okamoto, M. (1973): Distinctness of the Eigenvalues of a quadratic form in a multivariate sample, Ann. Stat. 1 (1973), 763-765
Okeke, F.I. (1998): The curvilinear datum transformation model, DGK, Reihe C, Heft Nr. 481.
Okeke, F. and F. Krumm (1998): Graph, graph spectra and partitioning algorithms in a geodetic network structural analysis and adjustment, Bolletino di Geodesia e Scienze Affini 57 (1998), 1-24
Oktaba, W. (1984): Tests of hypotheses for the general Gauss-Markov model. Biom. J. 26 (1984), 415-424.
Olkin, I. (1998): The density of the inverse and pseudo-inverse of a random matrix, Statistics and Probability Letters 38 (1998), 131-135
Olkin, J. (2000): The 70th anniversary of the distribution of random matrices: a survey, Technical Report No. 2000-06, Department of Statistics, Stanford University, Stanford 2000
Olkin, I. and S.N. Roy (1954): On multivariate distribution theory, Ann. Math. Statist. 25 (1954), 329-339
Olkin, I. and A.R. Sampson (1972): Jacobians of matrix transformations and induced functional equations, Linear Algebra Appl. 5 (1972), 257-276
Olkin, J. and J.W. Pratt (1958): Unbiased estimation of certain correlation coefficient, Annals Mathematical Statistics 29 (1958), 201-211
Olsen, A., Seely, J. and D. Birkes (1976): Invariant quadratic unbiased estimators for two variance components, Annals of Statistics 4 (1976), 878-890
Omre, H. (1987): Bayesian krigingmerging observations and qualified guess in kriging. Math Geol 19:25-39
Omre, H. and Halvorsen, K. (1989): The Bayesian bridge between simple and universal kriging. Math Geol 21:767-786
Ord, J.K. and S. Arnold (1997): Kendall's advanced theory of statistics, volume IIA, classical inference, Arnold Publ., 6th edition, London 1997
Ortega, J. and Rheinboldt, W. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.
Osborne, M.R. (1972): Some aspects of nonlinear least squares calculations, Numerical Methods for Nonlinear Optimization, ed. F.A. Lootsma, Academic Press, New York London 1972

Osborne, M.R. (1976): Nonlinear least squares the Levenberg algorithm revisited, J. Aust. Math. Soc. B19 (1976), 343-357
Osborne, M.R. and G.K. Smyth (1986): An algorithm for exponential fitting revisited, J. App. Prob. (1986), 418-430
Osborne, M.R. and G.K. Smyth (1995): A modified Prony algorithm for fitting sums of exponential functions, SIAM J. Sc. And Stat. Comp. 16 (1995), 119-138
Osiewalski, J. and M.F.J. Steel (1993): Robust Bayesian inference in elliptical regression models, J. Economatrics 57 (1993), 345-363
Ouellette, D.V. (1981): Schur complements and statistics, Linear Algebra Appl. 36 (1981), 187-295
Owens, W.H. (1973): Strain modification of angular density distributions, Techtonophysics 16 (1973), 249-261
Ozone, M.I. (1985): Non-iterative solution of the ϕ equations, *Surveying and Mapping* 45 (1985) 169-171.
Paciorek, C. (2007): Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectialGP package, J. Statist. Software 19 (2007) 1-38
Padmawar, V.R. (1998): On estimating nonnegative definite quadratic forms, Metrika 48 (1998), 231-244
Pagano, M. (1978): On periodic and multiple autoregressions, Annals of Statistics 6 (1978), 1310-1317
Paige, C.C. and M.A. Saunders (1975): Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal. 12 (1975), 617-629
Paige, C. and C. van Loan (1981): A Schur decomposition for Hamiltonian matrices, Linear Algebra and its Applications 41 (1981), 11-32
Pakes, A.G. (1999): On the convergence of moments of geometric and harmonic means, Statistica Neerlandica 53 (1999), 96-110
Pal, N. and W.K. Lim (2000): Estimation of a correlation coefficient: some second order decision - theoretic results, Statistics and Decisions 18 (2000), 185-203
Pal, S.K. and P.P. Wang (1996): Genetic algorithms for pattern recognition, CRC Press, Boca Raton 1996
Palancz, B. and J.L. Awange (2012): Application of Pareto optimality to linear models with errors-in-all-variables. J. Geo. (2011), doi: 10.1007/s00190-011-0536-1
Panazachos C. and Kiratzi A. (1992): A formulation for reliable estimation of active crustal deformation and its application to central Greece, Geophys. J. Int., 111, 424-432
Panda, R. et. al. (1989): Turbulenz in a randomly stirred fluid, Phys. Fluid, A1 (1989) 1045-1053
Papo, H. (1985): Extended Free Net Adjustment Constraints. Bulletin Geodesique 59, pp. 378-390.
Papo, H. (1986): Extended Free Net Adjustment Constraints. NOAA Technical Report NOS 119 NGS 37, September 1986, 16 p.
Papoulis, A. (1991): Probability, random variables and stochastic processes, McGraw Hill, New York 1991
Park, H. (1991): A parallel algorithm for the unbalanced orthogonal Procrustes problem, Parallel Computing 17 (1991), 913-923
Park, S.H., Kim Y.H. and H. Toutenberg (1992): Regression diagnostic for removing an observation with animating graphics, Statistical papers, 33 (1992), pp. 227-240
Parker, W.V. (1945): The characteristic roots of matrices, Duke Math. J. 12 (1945), 519-526
Parkinson, B.W. and Spilker, J.J. (eds) (1996): Global Positioning System: Theory and Applications. vol 1, American institute of Aeronautics and Astronautics, Washington, DC, 793p
Parthasaratky, K.R. (1967): Probability measures on metric spaces, Academic Press, New York London 1967
Parzen, E. (1962): On estimation of a probability density function and mode, Ann. Math. Statistics 33 (1962), 1065-1073
Passer, W. (1933): Über die Anwendung statischer Methoden auf den Ausgleich von Liniennetzen. OZfV 31, 66-71.

Patel, J.K. and C.B. Read (1982): Handbook of the normal distribution, Marcel Dekker, New York and Basel 1982
Patil, V.H. (1965): Approximation to the Behrens-Fisher distribution, Biometrika 52 (1965), 267-271
Patterson, H.D. and Thompson, R. (1971): Recovery of the inter-block information when block sizes are unequal. Biometrika, 58, 545-554 .
Paul, M.K. (1973): A note on computation of geodetic coordinates from geocentric (Cartesian) coordinates, *Bull. Geod.* No. 108 (1973)135-139.
Pazman, A. (1982): Geometry of Gaussian Nonlinear Regression - Parallel Curves and Confidence Intervals. Kybernetica, 18, 376-396.
Pazman, A. (1984): Nonlinear Least Squares - Uniqueness Versus Ambiguity. Math. Operationsforsch. u. Statist., ser. statist., 15, 323-336.
Pazman, A. and Müller, W.G. (2001): Optimal design of experiments subject to correlated errors. Stat Probab Lett 52:29-34
Pèazman, A. (1986): Foundations of optimum experimental design, Mathematics and its applications, D. Reidel, Dordrecht 1986
Pèazman, A. and J.-B. Denis (1999): Bias of LS estimators in nonlinear regression models with constraints. Part I: General case, Applications of Mathematics 44 (1999), 359-374
Pèazman, A. and W. Müller (1998): A new interpretation of design measures, in: MODA 5 - Advances in model-oriented data analysis and experimental design, Proceedings of the 5th International Workshop in Marseilles, eds. Atkinson, A.C., Pronzato, L. and H.P. Wynn, Physica-Verlag, Heidelberg 1998
Pearson, E.S. (1970): William Sealy Gosset, 1876-1937: Student as a statistician, Studies in the History of Statistics and Probalbility (E.S. Pearson and M.G. Kendall, eds.), Hafner Publ., 360-403, New York 1970
Pearson, E. S. and S. S. Wilks (1933): Methods of statistical analysis appropriate for k samples of two variables, Biometrika, 25, 353-378
Pearson, E.S. and H.O. Hartley (1958): Biometrika Tables for Statisticians Vol. 1, Cambridge University Press, Cambridge 1958
Pearson, K. (1905): The problem of the random walk, Nature 72 (1905), 294
Pearson, K. (1906): A mathematical theory of random migration, Mathematical Contributions to the Theory of Evolution, XV Draper's Company Research Memoirs, Biometrik Series III, London 1906
Pearson, K. (1931): Historical note on the distribution of the Standard Deviations of Samples of any size from any indefinitely large Normal Parent Population, Biometrika 23 (1931), 416-418
Peddada, S.D. and T. Smith (1997): Consistency of a class of variance estimators in linear models under heteroscedasticity, Sankhya 59 (1997), 1-10
Pleitgen, H and Saupe, D (1988): The science of fractal image, Springer Verlag, New York 1988
Peitgen, H-O. and Jurgens, H. and Saupe, D. (1992): Chaos and fractals, springer Verlag, 1992
Pelzer, H. (1971): Zur Analyse geodätischer Deformationsmessungen, Deutsche Geodätische Kommission, Akademie der Wissenschaften, Reihe C (164), München 1971
Pelzer, H. (1974): Zur Behandlung singulärer Ausgleichungsaufgaben, Z. Vermessungswesen 99 (1974), 181-194, 479-488
Pelzer, H. (1985): Geodiitische Netze , in Landes- und Ingenieurvermessung II. Wittwer, Stuttgart, 847p (in German)
Pena, D., Tiao, G.C., and R.S. Tsay (2001): A course in time series analysis, J. Wiley, New York 2001
Penev, P. (1978): The transformation of rectangular coordinates into geographical by closed formulas *Geo. Map. Photo*, 20 (1978) 175-177.
Peng, H.M., Chang, F.R. and Wang, L.S. (1999): Attitude determination using GPS carrier phase and compass data. In: Proceedings of ION NTM 99, pp 727-732
Penrose, R. (1955): A generalised inverse for matrices, Proc. Cambridge Phil. Soc. 51 (1955), 406-413

Penrose, R. (1956): On best approximate solution of linear matrix equations, Proc. Cambridge Philos. Soc. 52 (1956) pp. 17-19

Penny, K.I. (1996): Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis distance, in: Applied Statistics, ed. S.M. Lewis and D.A. Preece, J. Roy. Statist. Soc. 45 (1996), 73-81

Percival, D.B. and A.T. Walden (1993): Spectral analysis for physical applications, Cambridge, Cambridge University Press 1997

Percival, D.B. and A.T. Walden (1999): Wavelet methods for time series analysis, Cambridge University Press, Cambridge 1999

Perelmuter, A. (1979): Adjustment of free networks, *Bull. Geod.* 53 (1979) 291-295.

Perlman, M.D. (1972): Reduced mean square error estimation for several parameters, Sankhya series, B 34 (1972), pp. 89-92

Perrin, J. (1916): Atoms, Van Nostrand, New York 1916

Perron, F. and N. Giri (1992): Best equivariant estimation in curved covariance models, J. Multivar. Anal. 40 (1992), 46-55

Perron, F. and N. Giri (1990): On the best equivariant estimator of mean of a multivariate normal population, J. Multivar. Anal. 32 (1990), 1-16

Persson, C.G., (1980): MINQUE and related estimators for variance components in linear models. PhD thesis, Royal Institute of Technology, Stockholm, Sweden.

Petrov, V.V. (1975): Sums of independent random variables, Springer Verlag, Berlin-Heidelberg-New York 1975

Pfeufer, A. (1990): Beitrag zur Identifikation und Modellierung dynamischer Deformationsprozesse, Vermessungstechnik 38 (1990), 19-22

Pfeufer, A. (1993): Analyse und Interpretation von überwachungsmessungen - Terminologie und Klassifikation, Zeitschrift für Vermessungswesen 118 (1993), 19-22

Pick, M.(1985): Closed formulae for transformation of Cartesian coordinates into a system of geodetic coordinates, *Studia geoph. et geod.* 29 (1985) 653-666.

Pilz, J. (1983): Bayesian estimation and experimental design in linear regression models, Teubner-Texte zur Mathematik 55, Teubner, Leipzig 1983

Pilz, J. and Spock, G. (2008): Bayesian spatial sampling design. In: Proc. 18th Intl. Geostatics World Congeress (J. M. Oritz and X. Emery, Eds.), Gecamin Ltd., Santiago de Chile 2008, 21-30

Pilz, J.(1991): Bayesian estimation and experimental design in linear regression models. Wiley, New York

Pilz, J. Schimek, M.G. and Spock, G. (1997): Taking account of uncertainty in spatial covariance estimation. In: Baafi E, Schofield N (ed) Proceedings of the fifth international geostatistics congress. Kluwer, Dordrecht, pp. 302-313

Pilz, J. and Spock, G. (2006): Spatial sampling design for prediction taking account of uncertain covariance structure. In: Caetano M, PainhoM (ed) Proceedings of accuracy 2006. Instituto Geografico Portugues, pp. 109-118

Pilz, J. and Spock, G. (2008): Why do we need and how should we implement Bayesian kriging methods. Stoch Environ Res Risk Assess 22/5:621-632

Pincus, R. (1974): Estimability of parameters of the covariance matrix and variance components, Math. Operationsforschg. Statistik 5 (1974), 245-248

Pinheiro, J.C. and D.M. Bates (2000): Mixed-effects models in S and S-Plus, Statistics and Computing, Springer-Verlag, Heidelberg Berlin New York 2000

Piquet, J. (1999): Turbulent flows, models and physics, Springer Verlag, 761 pages, Berlin 1999

Pistone, G. and Wynn, H. P. (1996): Generalized confounding with Gröbner bases, *Biometrika* 83 (1996) 112-119.

Pitman, E.J.G. (1979): Some basic theory for statistical inference, Chapman and Hall, Boca Raton 1979

Pitmen, E.J.G. (1939): A note on normal correlation, Biometrika, 31, 9-12.

Pitman, J. and M. Yor (1981): Bessel processes and infinitely divisible laws, unpublished report, University of California, Berkeley

Plachky, D. (1993): An estimation-theoretical characterization of the Poisson distribution, Statistics and Decisions, Supplement Issue 3 (1993), 175-178

Plackett, R.L. (1949): A historical note on the method of least-squares, Biometrika 36 (1949), 458-460

Plackett, R.L. (1972): The discovery of the method of least squares, Biometrika 59 (1972), 239-251

Plato, R. (1990): Optimal algorithms for linear ill-posed problems yield regularization methods, Numer. Funct. Anal. Optim. 11 (1990), 111-118

Pohst, M. (1987): A modification of the LLL reduction algorithm, J. Symbolic Computation 4 (1987), 123-127

Poirier, D.J. (1995): Intermediate statistics and econometrics, The MIT Press, Cambridge 1995

Poisson, S.D. (1827): Connaissance des temps de l'annee 1827

Polasek, W. and S. Liu (1997): On generalized inverses and Bayesian analysis in simple ANOVA models, Student 2 (1997), 159-168

Pollock, D.S.G. (1979): The algebra of econometrics, Wiley, Chichester 1979

Pollock, D.S.G. (1999): A handbook of time series analysis, signal processing and dynamics, Academic Press, San Diego 1999

Polya, G. (1919): Zur Statistik der sphaerischen Verteilung der Fixsterne, Astr. Nachr. 208 (1919), 175-180

Polya, G. (1930): Sur quelques points de la thèeorie des probabilitèes, Ann. Inst. H. Poincare 1 (1930), 117-161

Polzehl, J. and S. Zwanzig (2004): On a symmetrized simulation extrapolation estimator in linear errors-in-variables mddels, Journal of Computational Statistics and Data Analysis.47 (2004) 675-688

Pope, A.J. (1972): Some pitfalls to be avoided in the iterative adjustment of nonlinear problems. Proceedings of the 38th Annual Meeting. American Society of Photogrammetry, Washington D. C., March 1972.

Pope, A.J. (1974): Two Approaches to Nonlinear Least Squares Adjustments. The Canadian Surveyor, Vol. 28, No.5, 663-669.

Pope, A.J. (1976): The statistics of residuals and the detection of outliers, NOAA Technical Report, NOW 65 NGS 1, U.S. Dept. of Commerce, Rockville, Md., 1976

Pope, A. (1982): Two approaches to non-linear least squares adjustments, *The Canadian Surveyor* <u>28</u> (1982) 663-669.

Popinski, W. (1999): Least-squares trigonometric regression estimation, Applicationes Mathematicae 26 (1999), 121-131

Pordzik, P.R. and Trenkler, G. (1996): MSE comparisons of restricted least squares estimators in linear regression model-revisited. Sankhya, Ser. B 58, 3 (1996), 352-359.

Portnoy, S.L. (1977): Robust estimation in dependent situations, Annals of Statistics, 5 (1977), pp. 22-43

Portnoy, S. and Koenker, R. (1997): The Gaussian hare and the Laplacian tortoise: computability of squared error versus absolute-error estimators, Statistical Science 12 (1997), 279-300

Powell, J.L. (1984): Least absolute deviations estimating for the censored regression model, J. of Econometrics, 25 (1984), pp. 303-325

Pratt, J.W. (1961): Length of confidence intervals, J. Am. Statist. Ass. 56 (1961), 549-567

Pratt, J.W. (1963): Shorter confidence intervals for the mean of a normal distribution with known variance, Ann. Math. Statist. 34 (1963), 574-586

Prentice, R.L. and L.P. Zhao (1991): Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, Biometrics, 47 (1991), pp. 825-839

Prenter, P.M. (1975): Splines and Variational Methods. John Wiley & Sons. New York.

Presnell, B. Morrison, S.P. and R.C. Littell (1998): Projected multivariate linear models for directional data, J. Am. Statist. Ass. 93 (1998), 1068-1077

Press, S.J. (1989): Bayesian statistics: Principles, models and applications, J. Wiley, New York 1989

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and B.P. Flannery (1992): Numerical Recipes in FORTRAN (2nd edition), Cambridge University Press, Cambridge 1992

Priestley, M.B. (1981): Spectral analysis and time series, Vols. 1 and 2, Academic Press, New York London 1981

Priestley, M.B. (1988): Nonlinear and nonstationary time series analysis, Academic Press, New York London 1988

Prony, R. (1795): Essai experimentale et analytique, Journal of Ecole Polytechnique (Paris) 1 (1795), 24-76

Pruscha, H. (1996): Angewandte Methoden der Mathematischen Statistik, Teubner Skripten zur Mathematischen Stochastik, Stuttgart 1996

Pugachev, V.S. (1965): Theory of random functions, Oxford 1965

Pugachev, V.S. and I.N. Sinitsyn (2002): Stochastic systems, Theory and applications, Russian Academy of Sciences 2002

Puntanen, S. (1986): Comments on " on necessary and sufficient condition for ordinary least estimators to be best linear unbiased estimators", J. of American Statistical Association, 40 (1986), p. 178

Puntanen, S. (1996): Some matrix results related to a partitioned singular linear model. Commun. Stat., Theory Methods 25, 2 (1996), 269-279.

Puntanen, S. (1997): Some further results related to reduced singular linear models. Commun. Stat., Theory Methods 26, 2 (1997), 375-385.

Puntanen, S. and Styan, G.P.H. (1989): The equality of the ordinary least squares estimator and the best linear unbiased estimator. Amer. Statist. 43 (1989), 153-16l.

Puntanen, S. and Scott, A.J. (1996): Some further remarks on the singular linear model. Linear Algebra Appl. 237/8 (1996), 313-327.

Puntanen, S., Styan, G.P.H. and H.J. Werner (2000): Two matrix-based proofs that the linear estimator Gy is the best linear unbiased estimator, J. Statist. Planning and Inference 88 (2000) 173-179

Pukelsheim, F. (1976): Estimating variance components in linear models, J. Multivariate Anal. 6 (1976), pp. 626-629

Pukelsheim, F. (1977): On Hsu's model in regression analysis, Math. Operations forsch. Statist. Ser. Statistics, 8 (1977), pp. 323-331

Pukelsheim, F. (1979): Classes of linear models, in: van Vleck, L.D. and S.R. Searle (eds.), pp. 69-83, cornell University/Ithaca 1979

Pukelsheim, F. (1980): Multilinear estimation of skewness and curtosis in linear models, Metrika, 27 (1980), pp. 103-113

Pukelsheim, F. (1981a): Linear models and convex geometry: Aspects of non-negative variance estimation, Math. Operationsforsch. u. Stat. 12 (1981), 271-286

Pukelsheim, F. (1981b): On the existence of unbiased nonnegative estimates of variance covariance components, Ann. Statist. 9 (1981), 293-299

Pukelsheim, F. (1990) : Optimal design of experiments , J. Wiley, New York 1990

Pukelsheim, F. (1993): Optimal design of experiments, J. Wiley, New York 1993

Pukelsheim, F. (1994): The three sigma rule, American Statistician 48 (1994), 88-91

Pukelsheim, F. and G.P. Styan (1979): Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model. Bull. Acad. Polonaise des Sci., Ser. Sci. Math., 27 (1979), pp. 327-330

Pukelsheim, F. and B. Torsney (1991): Optimal weights for experimental designs on linearly independent support points, The Annals of Statistics 19 (1991), 1614-1625

Pukelsheim, F. and W.J. Studden (1993): E-optimal designs for polynomial regression, Ann. Stat. 21 (1993), 402-415

Pynn, R. and Skjeltorp, A. (eds) Scaling Phenomena in Disordered Systems, Plenum Press New york (1985).

Qingming, G. and L. Jinshan (2000): Biased estimation in Gauss-Markov model, Allgemeine Vermessungsnachrichten 107 (2000), 104-108

Qingming, G., Yuanxi, Y. and G. Jianfeng (2001): Biased estimation in the Gauss-Markov model with constraints, Allgemeine Vermessungsnachrichten 108 (2001), 28-30

Qingming, G., Lifen, S., Yuanxi, Y. and G. Jianfeng (2001): Biased estimation in the Gauss-Markov model not full of rank, Allgemeine Vermessungsnachrichten 108 (2001), 390-393
Quintana-Orti, G., Sun, X. and C.H. Bischof (1998): A BLAS-3 version of the QR factorization with column pivoting, SIAM J. Sci. Comput. 19 (1998), 1486-1494
Raj, D. (1968): Sampling theory, Mc Graw-Hill Book Comp., Bombay 1968
Ralston, A. and Wilf, H. W. (1979): Mathematische Methoden fiir Digitalrechner 2. R. Oldenbourg Verlag, Miinchen.
Ramsey, J.O. and B.W. Silverman (1997): Functional data analysis, Springer-Verlag, Heidelberg Berlin New York 1997
Rao, B.L.S.P. (1997a): Weighted least squares and nonlinear regression, J. Ind. Soc. Ag. Statistics 50 (1997), 182-191
Rao, B.L.S.P. (1997b): Variance components, Chapman and Hall, Boca Raton 1997
Rao, B.L.S.P. and B.R. Bhat (1996): Stochastic processes and statistical inference, New Age International, New Delhi 1996
Rao, C.R. (1945): Generalisation of Markoff's Theorem and tests of linear hypotheses, Sankya 7 (1945), 9-16
Rao, C.R. (1948a): Test of significance in multivariate analysis, Biometrika 35 (1948a) 58-79
Rao, C.R. (1949): On the transformation useful in multivariate computations, Sankhya 9(1949) 251-253
Rao, C.R. of (1951): An asymptotic expansion of the distribution of Wilks Λ criterion Bull.Inst. Internat. Stat., 33, Part II, 177-180.
Rao, C.R. (1952a): Some theorems on Minimum Variance Estimation, Sankhya 12, 27-42
Rao, C.R. (1952b): Advanced statistical methods in biometric research, J. Wiley, New York 1952
Rao. C.R. (1965a): Linear Statistical Interference and ist Applications, J. Wiley, New York 1965
Rao, C.R. (1965b): The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, Biometrika 52 (1965), 447-458
Rao, C.R. (1965c): Linear Statistical Inference and Its Applications. John Wiley & Sons, New York.
Rao, C.R. (1967): Least squares theory using an estimated dispersion matrix and its application to measurement of signals, Procedure of the Fifth Barkeley Symposium, Barkeley 1967.
Rao, C.R. (1970): Estimation of heteroscedastic variances in linear models, J. Am. Stat. Ass. 65 (1970), 161-172
Rao, C.R. (1971a): Estimation of variance and covariance components - MINQUE theory, J. Multivar. Anal. 1 (1971), 257-275
Rao, C.R., (1971b): Estimation of variance components and application, North-Holland Series in statistics and probability, vol 3.
Rao, C.R. (1971c): Unified theory of linear estimation, Sankhya A33 (1971), 371-394
Rao, C.R. (1971d): Minimum variance quadratic unbiased estimation of variance components, J. Multivar. Anal. 1 (1971), 445-456
Rao, C.R. (1972): Estimation of variance and covariance components in linear models. J. Am. Stat. Ass. 67 (1972), 112-115
Rao, C.R. (1973a): Linear statistical inference and its applications, 2nd ed., J. Wiley, New York 1973
Rao, C.R. (1973b): Representation of best linear unbiased estimators in the Gauss-Markoff model with a singular dispersion matrix, J. Multivar. Anal. 3 (1973), 276-292
Rao, C.R. (1973c): Unified theory of least squares, Comm. Statist., 1 (1977), pp. 1-8
Rao, C.R. (1976): Estimation of parameters in a linear model, Ann. Statist. 4 (1976), 1023-1037
Rao, C.R. (1977): Prediction of future observations with special reference to linear models. In. Multivariate Analysis, Vol 4 (1977), pp. 193-208
Rao, C.R. (1978): Choice of the best linear estimators in the Gauss-Markov model with singular dispersion matrix, *Comm. Stat. Theory Meth.* A7 (13) (1978) 1199-1208.
Rao, C.R. (1979): Separation theorems for singular values of matrices and their applications in multivariate analysis, Journal of Multivariate Analysis 9, pp. 362-377.

Rao, C.R. (1984): Prediction of future observations in polynomial growth curve models, In Proceedings of the Indian Statistical Institute Golden Jubilee Intl. Conf. on Stat.: application and future directions, Indian Stat. Ins., pp. 512-520, Calcutta 1984

Rao, C.R. (1985): The inefficiency of least squares: extensions of the Kantorovich inequality, Linear algebra and its applications 70 (1985), 249-255

Rao, C.R. (1988a): Methodology based on the L1-norm in statistical inference, Sankhya 50, 289-313

Rao, C.R. (1988b): Prediction of future observations in growth curve models, J. of Statistical Science, 2, 434-471

Rao, C.R. (1989): A lemma on optimization of a matrix function and a review of the unified theory of linear estimation, ln: Y. Dodge (Ed.), Statistical Data Analysis and interference, pp. 337-418, Elsevier, Amsterdam 1989

Rao, C.R. and S.K. Mitra (1971): Generalized inverse of matrices and itsapplications, J. Wiley, New York 1971

Rao, C.R. and S.K. Mitra (1972): Generalized inverse of a matrix and its applications, in: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, Vol. 1, 601-620, University of California Press, Berkeley 1972

Rao, C.R. and J. Kleffe (1979): Variance and covariance components estimation and applications, Technical Report No. 181, Ohio State University, Dept. of Statistics, Columbus, Ohio.

Rao, C.R. and J. Kleffe (1988): Estimation of variance components and applications, North Holland, Amsterdam 1988

Rao, C.R. and L.C. Zhao (1993): Asymptotic normality of LAD estimator in censored regression models, Math. methods in Statistics, 2 (1993), pp. 228-239

Rao, C.R. and H. Toutenburg (1995): Linear models, least-squares and alternatives, Springer-Verlag, Heidelberg Berlin New York 1995

Rao, C.R. and R. Mukerjee (1997): Comparison of LR, score and Wald tests in a non-iiD setting, J. Multivar. Anal. 60 (1997), 99-110

Rao, C.R. and M.B. Rao (1998): Matrix algebra and its application to statistics and econometrics, World Scientific, Singapore 1998

Rao, C.R. and H. Toutenburg (1999): Linear models, Least squares and alternatives, 2nd ed., Springer-Verlag, Heidelberg Berlin New York 1999

Rao, C.R. and G.J. Szekely (2000): Statistics for the 21st century - Methodologies for applications of the future, Marcel Dekker, Basel 2000

Rao, P.S.R.S. and Y.P. Chaubey (1978): Three modifications of the principle of the MINQUE, Commun. Statist. Theor. Methods A7 (1978), 767-778

Ratkowsky, D.A. (1983): Nonlinear regression modelling, Marcel Dekker, New York

Ravi, V. and H.-J. Zimmermann (2000): Fuzzy rule based classification with Feature Selector and modified threshold accepting, European Journal of Operational Research 123 (2000), 16-28

Ravi, V., Reddy, P.J. and H.-J. Zimmermann (2000): Pattern classification with principal component analysis and fuzzy rule bases, European Journal of Operational Research 126 (2000), 526-533

Ravishanker, N. and D.K. Dey (2002): A first course in linear model theory, CRC Press, Boca Raton 2002

Rayleigh, L. (1880): On the resultant of a large number of vibrations of the same pitch and of arbitrary phase, Phil. Mag. 5 (1880), 73-78

Rayleigh, L. (1905): The problem of random walk, Nature 72 (1905), 318

Rayleigh, L. (1919): On the problem of random vibrations, and of random flights in one, two or three dimensions, Phil Mag. 37 (1919), 321-347

Rebai, S. Philip, H. and Taboada, A. (1992): Modern tectonic stress field in the Mediterranean region: evidence for variation instress directions at different scales, Geophys. J. Int., 110, 106-140

Reeves, J. (1998): A bivariate regression model with serial correlation, The Statistician 47 (1998), 607-615

Reich, K. (2000): Gauss' Schüler. Studierten bei Gauss und machten Karriere. Gauss' Erfolg als Hochschullehrer (Gauss's students: studied with him and were successful. Gauss's success as a university professor), Gauss Gesellschaft E.V.Göttingen, Mitteilungen Nr. 37, 33-62, Göttingen 2000

Reilly, W., Fredrich, G., Hein, G., Landau, H., Almazan, J. and Caturla, J. (1992): Geodetic determination of crustal defoemation across the Strait of Gibraltar, Geophys. J. Int., 111, 391-398

Relles, D.A. (1968): Robust regression by modified least squares, Ph.D. Thesis, Yale University, Yale 1968

Remmer, O. (1973): A Stability Investigation of Least Squares Adjustment by elements, Geodaetisk Institute Copenhagen, 1973

Remondi, B.W. (1984): Using the Global Positioning System (GPS) phase observable for relative geodesy: modelling, processing and results. Ph.D.Thesis, Center for Space Research, The University of Texas, Austin 1984

Ren, H. (1996): On the eroor analysis and implementation of some eigenvalue decomposition and singular value decomposition algorithms, UT-CS-96-336, LAPACK working note 115 (1996)

Rencher A.C. (2000): Linear models in statistics, J. Wiley, New York 2000

Renfer, J.D. (1997): Contour lines of L1-norm regression, Student 2 (1997), 27-36

Resnikoff, G.J. and G.J. Lieberman (1957): Tables of the noncentral t-distribution, Stanford University Press, Stanford 1957

Reubelt, T., Austen, G. and Grafarend, E.W. (2003): Harmonic analysis of the Earth's gravitational field by means of semi-continuous ephemeris of a low Earth orbiting (LEO) GPS-tracked satellitecase study CHAMP, J. Geodesy, 77, 257-278

Rham De, G. (1984): Differentiable Manifolds. Springer-Verlag, Berlin Heidelberg

Riccomagno, E., Schwabe, R. and H.P. Wynn (1997): Lattice-based optimum design for Fourier regression, Ann. Statist. 25 (1997), 2313-2327

Rice, J.R. (1969): The approximation of functions, Vol. II - Nonlinear and multivariate theory, Addison-Wesley, Reading 1969

Richards, F.S.G. (1961): A method of maximum likelihood estimation, J. Roy. Stat. Soc. B23 (1961), 469-475

Richardson, R. (1992): Ridge forces, absolute plate motions, and the intraplate stress field, J. geophys. Res., B97, 11739-11748

Richter, H. and V. Mammitzsch (1973): Methode der kleinsten Quadrate, Stuttgart 1973

Richter, B. (1986): Entwurf eines nichtrelativistischen geodätisch-astronomischen Bezugssystems, DGK, Reihe C, Heft Nr. 322.

Riedel, K.S. (1992): A Sherman-Morrison-Woodbury identity for rank augmenting matrices with application to centering, SIAM J. Matrix Anal. Appl. 13 (1992), 659-662

Riedwyl, H. (1997): Lineare Regression, Birkhäuser-Verlag, Basel Boston Berlin 1997

Richter, W.D. (1985): Laplace-Gauß integrals, Gaussian measure asymptotic behaviour and probabilities of moderate deviations, Z.f. Analysis und ihre Anwendungen 4 (1985), 257-267

Rinner, K. (1962): Über die Genauigkeit des räumlichen Bogenschnittes, *Zeitschrift für Vermessungswesen* 87 (1962) 361-374.

Rivest, L.P. (1982): Some statistical methods for bivariate circular data, J. Roy. Statist. Soc. B44 (1982), 81-90

Ritt, J.F. (1950): Differential algebra, AMS colloquium publications.

Rivest, L.P. (1988): A distribution for dependent unit vectors, Comm. Statistics A: Theory Methods 17 (1988), 461-483

Rivest, L. (1989): Spherical regression for concentrated Fisher-von Mises distributions, Annals of Statistics 17 (1989), 307-317

Roach, G.F. (1982): Green's Functions. Cambridge University Press, Cambridge.

Roberts, P.H. and H.D. Ursell (1960): Random walk on the sphere and on a Riemannan manifold, Phil. Trans. Roy. Soc. A252 (1960), 317-356

Robertson, H.P. (1940): The invariant theory of isotropic turbulence Proc. Cambridge Phil. Soc. 36 (1940) 209-223

Robinson, E.A. (1967): Multichannel time series analysis with digital programs, 298 pages, Holden Day, San Francisco 1967
Robinson, E.A. and Treitel, S. (1967): Principles of digital filtering, Geophysics 29 (1964), 395-404
Robinson, G.K. (1982): Behrens-Fisher problem, Encyclopedia of the Statistical Sciences, Vol. 1, J. Wiley, New York 1982
Rodgers, J.L. and W.A. Nicewander (1988): Thirteen ways to look at the correlation coefficient, the Maerican Statistician 42 (1988), 59-66
Rohde C.A. (1965): Generalized inverses of partitioned matrices. J. Soc. Ind. Appl. Math. 13 (1965), 1033-1035.
Rohde, C.A. (1966): Some results on generalized inverses, SIAM Rev. 8 (1966), 201-205
Romano, J.P. and A.F. Siegel (1986): Counterexamples in probability and statistics, Chapman and Hall, Boca Raton 1986
Romanowski, M. (1979): Random errors in observations and the influence of modulation on their distribution, K. Wittwer Verlag, Stuttgart 1979
Rosen, F.: The algebra of Mohammed Ben Musa, London: Oriental Translation Fund 1831
Rosen van, D. (1988): Moments for matrix normal variables, Statistics 19 (1988), 575-583
Rosen, J.B., Park, H. and J. Glick (1996): Total least norm formulation and solution for structured problems, SIAM J. Matrix Anal. Appl. 17 (1996), 110-126
Rosèen, K.D.P. (1948): Gauss's mean error and other formulae for the precision of direct observations of equal weight, Tätigkeitsbereiche Balt. Geod. Komm. 1944-1947, 38-62, Helsinki 1948
Rosenblatt, M. and Van Ness, J.W. (1965): Estimation of bispectrum, Ann. Math. Statist. 36 (1965), 1120-1136
Rosenblatt, M. (1966): Remarks on higher order spectra, in : Multivariate Analysis, P.R. Krishnaiah (ed.) Academic Press, New York 1966
Rosenblatt, M. (1971): Curve estimates, Ann. Math. Statistics 42 (1971), 1815-1842
Rosenblatt, M. (1997): Some simple remarks on an autoregressive scheme and an implied problem, J. Theor. Prob. 10 (1997), 295-305
Ross, G.J.S. (1982): Non-linear models, Math. Operationsforschung Statistik 13 (1982), 445-453
Rosenborck, H.H. (1960): An Automatic Method for finding the Greatest or Least Value of a Function. Computer Journal, 3, 175-184.
Ross, S.M. (1983): Stochastic processes, J. Wiley, New York 1983
Rotta, J.C. (1972): Turbulente Strömuugen, Teubner Verlag, 267 pages, Stuttgart 1972
Rousseeuw, P.J. (1984): Least Median of Squares Regression, Journal of the American Statistical Association, 79, pp. 871-880
Rousseeuw, P.J. and A.M. Leroy (1987): Robust regression and outlier detection, J. Wiley, New York 1987
Roy, T. (1995): Robust non-linear regression analysis, J. Chemometrics 9 (1995), 451-457
Rozanski, I.P. and R. Velez (1998): On the estimation of the mean and covariance parameter for Gaussian random fields, Statistics 31 (1998), 1-20
Rubin, D.B. (1976): Inference and missing data, Biometrica 63 (1976), pp. 581-590
Rueda, C., Salvador, B. and M.A. Fernèandez (1997): Simultaneous estimation in a restricted linear model, J. Multivar. Anal. 61 (1997), 61-66
Rueschendorf, L. (1988): Asymptotische Statistik, Teubner, Stuttgart 1988
Rummel, R. (1975): Zur Behandlung von Zufallsfunktionen und -folgen in der physikalischen Geodäsie, Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Report No. C 208, München 1975
Rummel, R. (1976): A model comparison in least-squares collocation. Bull Geod 50:181-192
Rummel, R. and K. P.Schwarz (1977): On the nonhomogenity of the global covariance function, Bull. Gèeodèesique 51 (1977), 93-103
Rummel, R. and Teunissen, P. (1982): A connection between geometric and gravimetric geodesy - some remarks on the role of the gravity field. Feestbundel ter gelegenheid van de 65ste verjaardag van Professor Baarda, Deel II, pp. 603-623, Department of GeodeticScience, Delft University

Runge, C. (1900): Graphische Ausgleichung beim Rückwätseinchneiden, *Zeitschrift für Vermessungswesen* 29 (1900) 581-588.
Ruppert, D. and R.J. Carroll (1980): Trimmed least squares estimation in the linear model, J. Am. Statist. Ass. 75 (1980), 828-838
Rutherford, D.E. (1933): On the condition that two Zehfuss matrices be equal, Bull. Amer. Math. Soc. 39 (1933), 801-808
Rutherford, A. (2001): Introducing Anova and Ancova - a GLM approach, Sage, London 2001
Rysavy, J. (1947): Higher geodesy. Jeska matice technicka, Praha, 1947 (in Czech).
Saalfeld, A. (1999): Generating basis sets of double differences, Journal of Geodesy 73 (1999), 291-297
Saastamoinen, J. (1973a): Contributions to the theory of atmospheric refraction. J. of Geodesy 46: 279-298
Saastamoinen, J. (1973b):Contribution to the theory of atmospheric refraction. Part II refraction corrections in satellite geodesy. Bulletin Godsique 107:13-34, 1973.
Sacks, J. and D. Ylvisaker (1966): Design for regression problems with correlated errors, Annals of Mathematical Statistics 37 (1966), 66-89
Sahai, H. (2000): The analysis of variance: fixed, random and mixed models, 778 pages, Birkhäuser-Verlag, Basel Boston Berlin 2000
Sahin, M., Cross, P.A. and P.C. Sellers. (1992): Variance components estimation applied to satellite laser ranging, Bulletin Geodesique, vol. 66, no. 3, p. 284-295.
Saichev, A.I. and W.A. Woyczynski (1996): Distributions in the physical and engineering sciences, Vol. 1, Birkäuser Verlag, Basel 1996
Saito, T. (1973): The non-linear least squares of condition equations, *Bull. Geod.* 110 (1973) 367-395.
Salmon, G. (1876): Lessons Introductory to modern higher algebra, Hodges, Foster and Co., Dublin 1876.
Samorodnitsky, G. and M.S. Taqqu (1994): Stable non-Gaussian random processes, Chapman and Hall, Boca Raton 1994
Sampson, P.D. and P. Guttorp (1992): Nonparametric estimation of nonstationary spatial covariance structure, J. Am. Statist. Ass. 87 (1992), 108-119
Sander, B. (1930): Gefugekunde und Gesteine, J. Springer, Vienna 1930
Sanséo, F. (1990): On the aliasing problem in the spherical harmonic analysis, Bull. Gèeodesique 64 (1990), 313-330
Sanséo, F. and G. Sona (1995): The theory of optimal linear estimation for continuous fields of measurements, Manuscripta Geodetica 20 (1995), 204-230
Sanso, F. (1973): An Exact Solution of the Roto-Translation Problem. Photogrammetria, Vol. 29, 203-216.
Sanso, F. (1986): Statistical methods in physical geodesy. In: Suenkel H (ed), Mathematical and numerical techniques in physical geodesy, Lecture Notes in Earth Sciences, vol. 7 Springer, pp 49-156
Sanso, F. and Tscherning, C.C. (2003): Fast spherical collocation: theory and examples. J Geod 77(1-2):101-112
Sastry, K. and Krishna, V. (1948) On a Bessel function of the second kind and Wilks Z-distribution. Proc Indian Acad Sci Series A 28:532-536
Sastry, S. (1999): Nonlinear systems: Analysis, stability and control, Springer-Verlag, Heidelberg Berlin New York 1999
Sathe, S.T. and H.D. Vinod (1974): Bound on the variance of regression coefficients due to heteroscedastic or autoregressive errors, Econometrica 42 (1974), 333-340
Saw, J.G. (1978): A family of distributions on the m-sphere and some hypothesis tests, Biometrika 65 (1978), 69-73
Saw, J.G. (1981): On solving the likelihood equations which derive from the Von Mises distribution, Technical Report, University of Florida, 1981

Sayed, A.H., Hassibi, B. and T. Kailath (1996), Fundamental inertia conditions for the minimization of quadratic forms in indefinite metric spaces, Oper. Theory: Adv. Appl., Birkhäuser-Verlag, Basel Boston Berlin 1996

Schach, S. and T. Schäfer (1978): Regressions- und Varianzanalyse, Springer-Verlag, Heidelberg Berlin New York 1978

Schafer, J.L. (1997): Analysis of incomplete multivariate data, Chapman and Hall, London 1997

Schaffrin, B. (1979): Einige ergänzende Bemerkungen zum empirischen mittleren Fehler bei kleinen Freiheitsgraden, Z. Vermessungsesen 104 (1979), 236-247

Schaffrin, B. (1981a): Some proposals concerning the diagonal second order design of geodetic networks, manuscripta geodetica 6 (1981) 303-326

Schaffrin, B. (1981b): Varianz-Kovarianz-Komponenten-Schatzung bei def Ausgleichung heterogener, Wiederhungsmessungen. Ph.D. Thesis, University of Bonn, Bonn, Germany.

Schaffrin, B. (1983a): Varianz-Kovarianz Komponentenschätzung bei der Ausgleichung heterogener Wiederholungsmessungen, Deutsche Geodätische Kommission, Report C 282, München, 1983

Schaffrin, B. (1983b): Model choice and adjustment techniques in the presence of prior information, Ohio State University Department of Geodetic Science and Surveying, Report 351, Columbus 1983

Schaffrin, B. (1983c): A note on linear prediction within a Gauβ - Markov model linearized with respect to a random approximation. 1st Int. Tampere Seminar on Linear Models and their Applications. Tampere/ Finland.

Schaffrin, B. (1984): Das geodätische Datum mit stochastischer Vorinformation. Habilitationsschrift Stuttgart.

Schaffrin, B. (1985a): The geodetic datum with stochastic prior information, Publ. C313, German Geodetic Commission, München 1985

Schaffrin, B. (1985b): On Design Problems in Geodesy using Models with Prior Information. Statistics & Decisions, Supplement Issue 2 443-453.

Schaffrin, B. (1985c): A note on linear prediction within a Gauss-Markov model linearized with respect to a random approximation. In: Proc. First Tampere Sem. Linear Models (1983), Eds.: T. Pukkila, S. Puntanen, Dept. of Math. Science/Statistics, Univ. of Tamp ere/Finland, Report No. A-138, 285-300.

Schaffrin, B. (1986): New estimation/prediction techniques for the determination of crustal deformations in the presence of geophysical prior information, Technometrics 130 (1986), pp. 361-367

Schaffrin, B. (1991): Generalized robustified Kalman filters for the integration of GPS and INS, Tech. Rep. 15, Geodetic Institute, Stuttgart Unversity 1991

Schaffrin, B. (1997): Reliability measures for correlated observations, Journal of Surveying Engineering 123 (1997), 126-137

Schaffrin, B. (1999): Softly unbiased estimation part1: The Gauss-Markov model, Linear Algebra and its Applications 289 (1999), 285-296

Schaffrin, B. (2001): Equivalent systems for various forms of kriging, including least-squares collocation, Zeitschrift für Vermessungswesen 126 (2001), 87-94

Schaffrin, B. (2005): On total least squares adjustment with constraints, in: A window on the future of Geodesy, F. Sanso (ed.), pp. 417-421., Springer Verlag, Berlin-Heidelberg-New York 2005

Schaffrin, B. (2007): Connecting the dots: The straight line case revisited, Deutscher Verein für Vermessungswesen (zfv2007), 132.Jg., 2007, pp. 385-394

Schaffrin, B. (2008): Minimum mean squared error (MSE) adjustment and the optimal Tykhonov-Phillips regularization parameter via reproducing best invariant quadratic uniformly unbiased estimates (repro-BIQUUE), J. Geodesy, 92 (2008) 113-121

Schaffrin, B., Grafarend, E. and Schmitt G. (1977): Kanonisches Design geodätischer Netze I. Manuscripta geodaetica 2, pp. 263-306.

Schaffrin, B., Grafarend, E. and G. Schmitt (1977): Kanonisches Design Geodätischer Netze I, Manuscripta Geodaetica 2 (1977), 263-306

Schaffrin, B., Grafarend, E. and G. Schmitt (1978): Kanonisches Design Geodätischer Netze II, Manuscripta Geodaetica 2 (1978), 1-22
Schaffrin, B., Krumm, F. and Fritsch, D. (1980): Positiv diagonale Genauigkeit 0p.timierun von Realnetzen Über den Komplementarltats-Algorlthmus In: Ingenieurvermessung 80, ed. Conzett R, Schmitt M,Schmitt H. Beitrage zum VIII. Internat. Kurs für Ingenieurvermessung, Zurich.
Schaffrin, B. and E.W. Grafarend (1982a): Kriterion-Matrizen II: Zweidimensionale homogene und isotrope geodätische Netze, Teil II a: Relative cartesische Koordinaten, Zeitschrift für Vermessungswesen 107 (1982), 183-194
Schaffrin, B. and E.W. Grafarend (1982b): Kriterion-Matrizen II: Zweidimensionale homogene und isotrope geodätische Netze. Teil II b: Absolute cartesische Koordinaten, Zeitschrift für Vermessungswesen 107 (1982), 485-493
Schaffrin, B. and E. Grafarend (1986): Generating classes of equivalent linear models by nuissance parameter elimination, manuscripta geodaetica 11 (1986), 262-271
Schaffrin, B. and E.W. Grafarend (1991): A unified computational scheme for traditional and robust prediction of random effects with some applications in geodesy, The Frontiers of Statistical Scientific Theory & Industrial Applications 2 (1991), 405-427
Schaffrin, B. and J.H. Kwon (2002): A Bayes filter in Friendland form for INS/GPS vector gravimetry, Geophys. J. Int. 149 (2002), 64-75
Schaffrin, B. and A. Wieser (2008): On weighted total least squares adjustment for linear regression, J. Geodesy, 82, (2008) 415-421
Schall, R., Dunne, T.T. (1988): A unified approach to outliers in the general linear model. Sankhya, Ser. B 50, 2 (1988), 157-167.
Scheffé, H. (1959): The analysis of variance, Wiley, New York 1959
Scheidegger, A.E. (1965): On the statistics of the orientation of bedding planes, grain axes and similar sedimentological data, U.S. Geol. Survey Prof. Paper 525-C (1965), 164-167
Schleider, D. (1982): Complex crustal strain approximation, Ph.D. dissertation, Reports of Department of Surveying Engineering of UNB No.91, Canada
Scheffe, H. (1943): On solutions of the Behrens-Fisher problem based on the t-distribution, Ann. Math. Stat., 14, 35-44
Scheffe, H. (1999): The analysis of variance. John Wiley & Sons, New York, 1999 (repr. of the 1959 orig.).
Schek, H.J. (1974): The Force Densities Method for Form Finding and Computation of General Networks. Computer Methods in Applied Mechanics and Engineering, 3, 115-134.
Schek, H.J. (1975): Least-Squares-Losungen und optimale Dampfung bei nichtlinearen Gleichungssystemen im Zusammenhang mit der bedingten Ausgleichung. Zeitschrift fiir Vermessungswesen, 2, 67-77.
Schek, H.J. and Maier, P. (1976): Nichtlineare Normalgleichungen zur Bestimmung der Unbekannten und deren Kovarianzmatrix, *Zeitschrift für Vermessungswesen* 101 (1976) 140-159.
Schetzen, M. (1980): The Volterra and Wiener theories of nonlinear systems, J. Wiley, New York 1980
Schick, A. (1999): Improving weighted least-squares estimates in heteroscedastic linear regression when the variance is a function of the mean response, J. Statist. Planning and Inference 76 (1999), 127-144
Schiebler, R. (1988): Giorgio de Chirico and the theory of relativity, Lecture given at Stanford University, Wuppertal 1988
Schmetterer, L. (1956): Einführung in die mathematische Statistik, Wien 1956
Schmidt, E. (1907): Entwicklung willkürlicher Funktionen, Math. Annalen 63 (1907), 433-476
Schmidt, K. (1996): A comparison of minimax and least squares estimators in linear regression with polyhedral prior information, Acta Applicandae Mathematicae 43 (1996), 127-138
Schmidt, K.D. (1996): Lectures on risk theory, Teubner Skripten zur Mathematischen Stochastik, Stuttgart 1996
Schmidt, P. (1976): Econometrics, Marcel Dekker, New York 1976

Schmidt-Koenig, K. (1972): New experiments on the effect of clock shifts on homing pigeons in animal orientation and navigation, Eds.: S.R. Galler, K. Schmidt-Koenig, G.J. Jacobs and R.E. Belleville, NASA SP-262, Washington D:C. 1972

Schmidt, V. (1999): Begrundung einer Meβanweisung für die Überwachungsmessungen am Maschinenhaus des Pumpspeicherwerkes Hohenwarthe, Diplomarbeit, TU Bergakademie Frieberg Inst. für Markscheidewesen und Geodasie, 1999

Schmidt, W.H. and S. Zwanzig (1984): Testing hypothesis in nonlinear regression for nonnormal distributions, Proceedings of the Conference on Robustness of Statistical Methods and Nonparametric Statistics, VEBDeutscher Verlag der Wissenschaften Berlin, 1984, 134 138.

Schmidt, W.H. and S. Zwanzig (1986):, Second order asymptotics in nonlinear regression. J. Multiv. Anal., VoL 18, No. 2, 187-215.

Schmidt, W.H. and S. Zwanzig (1986): Testing hypothesis in nonlinear regression for nonnormal distributions. Statistics, VoL 17, No.4, 483-503.

Schmitt, G. (1975): Optimaler Schnittwinkel der Bestimmungsstrecken beim einfachen Bogenschnitt, Allg. Vermessungsnachrichten 6 (1975), 226-230

Schmitt, G. (1977): Experiences with the second-order design problem in theoretical and practical geodetic networks, Proceedings International Symposium on Optimization of Design and Computation of Control Networks, Sporon 1977

Schmitt, G. (1977): Experiences with the second-order design problem in theoretical and practical geodetic networks, Optimization of design and computation of control networks. F. Halmos and J. Somogyi eds, Akadèemiai Kiadèo, Budapest (1979), 179-206

Schmitt, G. (1978): Gewichtsoptimierung bei Mehrpunkteinschaltung mit Streckenmessung, Allg. Vermessungsnachrichten 85 (1978), 1-15

Schmitt, G., Grafarend, E. and B. Schaffrin, (1978): Kanonisches Design Geodätischer Netze II. Manuscr. Geod. 1, 1-22.

Schmitt, G. (1979): Zur Numerik der Gewichtsoptimierung in geodätischen Netzen, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Report 256, München 1979

Schmitt, G. (1980): Second order design of a free distance network considering different types of criterion matrices, Bull. Geodetique 54 (1980), 531-543

Schmitt, G. (1985): Second Order Design, Third Order Design, Optimization and design of geodetic networks, Springer-Verlag, Heidelberg Berlin New York 1985, 74-121

Schmitt, G., Grafarend, E.W. and B. Schaffrin.(1977): Kanonisches Design geodätischer Netze I, manus-cripta geodaetica 2 (1977), 263-306

Schmitt, G., Grafarend, E.W. and B. Schaffrin (1978): Kanonisches Design geodätischer Netze II, manuscripta geodaetica 3 (1978), 1-22

Schmutzer, E. (1989): Grundlagen der Theoretischen Physik, Teil I, BI Wissenschaftsverlag, Mannheim

Schneeweiß, H. and H.J. Mittag (1986): Lineare Modelle mit fehlerbehafteten Daten, Physica-Verlag, Heidelberg 1986

Schock, E. (1987): Implicite iterative methods for the approximate solution of ill-posed problems, Bolletino U.M.I., Series 1-B, 7 (1987), 1171-1184

Schoenberg, I.J. (1938): Metric spaces and completely monotone functions, Ann. Math. 39 (1938), 811-841

Schön, S. (2003): Analyse und Optimierung geodätischer Messanordnungen unter besonderer Berücksichtigung des Intervallansatzes, DGK, Reihe C, Nr. 567, München, 2003

Schön, S. and Kutterer, H. (2006): Uncertainty in GPS networks due to remaining systematic errors: the interval approach, in: Journal of Geodesy, Vol. 80 no. 3, pp. 150-162, 2006.

Schönemann, P.H. (1996): Generalized solution of the orthogonal Procrustes problem, *Psychometrika* 31 (1996) 1-10.

Schönemann, S.D. (1969): Methoden der Öknometrie, Bd. 1, Vuhlen, Berlin 1969

Schönfeld P., Werner H.J. (1994): A note on C.R. Rao's wider definition BLUE in the general Gauss-Markov model. Sankhya, Ser. B 49, 1 (1987), 1-8.

Schott, J.R. (1997): Matrix analysis for statistics, J. Wiley, New York 1997

Schott, J.R. (1998): Estimating correlation matrices that have common eigenvectors, Comput. Stat. & Data Anal. 27 (1998), 445-459

Schouten, J.A. and J. Haantjes (1936): über die konforminvariante Gestalt der relativistischen Bewegungsgleichungen, in: Koningl. Ned. Akademie van Wetenschappen, Proc. Section of Sciences, Vol. 39, Noord-Hollandsche Uitgeversmaatschappij, Amsterdam 1936

Schouten, J.A. and Struik, D.J. (1938): Einfiihrung in die neueren Methoden der Differentialgeometrie. I/Il, Noordhoff, Groningen-Batavia.

Schreiber, O. (1882): Anordnung der Winkelbeobachtungen im GHttinger Basisnetz. lfV I, 129-161

Schroeder, M. (1991): Fractals, chaos, power laws, Freeman, New York 1991

Schultz, C. and G. Malay (1998): Orthogonal projections and the geometry of estimating functions, J. Statist. Planning and Inference 67 (1998), 227-245

Schultze, J. and J. Steinebach (1996): On least squares estimates of an exponential tail coefficient, Statistics & Decisions 14 (1996), 353-372

Schupler, B.R., Allshouse, R.L. and Clark, T.A. (1994): Signal Characteristics of GPS User Antennas. Navigation 4113: 177-295

Schuster, H.-F. and Forstner, W. (2003): Segmentierung, Rekonstruktion undDatenfusion bei der Objekterfassung mit Entfernungsdaten - ein Oberb/ick. Proceedings 2. Oldenburger 3DTage. Oldenburg

Schur, J. (1911): Bemerkungen zur Theorie der verschränkten Bilinearformen mit unendlich vielen Veränderlichen, J. Reine und Angew. Math. 140 (1911), 1-28

Schur, J. (1917): über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, J. Reine Angew. Math. 147 (1917), 205-232

Schwarz, G. (1978): Estimating the dimension of a model, The Annals Of Statistics 6 (1978), 461-464

Schwarz, H. (1960): Stichprobentheorie, Oldenbourg, München 1960

Schwarz, K. P. (1976): Least-squares collocation for large systems, Boll. Geodesia e Scienze Affini 35 (1976), 309-324

Schwarz, C.R. and Kok, J.J. (1993) Blunder detection and Data Snooping in LS and Robust Adjustments. J Surv Eng 119:127-136

Schwarze, V.S. (1995): Satellitengeodätische Positionierung in der relativistischen Raum-Zeit, DGK, Reihe C, Heft Nr.449.

Schwieger, W. (1996): An approach to determine correlations between GPS monitored deformation epochs, In: Proc. of the 8th International Symposium on Deformation Measurements, Hong Kong, 17-26, 1996.

Scitovski, R. and D. Jukièc (1996): Total least squares problem for exponential function, Inverse Problems 12 (1996), 341-349

Sckarosfsky, I.P. (1968): Generalized turbulence and space correlation and wave numberspectrum function pairs, Canadian J. of Physics 46 (1968) 2133-2153

Seal, H.L. (1967): The historical development of the Gauss linear model, Biometrika 54 (1967), 1-24

Searle, S.R. (1971a). Linear Models, J. Wiley, New York 1971

Searle, S.R. (1971b): Topics in variance components estimation, Biometrics 27 (1971), 1-76

Searle, S.R. (1974): Prediction, mixed models, and variance components, Reliability and Biometry (1974), 229-266

Searle, S.R. (1982): Matrix algebra useful for statistics, Wiley, New York 1982

Searle, S.R. (1994): Extending some results and proofs for the singular model. Linear Algebra Appl. 210 (1994), 139-15l.

Searle, S.R. and C.R. Henderson (1961): Computing procedures for estimating components of variance in the two-way classification, mixed model, Biometrics 17 (1961), 607-616

Searle, S.R., Casella, G. and C.E. McCulloch (1992): Variance components, J. Wiley, New York 1992

Seber, G.A.F. (1977): Linear regression analysis, J. Wiley, New York 1977

Seber, G.A.F. and Wild, C.J. (1989): Nonlinear Regression. John Wiley & Sons, New York.

Seeber, G. (1989): Satellitengeodiisie: - Grundlagen, Methoden und Anwendungen. Walter deGruyter, Berlin- New York, 489p

Seely, J. (1970): Linear spaces and unbiased estimation, Ann. Math. Statist. 41 (1970), 1725-1734

Seely, J. (1970): Linear spaces and unbiased estimation - Application to the mixed linear model, Ann. Math. Statist. 41 (1970), 1725-1734

Seely, J. (1971): Quadratic subspaces and completeness, Ann. Math. Statist. 42 (1971), 710-721

Seely, J. (1975): An example of an inquadmissible analysis of variance estimator for a variance component, Biometrika 62 (1975), 689

Seely, J. (1977): Minimal sufficient statistics and completeness, Sankhya, Series A, Part 2, 39 (1977), 170-185

Seely, J. (1980): Some remarks on exact confidence intervals for positive linear combinations of variance components, J. Am. Statist. Ass. 75 (1980), 372-374

Seely, J. and R.V. Hogg (1982): Unbiased estimation in linear models, Communication in Statistics 11 (1982), 721-729

Seely, J. and Y. El-Bassiouni (1983): Applying Wald's variance components test, Ann. Statist. 11 (1983), 197-201

Seely, J. and E.-H. Rady (1988): When can random effects be treated as fixed effects for computing a test statistics for a linear hypothesis?, Communications in Statistics 17 (1988), 1089-1109

Seely, J. and Y. Lee (1994): A note on the Satterthwaite confidence interval for a variance, Communications in Statistics 23 (1994), 859-869

Seely, J., Birkes, D. and Y. Lee (1997): Characterizing sums of squares by their distribution, American Statistician 51 (1997), 55-58

Seemkooei, A.A. (2001): Comparison of reliability and geometrical strength criteria in geodetic networks, Journal of Geodesy 75 (2001), 227-233

Segura, J. and A. Gil (1999): Evaluation of associated Legendre functions off the cut and parabolic cylinder functions, Electronic Transactions on Numerical Analysis 9 (1999), 137-146

Seidelmann, P.K. (ed) (1992): Explanatory Supplement to the Astronomical Almanac. University Science Books, Mill Valley, 752p

Selby, B. (1964): Girdle distributions on the sphere, Biometrika 51 (1964), 381-392

Sengupta D. (1995): Optimal choice of a new observation in a linear model. Sankhya, Ser. A 57, 1 (1995), 137-153.

Sengupta, A. and R. Maitra (1998): On best equivariance and admissibility of simultaneous MLE for mean direction vectors of several Langevin distributions, Ann. Inst. Statist. Math. 50 (1998), 715-727

Sengupta, D. and S.R. Jammalamadaka (2003): Linear models, an integrated approach, In series of Multivariate analysis 6 (2003)

Sering, R.J. (1980): Approximation theorems of mathematical statistics, J. Wiley, New York 1980

Shaban, A.M.M. (1994): ANOVA, MINQUE, PSD-MIQMBE, CANOVA and CMINQUE in estimating variance components, Statistica 54 (1994), 481-489

Shah, B.V. (1959): On a generalisation of the Kronecker product designs, Ann. Math. Statistics 30 (1959), 48-54

Shalabh (1998): Improved estimation in measurement error models through Stein rule procedure, J. Multivar. Anal. 67 (1998), 35-48

Shamir, G. and Zoback, M.D. (1992): Stress orientation profile to 3.5 km depth near the San Andreas fault at Cajon Pass, California, J. geophys. Res., B97, 5059-5080

Shao, Q.-M. (1996): p-variation of Gaussian processes with stationary increments, Studia Scientiarum Mathematicarum Hungarica 31 (1996), 237-247

Shapiro, S.S. and M.B. Wilk (1965): An analysis of variance for normality (complete samples), Biometrika 52 (1965), 591-611

Shapiro, S.S., Wilk, M.B. and M.J. Chen (1968): A comparative study of various tests for normality, J. Am. Statist. Ass. 63 (1968), 1343-1372

Shapiro, L.S. and Brady, M. (1995): Rejecting outliers and estimating errors in an orthogonal regression framework. Philos. Trans. R. Soc. Lond., Ser. A 350, 1694 (1995), 407-439.

Sheppard, W.F. (1912): Reduction of errors by means of negligible differences, Proc. 5th Int. Congress Mathematicians (Cambridge) 2 (1912), 348-384
Sheynin, O.B. (1966): Origin of the theory of errors, Nature 211 (1966), 1003-1004
Sheynin, O.B. (1979): Gauß and the theory of errors, Archive for History of Exact Sciences 20 (1979)
Sheynin, O. (1995): Helmert's work in the theory of errors, Arch. Hist. Exact Sci. 49 (1995), 73-104
Shiryayev, A.N. (1973): Statistical sequential analysis, Transl. Mathematical Monographs 8, American Mathematical Society, Providence/R.I. 1973
Shkarofsky, I.P. (1968): Generalized turbulence space-correlation and wave-number spectrum-function pairs, Canadian Journal of Physics 46 (1968), 2133-2153
Shorack, G.R. (1969): Testing and estimating ratios of scale parameters, J. Am. Statist. Ass. 64, 999-1013, 1969
Shrivastava, M.P. (1941): Bivariate correlation surfaces. Sci Cult 6:615-616
Shumway, R.H. and D.S. Stoffer (2000): Time series analysis and its applications, Springer-Verlag, Heidelberg Berlin New York 2000
Shut, G.H. (1958/59): Construction of orthogonal matrices and their application in analytical Photogrammetrie, *Photogrammetria* XV (1958/59) 149-162
Shwartz, A. and Weiss, A. (1995): Large deviations for performance analysis. Chapman and Hall, Boca Raton
Sibuya, M. (1960): Bivariate extreme statistics, Ann. Inst. Statist. Math. 11 (1960), 195-210
Sibuya, M. (1962): A method of generating uniformly distributed points on n-dimensional spheres, Ann. Inst. Statist. Math. 14 (1962), 81-85
Siegel, A.F. (1982): Robust Regression Using Repeated Medians Biometrika 69, pp. 242-244
Siegel, C.L. (1937): Über die analytische Theorie der quadratischen Formen, Ann. of Math. vol. 38, (1937), pp. 212-291
Sillard, P., Altamimi, Z. and C. Boucher (1998): The ITRF96 realization and its associated velocity field, Geophysical Research Letters 25 (1998), 3223-3226
Silvey, S.D. (1969): Multicollilinearlity and imprecise estimation, J. Roy. Stat. Soc., Series B 35 (1969), pp 67-75
Silvey, S.D. (1975): Statistical inference, Chapman and Hall, Boca Raton 1975
Silvey, S.D. (1980): Optimal design, Chapman and Hall, 1980
Sima, V. (1996): Algorithms for linear-quadratic optimization, Dekker, New York 1996
Simmonet, M. (1996): Measures and probabilities, Springer-Verlag, Heidelberg Berlin New York 1996
Simoncini, V. and F. Perotti (2002): On the numerical solution of and application to structural dynamics, SIAM J. Sci. Comput. 23 (2002), 1875-1897
Simonoff. J.S. (1996): Smoothing methods in statistics. Springer, Heidelberg Berlin New York
Sinai, Y.G. (1963): On properties of spectra of ergodic dynamical systems, Dokl. Akad. Nauk. SSSR 150 (1963)
Sinai, Y.G. (1963): On higher order spectral measures of ergodic stationary processes Theory Probability Appl. (USSR) (1963) 429-436
Singer, P., Ströbel, D,. Hördt, R., Bahndorf, J. and Linkwitz, K. (1993): Direkte Lösung des räumlichen Bogenschnitts, *Zeitschrift für Vermessungswesen* 118 (1993) 20-24
Singh, R. (1963): Existence of bounded length confidence intervals, Ann. Math. Statist. 34 (1963), 1474-1485
Sion, M. (1958): On general minimax theorems. Pac. J. Math. 8:171-176
Sirkova, L. and V. Witkovsky (2001): On testing variance components in unbalanced mixed linear model, Applications in Mathematics 46(2001), 191-213
Sjöberg, L.E. (1983a): Unbiased estimation of variance-covariance components in condition adjustment. Univ. of Uppsala, Institute of Geophysics, Dept. of Geodesy, Report No. 19, 223-237.
Sjöberg, L.E. (1983b): Unbiased estimation of variance-components in condition adjustment with unknowns-a MINQUE approach. Zeitschriftfur Vennessungen, 108(9), 382-387.

Sjöberg, L.E. (1984a): Non-negative variance component estimation in the Gauss-Helmert adjustment model. Manuscripta Geodaetica, 9, 247-280.

Sjöberg, L.E. (1984b): Least-Squares modification of Stokes' and Venin g-Meinez , formula by accounting for truncation and potential coefficients errors. Manuscripta Geodaetica, 9, 209-229.

Sjöberg, L.E. (1985): Adjustment and. variance components estimation with a singular covariance . matrix. Zeitschriftfur Vermessungen, 110(4),145-151.

Sjöberg, L.E. (1993): General Matrix calculus, adjustment, variance covariance components estimation. Lecture note, Royal Institute of Technology, Stockholm, Sweden.

Sjöberg, L.E. (1994): The Best quadratic minimum biased non-negative definite estimator for an additive two variance components model. Royal Institute of Technology, Stockholm, Sweden.

Sjöberg, L.E. (1995): The best quadratic minimum bias non-negative estimator for an additive two variance component model. Manuscripta Geodaetica (1995) 20:139-144.

Sjöberg, L.E. (1999): An efficient iterative solution to transform rectangular geocentric coordinates to geodetic coordinates, *Zeitschrift für Vermessungswesen* 124 (1999) 295-297.

Sjöberg, L.E. (2008): A computational scheme to model the geoid by the modified Stokes formula without gravity reductions, J. Geod., 74, 255-268.

Sjöberg, L.E. (2011): On the best quadratic minimum bias non-negative estimator of a two-variance component model, J. Geodetic Science 1(2011) 280-285

Skolnikoff, I.S. (1956): Mathematical Theory of Elasticity, McGraw-Hill, New York 1956

Slakter, M.J. (1965): A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit-tests with respect to validity, J. Am. Statist. Ass. 60 (1965), 854-858

Small, C.G. (1996): The statistical theory of shape, Springer-Verlag, Heidelberg Berlin New York 1996

Smith, A.F.M. (1973): A general Bayesian linear model, J. Roy. Statist. Soc. B35 (1973), 67-75

Smith, P.J. (1995): A recursive formulation of the old problem of obtaining moments from cumulants and vice versa, J. Am. Statist. Ass. 49 (1995), 217-218

Smith, T. and S.D. Peddada (1998): Analysis of fixed effects linear models under heteroscedastic errors, Statistics & Probability Letters 37 (1998), 399-408

Smyth, G.K. (1989): Generalized linear models with varying dispersion, J. Roy. Statist. Soc. B51 (1989), 47-60

Snedecor, G.W. and W.G. Cochran (1967): Statistical methods, 6th ed., Ames lowa State University Press 1967

Sneeuw, N. and R. Bun (1996): Global spherical harmonic computation by two-dimensional Fourier methods, Journal of Geodesy 70 (1996), 224-232

Solari, H.G., Natiello, M.A. and G.B. Mindlin (1996): Nonlinear dynamics, IOP, Bristol 1996

Soler, T. and Hothem, L.D. (1989): Important parameters used in geodetic transformations, *Journal of Surveying Engineering* 115 (1989) 414-417.

Soler, T. and van Gelder, B. (1991): On covariances of eigenvalues and eigenvectors of second rank symmetric tensors, Geophys. J. Int., 105, 537-546

Solm, B.Y. and Kim, G.B. (1997): Detection of outliers in weighted least squares regression. Korean J. Comput. Appl. Math. 4, 2 (1997), 441-452.

Solomon, P.J. (1985): Transformations for components of variance and covariance, Biometrica 72 (1985), 233-239

Somogyi, J. (1998): The robust estimation of the 2D-projective transformation, Acta Geod. Geoph. Hung. 33 (1998), 279-288

Song, S.H. (1999): A note on S2 in a linear regression model based on two-stage sampling data, Statistics & Probability Letters 43 (1999), 131-135

Soper, H.E. (1916): On the distributions of the correlation coefficient in small samples, Biometrika 11 (1916), 328-413

Soper, H.E., Young, A.W., Cave, B.M., Lee, A. and Pearsom, K. (1955) Biometrika 11 (1955) 328–415

Spanos, A. (1999): Probability theory and statistical inference, Cambridge University Press, Cambridge 1999

Späth, H. and G.A. Watson (1987): On orthogonal linear l1 approximation, Numerische Mathematik 51 (1987), 531-543
Spilker, J.J. (1996): Tropospheric Effects on GPS. In: Parkinson, B.W., Spilker, J.J. (eds) Global Positioning System: Theory and Applications, Volume 1, American Institute of Aeronautics and Astronautics, Washington DC, pp 517-546
Spivak, M. (1979): Differential Geometry I-V. Publish or Perish Inc., Berkeley.
Spöck, G. (1997): Die geostatistische Berucksichtigung von a-priori Kenntnissen uber die Trendfunktion und die Kovarianzfunktion aus Bayesscher, Minimax und Spektraler Sicht, master thesis, University of Klagenfurt
Spöck, G. (2008): Non-stationary spatial modeling using harmonic analysis. In: Ortiz, J.M., Emery, X. (ed) Proceedings of the eighth international geostatistics congress. Gecamin, Chile, pp. 389-398
Spöck, G. and J. Pilz (2008): Non-spatial modeling using harmonic analysis, VIII Int. Geostatistics congress, pp. 1-10, Santiago, 2008
Spöck, G. and J. Pilz (2009): Spatial modelling and design covariance- robust minimax prediction based on convex design ideas, Stoch. Environ. Res. Risk Assess (2009) 1-21
Sposito, V.A. (1982): On unbiased Lp regression. J. Am. Statist. Ass. 77 (1982), 652-653
Sprent, P. and N.C. Smeeton (1989): Applied nonparametric statistical methods, Chapman and Hall, Boca Raton, Florida 1989
Sprinsky, W.H. (1974): The design of special purpose horizontal geodetic control networks. Ph.D. thesis, Tha Ohio State Univ., Columbus.
Sprott, D.A. (1978): Gauss's contributions to statistics, Historia Mathematica 5 (1978), 183-203
Srivastava, A.K., Dube, M. and V. Singh (1996): Ordinary least squares and Stein-rule predictions in regression models under inclusion of some superuous variables, Statistical Papers 37 (1996), 253-265
Srivastava, A.K. and Shalabh, S. (1996): Efficiency properties of least squares and Stein-Rule predictions in linear regression models, J. Appl. Stat. Science 4 (1996), 141-145
Srivastava, A.K. and Shalabh, S. (1997): A new property of Stein procedure in measurement error model, Statistics & Probability Letters 32 (1997), 231-234
Srivastava, M.S. and D. von Rosen (1998): Outliers in multivariate regression models, J. Multivar. Anal. 65 (1998), 195-208
Srivastava, M.S., Rosen von, D. (2002): Regression models with unknown singular covariance matrix. Linear Algebra Appl. 354, 1-3 (2002), 255-273.
Srivastava, V.K. and Upadhyaha, S. (1975): Small-disturbance and large sample approximations in mixed regression estimation, Eastern Economic Journal, 2 (1975), pp. 261-265
Srivastava, V.K. and B. Rah (1979): The Existence of the Mean of the Estimator in Seemingly, Communication in Statistics-Theory and Methods, A8(7)(1979), pp. 713-717
Srivastava, V.K. and Maekawa, K. (1995): Efficiency properties of feasible generalized least squares estimators in SURE models under non-normal disturbances, J. Econometrics 66, pp. 99-121 (1995)
Stahlecker, P. and K. Schmidt (1996): Biased estimation and hypothesis testing in linear regression, Acta Applicandae Mathematicae 43 (1996), 145-151
Stahlecker, P., Knautz, H. and G. Trenkler (1996): Minimax adjustment technique in a parameter restricted linear model, Acta Applicandae Mathematicae 43 (1996), 139-144
Stam, A.J. (1982): Limit theorems for uniform distributions on spheres in high dimensional Euclidean spaces, J. Appl. Prob. 19 (1982), 221-229
Stark, E. and Mikhail, E. (1973): Least Squares and Non-Linear Functions. Photogrammetric Engineering, Vol. XXXIX, No.4, 405-412.
Stark, H. (ed) (1987): Image recovery: theory and application. Academic, New York
Staudte, R.G. and Sheather, S.J. (1990): In Robust Estimation and Testing. John Wiley & Sons: USA, 1990.
Steeb, W.H. (1991): Kronecker product of matrices and applications, B.J. Wissenschaftsverlag 1991
Steele, B.M.: A modified EM algorithm for estimation in generalized mixed models

Stefanski, L.A. (1989): Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models, Communications Statist. Theory Method. 18 (1989), 4335-4358

Stefansky, W. (1971): Rejecting outliers by maximum normal residual, Ann. Math. Statistics 42 (1971), 35-45

Stein, C. (1945): A two-sample test for a linear hypothesis whose power is independent of the variance, Ann. Math. Statistics 16 (1945), 243-258

Stein, C. (1950): Unbiased estimates with minimum variance, Ann. Math. Statist. 21 (1950), 406-415

Stein, C. (1959): An example of wide discrepancy between fiducial and confidence intervals, Ann. Math. Statist. 30 (1959), 877-880

Stein, C. (1964): Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean, Ann. Inst. Statist. Math. 16 (1964), 155-160

Stein, C. and A. Wald (1947): Sequential confidence intervals for the mean of a normal distribution with known variance, Ann. Math. Statist. 18 (1947), 427-433

Stein, M.L. (1999): Interpolation of spatial data: some theory for kriging. Springer, New York

Steinberg, G. (1994): A Kinematic Approach in the Analysis of Kfar Hnassi Network. Perelmuter Workshop on Dynamic Deformation Models, Haifa, Israel.

Steiner, F. and B. Hajagos (1999): A more sophisticated definition of the sample median, Acta Geod. Geoph. Hung. 34 (1999), 59-64

Steiner, F. and B. Hajagos (1999): Insufficiency of asymptotic results demonstrated on statistical efficiencies of the L1Norm calculated for some types of the supermodel fp(x), Acta Geod. Geoph.Hung. 34 (1999), 65-69

Steiner, F. and B. Hajagos (1999): Error characteristics of MAD-S (of sample medians) in case of small samples for some parent distribution types chosen from the supermodels fp(x) and fa(x), Acta Geod. Geoph. Hung. 34 (1999), 87-100

Steinmetz, V. (1973): Regressionsmodelle mit stochastischen Koeffizienten, Proc. Oper. Res. 2, DGOR Ann. Meet., Hamburg 1973, 95-104

Stenger, H. (1971): Stichprobentheorie, Physica-Verlag, Würzburg 1971

Stephens, M.A. (1963): Random walk on a circle, Biometrika 50 (1963), 385-390

Stephens, M.A. (1964): The testing of unit vectors for randomness, J. Amer. Statist. Soc. 59 (1964), 160-167

Stephens, M.A. (1969): Tests for randomness of directions against two circular alternatives, J. Am. Statist. Ass. 64 (1969), 280-289

Stephens, M.A. (1969): Test for the von Mises distribution, Biometrika 56 (1969), 149-160

Stephens, M.A. (1979): Vector correlations, Biometrika 66 (1979), 41-88

Stepniak, C. (1985): Ordering of nonnegative definite matrices with application to comparison of linear models, Linear Algebra And Its Applications 70 (1985), 67-71

Stewart, G.W. (1995): Gauss, statistics, and Gaussian elimination, Journal of Computational and Graphical Statistics 4 (1995), 1-11

Stewart, G.W. (1995): Afterword, in Translation: Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, pars prior-pars posterior-supplementum by Carl Friedrich Gauss Theory of the Combination of Observations Least Subject to Errors, Classics in Applied Mathematics, SIAM edition, 205-236, Philadelphia 1995

Stewart, G.W. (1977): On the perturbation of pseudo-inverses, projections and linear least squares, SIAM Review 19 (1977), 634-663

Stewart, G.W. (1992): An updating algorithm for subspace tracking, IEEE Trans. Signal Proc. 40 (1992), 1535-1541

Stewart, G.W. (1998): Matrix algorithms, Vol. 1: Basic decompositions, SIAM, Philadelphia 1998

Stewart, G.W. (1999): The QLP approximation to the singular value decomposition, SIAM J. Sci. Comput. 20 (1999), 1136-1348

Stewart, G.W. (2001): Matrix algorithms, Vol. 2: Eigen systems, SIAM, Philadelphia 2001

Stewart, G.W. and Sun Ji-Guang (1990): Matrix perturbation theory, Academic Press, New York London 1990

Stewart, K.G. (1997): Exact testing in multivariate regression, Econometric reviews 16 (1997), 321-352

Steyn, H.S. (1951): The Wishart distribution derived by solving simultaneous linear differential equations, Biometrika, 38, 470-472

Stigler, S.M. (1973): Laplace, Fisher, and the discovery of the concept of sufficiency, Biometrika 60 (1973), 439-445

Stigler, S.M. (1973): Simon Newcomb, Percy Daniell, and the history of robust estimation 1885-1920, J. Am. Statist. Ass. 68 (1973), 872-879

Stigler, S.M. (1977): An attack on Gauss, published by Legendre in 1820, Historia Mathematica 4 (1977), 31-35

Stigler, S.M. (1986): The history of statistics, the measurement of uncertainty before 1900, Belknap Press, Harvard University Press, Cambridge/Mass. 1986

Stigler, S.M. (1999): Statistics on the table, the history of statistical concepts and methods, Harvard University Press, Cambridge London 1999

Stigler, S.M. (2000): International statistics at the millennium: progressing or regressing, International Statistical Review 68 (2000), 2, 111-121

Stone, M. (1974): Cross-validatory choice and assessment of statistical predictions. J R Statist Soc B 36:111-133

Stopar, B. (1999): Design of horizontal GPS net regarding non-uniform precision of GPS baseline vector components, Bollettino di Geodesia e Scienze Affini 58 (1999), 255-272

Stopar, B. (2001): Second order design of horizontal GPS net, Survey Review 36 (2001), 44-53

Storm, R. (1967): Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle, Leipzig 1967

Stotskii, A.A. and Elgerad, K.G., Stoskaya, J.M. (1998): Structure analysis of path delay variations in the neutral atmosphere, Astr. Astrophys. Transact. 17 (1998) 59-68

Stoyan, D. and Stoyan, H. (1994): Fractals, Random Shapes and Point Fields, Chichester: John Wiley & Sons, 1994

Stoyanov, J. (1997): Regularly perturbed stochastic differential systems with an internal random noise, Nonlinear Analysis, Theory, Methods & Applications 30 (1997), 4105-4111

Stoyanov, J. (1998): Global dependency measure for sets of random elements: The Italian problem and some consequences, in: Ioannis Karatzas et al. (eds.), Stochastic process and related topics in memory of Stamatis Cambanis 1943-1995, Birkhäuser-Verlag, Basel Boston Berlin 1998

Strang, G. and Borre, K. (1997): Linear Algebra, Geodesy and GPS, Wellesley Cambridge Press, Wellesley 1997.

Strubecker, K. (1964): Differentialgeometrie I-III. Sammlung Goschen, Berlin.

Stuelpnagel, J. (1964): On the parametrization of the three-dimensional rotation group. SIAM Review, Vol. 6, No.4, 422-430.

Sturmfels, B. (1994): Multigraded resultant of Sylvester type, *Journal of Algebra* 163 (1994) 115-127.

Sturmfels, B. (1996): Gröbner bases and convex polytopes, American Mathematical Society, Providence 1996.

Sturmfels, B. (1998): Introduction to resultants, *Proceedings of Symposia in Applied Mathematics* 53 (1998) 25-39.

Stoyanov, J. (1999): Inverse Gaussian distribution and the moment problem, J. Appl. Statist. Science 9 (1999), 61-71

Stoyanov, J. (2000): Krein condition inprobabilistic moment problems, Bernoulli 6 (2000), 939-949

Srecok, A.J. (1968): On the calculation of the inverse of the error function, Math. Computation 22 (1968), 144-158

Ströbel, D. (1997): Die Anwendung der Ausgleichungsrechnung auf elastomechanische Systeme, Deutsche Geodätische Kommission, Bayerische Akademie der Wissenschaften, Report C478, München 1997

Stroud, A.H. (1966): Gaussian quadrature formulas, Prentice Hall, Englewood Cliffs, N.J. 1966

Stuart, A. and J.K. Ord (1994): Kendall's advanced theory of statistics: volume I, distribution theory, Arnold Publ., 6th edition, London 1997

Student: The probable error of a mean, Biometrika 6 (1908), 1-25

Stulajter, F. (1978): Nonlinear estimators of polynomials in mean values of a Gaussian stochastic process, Kybernetika 14 (1978), 206-220

Styan, G.P.H. (1973): Hadamard products and multivariate statistical analysis, Linear Algebra Appl., 6 (1973), pp. 217-240

Styan, G.P.H. (1983): Generalised inverses in: Encyclopedia of statistical sciences, Vol. 3, S. Kotz, N.L. Johnson and C.B. Read (eds.), pp. 334-337, wiley, New York, 1983

Styan, G.P.H. (1985): Schur complements and statistics, in: Proc. First International Tampere Seminar on Linear Statistical Models, T. Pukkila and S. Puntaten (eds.), pp. 37-75, Dept, of Math. Sciences, Univ. of Tampere, Finland, 1985

Subrahamanyan, M. (1972): A property of simple least squares estimates, Sankhya B34 (1972), 355-356

Sugaria, N. and Y. Fujikoshi (1969): Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence, Ann. Math. Stat. 40 (1969), 942-952

Sun, J.-G. (2000): Condition number and backward error for the generalized singular value decomposition, Siam J. Matrix Anal. Appl. 22 (2000), 323-341

Sun, W. (1994): A New Method for Localization of Gross Errors. Surv. Rev. 1994, 32, 344-358.

Sun, Z. and Zhao, W. (1995): An algorithm on solving least squares solution of rank-defective linear regression equations with constraints. Numer. Math., Nanjing 17, 3 (1995), 252-257.

Sünkel, H. (1984): Fourier analysis of geodetic networks. Erice Lecture Notes, this volume, Heidelberg.

Sünkel, H. (1999): Ein nicht-iteratives Verfahren zur Transformation geodätischer Koordinaten, *Öster. Zeitschrift für Vermessungswesen* 64 (1999) 29-33.

Svendsen, J.G.G. (2005): Some properties of decorrelation techniques in the ambiguity space. GPS Solutions. doi:10.1007/s10291-005-0004-6

Swallow, W.H. and S.R. Searle (1978): Minimum variance quadratic unbiased estimation (MIVQUE) of variance components, Technometrics 20 (1978), 265-272

Swamy, P.A.V.B. (1971): Statistical inference in random coefficient regression models, Springer-Verlag, Heidelberg Berlin New York 1971

Swamy, P.A.V.B. and J.A. Mehta (1969): On Theil's mixed regression estimators, J. Americ. Stst. Assoc., 64 (1969), pp. 273-276

Swamy, P.A.V.B. and J.A. Mehta (1977): A note on minimum average risks estimators for coefficients in linear models, Communications in Stat., Part A -Theory and Methods, 6 (1977), pp. 1181-1186

Sylvester, J.J. (1850): Additions to the articles, On a new class of theorems, and On Pascal's theorem, Phil. Mag. 37 (1850), 363-370

Sylvester, J.J. (1851): On the relation between the minor determinants of linearly equivalent quadratic functions, Phil. Mag. 14 (1851), 295-305

Szabados, T. (1996): An elementary introduction to the Wiener process and stochastic integrals, Studia Scientiarum Mathematicarum Hungarica 31 (1996), 249-297

Szasz, D. (1996): Boltzmann's ergodic hypothesis, a conjecture for centuries?, Studia Scientiarum Mathematicarum Hungarica 31 (1996), 299-322

Takos, I. (1999): Adjustment of observation equations without full rank, Bolletino di Geodesia e Scienze Affini 58 (1999), 195-208

Tanana, V.P. (1997): Methods for solving operator equations, VSP, Utrecht, Netherlands 1997

Tanizaki, H. (1993): Nonlinear filters - estimation and applications, Springer-Verlag, Heidelberg Berlin New York 1993

Tanner, A. (1996): Tools for statistical inference, 3rd ed., Springer-Verlag, Heidelberg Berlin New York 1996

Tarpey, T. (2000): A note on the prediction sum of squares statistic for restricted least squares, The American Statistician 54 (2000), 116-118
Tarpey, T. and B. Flury (1996): Self-consistency, a fundamental concept in statistics, Statistical Science 11 (1996), 229-243
Tasche, D. (2003): Unbiasedness in least quantile regression, in: R. Dutter, P. Filzmoser, U. Gather, P.J. Rousseeuw (eds.), Developments in Robust Statistics, 377-386, Physica Verlag, Heidelberg 2003
Tashiro, Y. (1977): On methods for generating uniform random points on the surface of a sphere, Ann. Inst. Statist. Math. 29 (1977), 295-300
Tatarski, V.J. (1961): Wave propagation in a turbulent medium Mc Graw Hill book Comp., New York 1961
Tate, R.F. (1959): Unbiased estimation functions of location and scale parameters, Ann. Math. Statist. 30 (1959), 341-366
Tate, R.F. and G.W. Klett (1959): Optimal confidence intervals for the variance of a normal distribution, J. Am. Statist. Ass.. 16 (1959), 243-258
Tatom, F.B. (1995): The relationship between fractional calculus and fractals. Fractals, 1995, 3, 217-229
Teicher, H. (1961): Maximum likelihood characterization of distribution, Ann. Math. Statist. 32 (1961), 1214-1222
Tennekes, H. and J.L. Lumley (1972): A First Course in Turbulence, MIT Press, 300 pages, Cambridge, MA
Teräsvirta, T. (1981): Some results on improving the least squares estimation of linear models by mixed estimators, Scandinavian Journal of Statistics 8 (1981), pp. 33-38
Teräsvirta, T. (1982): Superiority comparisons of homogeneous linear estimators. Communications in Statistics, All, pp. 1595-1601
Taylor, J.R. (1982): An introduction to error analysis, University Science Books, Sausalito 1982
Taylor, G.I. (1935): Statistical theory of turbulence. Proc. Roy. Soc. A 151, pp. 421-478.
Taylor, G.I. (1938): The spectrum of turbulence. Proceedings Royal Society of London A 164, pp. 476-490.
Taylor, G.I. (1938): Production and dissipation of vorticity in a turbulent fluid. Proceedings Royal Society of London A164, pp. 15-23.
Tesafikova, E. and Kubacek, L. (2003): Estimators of dispersion in models with constraints (demoprogram). Department of Algebra and Geometry, Faculty of Science, Palacky University, Olomouc, 2003.
Tesafikova, E. and Kubacek, L. (2004): A test in nonlinear regression models. Demo program. Department of Algebra and Geometry, Faculty of Science, Palacky University, Olomouc, 2004 (in Czech).
Tesafikova, E. and Kubacek, L. (2005): Variance components and nonlinearity. Demoprogram. Department of Algebra and Geometry, Faculty of Science, Palacky University, Olomouc, 2005.
Teunissen, P.J.G. (1985a): The geometry of geodetic inverse linear mapping and non-linear adjustment, Netherlands Geodetic Commission, Publications on Geodesy, New Series, Vol. 8/1, Delft 1985
Teunissen, P.J.G. (1985b): Zero order design: generalized inverses, adjustment, the datum problem and S-transformations, In: Optimization and design of geodetic networks, Grafarend, E.W. and F. Sanso eds., Springer-Verlag, Heidelberg Berlin New York 1985
Teunissen, P.J.G. (1987): The 1 and 2D Symmetric Helmert Transformation: Exact Non-linear Least-Squares Solutions. Reports of the Department of Geodesy, Section Mathematical and Physical Geodesy, No. 87.1, Delft
Teunissen, P.J.G. (1989a): Nonlinear inversion of geodetic and geophysical data: diagnosing nonlinearity, In: Brunner, F.K. and C. Rizos (eds.): Developments in four-dimensional geodesy, Lecture Notes in Earth Sciences 29 (1989), 241-264
Teunissen, P.J.G. (1989b): First and second moments of non-linear least-squares estimators, Bull. Geod. 63 (1989), 253-262

Teunissen, P.J.G. (1990): Non-linear least-squares estimators, Manuscripta Geodaetica 15 (1990), 137-150

Teunissen, P.J.G. (1991): Circular Letter to all Special Study Group 4.120 Members, 24.01.1991, nicht ver6:ffentlicht.

Teunissen, P.J.G. (1993): Least-squares estimation of the integer GPS ambiguities, LGR series, No. 6, Delft Geodetic Computing Centre, Delft 1993

Teunissen, P.J.G. (1995a): The invertible GPS ambiguity transformation, Manuscripta Geodaetica 20 (1995), 489-497

Teunissen, P.J.G. (1995b): The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation, Journal of Geodesy 70 (1995), 65-82

Teunissen, P.J.G. (1997): A canonical theory for short GPS baselines. Part I: The baseline precision, Journal of Geodesy 71 (1997), 320-336

Teunissen, P.J.G. (1997): On the sensitivity of the location, size and shape of the GPS ambiguity search space to certain changes in the stochastic model, Journal of Geodesy 71 (1997), 541-551

Teunissen, P.J.G. (1997): On the GPS widelane and its decorrelating property, Journal of Geodesy 71 (1997), 577-587

Teunissen, P.J.G. (1997): The least-squares ambiguity decorrelation adjustment: its performance on short GPS baselines and short observation spans, Journal of Geodesy 71 (1997), 589-602

Teunissen, P.J.G. (1998): Quality Control and GPS. In: Teunissen PJG, KIeusberg A(eds) GPSfor Geodesy, 2nd ed, Springer Verlag, Berlin, Hedilberg, pp. 271-318

Teunissen, P. (1998b): First and second moments of non-linear least-quares estimations, Bull. Geod. 63 (1989) 253-262

Teunissen, P.J.G. (1999a): The probability distribution of the GPS baseline for a class of integer ambiguity estimators. J Geod 73: 275-284

Teunissen, P.J.G. (1999b): An optimality property of the integer leastsquares estimator. J. Geod. 73:587-593

Teunissen, P.J.G. (2003): Theory of integer equivariant estimation with application to GNSS. J. Geod. 77:402-410

Teunissen, P.J.G. and Knickmeyer, E.H. (1988): Non-linearity and least squares, *CISM Journal ASCGC* 42 (1988) 321-330.

Teunissen, P.J.G. and Kleusberg, A. (1998): GPS for geodesy, 2nd enlarged edition. Springer, Heidelberg

Teunissen, P.J.G., Simons, D.G., and Tiberius, C.C.J.M. (2005): Probability and observation theory. Lecture Notes Delft University of Technology, 364 p

Teunissen, P.J.G. and A.R. Amiri-Simkooei (2008): Least-squares variance component estimation, Journal of Geodesy, 82 (2008): 65-82.

Teunissen, P.J.G. (2008):On a stronger-than-best property for best prediction, J. Geodesy 82 (2008) 165-175

Theil, H. (1963): On the use of incomplete prior information in regression analysis, J,. Amer. Stat. Assoc., 58 (1963), pp. 401-414

Theil, H. (1965): The analysis of disturbances in regression analysis, J. Am. Statist. Ass. 60 (1965), 1067-1079

Theil, H. (1971): Principles of econometrics, Wiley, New York 1971

Theil, H. and A.S. Goldberger (1961): On Pure and Mixed Statistical Estimation in Economics, International Economic Review, Vol. 2, 1961, p. 65

Theobald, C.M. (1974): Generalizations of Mean Square Error Applied to Ridge Regression, Journal of the Royal Statistical Society 36 (series B): 103-106.

Thompson, R. (1969). Iterative estimation of variance components for non-orthogonal data, Biometrics 25 (1969), 767-773

Thompson, W.A. (1955): The ratio of variances in variance components model, Ann. Math. Statist. 26 (1955), 325-329

Thompson, E. H. (1959a): A method for the construction of orthogonal matrices, *Photogrammetria* III (1959) 55-59.

Thompson, E. H. (1959b): An exact linear solution of the absolute orientation. *Photogrammetria* XV (1959) 163-179.
Thomson, D.J. (1982): Spectrum estimation and harmonic analysis, Proceedings Of The IEEE 70 (1982), 1055-1096
Thorand, V. (1990): Algorithmen zur automatischen Berechnung von Naherungskoordinaten in geodätischen Lagenetzen. Vermessungstechnik, 4, 120-124.
Thorpe, J.A. (1979): Elementary Topics in Differential Geometry. Springer-Verlag, New York.
Tychonoff, A.N. (1963): Solution of incorrectly formulated problems and regularization method, Dokl. Akad. Nauk. SSSR, 151 (1963), pp. 501-504
Tiberius, C.C.J.M., de Jonge, P.J. (1995): Fast positioning using the LAMBDA method. In: Proceedings DSNS-95, paper 30, 8 p
Tiberius, C.C.J.M., Teunissen, P.J.G. and de Jonge, P.J. (1997): Kinematic GPS: performance and quality control. In: Proceedings KIS97, pp. 289-299
Tiberius, C.C.J.M., Jonkman, N. and F. Kenselaar (1999): The Stochastics of GPS Observables. GPS World 10: 49-54
Tiberius, C.C.J.M. and F. Kenselaar (2000): Estimation of the stochastic model for GPS code and phase observables, Survey Review 35 (2000), 441-455
Tiberius, C.C.J.M. and F. Kenselaar (2003): Variance Component Estimation and Precise GPS Positioning: Case study, Journal of surveying engineering, 129(1): 11-18, 2003.
Tikhonov, A.N. (1963): Solution of incorrectly formulated problems arid regularization method. Dokl. Akad. Nauk. SSSR, 1 51, 501-504.
Tikhonov, A.N. and V.Y. Arsenin (1977): Solutions of ill-posed problems, J. Wiley, New York 1977
Tikhonov, A.N., Leonov, A.S. and A.G. Yagola (1998): Nonlinear ill-posed problems, Vol. 1, Appl. Math. and Math. Comput. 14, Chapman & Hall, London 1998
Tjöstheim, D. (1990): Non-linear time series and Markov chains, Adv. Appl. Prob. 22 (1990), 587-611
Tjur, T. (1998): Nonlinear regression, quasi likelihood, and overdispersion in generalized linear models, American Statistician 52 (1998), 222-227
Tobias, P.A. and D.C. Trinidade (1995): Applied reliability, Chapman and Hall, Boca Raton 1995
Tominaga, Y. and I. Fujiwara (1997): Prediction-weighted partial least-squares regression (PWPLS), Chemometrics and Intelligent Lab Systems 38 (1997), 139-144
Tong, H. (1990): Non-linear time series, Oxford University Press, Oxford 1990
Toranzos F.I. (1952): An asymmetric bell-shaped frequency curve, Ann. Math. Statist. 23 (1952), 467-469
Torge, W. and Wenzel, H.G. (1978): Dreidimensionale Ausgleichung des Testnetzes Westharz.Deutsche Geodattsche Kommission, Bayerische Akademie der Wissenschaften Report B234, München.
Torge, W. (2001): Geodesy, 3rd edn. Walter de Gruyter, Berlin
Toro-Vizearrondo, C. and T.D. Wallace (1969): A Test of the Mean Square Error Criterion for Restrictions in Linear Regression, Journal of the American Stst. Assoc., 63 (1969), pp. 558-572
Toutenburg, H. (1968): Vorhersage im allgemeinen linearen Regressionsmodell mit Zusatzinformation über die Koeffzienten, Operationsforschung Maths Statistik, Vol. 1, pp. 107-120, Akademic Verlag, Berlin 1968
Toutenburg, H. (1970a): Vorhersage im allgemeinen linearen Regressionsmodell mit stochastischen Regressoren, Math. Operationsforschg. Statistik 2 (1970), 105-116
Toutenburg, H. (1970b): Probleme linearer Vorhersagen im allgemeinen linearen Regressionsmodell, Biometrische Zeitschrift, 12: pp. 242-252
Toutenburg, H. (1970c): Über die Wahl zwischen erwartungstreuen und nichterwartungstreuen Vorhersagen, Operationsforschung Mathematische Statistik, vol. 2, pp. 107-118, Akademic Verlag, Berlin 1970
Toutenburg, H. (1973): Lineare Restriktionen und Modellwahl im allgemeinen linearen Regressionsmodell, Biometrische Zeitschrift, 15 (1973), pp. 325-342
Toutenburg, H. (1975): Vorhersage in linearen Modellen, Akademie Verlag, Berlin 1975

Toutenburg, H. (1976): Minimax-linear and MSE-linear estimators in generalized regression, Biometr. J. 18 (1976), pp. 91-104

Toutenburg, H. (1982): Prior Information in Linear Models, Wiley, New York.

Toutenburg, H. (1989a): Investigations on the MSE-superiority of several estimators of filter type in the dynamic linear model (i.e. Kalman model), Technical Report 89-26, Center for Multivariate Analysis, The Pennsylvania State University, State College

Toutenburg, H. (1989b): Mean-square-error-comparisons between restricted least squares, mixed and weighted mixed estimators, Forschungsbericht 89/12, Fachbereich Statistik, Universitat Dortmund, Germany

Toutenburg, H. (1996): Estimation of regression coefficients subject to interval constraints, Sankhya: The Indian Journal of Statistics A, 58 (1996), 273-282

Toutenburg, H. and Schaffrin, B. (1989): Biased mixed estimation and related problems, Technical Report, Universitat Stuttgart, Germany

Toutenburg, H. and Trenkler, G. (1990): Mean square error matrix comparisons of optimal and classical predictors and estimators in linear regression, Computational Statistics and Data Analysis 10: 297-305

Toutenburg, H., Trenkler, G. and Liski, E.P. (1992): Optimal estimation methods under weakened linear restrictions, Computational Statistics and Data Analysis 14: 527-536.

Toutenburg, H. and Shalabh (1996): Predictive performance of the methods of restricted and mixed regression estimators, Biometrical Journal 38: 951-959.

Toutenburg, H., Fieger, A. and Heumann, C. (1999): Regression modelling with fixed effects-missing values and other problems, in C. R. Rao and G. Szekely (eds.), Statistics of the 21st Century, Dekker, New York.

Townsend, A.A. (1976): The Structure of Turbulent Shear Flow, Second Edition, Cambridge University Press, 429 pages 1976

Townsend, E.C. and S.R. Searle (1971): Best quadratic unbiased estimation of variance components from unbalanced data in the l-way classification, Biometrics 27 (1971), 643-657

Trefethen, L.N. and D. Bau (1997): Numerical linear algebra, Society for Industrial and Applied Mathematics (SIAM), Philadelphia 1997

Trenkler, G. (1985): Mean square error matrix comparisons of estimators in linear regression, Communications in Statistics, Part A-Theory and Methods 14: 2495-2509

Trenkler, G. (1987): Mean square error matrix comparisons among restricted least squares estimators, Sankhya, Series A 49: 96-104

Trenkler, G. and Trenkler, D. (1983): A note on superiority comparisons of linear estimators, Communications in Statistics, Part A-Theory and Methods 17: 799-808

Trenkler, G. and Toutenburg, H. (1990): Mean-square error matrix comparisons between biased estimators an overview of recent results, Statistical Papers 31: 165-179

Troskie, C.G. and D.O. Chalton (1996): Detection of outliers in the presence of multicollinearity, in: Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, A.K. and V.L. Girko, 273-292, VSP, Utrecht 1996

Troskie, C.G., Chalton, D.O. and M. Jacobs (1999): Testing for outliers and influential observations in multiple regression using restricted least squares, South African Statist. J. 33 (1999), 1-40

Trujillo-Ventura, A. and Ellis, J.H. (1991): Multiobjective air pollution monitoring network design. Atmos Environ 25:469-479

Tscherning, C.C. (1978): Collocation and least-squares methods as a tool for handling gravity field dependent data obtained through space research techniques. Bull Geod 52:199-212

Tseng, Y.Y. (1936): The characterisitic value problem of Hermitian functional operators in a non-Hilbert space, Ph.D. Thesis, Univeristy of Chicago, Chicago 1936

Tseng, Y.Y. (1949a): Sur les solutions des èquations opèratrices fonctionnelles entre les espaces enitaires, C.R. Acad. Sci. Paris, 228 (1949), pp. 640-641

Tseng, Y.Y. (1949b): Generalized inverses of unbounded operators between two arbitrary spaces, Dokl. Acas. Nauk SSSR (new series) 67 (1949), pp. 431-434

Tseng, Y.Y. (1949c): Properties and classification of generalized inverses of closed operators, Dokl. Acad. Nauk SSSR (new series) 67 (1949), pp. 607-610

Tseng, Y.Y. (1956): Virtual solutions and general inversions, Uspehi. Mat. Nauk (new series) 11, (1956), pp. 213-215

Tsimikas, J.V. and J. Ledolter (1997): Mixed model representation of state space models: New smoothing results and their application to REML estimation, Statistica Sinica 7 (1997), 973-991

Tu, R. (1996): Comparison between confidence intervals of linear regression models with and without restriction. Commun. Stat., Theory Methods 28, 12 (1999), 2879-2898.

Tufts, D.W. and R. Kumaresan (1982): Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood, Proc. of IEEE Special issue on Spectral estimation 70 (1982), 975-989

Tukey, J.W. (9159): An introduction to the measurement of spectra, in: Grenander U.: Probability and statistics, New York

Turkington, D. (2000): Generalised vec operators and the seemingly unrelated regression equations model with vector correlated disturbances, Journal of Econometrics 99 (2000), 225-253

Ulrych, T.J. and R.W. Clayton (1976): Time series modelling and maximum entropy, Phys. Earth and Planetary Interiors 12 (1976), 188-200

Vainikko, G.M. (1982): The discrepancy principle for a class of regularization methods, USSR. Comp. Math. Math. Phys. 22 (1982), 1-19

Vainikko, G.M. (1983): The critical level of discrepancy in regularization methods, USSR. Comp. Math. Math. Phys. 23 (1983), 1-9

Van Der Waerden, B.L. (1950): Modern Algebra, 3rd Edition, F. Ungar Publishing Co., New York 1950.

Van Huffle, S. (1990): Solution and properties of the restricted total least squares problem, Proceedings of the International Mathematical Theory of Networks and Systems Symposium (MTNS 1989), 521-528

Vanicek, P. and E.W. Grafarend (1980): On the weight estimation in leveling, National Oceanic and Atmospheric Administration, Report NOS 86, NGS 17, Rockville 1980

Vanicek, P. and Krakiwsky, E.J. (1982): Geodesy: The concepts. North-Holland Publishing Company, Amsterdam-New York-Oxford 1982.

Van Loa, C.F. (1978): Computing integrals involving the matrix exponential. IEEE Trans. Autom. Control AC- 23: 395-404

Van Mierlo, J. (1988): Rückwärtschnitt mit Streckenverhältnissen, *Algemain Vermessungs Nachrichten* 95 (1988) 310-314.

Van Montfort, K. (1988): Estimating in structural models with non-normal distributed variables: some alternative approaches, Leiden 1988

Van Ness, J.W. (1965): Asymptotic normality of bispectral estimates, Technical report No 14, Dept. Statistics, Stanford University 1965

Vasconcellos, K.L.P. and M.C. Gauss (1997): Approximate bias for multivariate nonlinear heteroscedastic regressions, Brazilian J. Probability and Statistics 11 (1997), 141-159

Vasconcelos, W.V. (1998): Computational methods in commutative algebra and algebraic geometry, Springer-Verlag, Berlin-Heidelberg 1998.

Ventsel, A.D. and M.I. Freidlin (1969): On small random perturbations of dynamical systems, Report delivered at the meeting of the Moscow Mathematical Society on March 25, 1969, Moscow 1969

Ventzell, A.D. and M.I. Freidlin (1970): On small random perturbations of dynamical systems, Russian Math. Surveys 25 (1970), 1-55

Verbeke, G. and G. Molenberghs (2000): Linear mixed models for longitudinal data, Springer-Verlag, Heidelberg Berlin New York 2000

Vernik, L. and Zoback, M.D. (1992): Estimation of ma.xjmum horizontal principal stress magnitude from stress-induced well bore breakouts in the Cajon Pass scientific research borehole, J. geophys. Res., B97, 5109-5119

Vernizzi, A., Goller, R. and P. Sais (1995): On the use of shrinkage estimators in filtering extraneous information, Giorn. Econ. Annal. Economia 54 (1995), 453-480

Veroren, L.R. (1980): On estimation of variance components, Statistica Neerlandica 34 (1980), 83-106

Vetter, M. (1992): Automatische Berechnung zweidimensionaler Naherungskoordinaten - Konzeption und Realisierung im Programm AURA (Automatische Berechnung und Revision approximativer Koordinaten). Allgemeine Vermessungs-N achrichten, 99. Jg., Heft 6, 245-255.

Vichi, M. (1997): Fitting L2 norm classification models to complex data sets, Student 2 (1997), 203-213

Viertl, R. (1996): Statistical Methods for Non-Precise DATA. CRC Press, Boca Raton, New York, London and Tokyo, 1996

Vigneau, E., Devaux, M.F., Qannari, E.M. and P. Robert (1997): Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration, J. Chemometrics 11 (1997), 239-249

Vincenty, T. (1978): Vergleich zweier Verfahren zur Berechnung der geodätischen Breite und Höhe aus rechtwinkligen koorninaten, *Allgemeine Vermessungs-Nachrichten* 85 (1978) 269-270.

Vincenty, T. (1980): Zur räumlich-ellipsoidischen Koordinaten-Transformation, *Zeitschrift für Vermessungswesen*, 105 (1980) 519-521.

Vinod, H.D. and L.R. Shenton (1996): Exact moments for autoregressive and random walk models for a zero or stationary initial value, Econometric Theory 12 (1996), 481-499

Vinograde, B. (1950): Canonical positive definite matrices underinternal linear transformations, Proc. Amer. Math. Soc.1 (1950), 159-161

Vogel, M. and Van Mierlo, J. (1994): Deformation Analysis of the Kfar-Hanassi Network. Perelmuter Workshop on Dynamic Deformation Models, Haifa, Israel.

Vogler, C.A. (1904): Zeitschrift für Vermessungswesen 23 (1904), pp. 394-402

Voinov, V.G. and M.S. Nikulin (1993): Unbiased estimators and their applications, volume 1: univariate case, Kluwer, Academic Publishers, Dordrecht 1993

Voinov, V.G. and M.S. Nikulin (1993): Unbiased estimators and their applications, volume 2: multivariate case, Kluwer, Academic Publishers, Dordrecht 1993

Volaufova, J. (1980): On the confidence region of a vector parameter. Math. Slovaca 30 (1980), 113-120.

Volterra, V. (1930): Theory of functionals, Blackie, London 1930

Von Mises, R. (1981): über die Ganzzahligkeit der Atomgewichte und verwandte Fragen, Phys. Z. 19 (1918), 490-500

Wackernagel, H. (1995): Multivariate geostatistics. Springer, Heidelberg

Wackernagel, H. (2003): Multivariate geostatistics. Springer-Verlag, Berlin, 3rd edition.

Waerden Van Der, B.L. (1967): Algebra II. Springer-Verlag, Berlin.

Wagner, H. (1959): Linear programming techniques for regression analysis, J. Am. Statist. Ass. 56 (1959), 206-212

Wahba, G. (1975): Smoothing noisy data with spline functions, Numer. Math. 24 (1975), 282-292

Wald, A. (1939): Contributions to the theory of statistical estimation and testing hypotheses, Ann. Math. Statistics 10 (1939), 299-326

Wald, A. (1945): Sequential tests for statistical hypothesis, Ann. Math. Statistics 16 (1945), 117-186

Walder, O. (2006): Forpflanzung der Un scharfe von Messdaten auf abgeleitete differential geometrische Grossen, PFG 6 (2006), 491-499

Walder, O. (2007a): An application of the fuzzy theory in surface interpolation and surface deformation analysis, Fuzzy sets and systems 158 (2007), 1535-1545

Walder, O. (2007b): On analysis and forecasting of surface movement and deformation: Some AR models and their application, Ace. Verm, Nachr., 3 (2007), 96-100

Walder, O. and M. Buchroithner (2004): A method for sequential thinning of digital raster terrain models, PF G 3, (2004), 215-221

Walk, H. (1967a): Randeigenschaften zufalliger P otenzreihen und eine damit zusamnlenhangende Verrillgemeirierung des Satzes von Fatou und Nevanlinna (mit K. Hinderer). Math. Ann. 172, 94-104 (1967).

Walk, H. (1967b): Symmetrisierte und zentrierte Folgen von Zufallsvariablen. Math. Z. 102, 44-55 (1967).
Walk, H. (1968): Über das Randverhalten zufalliger Potenzreihen. 1. reine angew. Math. 230, 66-103 (1968).
Walk, H. (1969a): Approximation durch Folgen linearer positiver Operatoren. Arch. Math. 20, 398-404 (1969).
Walk, H. (1969b): Wachsumsverhalten zufalliger Potenzreihen. Z. Wahrscheinlichkeitstheorie verw. Gebiete 12, 293-306 (1969)
Walk, H. (1970a): Bemerkungen tiber multiplikative Systeme. Math. Ann. 186, 36-44 (1970)
Walk, H. (1970b): Approximation unbeschrankter Funktionen durch lineare positive Operatoren. Habilitationsschrift, Dniv. Stuttgart, Juni 1970.
Walk, H. (1972): Konvergenz- und Guteaussagen für die Approximation durch Folgen linearer positiver - Operatoren (mit M. W. Muller). Proceedings of the International Conference on Constructive Function Theory, Varna, May 1970, 221-233 (1972).
Walk, H. (1973): Zufallige Potenzreihen mit multiplikativ abhangigen Koeffizienten. Buletinul Institutului Politehnic din Jasi 19 (23), 107-114 (1973).
Walk, H. (1974): A generalization of renewal processes (mit K. Hinderer). European Meeting of Statisticians, Budapest 1972. Colloquia Mathematica Societatis Janos Bolyai 9, 315-318 (1974).
Walk, H. (1977a): Approximation von Ableitungen unbeschrankter Funktionen durch line are Operatoren. Mathematica - Revue d'Analyse Numerique et de Theorie de l'Approximation 6, 99-105 (1977).
Walk, H. (1977b): An invariance principle in stochastic approximation. Recent Developments in Statistics (eds. J. R. Barra et al.), 623-625 (1977). North Holland, Amsterdam
Walk, H. (1979): Sequential estimation of the solution of an integral equation in filtering theory. In: Stochastic Control Theory and Stochastic Differential Systems (eds. M. Kohlmann, W. Vogel), 598-605. Springer-Verlag, Berlin, 1979
Walk, H. (1980): A functional central limit theorem for martingales in C(K) and its application to sequential - estimates.J. reine angew. Math. 314 (1980), 117-135
Walk, H. (1983): Stochastic iteration for a constrained optimization problem. Commun. Statist. - Sequential Analysis 2 (1983-84), 369-385
Walk, H. (1985): On recursive estimation of the mode, Statistics and Decisions 3 (1985), 337-350.
Walk, H. (1988): Limit behavior of stochastic approximation processes. Statistics and Decisions 6 (1988), 109-128.
Walk, H. (1992a): Stochastic Approximation and Optimization of Random Systems (mit L. Ljung und G. Pflug) (113 S.). Birkhauser, Basel, 1992
Walk, H. (1992b): Stochastic Approximation and Optimization of Random Systems (mit L. Ljung und G. Pflug) (113 S.). Birkhauser, Basel, 1992.
Walk, H. (1998): Weak and strong universal consistency of semi-recursive partitioning and kernel regression estimates (mit Gyorfi L, Kohler M). Stat. Decis. 16 (1998) 1-18
Walk, H. (2001): Strong universal point wise consistency of recursive regression estimates. Ann. Inst. Statist. Math. 53 (2001), 691-707.
Walk, H. (2002): Almost sure convergence properties of Nadaraya-Watson regression estimates. In: Modeling Uncertainty: An Examination of its Theory, Methods and Applications (eds. M. Dror, P. L'Ecuyer, F. Szidarovszky), 201-223 (2002). Kluwer Academic Publishers, Dordrecht.
Walk, H. (2003): The estimation problem of minimum mean' squared error (mit L. Devroye, L. Gyorfi und D. Scnafer). Statistics and Decisions 21 (2003), 15-28
Walk, H. (2005): Strong laws of large numbers by elementary Tauberian arguments. Monatsh. Math. 144 (2005), 329-346.
Walk, H. (2006): Rates of convergence for partitioning and nearest neighbor regression estimates with unbounded data (mit M. Kohler und A. Krzyzak). 1. Multivariate Anal. 97 (2006), 311-323
Walk, H. (2008a): A universal strong law of large numbers for conditional expectations via nearest neighbors. J. Multivariate Anal. 99 (2008), 1035-1050.

Walk, H. (2008b): Non parametric nearest neighbor based empirical portfolio selection strategies (mit L. Gyorfi und F. Udina). Statistics and Decisions 26 (2008), 145-157 .

Walk, H. (2009a): Optimal global rates of convergence for nonparametric regression with unbounded data (mit M. Kohler und A. Krzyzak). 1. Statist. Planning Inference 139 (2009), 1286-1296.

Walk, H. (2009b): Strong laws of large numbers and nonparametric estimation. Universitat Stuttgart, Fachbereich Mathematik, Preprint 2009-003. In: Recent Developments in Applied Probability and Statistics (eds. L. Devroye, B. Karasozen, M. Kohler, R. Korn), 183-214 (2010). Physica-Verlag, Heidelberg

Walk, H. (2010): Strong consistency of kernel estimates of regression function under dependence. Statist. Probab. Letters 80 (2010), 1147-1156 .

Walker, J.S. (1996): Fast Fourier transforms, 2nd edition, CRC Press, Boca Raton 1996

Walker, P.L. (1996): Elliptic functions, J. Wiley, New York 1996

Walker, S. (1996): An EM algorithm for nonlinear random effect models, Biometrics 52 (1996), 934-944

Wallace, D.L. (1980): The Behrens-Fisher and Fieller-Creasy problems, in: R.A. Fisher: an appreciation, Fienberg and Hinkley, eds, Springer-Verlag, Heidelberg Berlin New York 1980, 117-147

Wan, A.T.K. (1994): The sampling performance of inequality restricted and pre-test estimators in a misspecified linear model, Austral. J. Statist. 36 (1994), 313-325

Wan, A.T.K. (1994): Risk comparison of the inequality constrained least squares and other related estimators under balanced loss, Econom. Lett. 46 (1994), 203-210

Wan, A.T.K. (1994): The non-optimality of interval restricted and pre-test estimators under squared error loss, Comm. Statist. A - Theory Methods 23 (1994), 2231-2252

Wan, A.T. (1999): A note on almost unbiased generalized ridge regression estimator under asymmetric loss, J. Statist. Comput. Simul. 62 (1999), 411-421

Wan, A.T.K. and K. Ohtani (2000): Minimum mean-squared error estimation in linear regression with an inequality constraint, J. Statist. Planning and Inference 86 (2000), 157-173

Wang, C.C. (1970): A new representation theorem for isotropic functions: an answer to Professor G.F. Smith's criticism of my papers on representations for isotropic functions. Arch. Rat. Mech. Anal 36.

Wang, C.C. (1971): Corrigendum to my recent papers on "Representations for isotropic functions", Arch. Rational Mech. Anal. 43 (1971) 392-395

Wang H. and D. Suter (2003): Using Symmetry in Robust Model Fitting, Pattern Recognition Letters (PRL), Vol. 24, No.16, pp. 2953-2966

Wang, J. (1996): Asymptotics of least-squares estimators for constrained nonlinear regression, Annals of Statistics 24 (1996), 1316-1326

Wang, J. (1999): Stochastic Modeling for Real-Time Kinematic GPS/GLONASS Positioning. Journal of The Institute of Navigation 46: 297-305

Wang, J. (2000): An approach to GLONASS ambiguity resolution, Journal of Geodesy 74 (2000), 421-430

Wang, J. and Satirapod, C. and Rizos, C.(2002): Stochastic assessment of GPS carrier phase measurements for precise static relative positioning, Journal of Geodesy, 76:95-104, 2002.

Wang, M. C. and G.E. Uhlenbeck (1945): On the theory of the Brownian motion II, Review of Modern Physics 17 (1945), 323-342

Wang, N., Lin, X. and R.G. Guttierrez (1998): Bias analysis and SIMEX approach in generalized linear mixed measurement error models, J. Amer. Statist. Ass., 93, (1998), 249-261

Wang, N., Lin, X. and R.G. Guttierrez (1999): A bias correction regression calibration approach in generalized linear mixed measurement error models, Commun.Statist. Theory Meth. 28 (1999), 217-232

Wang, Q-H. and B-Y. Jing (1999): Empirical likelihood for partial linear models with fixed designs, Statistics & Probability Letters 41 (1999), 425-433

Wang, S. and Uhlenbeck, G.I. (1945): On the theory of the Brownian motion II, Rev. Modern Phys. 17 (1945), 323-342

Wang, S.G. and Chow, S.C. (1993): Advanced linear models. Theory and applications. Marcel Dekker, Inc., New York, 1993.
Wang, S.G. and Ma, W.Q. (2002): On exact tests of linear hypothesis in linear models with nested error structure. J. Stat. Plann. Inference 106, 1-2 (2002), 225-233.
Wang, T. (1996): Cochran Theorems for multivariate components of variance models, Sankhya: The Indian Journal of Statistics A, 58 (1996), 238-342
Ward, P. (1996): GPS Satellite Signal Characteristics. In: Kaplan, E.D. (ed.): Understanding GPS-Principles and Applications. Artech House Publishers, BostonILondon, 1996
Wassel, S.R. (2002): Rediscovering a family of means, Mathematical Intelligencer 24 (2002), 58-65
Waterhouse, W.C. (1990): Gauss's first argument for least squares, Archive for the History of Exact Sciences 41 (1990), 41-52
Watson, G.S. (1983): Statistics on spheres, J. Wiley, New York 1983
Watson, G.S. (1956): Analysis of dispersion on a sphere, Monthly Notices of the Royal Astronomical Society, Geophysical Supplement 7 (1956), 153-159
Watson, G.S. (1956): A test for randomness of directions, Monthly Notices of the Royal Astronomical Society Geophysical Supplement 7 (1956), 160-161
Watson, G.S. (1960): More significance tests on the sphere, Biometrika 47 (1960), 87-91
Watson, G.S. (1961): Goodness-of-fit tests on a circle, Biometrika 48 (1961), 109-114
Watson, G.S. (1962): Goodness-of-fit tests on a circle-II, Biometrika 49 (1962), 57-63
Watson, G.S. (1964): Smooth regression analysis, Sankhya: The Indian Journal of Statistics: Series A (1964), 359-372
Watson, G.S. (1965): Equatorial distributions on a sphere, Biometrika 52 (1965), 193-201
Watson, G.S. (1966): Statistics of orientation data, Jour. of Geology 74 (1966), 786-797
Watson, G.S. (1967): Another test for the uniformity of a circular distribution, Biometrika 54 (1967), 675-677
Watson, G.S. (1967): Some problems in the statistics of directions, Bull. of I.S.I. 42 (1967), 374-385
Watson, G.S. (1968): Orientation statistic in the earth sciences, Bull of the Geological Institutions of the Univ. of Uppsala 2 (1968), 73-89
Watson, G.S. (1969): Density estimation by orthogonal series, Ann. Math. Stat. 40 (1969): Density estimation by orthogonal series, Ann. Math. Stat. 40 (1969), 1469-1498
Watson, G.S. (1970): The statistical treatment of orientation data, Geostatistics - a colloquium (ed. D.F. Merriam), Plenum Press, New York 1970, 1-10
Watson, G.S. (1974): Optimal invariant tests for uniformity, Studies in Probability and Statistics, Jerusalem Academic Press (1974), 121-128
Watson, G.S. (1981) : The Langevin distribution on high -dimensional spheres, J. Appl. Statis. 15 (1981) 123-130
Watson, G.S. (1982) : Asymptotic spectral analysis of cross - product matrices, Tech. Report Dep.Statist., Princeton University 1982
Watson, G.S: (1982): The estimation of palaeomagnetic pole positions, Statistics in Probability: Essay in hohor of C.R. Rao, North-Holland, Amsterdam and New York
Watson, G.S. (1982): Distributions on the circle and sphere, in: Gani, J. and E.J. Hannan (eds.): Essays in Statistical Science, Sheffield, Applied Probability Trust, 265-280 (J. of Applied Probability, special volume 19A)
Watson, G.S. (1983) : Large sample theory for distributions on the hypersphere with rotational symmetry, Ann. Inst.Statist. Math. 35 (1983) 303-319
Watson, G.S. (1983): Statistics on Spheres. Wiley, New York
Watson, G.S. (1984) : The theory of concentrated Langevin distributions, J. Multivariate Analysis 14 (1984) 74-82
Watson, G.S. (1986): Some estimation theory on the sphere, Ann. Inst. Statist. Math. 38 (1986), 263-275
Watson, G.S. (1987): The total approximation problem, in: Approximation theory IV, eds. Chui, C.K. et al., 723-728, Academic Press New York London 1987

Watson, G.S. (1988): The Langevin distribution on high dimensional spheres, J. Applied Statistics 15 (1988), 123-130
Watson, G. (1998): On the role of statistics in polomagnetic proof of continental drift, Canadian J. Statistics 26 (1998), 383-392
Watson, G.S. and E.J. Williams (1956): On the construction of significance tests on the circle and the sphere, Biometrika 43 (1956), 344-352
Watson, G.S. and E. Irving (1957): Statistical methods in rock magnetism, Monthly Notices Roy. Astro. Soc. 7 (1957), 290-300
Watson, G.S. and M.R. Leadbetter (1963): On the estimation of the probability density-I, Ann. Math. Stat 34 (1963), 480-491
Watson, G.S. and S. Wheeler (1964): A distribution-free two-sample test on a circle, Biometrika 51 (1964), 256
Watson, G.S. and R.J. Beran (1967): Testing a sequence of unit vectors for serial correlation, Jour. of Geophysical Research 72 (1967), 5655-5659
Watson, G.S., Epp, R. and J.W. Tukey (1971): Testing unit vectors for correlation, J. of Geophysical Research 76 (1971), 8480-8483
Webster, R., Atteia, O. and Dubois, J-P. (1994): Coregionalization of trace metals in the soil in the Swiss Jura. Eur J Soil Sci 45:205-218
Wedderburn, R. (1974): Quasi-likelihood functions, generalized linear models, and the Gauß-Newton method, Biometrica 61 (1974), 439-447
Wedderburn, R. (1976): On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models, Biometrika 63: 27-32
Wei, B.C. (1998): Exponential family: nonlinear models, Springer-Verlag, Heidelberg Berlin New York 1998
Wei, B.-C. (1998): Testing for varying dispersion in exponential family nonlinear models, Ann. Inst. Statist. Math. 50 (1998), 277-294
Wei, M. (1997): Equivalent formulae for the supremum and stability of weighted pseudoinverses, Mathematics of Computation 66 (1997), 1487-1508
Wei, M. (2000): Supremum and stability of weighted pseudoinverses and weighted least squares problems: analysis and computations, Nova Science Publisher, New York 2000
Wei, M. (2001): Supremum and stability of weighted pseudoinverses and weighted least squares problems analysis and computations, Nova Science Publishers, New York 2001
Wei, M. and A.R. De Pierro (2000): Upper perturbation bounds of weighted projections, weighted and constrained least squares problems, SIAM J. Matrix Anal. Appl. 21 (2000), 931-951
Wei, M. and A.R. De Pierro (2000): Some new properties of the equality constrained and weighted least squares problem, Linear Algebra and its Applications 320 (2000), 145-165
Weibull, M. (1953): The distribution of t- and F-statistics and of correlation and regression coefficients in stratified samples from normal populations with different means, Skand. Aktuarietidskr., 36, Supple., 1-106.
Weisberg, S. (1985): Applied Linear Regression, 2nd ed., Wiley, New York
Weiss, J. (1993): Resultant methods for the inverse kinematics problem, Computational Kinematics (Eds.) J. Angeles et al., Kluwer Academic Publishers, Netherlands 1993.
Weiss, G. and R. Rebarber (2000): Optimizability and estimability for infinite-dimensional linear systems, Siam J. Control Optim. 39 (2000), 1204-1232
Weisstein, E.W. (1999): Legendre Polynomial, CRC Press LLC, Wolfram Research Inc. 1999
Wellisch, S. (1910): Theorie und Praxis der Ausgleichungsrechnung Band 2: Probleme der Ausgleichungsrechnung, 46-49, Kaiserliche und königliche Hof-Buchdruckerei und Hof-Verlags-Buchhandlung, Carl Fromme, Wien und Leipzig 1910
Wellner, J. (1979): Permutation tests for directional data, Ann. Statist. 7 (1979), 924-943
Wells, D.E., Lindlohr, W., Schaffrin, B. and E. Grafarend (1987): GPS design: Undifferenced carrier beat phase observations and the fundamental difference theorem, University of New Brunswick, Surveying Engineering, Technical Report Nr. 116, 141 pages, Fredericton/Canada 1987

Welsch, W. (1982): Zur Beschreibung des homogenen Strains oder Einige Betrachtungen zur affinen Transformation, Zeitschrijt für Vermessungswesen 107(1982), Heft 5, 8.173-182

Welsch, W., Schlemmer, H. and Lang, M., (Hrsg.) (1992): Geodätische Me:Bverfahren im Maschinenbau. Schriftenreihe des Deutschen Vereins fiir Vermessungswesen, Verlag Konrad Wittwer, Stuttgart.

Welsh, A.H. (1987): The trimmed mean in the linear model, Ann. Statist. 15 (1987), pp. 20-36

Welsh, A.H. (1996): Aspects of statistical inference, J. Wiley, New York 1996

Wenzel, H.G. (1977): Zur Optimierung von Schwerenetzen, Z. Vermessungswesen 102 (1977), 452-457

Werkmeister, P. (1916): Graphische Ausgleichung bei trigonometrischer Punktbestimmung durch Einschneiden, Zeitschrift für Vermessungswesen 45 (1916), 113-126

Werkmeister, P. (1916): Trigonometrische Punktbestimmung durch einfaches Einschneiden mit Hilfe von Vertikalwinkeln, *Zeitschrift für Vermessungswesen* 45 (1916) 248-251.

Werkmeister, P. (1920): Über die Genauigkeit trigonometrischer Punktbestimmungen, *Zeitschrift für Vermessungswesen* 49 (1920) 401-412, 433-456.

Werner, D. (1913): Punktbestimmung durch Vertikalwinkelmessung, *Zeitschrift für Vermessungswesen* 42 (1913) 241-253.

Werner, H.J. (1985): More on BL U estimation in regression models with possibly singular covariances. Linear Algebra Appl. 67 (1985), 207-214.

Werner, H.J. and Yapar, C. (1995): More on partitioned possibly restricted linear regression. In: New trends in probability and statistics. Multivariate statistics and matrices in statistics. (Proceed. 5th Conf. Tartu - Plihajarve, Estonia, May 23-28, 1994, E.-M. Tiit et al., ed.), Vol. 3, 1995, Utrecht, pp. 57-66.

Werner, H.J. and Yapar, C. (1996): A BLUE decomposition in the general linear regression model. Linear Algebra Appl. 237/238 (1996), 395-404.

Wernstedt, J. (1989): Experimentelle Prozeßanalyse, Oldenbourg Verlag, München 1989

Wess, J. (1960): The conformal invariance in quantum field theory, in: Il Nuovo Cimento, Nicola Zanichelli (Hrsg.), Vol. 18, Bologna 1960

Wetzel, W., Jöhnk, M.D. and P. Naeve (1967): Statistische Tabellen, de Gruyter, Berlin 1977

Winkler, F. (1996): A polynomial algorithm in computer algebra, Springer-Verlag, Wien 1996.

Whittaker, E.T. and G. Robinson (1924): The calculus of observations, Blackie, London 1924

Whittle, P. (1952): On principal components and least square methods of factor analysis, Skand. Aktuarietidskr., 35, 223-239.

Whittle, P. (1954): On Stationary Processes in the Plane. Biometrika 41, 434-449.

Whittle, P. (1963): Prediction and regulation, D. van Nostrand Co., Inc., Princeton 1963

Whittle, P. (1963): Stochastic processes in several dimensions, Bull. Inst. Int. Statist. 40 (1963), 974-994

Whittle, P. (1973): Some general points in the theory of optimal experimental design, J. Roy. Statist.-Soc. B35 (1973), 123-130

Wickerhauser, M.V. (1996): Adaptive Wavelet-Analysis, Theorie und Software, Vieweg & Sohn Verlag, Braunschweig/Wiesbaden 1996

Wiener, N. (1923): Differential Space, J. Math. and Phys. 2 (1923) 131-174

Wiener, N. (1949): Extrapolation, interpolation and smoothing of stationary time series, New York 1949

Wiener, N. (1958): Non linear problems in random theory, MIT - Press 1958

Wiener, N. (1964): Selected papers of N. Wiener, MIT - Press 1964

Wiener, N. (1968): Kybernetik, rororo-Taschenbuch, Hamburg 1968

Wieser, A.; Brunner, F.K. (2001): An Extended Weight Model for GPS Phase Observations. Earth, Planets and Spac 52:777-782

Wieser, A.; Brunner, F.K. (2001): Robust estimation applied to correlated GPS phase observations. Presented at First International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS, Zurich, Switzerland, 2001

Wieser, A. (2002): Robust and fuzzy techniques for parameter estimation and quality assessment in GPS. Reihe 'Ingenieurgeodasie - TV Graz', J- Shaker Verlag, Aachen (English)

Wieser, A. and Gaggl, M. and Hartinger, H. (2005): Improved positioning accuracy with high sensitivity GNSS receivers and SNR aided integrity monitoring of pseudo-range observations. In: Proc ION GNSS 2005, September 13-16, Long Beach, CA: 1545-1554

Wigner, E.P. (1958): On the distribution of the roots of certain symmetric matrices, Ann. Math. 67 (1958)

Wilcox, R.R. (1997): Introduction to robust estimation and hypothesis testing, Academic Press, San Diego 1997

Wilcox, R.R. (2001): Fundamentals of modern statistical methods, Springer-Verlag, Heidelberg Berlin New York 2001

Wilders, P. and E. Brakkee (1999): Schwarz and Schur: an algebraical note on equivalence properties, SIAM J. Sci. Comput. 20 (1999), 2297-2303

Wilkinson, J. (1965): The algebraic eigenvalue problem, Clarendon Press, Oxford 1965

Wilks, S.S. (1932a): Certain generalizations in the analysis of variance, Biometrika 24, 471-494

Wilks, S.S. (1932b): On the sampling distribution of the multiple correlation coefficient, Ann. Math. Stat., 3, 196-203

Wilks, S.S. (1932c): Moments and distributions of estimates of population parameters from fragmentary samples, Annals of Mathematical Statistics 3: 163-195

Wilks, S.S. (1934): Moment-generating operators for determinants of product moments in samples from a normal system, Ann. of Math 35, 312-340

Wilks, S.S. (1962): Mathematical statistics, J. Wiley, New York 1962

Williams, E.J. (1963): A comparison of the direct and fiducial arguments in the estimation of a parameter, J. Roy. Statist. Soc. B25 (1963), 95-99

Wimmer, G. (1995): Properly recorded estimate and confidence regions obtained by an approximate covariance operator in a special nonlinear model, Applications of Mathematics 40 (1995), 411-431

Wimmer, H. (1981a): Ein Beitrag zur Gewichtsoptimierung geodätischer Netze, Deutsche Geodätische Kommission, München, Reihe C (1981), 269

Wimmer, H. (1981b): Second-order design of geodetic networks by an iterative approximation of a given criterion matrix, in: Proc. of the IAG Symposium on geodetic networks and computations, R. Sigle ed., Deutsche Geodätische Kommission, München, Reihe B, Nr. 258 (1981), Heft Nr. III, 112-127

Wimmer, H. (1982): Ein Beitrag zur gewitchs optimierung geodetischer Netu Report C 269, Deutsche Geodestiche Kommission, Bayerische Akademic der Wissenschaften, München 1982

Wishart, J. (1928): The generalized product moment distribution in samples from a normal multivariate population, Biometrika 20 (1928), 32-52

Wishart, J. (1931): The mean, and second moment coefficient of the multiple correlation coefficient in samples from a normal population, Biometrika, 22, 353-361.

Wishart, J. (1933): The generalized product moment distribution in a normal system, Proc. Camb. Phil. Soc., 29, 260-270

Wishart, J. (1948a): Proofs of the distribution law of the second order moment statistics, Biometrika, 35, 55-57.

Wishart, J. (1948b): Test of homogeneity of regression coefficients, and its application in the analysis of covariance, presented to the Colloque International de Calcul des Probabilites et de Statistique Mathematique, Lyon.

Wishart, J. (1955): Multivariate analysis, App. Stat., 4, 103-116

Wishart, J. and M.S. Bartlett (1932): The distribution of second order moment statistics in a normal system, Proc. Camb. Phil. Soc., 28, 455-459.

Wishner, R., Tabaczynski, J. and M. Athans (1969): A comparison of three non-linear filters, Automatica 5 (1969), 487-496

Witkovsky, V. (1996): On variance-covariance components estimation in linear models with AR(1) disturbances. Acta Math. Univ. Comen., New Ser. 65, 1 (1996), 129-139.

Witkovsky, V. (1998): Modified minimax quadratic estimation of variance components, Kybernetika 34 (1998), 535-543

Witting, H. and G. Nölle (1970): Angewandte Mathematische Statistik, Teubner Verlag, Stuttgart 1970
Wold, S., Wold, H.O., Dunn, W.J. and Ruhe, A. (1984): The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, SIAM Journal on Scientific and Statistical Computing 5: 735-743
Wolf, H. (1961): Das Fehlerfortpflanzungsgesetz mit Guedern II, ZfV
Wolf, H. (1968): Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, Ferdinand Dümmlers Verlag, Bonn 1968
Wolf, H. (1973): Die Helmert-Inverse bei freien geodätischen Netzen, Z. Vermessungswesen 98 (1973), 396-398
Wolf, H. (1975): Ausgleichungsrechnung I, Formeln zur praktischen Anwendung, Duemmler, Bonn 1975
Wolf, H. (1976): The Helmert block method, its origin and development, Proc. 2nd Int. Symp. on problems related to the Redefinition of North American Geodetic Networks, 319-326, Arlington 1976
Wolf, H. (1977): Eindeutige und mehrdeutige Geodätische Netze. Adh. Braunschw. Wiss. Ges., Band 28, G6ttingen, FRG, 1977.
Wolf, H. (1980a): Ausgleichungsrechnung II, Aufgaben und Beispiel zur praktischen Anwendung, Duemmler, Bonn 1980
Wolf, H. (1980b): Hypothesentests im Gauß-Helmert-Modell, Allg. Vermessungsnachrichten 87 (1980), 277-284
Wolf, H. (1997): Ausgleichungsrechnung I und II, 3. Au age, Ferdinand Dümmler Verlag, Bonn 1997
Wolkowicz, H. and G.P.H. Styan (1980): More bounds for eigenvalues using traces, Linear Algebra Appl. 31 (1980), 1-17
Wolter, K.H. and Fuller, W.A. (1982), Estimation of the quadratic errors-in-variables model, Biometrika 69 (1982), 175-182
Wong, C.S. (1993): Linear models in a general parametric form, Sankhya 55 (1993), 130-149
Wong, W.K. (1992): A unified approach to the construction of minimax designs, Biometrika 79 (1992), 611-620
Wood, A. (1982): A bimodal distribution for the sphere, Applied Statistics 31 (1982), 52-58
Woolson, R.F. and W.R. Clarke (1984): Analysis of categorical incomplete data, J. Roy. Statist. Soc. A147 (1984), 87-99
Worbs, E. (1955): Carl Friedrich Gauß, ein Lebensbild, Leipzig 1955
Wu, C.F. (1986): Jackknife, bootstrap and other resampling methods in regression analysis, Annals of Statistics 14: 1261-1350
Wu, C.F. (1980): On some ordering properties of the generalized inverses of non negative definite matrices, Linear Algebra Appl., 32 (1980), pp. 49-60
Wu, C.F.J. (1981): Asymptotic theory of nonlinear least squares estimation, Ann. Stat. 9 (1981), 501-513
Wu, I.T., Wu, S.C., Hajj, G.A., Bertiger, W.I. and Lichten, S.M. (1993): Effects of antenna orientation on GPS carrier phase. Manuscripta Geodaetica 18: 91-98
Wu, Q. and Z. Jiang (1997): The existence of the uniformly minimum risk equivariant estimator in Sure model, Commun. Statist. - Theory Meth. 26 (1997), 113-128
Wu, W. (1984): On the decision problem and mechanization of the theorem proving elementary geometry, *Scientia Sinica* 21 (1984) 150-172.
Wu, Y. (1988): Strong consistency and exponential rate of the “minimum L1-norm” estimates in linear regression models, Computational Statistics and Data Analysis 6: 285-295
Wunsch, G. (1986): Handbuch der Systemtheorie, Oldenbourg Verlag, München 1986
Wyss, M., Liang, B.Y., Tanagawa, W. and Wu, X. (1992): Comparison of orientations of stress and strain tensors based on fault plane solutions in Kaoiki, Hawaii, J. geophys. Res., 97, 4769-4790
Xi, Z. (1993): Iterated Tikhonov regularization for linear ill-posed problems, Ph.D. Thesis, Universität Kaiserslautern, Kaiserslautern 1993

Xu, C., Ding, K., Cai, J. and Grafarend, E. (2009): Method for determining weight scale factor parameter in joint inverse problem of geodesy, Journal of Geodynamics, Vol. 47, 39-46. doi:10.1016/j.jog.2008.06.005
Xu, P. and Rummel, R. (1994): Generalized ridge regression with applications in determination of potential fields, Manuscripta. Geodetica, 20, pp. 8-20
Xu, P.L. (1986): Variance-covariance propagation for a nonlinear function, J. Wuhan Techn. UnL Surv. Mapping, 11, No. 2, 92-99
Xu, P.L. (1989a): On the combined adjustment of gravity and levelling data, Ph.D. diss., Wuhan Tech. Univ. Surv. Mapping
Xu, P.L. (1989b): On robust estimation with correlated observations, Bull. Gèeodesique 63 (1989), 237-252
Xu, P.L. (1989c): Statistical criteria for robust methods, ITC Journal, 1 (1989), pp. 37-40
Xu, P.L. (1991): Least squares collocation with incorrect prior information, Z. Vermessungswesen 116 (1991), 266-273
Xu, P.L. (1992): The value of minimum norm estimation of geopotential fields, Geophys. J. Int. 111 (1992), 170-178
Xu, P.L. (1993): Consequences of Constant Parameters and Confidence Intervals of Robust Estimation, Boll. Geod. Sci. Affini, 52, pp. 231-249
Xu, P.L. (1995): Testing the hypotheses of non-estimable functions in free net adjustment models, manuscripta geodaetica 20 (1995), 73-81
Xu, P.L. (1996): Statistical aspects of 2-D symmetric random tensors, Proc. FIG int. Symp.
Xu, P.L. (1999a): Biases and accuracy of, and an alternative to, discrete nonlinear filters, Journal of Geodesy 73 (1999), 35-46
Xu, P.L. (1999b): Spectral theory of constrained second-rank symmetric random tensors, Geophys. J. Int. 138 (1999), 1-24
Xu, P.L. (2001): Random simulation and GPS decorrelation, Journal of Geodesy 75 (2001), 408-423
Xu, P.L. (2002a): Isotropic probabilistic models for directions, planes and referential systems, Proc. Royal Soc. London A458 (2002), 2017-2038
Xu, P.L. (2002b): A Hybrid Global Optimization Method: One-dimensional Case , J. comput. appl. Math., 147, pp. 301-314.
Xu, P.L. (2003): A Hybrid Global Optimization Method: The Multi-dimensional Case, J. comput. appl. Math., 155, pp. 423-446
Xu, P.L. (2005): Sign-constrained Robust Least Squares, Subjective Breakdown Point and the Effect of Weights of Observations, J. of Geodesy, 79 (2005), pp. 146-159
Xu, P.L. (2006) : Voronoi cells, probabilistic bounds and hypothesis testing in mixed integer linear models, IEEE Transactions on Information Theory 52 (2006) 3122-3138
Xu, P.L. and E.W. Grafarend (1996a): Statistics and geometry of the eigenspectra of three-dimensional second-rank symmetric random tensors, Geophysical Journal International 127 (1996), 744-756
Xu, P.L. and E.W. Grafarend (1996b): Probability distribution of eigenspectra and eigendirections of a twodimensional, symmetric rank two random tensor, Journal of Geodesy 70 (1996), 419-430
Xu, P.L., Shen, Y., Fukuda, Y. and Liu, Y. (2006): Variance component estimation in linear inverse ellipsoid models, J. Geodesy 80 (2006) 69-81
Xu, P.L., Liu, Y., Shen, Y. and Fukuda, Y. (2007): Estimability analysis of variance ad covariance components. J. Geodesy 81 (2007) 593-602
Xu, Z.H., Wang, S.Y., Huang, Y.R. and Gao, A. (1992): Tectonic stress field of China inferred from a large number of small earthquakes, J. geophys. Res., B97, 11867-11877
Yadrenko, M.I. (1987): Correlation theory of stationary and related random functions, Vol I, Basic results, Springer Verlag, New York 1987
Yaglom, A.M. (1961): Second-order homogeneous random fields in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 593-622, University of California Press, Berkeley 1961

Yaglom, A.M. (1962): An introduction to the theory of stationary random functions. Dover New York, 1962.
Yaglom, A.M. (1987): Correlation theory of stationary and related random functions. Springer, New York
Yakhot, V. et. al (1989): Space-time correlations in turbulence: kinematical versus dynamical effects, Phys. Fluids A1 (1989) 184-186
Yancey, T.A., Judge, G.G., and Bock, M.E. (1974): A mean square error test when stochastic restrictions are used in regression, Communications in Statistics, Part A-Theory and Methods 3: 755-768
Yang, H. (1996): Efficiency matrix and the partial ordering of estimate, Commun. Statist. - Theory Meth. 25(2) (1996), 457-468
Yang, Y. (1994): Robust estimation for dependent observations, Manuscripta geodaetica 19 (1994), 10-17
Yang, Y. (1999): Robust estimation of geodetic datum transformation, Journal of Geodesy 73 (1999), 268-274
Yang, Y. (1999): Robust estimation of systematic errors of satellite laser range, Journal of Geodesy 73 (1999), 345-349
Yang, Y., Song, L. and Xu, T. (2001): Robust Estimator for the Adjustment of Correlated GPS Networks. Presented at First International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS, Zurich, Switzerland, 2001.
Yang, Y., Song, L. and Xu, T. (2002): Robust parameter estimation for geodetic correlated observations. Acta Geodaetica et Cartographica Sinica, 31(2), pp. 95-99
Yazji, S. (1998): The effect of the characteristic distance of the correlation function on the optimal design of geodetic networks, Acta Geod. Geoph. Hung. 33 (2-4) (1998), 215-234
Ye, Y. (1997): Interior point algorithms: Theory and analysis, J. Wiley, New York 1997
Yeh, A.B. (1998): A bootstrap procedure in linear regression with nonstationary errors, The Canadian J. Stat. 26 (1998), 149-160
Yeung, M.-C. and T.F. Chan (1997): Probabilistic analysis of Gaussian elimination without pivoting, SIAM J. Matrix Anal. Appl. 18 (1997), 499-517
Ylvisaker, D. (1977): Test resistance, J. Am. Statist. Ass. 72, 551-557, 1977
Yohai, V.J. (1987): High breakdown point and high efficiency robust estimates for regression, Ann. Stat., 15 (1987), pp. 642-656
Yohai, V.J. and R.H. Zamar (1997): Optimal locally robust M-estimates of regression, J. Statist. Planning and Inference 64 (1997), 309-323
Yor, M. (1992): Some aspects of Brownian motion, Part I: Some special functionals, Birkhäuser-Verlag, Basel Boston Berlin 1992
Yor, M. (1992): Some aspects of Brownian motion, Part II: Some recent martingale problems, Birkhäuser-Verlag, Basel Boston Berlin 1997
You, R.J.(2000): Transformation of Cartesian to geodetic coordinates without iterations, *Journal of Surveying Engineering* 126 (2000) 1-7.
Youssef, A.H.A. (1998): Coefficient of determination for random regression model, Egypt. Statist. J. 42 (1998), 188-196
Yu, Z.C. (1992): A generalization theory of estimation of variance-covariance components, Manusc. Geodetica, 17, 295-301
Yu, Z.C. (1996): A universal formula of maximum likelihood estimation of variance-covariance components, Journal of Geodesy 70 (1996), 233-240
Yu. Z. and M. Li (1996): Simultaneous location and evaluation of multi-dimensional gross errors, Wuhan Tech. University of Surveying and mapping, Science reports, 1, (1996), pp. 94-103
Yuan, K.H. and P.M. Bentler (1997): Mean and covariance structure analysis - theoretical and practical improvements, J. Am. Statist. Ass. 92 (1997), 767-774
Yuan, K.H. and P.M. Bentler (1998): Robust mean and covariance structure analysis, British Journal of Mathematical and Statistical Psychology (1998), 63-88

Yuan, Y. (1999): On the truncated conjugate gradient method, Springer-Verlag, Heidelberg Berlin New York

Yuan, Y. (2000): On the truncated conjugate gradient method, Math. Prog. 87 (2000), 561-571

Yunck, T.P. (1993): Coping with the Atmosphere and Ionosphere in Precise Satellite and Ground Positioning In: Environmental Effects on Spacecraft Positioning and Trajectories, Geophysical Monograph 73, WGG, Vol l3, pp 1-16

Yusuf, S., Peto, R., Lewis, J., Collins, R. and P. Sleight (1985): Beta blockade during and after myocardial infarction: An overview of the randomized trials, Progress in Cardiovascular Diseases 27 (1985), 335-371

Zabell, S. (1992): R.A. Fisher and the fiducial argument, Statistical Science 7 (1992), 369-387

Zackin, R., de Gruttola, V. and N. Laird (1996): Nonparametric mixed-effects for repeated binary data arising in serial dilution assays: Application to estimating viral burden in AIDS, J. Am. Statist. Ass. 91 (1996), 52-61

Zacks, S. (1971): The theory of statistical inference, J. Wiley, New York 1971

Zadeh, L. (1965): Fuzzy sets, Information and Control 8 (1965), 338-353

Zaiser, J. (1986): Begründung Beobachtungsgleichungen und Ergebnisse für dreidimensionales geometrisch-physikalishes Modell der Geodäsie, *Zeitschrift für Vermessungswesen* 111 (1986) 190-196.

Zavoti J. and T. Jansco (2006): The solution of the 7 parameter Datum transformation problem with and withour the Gröbner basis, Acta. Geod. Geophy. Hung. 41 (2006) 87-100

Zèavoti, J. (1999): Modified versions of estimates based on least squares and minimum norm, Acta Geod. Geoph. Hung. 34 (1999), 79-86

Zehfuss, G. (1858): Über eine gewisse Determinante, Zeitschrift für Mathematik und Physik 3 (1858), 298-301

Zellner, A. (1962): An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 57: 348-368

Zellner, A. (1971): An introduction to Bayesian inference in econometrics, J. Wiley, New York 1971

Zha, H. (1995): Comments on large least squares problems involving Kronecker products, SIAM J. Matrix Anal. Appl. 16 (1995), 1172

Zhan, X. (2000): Singular values of differences of positive semi-definite matrices, Siam J. Matrix Anal. Appl. 22 (2000), 819-823

Zhang, B.X., Liu, B.S. and Lu, C.Y. (2004): A study of the equivalence of the BLUEs between a partitioned singular linear model and its reduced singular linear models. Acta Math. Sin., Engl. Ser. 20, 3 (2004), 557-568.

Zhang, Y. (1985): The exact distribution of the Moore-Penrose inverse of X with a density, in: Multivariate Analysis VI, Krishnaiah, P.R. (ed), 633-635, Elsevier, New York 1985

Zhang, S. (1994): Anwendung der Drehmatrix in Hamilton normierten Quaternionenen bei der Bündelblock Ausgleichung, *Zeitschrift für Vermessungswesen* 119 (1994) 203-211.

Zhang, J., Zhang, K. and Grenfell, R. (2004a): On the relativistic Doppler Effects and high accuracy velocity estimation using GPS. In Proc The 2004 International Symposium on GNSSIGPS, December 6-8, Sydney, Australia

Zhang, J., Zhang, K., Grenfell, R., Li, Y. and Deakin, R. (2004b): Real-Time DopplerlDoppler Rate Derivation for Dynamic Applications. In Proc The 2004 International Symposium on GNSSIGPS, December 6-8, Sydney, Australia

Zhang, J., Zhang, K., Grenfell, R. and Deakin, R. (2006a): GPS satellite velocity and acceleration Determination using the Broadcast Ephemeris. The Journal of Navigation 59: 293-305

Zhang, J., Zhang, K., Grenfell, R. and Deakin, R. (2006b): Short note: On the relativistic Doppler effect for precise velocity determination using GPS. Journal of Geodesy 80: 104-110

Zhang, J.Z., Chen, L.H. and N.Y. Deng (2000): A family of scaled factorized Broyden-like methods for nonlinear least squares problems, SIAM J. Optim. 10 (2000), 1163-1179

Zhang, Z. and Y. Huang (2003): A projection method for least squares problems with a quadratic equality constraint, SIAM J. Matrix Anal. Appl. 25 (2003), 188-212

Zhao, L.P., and Prentice, R.L. (1990): Correlated binary regression using a generalized quadratic model, Biometrika 77: 642-648

Zhao, L.P., Prentice, R.L., and Self, S.G. (1992): Multivariate mean parameter estimation by using a partly exponential model, Journal of the Royal Statistical Society, Series B 54: 805-811

Zhao, Y. and S. Konishi (1997): Limit distributions of multivariate kurtosis and moments under Watson rotational symmetric distributions, Statistics & Probability Letters 32 (1997), 291-299

Zhdanov, M.S. (2002): Geophysical inverse theory and regularization problems, Methods in Geochemistry and Geophysics 36, Elsevier, Amsterdam Boston London 2002

Zhenhua, X. (1993): Iterated Tikhonov regularization for linear ill-posed problem, Ph.D.Thesis University of Kaiserslautern, Kaiserslautern 1993

Zhen-Su, S., Jackson, E. and Orszag, S. (1990): Intermittency of turbulence, in: The Legacy of John von Neumann (J. Glimm, J. Impagliazzo, I. Singer eds.) Proc. Symp. Pure Mathematics, Vol. 50, 197-211, American Mathematical Society, Providence, Rhode Island 1990

Zhou, J. (2001): Two robust design approaches for linear models with correlated errors, Statistica Sinica 11 (2001), 261-272

Zhou, K.Q. and S.L. Portnoy (1998): Statistical inference on hetero-skedastic models based on regression quantiles, J. Nonparametric Statistics 9 (1998), 239-260

Zhou, L.P. and Mathew, T. (1993): Combining independent tests in linear models. J. Am. Stat. Assoc. 88, 422 (1993), 650-655.

Zhu, J. (1996): Robustness and the robust estimate, Journal of Geodesy 70 (1996), 586-590

Zhu, Z. and Stein, M.L. (2006): Spatial sampling design for prediction with estimated parameters. J Agricult Biol Environ Stat 11:24-49

Ziegler, A., Kastner, C. and M. Blettner (1998): The generalised estimating equations: an annotated bibliography, Biometrical Journal 40 (1998), 115-139

Zimmermann, H.-J. (1991): Fuzzy set theory and its applications, 2nd ed., Kluwer Academic Publishers, Dordrecht 1991

Zippel, R. (1993): Effective polynomial computation, Kluwer Academic Publications, Boston 1993.

Zolotarev, V.M. (1997): Modern theory of summation of random variables, VSP, Utrecht 1997

Zmyslony R. (1980): A characterization of best linear unbiased estimators in the general linear model. In: Mathematical statistics and probability theory. (Proceed. 6th Int. Conf., Wisla, Poland, 1978, Lect. Notes Stat. 2), 1980, pp. 365-373.

Zoback, M.D. (1992a): Introduction to special section on the Cajon Pass scientific drilling project, J. geophys. Res., B97, 4991-4994

Zoback, M.D. (1992b): First and second-order patterns of stress in the lithosphere: the World Stress Map Project, J. geophys. Res., B97, 11703-11728

Zoback, M.D. (1992c): Stress field constraints on intraplate seismicityin eastern North America, J. geophys. Res., B97, 11761-11782

Zoback, M.D. and Healy, J. (1992): In-situ stress measurements to 3.5 km depth in the Cajon Pass scientific research borehole: Implications for the mechanics of crustal faulting, J. geophys. Res., B97, 5039-5057

Zumberge, J.F. and Bertiger, W.I. (1996): Ephemeris and Clock Navigation Message Accuracy. In: Parkinson B.W., Spilker J.J. (eds) Global Positioning System: Theory and Applications, Volume 1, American Institute of Aeronautics and Astronautics, Washington DC, pp. 585-599

Zurmühl, R. and S. Falk (1984): Matrizen und ihre Anwendungen, Teil 1: Grundlagen, 5.ed., Springer-Verlag, Heidelberg Berlin New York 1984

Zurmühl, R. and S. Falk (1986): Matrizen und ihre Anwendungen, Teil 2: Numerische Methoden, 5.ed., Springer-Verlag, Heidelberg Berlin New York 1986

Zvara, K. (1989): Regression Analysis. Academia, Praha, 1989 (in Czech).

Zwanzig, S. (1980): The choice of approximative models in nonlinear regression, Statistics, Vol. 11, No. 1, 23-47.

Zwanzig, S. (1983): A note to the paper of Ivanov and Zwanzig on the asymptotic expansion of the least squares estimator. Statistics, Vol. 14, No. 1, 29-32.

Zwanzig, S. (1985): A third order comparison of least squares, jacknifing and cross validation for error variance estimation in nonlinear regression. Statistics, VoL 16, No. 1, 47-54.

Zwanzig, S. (1987): On the consistency of the least squares estimator in nonlinear errors-in-variables models, Folia Oeconomia, Acta Universitatis Lodziensis. 8 poo

Zwanzig, S. (1989): On an asymptotic minimax result in nonlinear errors-in-variables models, Proceedings of the Fourth Prague Symposium on Asymptotic Statistics, Prague 1989, 549-558.

Zwanzig, S. (1990): Discussion to A. Pazman: Small-sample distributional properties of nonlinear regression estimator (a geometric approach). Statistics Vol. 21, No. 3, 361-363.

Zwanzig, S. (1991): Least squares estimation in nonlinear functional relations. Proceedings Probastat 91, Bratislava, 171-177.

Zwanzig, S. (1994): On adaptive estimation in nonlinear regression. Kybemetika, Vol. 30, No. 3, 359-367.

Zwanzig, S. (1997): On L1-norm estimators in nonlinear regression and in nonlinear error-in-variables models, in L1-Statistical Procedures and Related Topics IMS Lecture Notes-Monographs Series. Vol 31, 1997, 101-118.

Zwanzig, S. (1998a): Application of Hipparcos dates: A new statistical method for star reduction. Proceedings of Journees 1997, Academy of Sciences of the Czech Republic, Astronomical Institute, 142-144.

Zwanzig, S. (1998b): Experimental design for nonlinear error-in-variables models,Proceedings of the Third St.-Petersburg Workshop on Simulation, 225-230.

Zwanzig, S. (1999): How to adjust nonparametric regression estimators to models with errors in the variables, Theory of Stochastic Processes. 5 (21) 272-277.

Zwanzig, S. (2000): Experimental design for nonlinear error-in-variables models, in Advanced in Stochastc Simulation Methods Ed. Balakrishan, Melas, Ermakov

Zwanzig, S. (2007): Local linear estimation in nonparametric errors-in-variables models. Theory of Stochastic Processes, VoL 13, 316-327

Zwet van, W.R. and J. Osterhoff (1967): On the combination of independent test statistics, Ann. Math. Statist. 38 (1967), 659-680

Zwinck, E. (ed) (1988): Christian Doppler- Leben und Werk; Der Doppler effekt. Schriftenreihe des Landespressebiiros / Presse- und Informationszentrum des Bundeslandes Salzburg, Sonderpublikationen 76, 140p (in German)

Zyskind, G. (1967): On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. Ann. Math. Stat. 38 (1967), 1092-1109.

Zyskind, G. (1969): On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure, SIAM J. Appl. Math. 17 (1969), 1190-1202

Zyskind, G. (1975): Error structures. Projections and conditional inverses in linear model theory. In A survey of statistical design and linear models, J.W. Srivastava (ed.), pp. 647-663, North Holland, Amsterdam 1975

Index

E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models*,
Springer Geophysics, DOI 10.1007/978-3-642-22241-2,

© Springer-Verlag Berlin Heidelberg 2012