Lecture Notes in Earth Sciences 95

Athanasios Dermanis Armin Grün
Fernando Sansò (Eds.)

Geomatic Methods for the Analysis of Data in the Earth Sciences

Springer

Lecture Notes in Earth Sciences 95

Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Athanasios Dermanis Armin Grün
Fernando Sansò (Eds.)

Geomatic Methods for the Analysis of Data in the Earth Sciences

With 64 Figures

Springer

Editors

Professor Dr. Athanasios Dermanis
The Aristotle University of Thessaloniki
Department of Geodesy and Surveying
University Box, 503
54006 Thessaloniki, Greece

E-mail: *dermanis@topo.auth.gr*

Professor Fernando Sansò
Politecnico di Milano
Dipartimento di Ingegneria Idraulica,
Ambientale e del Rilevamento
Piazza Leonardo da Vinci, 32
20133 Milano, Italy

E-mail: *fsanso@ipmtf4.topo.polimi.it*

Professor Dr. Armin Grün
ETH Hönggerberg
Institute of Geodesy and Photogrammetry
Hil D 47.2
8093 Zurich, Switzerland

E-mail: *agruen@geod.ethz.ch*

"For all Lecture Notes in Earth Sciences published till now please see final pages of the book"

Library of Congress Cataloging-in-Publication Data
Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Geomatic methods for the analysis of data in the earth sciences /
Athanasios Dermanis ... (ed.). - Berlin; Heidelberg; New York; Barcelona;
Hong Kong; London; Milan; Paris; Singapore; Tokyo: Springer, 2000
(Lecture notes in earth sciences; 95)
ISBN 3-540-67476-4

ISSN 0930-0317
ISBN 3-540-67476-4 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera ready by editors
Printed on acid-free paper SPIN: 10768074 32/3130-543210

PREFACE

There has been a time when statistical modeling of observation equations was clear in disciplines like geodesy, geophysics, photogrammetry and practically always based on the conceptual arsenal of least square theory despite the different physical realities and laws involved in their respective observations.

A little number of (very precise) observations and an even smaller number of parameters to model physical and geometrical laws behind experimental reality, have allowed the development of a neat line of thought where "errors" were the only stochastic variables in the model, while parameters were deterministic quantities related only to the averages of the observables. The only difficulty there was to make a global description of the manifold of mean values which could, as a whole, be a very complicated object on which finding the absolute minimum of the quadratic functional[1] could be a difficult task, for a general vector of observations. This point however was mostly theoretical, since the accuracy of observations and the strong belief in the deterministic model were such that only a very small part of the manifold of the means was really interested in the minimization process and typically the non-linearity played a minor part in that.

The enormous increase of available data with the electronic and automatic instrumentation, the possibility of expanding our computations in the number of data and velocity of calculations (a revolution which hasn't yet seen a moment of rest) the need of fully including unknown fields (i.e. objects with infinitely many degrees of freedom) among the "parameters" to be estimated have reversed the previous point of view. First of all any practical problem with an infinite number of degree of freedom is underdetermined; second, the discrepancy between observations and average model is not a simple noise but it is the model itself that becomes random; third, the model is refined to a point that also factors weakly influencing the observables are included, with the result that the inverse mapping is unstable. All these factors have urged scientists in these disciplines to overcome the bounds of least squares theory (namely the idea of "minimizing" the discrepancies between observations and one specific model with a smaller number of parameters) adopting (relatively) new techniques like Tikhonov regularization, Bayesian theory, stochastic optimization and random fields theory to treat their data and analyze their models.

Of course the various approaches have been guided by the nature of the fields analyzed and the physical laws underlying the measurements in different disciplines (e.g. the field of elastic waves in relation to the elastic parameters and their discontinuities in the earth, the gravity field in relation to the earth mass density and the field of gray densities and its discontinuities within digital images of the earth in relation to the earth's surface and its natural or man-made coverage).

So, for instance, in seismology, where 1% or even 10% of relative accuracy is acceptable, the idea of random models/parameters is widely accepted and conjugated with other methods for highly non-linear phenomena, as the physics of elastic wave propa-

---

[1] Note that in least squares theory the target function is quadratic in the mean vector, not in the parameter vector.

gation in complex objects like the earth dictates. In geodesy deterministic and stochastic regularization of the gravity field is used since long time while non-linearity is typically dealt with in a very simple way, due to the substantial smoothness of this field; in image analysis, on the contrary, the discontinuities of the field are even more important than the continuous "blobs", however these can be detected with non-convex optimization techniques, some of which are stochastic and lead naturally to a Bayesian interpolation of the field of gray densities as a Markov random field.

The origin of the lecture notes presented here, is the IAG International Summer School on "Data Analysis and the Statistical Foundations of Geomatics", which took place in Chania, Greece, 25-30 May 1998 and was jointly sponsored by the International Association of Geodesy and the International Society of Photogrammetry and Remote Sensing. According to the responses of the attendees (who were asked to fill a questionnaire) the School has been a great success from both the academic and organizational point of view. In addition to the above mentioned scientific organizations we would also like to thank those who contributed in various ways: The Department of Geodesy and Surveying of The Aristotle University of Thessaloniki, the Department of Mineral Resources Engineering of Technical University of Crete, the Mediterranean Agronomic Institute of Chania, in the premises of which the school took place, the excellent teachers, the organizing committee and especially Prof. Stelios Mertikas who took care of the local organization.

This school represents a first attempt to put problems and methods developed in different areas one in front of the other, so that people working in various disciplines could get acquainted with all these subjects. The scope is to attempt tracking a common logical structure in data analysis, which could serve as a reference theoretical body driving the research in different areas.

This work has not yet been done but before we can come so far we must find people eager to look into other disciplines; so this school is a starting point for this purposes and hopefully others will follow.

In any case we believe that whatever will be the future of this attempt the first stone has been put into the ground and a number of young scientists have already had the opportunity and the interest to receive this widespread information. The seed has been planted and we hope to see the tree sometime in the future.

The editors

CONTENTS

LIST OF CONTRIBUTORS

Athanasios Dermanis
Department of Geodesy and Surveying
The Aristotle University of Thessaloniki
University Box 503, 54006 Thessaloniki
Greece

e-mail: dermanis@topo.auth.gr

Wolfgang Förstner
Institut of Photogrammetry, Bonn University
Nussallee 15, D-53115 Bonn
Germany

http//:www.ipb.uni-bonn.de

e-mail: wf@ipb.uni-bonn.de

Pascal Fua
Computer Graphics Lab (LIG), Swiss Federal Institute of Technology
CH-1015 Lausanne
Switzerland

e-mail: fua@lig.di.epfl.ch

Armin Grün
Institute of Geodesy and Photogrammetry, ETH Hönggerberg
HIL D 47.2, CH-8093 Zürich
Switzerland

e-mail: agruen@geod.ethz.ch

Haihong Li
Institute of Geodesy and Photogrammetry, ETH Hönggerberg
HIL D 47.2, CH-8093 Zürich
Switzerland

Fabio Rocca
Dipartimento di Elettronica ed Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano
Italy

http://www.elet.polimi.it

e-mail: Fabio.Rocca@elet.polimi.it

R. Rummel
Institut für Astronomische und Physikalische Geodäsie
Technische Universität München
Arcisstrasse 21, D-80290 Münhen
Germany

e-mail: rummel@step.iapg.verm.tu-muenchen.de

F. Sansò
Dipartimento di Ingegneria Idraulica, Ambientale e del Rilevamento
(Sezione Rilevamento), Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano
Italy

e-mail: fsanso@ipmtf4.topo.polimi.it

R. Snieder
Department of Geophysics, Utrecht University
P.O. Box 80.021, 3508 TA Utrecht
The Netherlands

e-mail: snieder@geo.uu.nl

J. Trampert
Department of Geophysics, Utrecht University
P.O. Box 80.021, 3508 TA Utrecht
The Netherlands

Stefano Valle
Dipartimento di Elettronica ed Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano
Italy

e-mail: Stefano.Valle@elet.polini.it

Luigi Zanzi
Dipartimento di Elettronica ed Informazione, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano
Italy

e-mail; Luigi.Zanzi@elet.polimi.it

An overview of data analysis methods in geomatics

A. Dermanis, F. Sansò and A. Grün

Every applied science is involved in some sort of data analysis, where the examination and further processing of the outcomes of observations leads to answers about some characteristics of the physical reality.

There are fields where the characteristics sought are of a qualitative nature, while observed characteristics are either qualitative or quantitative. We will be concerned here with the analysis of numerical data, which are the outcomes of measurements, to be analyzed by computational procedures. The information sought is of spatial context related to the earth, in various scales, from the global scale of geophysics to, say, the local scale of regional geographical analysis. The traditional type of information to be extracted from the data is of quantitative nature, though more modern applications extend also to the extraction of qualitative information.

The classical problems deal with the determination of numerical values, which identify quantitative characteristics of the physical world. Apart from value determination, answering the question of *"how much"* (geophysics, geodesy, photogrammetry, etc.), spatial data analysis methods are also concerned with the questions of *"what"* and *"where"*, i.e., the identification of the nature of an object of known position (remote sensing, image analysis) and the determination of the position of known objects (image analysis, computer vision).

The most simple problems with quantitative data and unknowns, are the ones modeled in a way that is consistent and well determined, in the sense that to each set of data values correspond to a unique set of unknown values. This is definitely not a case of particular interest and hardly shows up in the analysis of spatial data. The first type of problems to present some challenge to the data analysis, have been *overdetermined problems*, where the number of data values exceeds the number of unknowns, with immediate consequence the lack of consistency. Any set of parameter values does not reproduce in general the actual data and the differences are interpreted as "observational errors", although they might reflect modeling errors, as well. The outcome of the study of such data problems has been the "theory of errors" or the "adjustment of observations".

Historically, the treatment of overdetermined problems is associated with the *method of least squares* as devised by Gauss (and independently by Legendre) and applied to the determination of orbits in astronomy. Less known – at least outside the geodetic community – are the geodetic applications for the adjustment of a geodetic network in the area of Hanover, where Gauss had the ambition to test the Euclidean nature of space, by checking whether the angles of a triangle sum up to 180^o or not. Of course, even the most advanced modern measurement techniques are not sufficiently accurate to settle such a problem. However, such an application shows the importance and relevance of observational accuracy, which has always been a main concern of geodetic methodology and technology. Although least square methods found a wide spectrum of applications, in all type of scientific fields, they have had a special place in geodesy, being the hart of geodetic data analysis methods. It is therefore of no surprise that, in the context of studying such problems, the concept of the

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 1–16, 2000.

generalized inverse of a matrix has been independently (re)discovered in geodesy, preceding its revival and study in applied mathematics.

This brings us to the fact that overdetermined problems are, in modern methodology *"inverse problems"*. The study of unknown spatial functions, such as the density of the earth in geophysics, or its gravity potential in geodesy, necessitated the consideration of inverse problems, which are not only overdetermined but also *underdetermined*. Functions are in general objects with an infinite number of degrees of freedom. Their proper representation requires an infinite number of parameters, in theory, or at least a large number of parameters, in practice. Thus, the number of unknowns exceeds the number of data and the consistency problem is been overtaken by the uniqueness problem. An optimization criterion is needed for the choice of a single solution, out of many possible, similar in a sense to the least squares criterion, which solves the consistency problem (lack of solution existence) by choosing an "optimal" set of consistent adjusted observations, out of many possible.

In general an inverse problem is described by an equation of the abstract form

$$y=f(x)\ , \tag{1}$$

where f is a known mapping and a known value b for y. The object is to construct a reasonable inverse mapping g, which maps the data b into an estimate

$$\hat{x}=g(b) \tag{2}$$

of the unknown x. In the most general case, neither the existence or the uniqueness of a solution to the equation $y=f(x)$ is guaranteed. The model consists of the choice of the known mapping f, as well as, the function spaces X and Y where unknown and data belong: $x\in X$, $y\in Y$, $b\in Y$. In practical applications, where we have to treat a finite number n of discrete data values, Y is R^n, or R^n equipped with some additional metric structure. More general space types for Y appear in theoretical studies, related to data analysis problems, where also the limiting case of continuous type observations is considered. The mapping f may vary from a simple algebraic mapping to more general mappings involving differential or integral operators. Differential equations that arise in the modeling of a particular physical process are related not only to f, but also to the choice of the domain space X. For example, the Laplace differential equation for the attraction potential of the earth leads to modeling X as the space of functions harmonic outside the earth and regular (vanishing) at infinity.

The mapping g solves both the uniqueness and the existence (consistency) problem by implicitly replacing the data b with a set of consistent (adjusted) data

$$\hat{y}=f(\hat{x})=f\big(g(b)\big)=(f\circ g)(b)\,. \tag{3}$$

An estimate of the observation errors $v=b-y=b-f(x)$ follows implicitly from

$$\hat{v}=b-\hat{y}=b-(f\circ g)(b)=(id_Y-f\circ g)(b) \tag{4}$$

where id_Y is the identity mapping in Y. The estimate is related to the unknown by

$$\hat{x}=g(b)=g(y+v)=g\big(f(x)+v\big) \tag{5}$$

with a respective estimation error

$$e=\hat{x}-x=g\big(f(x)+v\big)-x \tag{6}$$

In the particular case where g, along with f, is linear, the above equations take the form

$$\hat{x}=(g\circ f)(x)+g(v), \qquad e=(g\circ f-id_X)(x)+g(v). \tag{7}$$

where id_X is the identity mapping in X.

The choice of g should be such that $\hat{v}$ and in particular $e=\hat{x}-x$ are made as small as possible. This means that a way of measuring the magnitude of elements in the spaces X and Y (typically a norm) must be introduced in a reasonably justifiable way. Such a justification is provided by probabilistic tools, where a "probable" or statistical behavior of the errors v and the unknown x, is assumed to be known or provided by independent procedures (sampling methods).

Independently of any such justification, the inverse problem is solved by considering the spaces X and Y to be normed spaces, in one of two ways:

(a) apply

$$\|b-\hat{y}\|_Y=\min_{y\in R(f)}\|b-y\|_Y, \tag{8}$$

followed by

$$\|\hat{x}-x_0\|_X=\min_{x\in X,\, f(x)=\hat{y}}\|x-x_0\|_X \tag{9}$$

(b) apply

$$\|b-f(\hat{x})\|_Y^2+\alpha\|\hat{x}-x_0\|_X^2=\min_{x\in X}\Big\{\|b-f(x)\|_Y^2+\alpha\|x-x_0\|_X^2\Big\}. \tag{10}$$

Above $\|\ \|_X$ and $\|\ \|_Y$ are the norms in X and Y, respectively, $R(f)$ is the range of f, i.e.,

$$R(f)=\{y\in Y \mid y=f(x), \text{ for some } x\in X\}, \tag{11}$$

$\alpha>0$ is a known constant and $x_0 \in X$ is a known a priori estimate of x, which can always be made equal to zero by replacing an original model $y=f^*(x^*)$ by the model $y=f(x)\equiv f^*(x_0+x)$ with $x=x^*-x_0$.

The approach (b) is known as the *Tikhonov regularization method*, where α is the *regularization parameter.*

In addition to the *overdetermined-underdetermined problem*, where $b\notin R(f)$ and for $\hat{y}\in R(f)$ the equation $\hat{y}=f(x)$ has more than one solution, Tikhonov's approach may also be applied to the solution of the *underdetermined problem*, where $b\in R(f)=Y$ and the equation $b=f(x)$ has more than one solution. In fact the latter is the problem that actually arises in practice, where Y is always a normed version of R^n. When the *stepwise approach* (a) is applied to the underdetermined problem, the first step is skipped (since obviously $\hat{y}=b$) and the second step becomes

$$\|\hat{x}-x_0\|_X = \min_{x\in X,\, f(x)=b} \|x-x_0\|_X . \tag{12}$$

As a consequence $f(\hat{x})=b$ leading to the error estimate $\hat{v}=0$, despite the fact that observational errors are unavoidable and thus $v\neq 0$. On the contrary the Tikhonov regularization divides, in this case, the "inconsistencies" of the problem between the errors v and the discrepancies $x-x_0$ of the solution from its prior estimate, in a balanced way, which is governed by the choice of the regularization parameter α.

By the way the choice of α is not a problem independent of the *problem of choice of norm* $\|\ \|_X$: the regularization parameter can be incorporated into the norm definition by replacing an initial norm $\|\ \|_{0,X}$ with the equivalent norm $\|\ \|_X=\sqrt{\alpha}\,\|\ \|_{0,X}$.

To trace the development of inverse methods for data analysis with quantitative data and quantitative unknowns, we must return to the classical *overdetermined problem*, where $b\notin R(f)$ and for $\hat{y}\in R(f)$ the equation $\hat{y}=f(x)$ has a unique solution. In this case the stepwise method (a) and the Tikhonov regularization method (b) may be identified by neglecting the second unnecessary step in (a) and choosing $\alpha=0$ in (b). Thus we must apply either

$$\text{(a*)}\quad \|b-\hat{y}\|_Y^2 = \min_{y\in R(f)} \|b-y\|_Y^2 , \tag{13}$$

followed by the determination of the unique solution $\hat{x}$ of $\hat{y}=f(x)$, or apply directly

$$\text{(b*)}\quad \|b-f(\hat{x})\|_Y^2 = \min_{x\in X} \|b-f(x)\|_Y^2 . \tag{14}$$

The overdetermined problem is typically finite dimensional, where with n observations and $m<n$ unknowns, X is a normed version of R^n and Y can be identified with

R^m. In geodesy and photogrammetry such problems involving finite-dimensional spaces X and Y are sometimes characterized as "full rank models", as opposed to the "models without full rank" which are simultaneously overdetermined and underdetermined.

Among the possible norm definitions the choice that proved more fruitful has been the one which is implied by an inner product

$$(y,z)_Y = \mathbf{y}^T\mathbf{P}\mathbf{z}\,, \qquad \|y\|_Y = \sqrt{(y,y)_Y} = \sqrt{\mathbf{y}^T\mathbf{P}\mathbf{y}}\,, \tag{15}$$

where y and z are represented by $n\times 1$ matrices $\mathbf{y}$ and $\mathbf{z}$, respectively, while $\mathbf{P}$ is an $n\times n$ positive definite weight matrix.

In this case the solution to the inverse problem is a least squares solution $\hat{\mathbf{x}}$ resulting from the minimization of the "weighted sum of squares" $\mathbf{v}^T\mathbf{P}\mathbf{v}=\min$, of the errors $\mathbf{v}=\mathbf{b}-f(\mathbf{x})$

$$\big(\mathbf{b}-f(\hat{\mathbf{x}})\big)^T\mathbf{P}\big(\mathbf{b}-f(\hat{\mathbf{x}})\big)=\min_{\mathbf{x}\in R^m}\big(\mathbf{b}-f(\mathbf{x})\big)^T\mathbf{P}\big(\mathbf{b}-f(\mathbf{x})\big). \tag{16}$$

An open problem is the choice of the weight matrix $\mathbf{P}$, except for the case of observations of the same type and accuracy where the choice $\mathbf{P}=\mathbf{I}$ was intuitively obvious. This problem has been resolved by resorting to probabilistic reasoning, as we will see below.

In the nonlinear case the least squares solution $\hat{\mathbf{x}}$ can be found only by a numerical procedure which makes use of the given particular value $\mathbf{b}$. A computational procedure of an iterative nature can be used in order to minimize the distance ρ of $\mathbf{b}$ from the curved manifold $M=R(f)$ with the same dimension m as X. The unknowns $\mathbf{x}$ serve in this case as a set of curvilinear coordinates for M. The knowledge of an approximate value $\mathbf{x}^0$ of $\mathbf{x}$ and the corresponding "point" $\mathbf{y}^0=f(\mathbf{y}^0)\in M$ is sufficient for the determination of a local minimum $\rho(\mathbf{b},\hat{\mathbf{y}})$ of the distance

$\rho(\mathbf{b},\mathbf{y})=\sqrt{(\mathbf{b}-\mathbf{y})^T\mathbf{P}(\mathbf{b}-\mathbf{y})}$, $\mathbf{y}\in M$, with $\hat{\mathbf{y}}$ in a small neighborhood of $\mathbf{y}^0$.

In this sense, we do not have a general solution to the inverse problem: a mapping g, which would map any data vector $\mathbf{b}$ into the least squares solution $\hat{\mathbf{x}}=g(\mathbf{b})$, has not been determined. The determination of such a mapping is possible only in the special case where f is a linear mapping represented by an $n\times m$ matrix $\mathbf{A}$. The well known least squares inverse mapping g is represented by a matrix $\mathbf{A}^-=(\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}\mathbf{P}$ which provides the least squares solution $\hat{\mathbf{x}}=g(\mathbf{b})=\mathbf{A}^-\mathbf{b}$ for any value of the data $\mathbf{b}$. It turns out that, as expected, the point $\hat{\mathbf{y}}=\mathbf{A}\hat{\mathbf{x}}=\mathbf{A}\mathbf{A}^-\mathbf{b}$ is the orthogonal projection of $\mathbf{b}$ on the linear manifold M. Indeed the operator $p=f\circ g$ represented by the matrix $\mathbf{P}_M=\mathbf{A}\mathbf{A}^-$, is a projection operator from X to its linear subspace M.

The linear problem has also allowed a probabilistic approach to the inversion (estimation) problem, which turned out to provide a solution to the "weight choice" prob-

lem of least squares. The observational errors $\mathbf{v}$ are modeled as (outcomes of) random variables with zero means $E\{\mathbf{v}\}=\mathbf{0}$ and covariance matrix $E\{\mathbf{v}\mathbf{v}^T\}=\mathbf{C}$, so that the observations $\mathbf{b}$ are also (outcomes of) random variables with means their "true" values $E\{\mathbf{b}\}=\mathbf{y}$ and the same covariance matrix $E\{(\mathbf{b}-\mathbf{y})(\mathbf{b}-\mathbf{y})^T\}=\mathbf{C}$. For any linear function of the parameters $q=\mathbf{a}^T\mathbf{x}$, an estimate is sought, which is a linear function of the available data $\hat{q}=\mathbf{d}^T\mathbf{b}$, such that the *mean square estimation error*

$$E\{(\hat{q}-q)^2\}=\mathbf{d}^T\mathbf{C}\mathbf{d}+\mathbf{x}^T\left[\mathbf{A}^T\mathbf{d}-\mathbf{a}\right]\left[\mathbf{A}^T\mathbf{d}-\mathbf{a}\right]^T\mathbf{x}=$$

$$=\mathbf{d}^T(\mathbf{C}+\mathbf{A}\mathbf{x}\mathbf{x}^T\mathbf{A}^T)\mathbf{d}-2(\mathbf{a}^T\mathbf{x})\mathbf{x}^T\mathbf{A}^T\mathbf{d}+(\mathbf{a}^T\mathbf{x})^2=\phi(\mathbf{d}) \tag{17}$$

is minimized among all *uniformly unbiased linear estimates*, i.e. those which satisfy the condition $E\{\hat{q}\}=\mathbf{d}^T\mathbf{A}\mathbf{x}=E\{q\}=\mathbf{a}^T\mathbf{x}$ for any value of $\mathbf{x}$. Consequently, one has to find the value $\mathbf{d}$ which minimizes the quadratic expression $\phi(\mathbf{d})$ under the side condition $\mathbf{A}^T\mathbf{d}-\mathbf{a}=\mathbf{0}$. Application of the method of Lagrange multipliers leads to the optimal value

$$\mathbf{d}=\mathbf{C}^{-1}\mathbf{A}(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{a} \tag{18}$$

and the Best Linear (uniformly) Unbiased Estimate (BLUE)

$$\hat{q}=\mathbf{a}^T(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{b} \tag{19}$$

and after separate application to each component x_k of $\mathbf{x}$ to the BLUE of the parameters

$$\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{C}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}^{-1}\mathbf{b} \tag{20}$$

This estimate which is optimal in a probabilistic sense (Best = minimum mean square estimation error) can be identified with the least squares estimate with the particular choice $\mathbf{P}=\mathbf{C}^{-1}$ for the weight matrix. This classical result is essentially the *Gauss-Markov theorem*, where (in view of the obvious fact that the least squares estimate is independent of any positive multiplier of the weight matrix) it is assumed that $\mathbf{C}=\sigma^2\mathbf{Q}$, with $\mathbf{Q}$ known and σ^2 unknown, while $\mathbf{P}=\mathbf{Q}^{-1}$.

This choice is further supported by the fact that an other probabilistic method, the maximum likelihood method, yields the same estimate under the additional assumption that the observational errors follow the Gaussian distribution.

Examples of such overdetermined problems are the determination of the shape of a geodetic network from angle and distance observations, and the determination of the ground point coordinates from observations of image coordinates on photographs in analytical photogrammetry.

Even in this case the need to apply weights to some of the parameters was felt, especially for the "stabilization" of the solution. Since weighting and random character are interrelated through the Gauss-Markov Theorem, the weighted parameters were implicitly treated as random quantities with means their respective approximate values introduced for the linearization of the model.

Unfortunately this result cannot be extended to the *non-linear model.* Linearity is essential for the application of the principle of uniform unbiased estimation, which makes the optimal estimate independent of the unknown true values $\mathbf{x}$ of the parameters. This can be easily seen in the second term of eq. (17), where the condition for uniformly unbiased estimate $\mathbf{A}^T\mathbf{d}-\mathbf{a}=\mathbf{0}$ makes the mean square error independent of $\mathbf{x}$. To get a geometric insight into the situation, consider the sample points of the random variable $\mathbf{b}=\mathbf{y}+\mathbf{v}=\mathbf{A}\mathbf{x}+\mathbf{v}$ as a "cloud" of point masses having as "center of mass" an unknown point $\mathbf{y}=E\{\mathbf{b}\}$ on the linear manifold M. The orthogonal projection $\mathbf{P}_M=\mathbf{A}\mathbf{A}^-$ maps each sample point of $\mathbf{b}$ into a corresponding sample point of $\hat{\mathbf{y}}=\mathbf{P}_M\mathbf{b}$ in a such a way that the center of mass is preserved! indeed the resulting from the projection sample points of $\hat{\mathbf{y}}$ have center of mass $E\{\hat{\mathbf{y}}\}=\mathbf{P}_M E\{\mathbf{b}\}=$ $=\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x}=\mathbf{A}\mathbf{x}=\mathbf{y}$. When the manifold M is curved, there is in general no way to construct a mapping from Y to M with the property of preserving the center of mass of sample points.

The need to model unknowns also as random quantities, became obvious when geodesists were confronted with an underdetermined problem, namely that of the determination of the gravity filed of the earth from discrete gravity observations at points on the earth surface. The unknown potential function is a mathematical object with infinite degrees of freedom and its faithful representation requires an infinity (in practice very large) number of parameters, such as the coefficients of its expansion in spherical harmonics. Assigning random character to these representation parameters means that the function itself is modeled as a random function, i.e., as a stochastic process. Spatial random functions are usually called *random fields*, and their study became relevant for applications in many earth sciences. The first steps in the direction of producing a reasonable "optimal" estimate of the unknown function, and indeed independently of its parameterization by any specific set of parameters, was based on methods developed for stochastic processes with time as their domain of definition, originating in communication engineering for the treatment of signals. The applicability of these estimation, or rather *prediction*, methods was so successful that the word "signal" (with a specific original meaning) has been eventually used for all types of physical processes.

The value of a random field at any particular point is a random variable that is correlated with the observables (observed quantities before the effect of random errors) which are random variables related to the same random field. The problem of the spatial function determination can be solved, in this context, by applying the method of *minimum mean square error linear prediction* of (the outcomes of) a random variables z from the known (outcomes of) another set of random variables $\mathbf{b}$, when both sets are correlated. This method of prediction can be characterized as a second order

method since it uses only up to second order statistics of the random variables, namely their means

$$\mathbf{m}_\mathbf{b} = E\{\mathbf{b}\}\,,\ m_z = E\{z\}\,, \tag{21}$$

their covariances

$$\mathbf{C}_\mathbf{b} = E\{(\mathbf{b}-\mathbf{m}_\mathbf{b})(\mathbf{b}-\mathbf{m}_\mathbf{b})^T\}\,, \qquad \sigma_z^2 = E\{(z-m_z)^2\} \tag{22}$$

and their cross-covariances

$$\mathbf{c}_{\mathbf{b}z} = E\{(\mathbf{b}-\mathbf{m}_\mathbf{b})(z-m_z)\}\,, \qquad \mathbf{c}_{z\mathbf{b}} = \mathbf{c}_{\mathbf{b}z}^T\,. \tag{23}$$

The optimal estimate of any unobserved random variable z is given by a linear function $\hat{z}=\mathbf{d}^T\mathbf{b}+\kappa$ of the observed random variables $\mathbf{b}$, where the parameters $\mathbf{d}$ and κ are chosen in such a way that the mean square error of prediction

$$E\{(\hat{z}-z)^2\} = \mathbf{d}^T\mathbf{C}_{\mathbf{bb}}\mathbf{d} - 2\mathbf{d}^T\mathbf{c}_{\mathbf{b}z} + \sigma_z^2 + [\kappa + \mathbf{d}^T\mathbf{m}_\mathbf{b} - m_z]^2 =$$

$$= \mathbf{d}^T(\mathbf{C}_{\mathbf{bb}} + \mathbf{m}_\mathbf{b}\mathbf{m}_\mathbf{b}^T)\mathbf{d} + 2[(\kappa - m_z)\mathbf{m}_\mathbf{b} - \mathbf{c}_{\mathbf{b}z}]^T\mathbf{d} + \sigma_z^2 + (\kappa - m_z)^2 = \phi(\mathbf{d}) \tag{24}$$

is minimized under the condition that the prediction is unbiased, i.e. ,

$$E\{\hat{z}\} = \mathbf{d}^T\mathbf{m}_\mathbf{b} + \kappa = E\{z\} = m_z \tag{25}$$

The minimization of the quadratic expression $\phi(\mathbf{d})$ under the side condition $\mathbf{m}_\mathbf{b}^T\mathbf{d} + \kappa - m_z = 0$ yields the values

$$\mathbf{d} = \mathbf{C}_{\mathbf{bb}}^{-1}\mathbf{c}_{\mathbf{b}z}\,, \qquad \kappa = m_z - \mathbf{c}_{\mathbf{b}z}^T\mathbf{C}_{\mathbf{bb}}^{-1}\mathbf{m}_\mathbf{b} \tag{26}$$

so that the minimum mean square error unbiased linear prediction becomes

$$\hat{z} = m_z + \mathbf{c}_{\mathbf{b}z}^T\mathbf{C}_{\mathbf{bb}}^{-1}(\mathbf{b}-\mathbf{m}_\mathbf{b}) \tag{27}$$

A straightforward extension to a vector of predicted variables z follows from the separate prediction of each component z_k and has the similar form

$$\hat{\mathbf{z}} = \mathbf{m}_\mathbf{z} + \mathbf{C}_{\mathbf{z}\mathbf{b}}\mathbf{C}_{\mathbf{bb}}^{-1}(\mathbf{b}-\mathbf{m}_\mathbf{b})\,. \tag{28}$$

This prediction method can be directly applied when the observables $\mathbf{y}$ are the values of functionals of the relevant random field x (i.e. real valued quantities depending on the unknown spatial function) which are usually linear or forced to become linear

through linearization. If $z=x(P)$ is the value of the field at any point of its domain of definition the point-wise prediction $\hat{z}=\hat{x}(P)$ provides virtually an estimate $\hat{x}$ of the unknown field x. The presence of additional noise $\mathbf{n}$ with $E\{\mathbf{n}\}=\mathbf{0}$ yields the observations $\mathbf{b}=\mathbf{y}+\mathbf{n}$ with mean $\mathbf{m_b}=\mathbf{m_y}$ and covariance matrices $\mathbf{C_{bb}}=\mathbf{C_{yy}}+\mathbf{C_{nn}}$, $\mathbf{C_{zb}}=\mathbf{C_{zy}}$. Consequently the prediction algorithm becomes

$$\hat{\mathbf{z}}=\mathbf{m_z}+\mathbf{C_{zy}}(\mathbf{C_{bb}}+\mathbf{C_{nn}})^{-1}(\mathbf{b}-\mathbf{m_y}). \tag{29}$$

The applicability of the method presupposes that all the relevant covariances can be derived in a mathematically consistent way from the covariance function of the random field, which could be chosen in a meaningful way. These assumptions are not trivial and they pose interesting mathematical questions. The assumptions of homogeneity of the random field (geostatistics – ore estimation) or of both homogeneity and isotropy (geodesy – gravity field determination) are proven to be necessary for solving such problems in a reasonable way.

The minimum mean square error linear prediction method is used in geodesy under the (somewhat misleading) name *"collocation"*. A variant of the same method is used in geostatistics, for the prediction of ore deposits, under the name *"Krieging"*. The main difference between collocation and Krieging is that in the latter optimal prediction is sought in the class of strictly linear predictors of the form $\hat{z}=\mathbf{d}^T\mathbf{b}$ instead of the class $\hat{z}=\mathbf{d}^T\mathbf{b}+\kappa$ used in collocation.

It is interesting to note that the *duality*, which appears in the Gauss-Markov theorem, between the deterministic least squares principle and the probabilistic principle of minimum mean square estimation error, finds an analogue in the problem of the estimation of an unknown spatial function modeled as a random process. The solution (29) can also be derived in a deterministic way, by applying an optimization criterion of Tikhonov type

$$\mathbf{n}^T\mathbf{P}\mathbf{n}+\|x\|_H=\min \tag{30}$$

where $\|x\|_H$ is the norm of the function x which is modeled to belong to a *Hilbert space H* with a *reproducing kernel k*.

This roughly means that H is an infinite dimensional function space with an inner product, having the mathematical properties of separability and completeness, and possessing a two point function $k(P,Q)$ with the reproducing property

$$\left(k^P, f\right)_H = f(P), \tag{31}$$

$k^P(\cdot)=k(P,\cdot)$ being the function resulting by fixing the point P in $k(P,Q)$. The duality is now characterized, in addition to $\mathbf{P}=\mathbf{C_{nn}^{-1}}$, by the equality $k(P,Q)=C(P,Q)$ of the reproducing kernel $k(P,Q)$ with the covariance function $C(P,Q)$ of x, defined by

$C(P,Q)=E\{[(x(P)-m(P)[(x(Q)-m(Q)]\}$, where $m(P)=E\{x(P)\}$ is the mean function of x. As a result, this duality solves the problem of choice of norm for the function x. Under the simplifying assumptions of homogeneity and isotropy, it allows the estimation of the covariance function from the available observations with the introduction of one more assumption, that of the identification of averages over outcomes with averages over the domain of definition (covariance ergodicity).

The treatment of the unknown mean function m poses also a problem. It is usually treated as equal to a known model function m_0, which is subtracted from the function x which is thereof replaced by a zero mean random field $\delta x=x-m_0$. An additional trend $\delta m=m-m_0$ can be modeled to depend on a set of unknown parameters $\mathbf{a}=[a_1 a_2 \ldots a_s]^T$, e.g. the coefficients of a polynomial or trigonometric series expansion, which are estimated from the available data, either a priori or simultaneously with the prediction (mixed model). Usually only a constant $\delta m=\overline{m}$ is estimated as the mean of the available data, or at most a linear trend such as $\delta m=a_0+a_1 x+a_2 y$ in the planar case. The problem is that an increasing number s of parameters absorbs increasing amount of information from the data leaving little to be predicted, in fact $\delta x=0$ when $s=n$, n being the number of observations.

Again this statistical justification of the norm choice presupposes linearity of the model $y=f(x)$. This means that each component y_k of the observables $\mathbf{y}$ must be related to the unknown function through $y_k=f_k(x)$ where f_k is a linear functional (mapping of a function to a real number). In fact it should be a continuous (bounded) linear functional and the same holds true for the any functional f_z for the corresponding quantity $z=f_z(x)$ to be predictable. As in the case with a finite-dimensional unknown, linearity refers to both the mathematical model $y=f(x)$ and the mapping $\hat{z}=h(\mathbf{b})$ from the erroneous data $\mathbf{b}$ into the estimate/prediction of the unknown or any quantity z related linearly to the unknown.

Apart from the linearity we are also within a *"second order theory"* since only means and covariances are involved, without requiring the complete knowledge of the probability distribution of the relevant random variables and random fields.

The introduction of a stochastic model for the unknown (finite or infinite dimensional) finds justification also within the framework of Bayesian statistics. We should distinguish between the more general Bayesian point of view and the "Bayesian methods" of statistics. The Bayesian spirit calls for treating all unknown parameters as random variables having a priori statistical characteristics which should be revised with the evidence provided by the observations. Using the standard statistical terminology, the linear Gauss-Markov model is replaced by either a linear *mixed model*, where only some of the unknown parameters are treated as random variables, or by a linear *random effects model* with all parameters random. These models and their corresponding solutions for estimation and/or prediction, cover only the case of a finite-dimensional unknown, but the treatment of an infinite-dimensional one, in the above case of a spatial function modeled as a random process, may well be considered a Bayesian approach.

Bayesian methods, in the narrow sense, extend outside the bounds of a second order theory because they are based on knowledge of the distributions (described by probability density functions) of the random variables. In this aspect they are similar to the maximum likelihood estimation method for linear models with deterministic parameters. Furthermore they primarily aim not to estimation (prediction) itself, but to the determination of the a posteriori distribution $p(\mathbf{x}|\mathbf{y})=p_{\mathbf{x}|\mathbf{y}}(\mathbf{x},\mathbf{y})$ of the unknowns $\mathbf{x}$, based on their prior distribution $p(\mathbf{x})=p_{\mathbf{x}}(\mathbf{x})$, the (conditional on the values of the unknowns) distribution of the observations $p(\mathbf{y}|\mathbf{x})=p_{\mathbf{y}|\mathbf{x}}(\mathbf{x},\mathbf{y})$ and the actual outcomes of the observations $\mathbf{y}$. The a posteriori distributions are provided by the famous Bayes formula

$$p(\mathbf{x}|\mathbf{y})=\frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{\int p(\mathbf{y}|\xi)\,p(\xi)\,d\xi}. \tag{32}$$

The function $p(\mathbf{y}|\mathbf{x})=p_{\mathbf{y}|\mathbf{x}}(\mathbf{x},\mathbf{y})$ when viewed as a function of $\mathbf{x}$ only, with $\mathbf{y}$ taking the observed values, is in fact the likelihood function $l(\mathbf{x})=p(\mathbf{y}|\mathbf{x})$ of the maximum likelihood method. Estimation in the Bayesian methodology is a by-product of the determination of the a posteriori distribution $p(\mathbf{x}|\mathbf{y})$, the *maximum a posteriori estimate* $\hat{\mathbf{x}}$ provided by

$$p(\hat{\mathbf{x}}|\mathbf{y})=\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \tag{33}$$

This should be compared to the (different) classical maximum likelihood estimate

$$p(\mathbf{y}|\hat{\mathbf{x}})=\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})\,. \tag{34}$$

where the likelihood function $l(\mathbf{x})=p(\mathbf{x},\mathbf{y})=p(\mathbf{y}|\mathbf{x})$ is identical in form to the distribution $p(\mathbf{y}|\mathbf{x})$ but $\mathbf{x}$ is now unknown, while the unknown $\mathbf{y}$ is fixed to its known observed value (sample).

The classical case of completely unknown parameters is incorporated in the Bayesian scheme with the use of non-informative prior distributions, which assign the same probability to all unknowns. In agreement with the Gauss-Markov setup, a constant factor σ^2 of the covariance matrix of the observations is included in the unknowns. Again the use of basic assumptions is not dictated by the physics of the problem, but rather by computational convenience: prior distributions for $\mathbf{x}$ and σ^2 are matched with the distribution of the observations $p(\mathbf{y}|\mathbf{x},\sigma^2)$, in pairs which lead to a convenient computationally tractable posterior distribution $p(\mathbf{x},\sigma^2|\mathbf{y})$. This drawback is similar to the use of the Gaussian distribution in the maximum likelihood method, or the choice to minimize $E\{e^2\}$ where e is the estimation or the prediction error, instead of, say, $E\{|e|\}$. In the framework of Statistical Decision Theory $e^2(\mathbf{x})$

is a particular choice of a *loss function* $l(\mathbf{x})$, while estimation is based on the minimization of the corresponding *risk function* $r(\mathbf{x})=E\{l(\mathbf{x})\}$.

The introduction of probabilistic (or statistical) approaches to inverse problems has its own merits and should not be viewed as merely a means of choosing the relevant norms. The most important aspect is the fact that the estimate $\hat{x}=g(b)$, being a function of the random data b, it is itself random with a distribution that can be in principle derived from the known distribution of b. In reality the distribution of $\hat{x}$ can be effectively determined only in the simpler case where the inverse mapping g is linear and furthermore the data b follow the normal distribution. This explains the popularity of the normal distribution even in cases where there is no physical justification for its use. The knowledge of the distribution of $\hat{x}$ allows a *statistical inference* about the unknown x. This includes the construction of confidence regions around $\hat{x}$ where x should belong with a given high probability. Even more important is the possibility to distinguish (in the sense of determining which is more probable) between alternative models in relation to the same data. Usually the original model $f:X\rightarrow Y$ is compared with an alternative model $f':X'\rightarrow Y$, where f' is the restriction of f to a subset $X'\subset X$, defined by a means of constrains $h(x)=0$ on the unknown x. In practice, within the framework of the linear (or linearized) approach, a linear set of constrains $\mathbf{Hx}-\mathbf{z}=\mathbf{0}$ is used, which allows the testing of the *general hypothesis* $\mathbf{Hx}=\mathbf{z}$.

Along with above class of inverse problems, concerned with the determination of numerical values corresponding to a quantitative unknown, problems of a much different type arose in the group of disciplines that we now call *geomatics*. The first such example comes from photogrammetry, or to be specific from the closely associated field of photointerpretation. The interpretation of photographs was a discipline where both the "unknowns" sought and the methods of analysis were strictly qualitative. The possibility of computational treatment has been a byproduct of technical improvements that led to the use of digital (or digitized) photography. The possibility to treat photographs as a set of numerical data to be processed computationally for the determination of, more or less, the same qualitative unknowns, caused such important developments that the new field of remote sensing came into being. Of course the methodology and applications of remote sensing span a much wider range of disciplines, but it was the photogrammetric world that has been greatly influenced and undergone a deep transformation in its interests, as the change of names and content of scientific societies and relevant journals demonstrates.

If we attempt to express the remote sensing problem in the old terminology of inverse problems, we deal again with observations $\mathbf{b}$ of the intensity values of a pixel in a number of spectral bands, which depends on the physical identity x of the depicted object. We have though two essential differences: the first is that x is not quantitative, but qualitative, taking one of possible discrete values $\omega_1,\omega_2,\ldots,\omega_s$ that correspond to classes of physical objects. We can formally write $x\in X$ with $X=\{\omega_1,\omega_2,\ldots,\omega_s\}$. The second is that the mapping f of the model $\mathbf{b}=f(x)$ is not

a deterministic but rather a random mapping. Indeed for any specific class ω_k the value $\mathbf{b}=f(\omega_k)$ is not a fixed number but a random variable. Even in non-statistical methods of remote sensing, $\mathbf{b}$ for a given class ω_k has a variable value, due to the fact that a collection of varying objects have been categorized as belonging to the same class ω_k.

The usual approach to statistical pattern recognition or statistical classification, treats the data $\mathbf{b}$ as outcomes from one of distinct (vector) random variables corresponding to respective object classes $\omega_1,\omega_2,\ldots,\omega_s$. This is a typical statistical discrimination problem, i.e., the determination of which statistical population, out of a given set, comes a particular observed sample. However we can reduce the problem to our inverse problem terminology by making use of the mean vectors $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_s$ of the distributions corresponding to the respective classes (values of x) $\omega_1,\omega_2,\ldots,\omega_s$. The mapping $f:X\rightarrow Y$ is trivially defined in this case by

$$f(\omega_k)=\mathbf{y}_k, \quad (k=1,2,\ldots,s) \tag{35}$$

while the range of f

$$R(f)=\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_s\}\in Y \tag{36}$$

is not a subspace but rather consists of a set of discrete set of isolated points in the space of spectral values Y where the observed data belong ($\mathbf{b}\in Y$). Y is a finite dimensional space, essentially R^n, where n is the number of available spectral bands. The observed data $\mathbf{b}$ differ from the values $\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_s\}$, that is $\mathbf{b}\notin R(f)$ not so much because of errors related to the observation process (sensor performance, illumination and atmospheric conditions, etc.) but mainly due to the variation of the actually observed physical object from the corresponding artificial prototype ω_k of the class to which it belongs. Since the inversion of f seen as a function $f:X\rightarrow R(f):\{\omega_1,\ldots,\omega_s\}\rightarrow\{\mathbf{y}_1,\ldots,\mathbf{y}_s\}$ is trivial, the only remaining part is the construction of the mapping $p:Y\rightarrow R(f)$, which can hardly be called a projection any more. As in the usual case of overdetermined (but not simultaneously underdetermined) problems, we can get the answer sought by applying a minimization principle similar to (8). However, the distance of $\mathbf{b}$ from each point $\mathbf{y}_i$ has to be measured in a different way, at least when a statistical justification is desirable, since each $\mathbf{y}_i$ is the mean of a different probability distribution with different probability density function $p_i(\mathbf{y})$ and thus different covariance matrix $\mathbf{C}_i$. One solution is to use only up to second order statistics $(\mathbf{y}_i,\mathbf{C}_i,i=1,\ldots,s)$ and to determine the "optimal" $\mathbf{y}_k$ by applying the minimization principle

$$(\mathbf{b}-\mathbf{y}_k)^T\mathbf{C}_k^{-1}(\mathbf{b}-\mathbf{y}_k)=\min_i\left\{(\mathbf{b}-\mathbf{y}_i)^T\mathbf{C}_i^{-1}(\mathbf{b}-\mathbf{y}_i)\right\}. \tag{37}$$

In order to obtain a probabilistic justification of the above choice we must get out of the limits of the second order approach and resort to the maximum likelihood method where the likelihood function of the unknown x is $l(\omega_i)=p_i(\omega_i|\mathbf{b})=p_i(\mathbf{y}|\omega_i)|_{\mathbf{y}=\mathbf{b}}$. Due to the relevant one-to-one correspondence we may replace ω_i with $\mathbf{y}_i$ and the optimal $\mathbf{y}_k$ is derived from $l(\mathbf{y}_k)=\max_i l(\mathbf{y}_i)$ or explicitly from

$$p_k(\mathbf{y}_k|\mathbf{b})=\max_i p_i(\mathbf{y}_i|\mathbf{b})\,. \tag{38}$$

The additional assumption that the relevant random variables are normally distributed, $p_i(\mathbf{y}|\omega_i)=k_i\exp\{\frac{1}{2}(\mathbf{y}_i-\mathbf{y})^T\mathbf{C}_i(\mathbf{y}_i-\mathbf{y})\}$, does not lead to the choice (37), due to the presence of $k_i=k(\mathbf{C}_i)=[(2\pi)^n|\mathbf{C}_i|]^{-1/2}$, but to the slightly different result

$$\ln(|\mathbf{C}_k|)+(\mathbf{b}-\mathbf{y}_k)^T\mathbf{C}_k^{-1}(\mathbf{b}-\mathbf{y}_k)=\min_i\{\ln(|\mathbf{C}_i|)+(\mathbf{b}-\mathbf{y}_i)^T\mathbf{C}_i^{-1}(\mathbf{b}-\mathbf{y}_i)\}. \tag{39}$$

The solution (38) is a special case of the more general Bayesian approach where the unknown object class x is also assumed to be random with distribution determined by the a-priori known probabilities $p(\omega_i)$, $i=1,2,\ldots,s$, of occurrence for each class. This results to the Bayesian classification to the class ω_k

$$p(\omega_k|\mathbf{b})=\max_i p(\omega_i|\mathbf{b})\,,\qquad p(\omega_i|\mathbf{y})=\frac{p(\mathbf{y}|\omega_i)p(\omega_i)}{\sum_k p(\mathbf{y}|\omega_k)p(\omega_k)} \tag{40}$$

where the integral in the denominator of (32) has been replaced by summation over the discrete set of values of the unknown x.

The above particular point of view on remote sensing (or pattern recognition in a more general context) is somewhat unorthodox, but it has the advantage of a unification under the more wide "umbrella" of inverse problems. In addition to a common approach to the originally different problems of determining "how much" (quantitative unknown) and "what" (qualitative unknown), it smoothes the way for a similar consideration of the problem of determining the "where" of a given object (feature extraction), always departing from a set of available numerical (quantitative) data.

To make the relevance a little more obvious let us first consider the problem of determining the parameters ($\mathbf{y}_i$ and $\mathbf{C}_i$) of the relevant distributions, which can be solve by resorting to sampling, exactly as in the case of the "parameter estimation" solution to inverse problems. In this case the sampling procedure based on available data (pixels) belonging to known classes is called *training*.

On the other side of the statistical supervised (trained) classification lies the deterministic unsupervised classification, where clustering techniques identify clusters of neighboring (in spectral space *Y*) pixels to which correspond respective classes, which are eventually identified through pixels of known class. These clusters have their own

mean values and covariances, so that statistical characteristics are present, at least implicitly.

We may see training as a procedure during which the general algorithm undergoes a training procedure during which it "learns" how to better adopt itself to a particular application. This idea of learning, in a sense much more general than the statistical concept of sampling, is crucial in more modern pattern recognition (classification) methods, such as the so called "neural networks". These are sets of interacting nodes, each one having input and output to the others. The node is essentially a flexible algorithm, which to a given input produces a specific output. The flexibility of the algorithm lies in its ability of self-modification under inputs of known class so that a correct output is produced. This modification under known inputs has the character of learning similar to the learning involved in the classification training, or even in the sampling procedures for the estimation of values for means and covariance, which make the general algorithms adapt to a specific application in the case of best linear unbiased parameter estimation or minimum mean square error prediction, which are statistical solutions to inverse problems. This learning aspect brings the various fields of geomatics closer to relevant fields of computer science, such as learning machines, artificial intelligence, expert systems, automation in image analysis, computer vision, etc.

Apart from the pixel-wise treatment of images, it is possible - and in certain applications necessary – to treat the whole digital image as a set of discrete observations on a grid corresponding to an underlying continuous spatial function, which may be also modeled as a random field. However we are primarily interested not on the determination of the spatial function, but rather to its discontinuities, and most of all on the qualitative interpretation of these discontinuities. This brings as to development of new techniques in traditional photogrammetry for the automatization of the problem of the point-by-point image matching (image correlation), which has been traditionally achieved through stereoscopic vision. Independent developments in computer science (computer vision, artificial intelligence) became of relevance for the solution of photogrammetric problems. Another problem of relevance to image correlation, but also with stand-alone importance, is the automated extraction of features from images. Feature extraction has its particular characteristics but bears also some resemblance to more traditional inverse problems, if we thing of the unknown as the location (the "where") of a known object or rather of an object belonging to a known specific class. The data in this case are primarily the intensity values of a digital image, but a preprocessing is necessary to provide derived data which are more appropriate to the formulation and solution of the "where" problem. the maxima of gradient values produced from the original intensity values provide a mean of identifying lines of separation in the image, which correspond to the "outlines" of the depicted objects. Line segments, converted from raster to vector form, may be combined into sets which may be examined for resemblance to prototype line sets corresponding to known object outlines. Thus one may identify objects as land parcels, roads, buildings, etc.

Again the problem is not foreign to a probabilistic point of view. The appearance of line sets on the image does vary in a more or less random way from the prototype of the class. This variation is due to the variation of the actual depicted objects from

their corresponding prototypes, but one has to take into account the variations due to the different perspectives from which the object may be depicted on the image.

The solution to the problem involves again a minimization problem where a measure of the difference between the data and the prototype is to be minimized in order to assign the "correct" prototype to the data from out of a set of given prototypes.

The main difference is that the "space" Y, where the data and the representatives of the prototypes ($R(f)$!) belong are more complex, than the space of the corresponding numerical data which define the relevant (vectorized) line segments. Variation of "position" in this space involves more than variation in the above numerical values. Transformations of the configuration of the "observed" feature which correspond to different perspectives or object variations within the class, must be taken into account. These variations involve variations in the angles between line segments or even disappearance of segments due to particular perspectives (think of the identification of buildings from their straight edges).

The complexity of the problem leads to various solution approaches, which have a somewhat heuristic character, in the sense that they are not always derived from a well defined general problem formulation and solution principle.

Of course we are dealing in this case with a field which is under continuous development, in contrast to the standard inverse problems where the solution methods and techniques have reached a state of maturity (not always a term with a positive meaning!) and allow their examination under a theoretically secure point of view. On the contrary, when it comes to the problem of automatic feature extraction and other similar problems of image analysis, it is difficult (and even dangerous) to theorize.

The future development and the success of particular approaches in repeated applications will also settle the theoretical aspects. We have only made an attempt here to have a unifying view through the formalism of inverse problems, stretching the relevance of probabilistic-statistical methods in achieving an optimal solution.

It remains to be seen whether all various fields of data analysis problems in geomatic applications can be truly unified (at least in theory), or whether they are separate problems tied together under the academic umbrella of the University curriculums of Surveying Departments or formerly Surveying and now Departments of Geomatics, as the trend seems to be.

Data analysis methods in geodesy

Athanasios Dermanis

Department of Geodesy and Surveying
The Aristotle University of Thessaloniki

Reiner Rummel

Institute of Astronomical and Physical Geodesy
Technical University of Munich

1 Introduction

"Geodesy" is a term coined by the Greeks in order to replace the original term "geometry", which had meanwhile lost its original meaning of "earth or land measuring" (surveying) and acquired the new meaning of an abstract "theory of shapes". Aristotle tells us in his "Metaphysics" that the two terms differ only in this respect: "Geodesy refers to things that can be sensed, while geometry to things that they cannot". Many centuries afterwards the word geodesy was set in use anew, to denote the determination of the shape of initially parts of the earth surface and eventually, with the advent of space methods, the shape of the whole earth. Thus it remained an applied science, while facing at the same time significant and challenging theoretical problems, in both physical modeling and data analysis methodology.

From early times the relevance of the gravity field to the determination of shape has been evident once applications exceeded the bounds of small areas where the direction of the gravity field can be considered as constant within the bounds of the achieved observational accuracy. Shape can be derived from location (or relative location to be more precise) and the description of location by coordinates is not uniform on the surface of the earth where there exists locally a distinct direction: the vertical direction of the gravity force. Thus "height" is distinguished from the other "horizontal" coordinates.

The relevance of the gravity field manifests itself in two ways: from the need to use heights and from the very definition of "shape". Of course, it is possible (and now a days even practically attainable) to consider the shape of the natural surface of the earth and furthermore to treat location (as a means of defining shape) in a uniform way, e.g. by a global cartesian coordinate system. However, even in ancient times it was obvious that there is another type of "shape" definition, the one that was implicit in disputes about the possibility of the earth being either flat or spherical. In this notion of shape the mountains are "removed" as a type of vertical deviations from true shape. This concept took a specific meaning by considering the shape of the geoid, which is one of the equipotential surfaces of the earth. In fact it is the one that would coincide with the sea-level of an idealized uniformly rotating earth (both in direction and speed), without the influence of external and internal disturbing forces, such as

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 17–92, 2000.

the attraction of the moon, sun and planets, winds, currents, variations in atmospheric pressure and sea-water density, etc.

Even if one would insist to separate the "geometric" problem from the "physical" one, by sticking to the determination of the shape of the earth surface, he would eventually come to the problem that there is a need for a physical rather than a geometric definition of height. And this is so, not only for the psychological reasons which relate "up" and "down" to the local horizon (plane normal to the direction of the gravity vector), but for some more practical ones: In most applications requiring heights, one is hardly concerned with how far a point is from the center of the earth or from a spherical, or ellipsoidal, or even a more complicated reference-surface. What really matters is the direction of water flow: a point is higher from another if water can flow from the first to the second. This means that we need a "physical" rather than a geometrical definition of height h which is a monotonically decreasing function of the potential W, i.e. such that $W_P > W_Q \Leftrightarrow h_P < h_Q$. (Note that potential in geodesy has opposite sign that in physics, thus vanishing at infinite distance from the earth.)

The above choice between pure geometry and physics was made possible with the development of space techniques, but even so (in most observation techniques) the gravity field is still present, as the main driving force that shapes the orbits of the observed satellites. In the historical development of geodesy, the presence of the gravity field has been unavoidable for practical reasons due to the type of ground based observations. Geometric type of observations (angles and distances) determine horizontal relative position with sufficient accuracy, but they are very weak in the determination of the vertical component, as a result of the disturbing influence of atmospheric refraction. One had to rely to separate leveling observations, which produce increments in the local direction of the vertical that could not be added directly in a meaningful way to produce height differences. They can be added only after being converted to potential differences utilizing knowledge of the local value of gravity (modulus of the gravity vector). Thus the determination of the gravity field of the earth became, from the very beginning, an integral part of geodesy. Let us note that in view of the Dirichlet principle of potential theory (potential is uniquely defined by its values on a boundary surface), the determination of the external gravity field of the earth coincides with the determination of the shape of the geoid.

The determination of the shape of the earth surface has been associated with the iterative densification of points, starting with fundamental control networks. The determination of the shape of independent national or even continental networks left unsolved the problem of relating them to each other or to the earth as a whole. The use of connecting observations was not possible over the oceans and had to wait for the advent of space techniques. However an element of network location, that of orientation, could be determined by astronomical observations, which have already played a crucial role for determining "position" in navigation. Such observations could determine the direction of the local vertical with respect to the stellar background (an inertial reference frame with orientation but no position) and finally to the earth itself, provided that the orientation of the earth could be independently determined as a function of time. In addition the determined vertical directions provided an

additional source of information for the determination of the gravity field, the basic source being gravity observations using gravimeters.

Thus the rotation of the earth has been another "unknown" function that entered geodetic methodology, although, unlike the gravity potential function, it was not included in the objectives of geodesy. Its determination was based mainly on theory and was realized outside the geodetic discipline.

The enormous leap from ground-based to space techniques and the improvement of observational accuracy resulted into profound changes in both geodetic practice and theory. First of all the traditional separation in "horizontal" and "vertical" components is not strictly necessary any more and geodesy becomes truly three-dimensional. Furthermore the earth cannot be considered a rigid body any more, even if periodic variations (e.g. tides) are independently computed and removed. The shape of the earth surface and the gravity field of the earth must be now considered as functions, not only of shape, but also of time. Thus geodesy becomes four-dimensional.

Another important aspect is that the theoretical determination of earth rotation is not sufficient in relation to observational accuracy, and an independent empirical determination must be made from the analysis of the same observations that are used for geodetic purposes. As in the case of the abandoned astronomical observations, earth rotation is present in observations carried from points on the earth, because it relates an earth-attached reference frame to an inertial frame, in which Newton's laws hold and determine the orbits of satellites, or to an inertial frame to which the directions of radio sources involved in VLBI observations refer. Therefore the determination of earth rotation should be formally included in the definition of the object of geodesy.

The old choice between geometry and physics has been replaced now with a choice between geo-kinematics and geodynamics. Geo-kinematics refers to the determination of the temporal variation of the shape and orientation of the earth from observations alone without reference to the driving forces. (It could be compared to the determination of the gravity field from observations alone without reference to the real cause, the distribution of density within the earth.)

Geodynamics relies in addition to the observational evidence to models that take into account the relevant driving forces, such as the equations of motion of an elastic earth and the dynamic equations of earth rotation. In this respect earth rotation and deformation interact and cannot be treated separately. In a similar way temporal variations of the gravity field are interrelated with deformation. To the concept of deformation one should add the self-induced oscillations of a pulsating earth.

Geo-kinematics might become a geodetic field, but geodynamics will remain an interdisciplinary field involving different disciplines of geophysics.

2 The art of modeling

A model is an image of reality, expressed in mathematical terms, in a way, which involves a certain degree of abstraction and simplification.

The characteristics of the model, i.e., that part of reality included in it and the degree of simplification involved, depends on a particular purpose which is to use certain

observations in order to predict physical phenomena or parameters without actually having to observe them.

The models consists of an adopted set of mathematical relations between mathematical objects (e.g. parameters or functions) some of which may be observed by available measurement techniques, so that all other objects can be predicted on the basis of the specific values of the observables.

A specific model involves: a set of objects to be observed (observables), a set of objects (unknowns) such that any other object of the model can be conveniently predicted when the unknowns are determined. We may schematically denote such a specific model by a set of mathematical relations f connecting unknowns x and observables y, $f(x,y)=0$, or more often of the straightforward form $y=f(x)$. We may even select the unknowns to be identical with the observables, e.g. conditions equations ($x=y$, $f(y)=0$) in network adjustment, but in general to have a more economic description by not using "more" unknowns than what is really needed for the prediction of other objects $z=g(x)$.

When measurements are carried out the resulting observations do not fit the mathematical model, a result which is a direct consequence of the abstraction and simplifications present in the model, in relation to the complexity of reality. Thus one has to find a balance between the economy of the model and the fit of the observations to the observables. The discrepancy of this fit is usually described in a statistical manner. The degree of fit is a guiding factor in deciding on the formulation of the proper model. The standard approach is to label the discrepancies $v=b-y$ between observations b and observables y, as observational errors and to view them as random variables. Their behavior is described only in the average by probabilistic tools. In this way our final mathematical model consists of a functional part $y=f(x)$ or rather $b=f(x)+v$ and a stochastic model which consists of the probability distribution of the errors v, or at least some of its characteristics (zero mean and covariance).

Depending on the particular problem, the complexity of the model may vary from very simple such as the geometry of plane triangles to very complex as the mathematical description of the dynamics of the interaction of the solid earth with atmosphere, oceans and ice.

This may imply either discrete models involving a finite number of unknowns (geodetic networks) or continuous models, which in principle involve an infinite number of parameters (gravity field).

These basic ideas about modeling were realized, within the geodetic discipline at least, long ago. Bruns (1876) for example has given thorough consideration to the modeling problem in relation to the geodetic observational techniques available at his time.

From his analysis one can derive three principle geodetic objectives where varying available observations call for corresponding models of progressing complexity:

a. Measurements that determine the *geometric shape* of a small segment, or a large part of the earth's surface, or of the earth as a whole. This is the famous Bruns *polyhedron*, or in more modern language a *form element*, as introduced

by Baarda (1973), or a geometric object. If all measurements that define such a form element have been carried out, i.e. if no configuration defect exists, it may be visualized as a wire skeleton. It is typical that this wire frame is uniquely defined in its "internal" geometric shape, but it may be turned or shifted, as we like. The shape may be more rigid in some parts or in some internal directions than in others, depending on the strength, i.e. precision of the incorporated measurements. Typical measurements that belong to this first category would be angles and distances.

b. A second category of measurements, astronomical latitude, longitude and azimuth, provide the *orientation* of geometric form elements *in earth space*, after transformation of their original orientation with respect to fixed stars. They allow orienting form elements relative to each other or absolutely on the globe. They do not allow however to fix them in absolute position.

c. Finally there exists a third group of measurements that in their essence are employed in order to determine the *difference in gravity potential* between the vertices of the form elements. With this third category the direction and intensity of the flow of water is decided upon. Only these measurements tell us which point is up and which one is down not in a geometric but in a meaningful physical sense. They are derived from the combination of leveling and gravity.

From an analysis of the measurement techniques at their precision he deduces two fundamental limitations of geodesy at his time. First, zenith distances, which are subject to large distortions caused by atmospheric refraction, did not allow a geometrically strong determination of the polyhedron in the vertical direction. Second, oceans were not accessible to geodetic observations and therefore global geometric coverage was not attainable.

In our opinion, it is to the same limitations that one should attribute the intense development of geodetic boundary value problems as an indirect means of strengthening position in the vertical direction. It was only after a period of silence that Hotine (1969) revived the idea of three-dimensional geodesy.

Meanwhile, space techniques have changed the face of geodesy and the two fundamental limitations identified by Bruns do not hold any more. The global polyhedron is a reality and the shape of the oceans is determined at least as accurate and dense as the continents.

Nevertheless, the above three objectives are still the three fundamental geodetic tasks, with certain modifications in model complexity, which the increase of observational accuracy made possible.

Today measurement precision has reached 10^{-8} to 10^{-9}, relatively. This means for example that distances of 1000 km can be measured accurately to 1 cm, variations to the length of day to 1 msec, or absolute gravity (9.8 m/s^2) to 10^{-8} m/s^2 or 1 μgal. At this level of precision the earth is experienced as a deformable and pulsating body, affected by the dynamics of the earth's interior, ice masses, circulating oceans, weather and climate and sun moon and planets.
This implies for the three objectives discussed above:

a. Rigid geometric form elements become deformable: *geo-kinematics*, reflecting phenomena such as plate motion, inter-plate deformation, subsidence, sea level rise or fall, post-glacial uplift, or deformation of volcanoes.

b. Irregularities in polar motion, earth rotation and nutation occur in response to time variable masses and motion, inside the earth, as well as, between earth system components.

c. The analysis of potential differences between individual points has changed to the detailed determination of the gravity field of the earth as a whole including its temporal variations.

In short one could say that in addition to the three spatial dimensions and gravity, geodesy has conquered one more dimension: time. In addition it has extended from its continental limitations to the oceans.

Helmert (1880, p. 3) once stated that in the case of the topographic relief there are no physical laws that govern its shape, in a way that would allow one to infer one part from another, which means that they are not susceptible to modeling in the above sense. Therefore topography is determined by taking direct measurements of all its characteristic features.

On the other hand, the global shape of the earth as determined by gravitation and earth rotation is governed by physical laws which are simple enough to allow a general, albeit approximate, mathematical description.

Over the years the earth's gravity field became known to such great detail that one may rightly argue that further progress in its representation can achieved only by direct measurement.

As a consequence, local techniques would have to replace or complement global gravity representation methods. On the other hand representation of topographic features may be more and more supplemented by methods that provide a description by means of the underlying physical laws, such as ocean circulation models describing and predicting dynamic ocean topography, or plate tectonic models forecasting the temporal change of topographic features. Dynamic satellite orbit computation is a geodetic example that beautifully explains the interaction between direct measurement and modeling by physical laws and how this interaction changed during the past thirty years.

Let us now give the basic characteristics of model building as it applies to geodesy. It seems to be a fundamental strength of geodetic practice that the task of determining unknown parameters or functions is intrinsically connected with the determination of a complementing measure of accuracy. From the need to fulfill a certain accuracy requirement, a guideline is derived for the choice of the observations, the functional and the stochastic model, which we will discuss next.

(a) *Observables* y are model parameters, which we choose to measure on the basis of their informational content, and of course the availability of instrumentation. The numerical output of the *measurement* or "observation" process are the *observations* b, which being a mapping from the real word to numbers, they do not coincide with the corresponding observables which exist only in the model world. The observations are the link between the model world and the real world that it as-

pires to describe. The choice of observables has also a direct effect on the model because the measuring process is a part of the real world that needs its counterpart in the model world.

(b) *Functional models* are mathematical relations between the observables and the chosen unknown parameters. Its purpose is to select from the large arsenal geometric and physical laws the specific mathematical set of equations that best describes the particular problem. Examples are plane or spatial point networks, satellite orbits, earth rotation, gravity field representation. With each of the chosen models goes an appropriate choice of parametrization.

(c) *Stochastic models* are introduced to deal with the misclosure between observations and observables. These misclosures have two parts stemming from different causes: discrepancies between reality and functional model (modeling or systematic errors) and discrepancies between observables and actually observed quantities, due to imperfect instrument behavior. These discrepancies cannot be described directly but only in their average behavior within an assembly of virtual repetitions under seemingly identical conditions. The observations at hand which are concrete numbers and thus quite deterministic, are modeled to be samples of corresponding random variables. Unfortunately they are also called "observations" and they are denoted by the same symbols, as a matter of standard practice. One has to distinguish the different concepts from the context. Thus a stochastic model consists of the following

(1) $\mathbf{b}$ a large ensemble of repetitions.

(2) The first two moments (mean and covariance) which provide full description of the stochastic behavior. Usually the normal distribution is assumed

$$\mathbf{b} \sim N(\mathbf{y},\mathbf{C}) \quad \Leftrightarrow \quad f_{\mathbf{b}}(\mathbf{b}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}|}} \exp\{-\frac{1}{2}(\mathbf{b}-\mathbf{y})^T \mathbf{C}^{-1}(\mathbf{b}-\mathbf{y})\} . \qquad (2.1)$$

(3) $\mathbf{b}=\mathbf{y}+\mathbf{v}$

$$E\{\mathbf{b}\}=\mathbf{y} , \; E\{(\mathbf{b}-\mathbf{y})(\mathbf{b}-\mathbf{y})^T\}=\mathbf{C} \quad \Leftarrow \quad E\{\mathbf{v}\}=\mathbf{0} , \; E\{\mathbf{v}\mathbf{v}^T\}=\mathbf{C} . \qquad (2.2)$$

Much could be added to these three assumptions. Let us add the following two remarks:

In geodetic practice hardly ever many repetitions of the measurements are carried out. A registration describing the specific circumstances of the measurements is helpful to relate the stochastic behavior of the field sample, to a large set of empirical distribution functions build up in the laboratory/calibration experiments (e.g. by the instrument manufacturer).

Assumption (3) establishes in a unique manner the connection between observations and observables.

Often the magnitude of the errors is too large to be acceptable. In this case, two options are available. Either one improves the functional model, by bringing the observations closer to the observables by reductions; this is the process of applying a

known correction to the original observations (atmospheric corrections, clock corrections, relativistic corrections). Or the functional model is extended so that the observables come closer to the observations, that is closer to the real world.

Typically the (finite-dimensional) functional model is not linear $\mathbf{y}=\mathbf{f}(\mathbf{x})$ which apart from the obvious numerical difficulties, has the disadvantage that no consistent probabilistically optimal estimation theory exists for parameter determination. On the other hand a set of good approximate values $\mathbf{x}^0$ is usually available for the unknown parameters. In this case a satisfactory linear approximation can be used, based on Taylor expansion to the first order

$$\mathbf{y}=f(\mathbf{x}^0)+\left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}=x^0}(\mathbf{x}-\mathbf{x}^0)\equiv\mathbf{y}^0+\mathbf{A}\delta\mathbf{x}, \qquad \delta\mathbf{b}=\mathbf{b}-\mathbf{y}^0=\mathbf{A}\delta\mathbf{x}+\mathbf{v} \tag{2.3}$$

or simply $\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{v}$, with notational simplification.

An other unique feature of geodetic modeling is the use of unknown parameters (coordinates) which describe more (shape and position) that the observations can really determine (shape). The additional information is introduced in an arbitrary way (datum choice) but one has to be careful and restrict the prediction of other parameters $\mathbf{z}=\mathbf{g}(\mathbf{x})$ to those, which remain unaffected by this arbitrary choice.

3 Parameter estimation as an inverse problem

Let us assume that we have a linear(ized) finite-dimensional model of the form $\mathbf{y}=\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a known $n\times m$ matrix, $\mathbf{x}$ is the $m\times 1$ vector of unknown parameters and $\mathbf{y}$ is the $n\times 1$ vector of unknown observables, to which a known $n\times 1$ vector $\mathbf{b}$ of available observations corresponds. The objective of a (linear) estimation procedure is to construct an "optimal" inverse mapping represented by an $m\times n$ inverse matrix $\mathbf{G}=\mathbf{A}^-$, which maps the data into an estimate of the unknown parameters $\hat{\mathbf{x}}=\mathbf{G}\mathbf{b}=\mathbf{A}^-\mathbf{b}$. (Note that we avoid writing down the model in the form of the equation $\mathbf{b}=\mathbf{A}\mathbf{x}$, which is not satisfied in the general case.)

The choice of the parameters $\mathbf{x}$ is a matter of convenience and any other set $\mathbf{x}'=\mathbf{S}^{-1}\mathbf{x}$, where $\mathbf{S}$ is a non-singular ($|\mathbf{S}|\neq 0$) $m\times m$ matrix is equally acceptable. In a similar way we may replace the original set of observables and observations with new sets $\mathbf{y}'=\mathbf{T}^{-1}\mathbf{y}$, $\mathbf{b}'=\mathbf{T}^{-1}\mathbf{b}$, where $\mathbf{T}$ is a non-singular ($|\mathbf{T}|\neq 0$) $n\times n$ matrix. In fact in many situations we do not use the original observations but such transformed "synthetic" observations.

With respect to the transformed quantities the model takes the form

$$\mathbf{y}'=\mathbf{T}^{-1}\mathbf{y}=\mathbf{T}^{-1}\mathbf{A}\mathbf{x}=\mathbf{T}^{-1}\mathbf{A}\mathbf{S}\mathbf{S}^{-1}\mathbf{x}=\mathbf{T}^{-1}\mathbf{A}\mathbf{S}\mathbf{x}'=\mathbf{A}'\mathbf{x}' \tag{3.1}$$

where

$$\mathbf{A}'=\mathbf{T}^{-1}\mathbf{AS}\,, \qquad \mathbf{A}=\mathbf{TA}'\mathbf{S}^{-1}\,. \tag{3.2}$$

One should respect that the estimation procedure is not affected by these choices and has the invariance property

$$\hat{\mathbf{x}}'=\mathbf{G}'\mathbf{b}'=\mathbf{S}^{-1}\hat{\mathbf{x}}=\mathbf{S}^{-1}\mathbf{Gb}=\mathbf{S}^{-1}\mathbf{GTT}^{-1}\mathbf{b}=\mathbf{S}^{-1}\mathbf{GTb}' \quad \Rightarrow \quad \mathbf{G}'=\mathbf{S}^{-1}\mathbf{GT}\,. \tag{3.3}$$

This means that the inverse $\mathbf{G}$ must be constructed in such a way that whenever $\mathbf{A}$ transforms into $\mathbf{A}'=\mathbf{T}^{-1}\mathbf{AS}$, $\mathbf{G}$ transforms accordingly into $\mathbf{G}'=\mathbf{S}^{-1}\mathbf{GT}$.

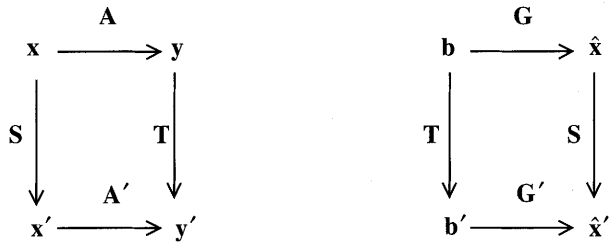


Fig. 1: Invariance property of the linear estimator $G:Y\to X:b\to\hat{x}$ for the model $A:X\to Y:x\to y$, $b=y+v\in Y$, with respect to different representations in both X and Y, connected by non-singular transformation matrices $\mathbf{S}$ and $\mathbf{T}$.

We could also add translations to the above transformations, but this will introduce more complicated equations, which would have nothing to offer to the essential point that we are after.

When the spaces X and Y of the model ($\mathbf{x}$ and $\mathbf{y}$ being representations of $x\in X$ and $y\in Y$, respectively) have metric properties described by the inner products

$$(x_a, x_b)=\mathbf{x}_a^T\mathbf{P}\,\mathbf{x}_b\,, \qquad (y_a, y_b)=\mathbf{y}_a^T\mathbf{Q}\,\mathbf{y}_b \tag{3.4}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are the respective metric (weight) matrices, these must transform into

$$\mathbf{Q}'=\mathbf{S}^T\mathbf{QS}\,, \qquad \mathbf{P}'=\mathbf{T}^T\mathbf{PT}\,, \tag{3.5}$$

if the metric properties are to remain unchanged, e.g.

$$(y_\alpha, y_\beta)=\mathbf{y}'^T_\alpha\mathbf{P}'\mathbf{y}'_\beta=(\mathbf{T}^{-1}\mathbf{y}_\alpha)^T(\mathbf{T}^T\mathbf{PT})(\mathbf{T}^{-1}\mathbf{y}_\beta)=$$

$$=\mathbf{y}_\alpha^T\mathbf{T}^{-T}\mathbf{T}^T\mathbf{PTT}^{-1}\mathbf{y}_\beta=\mathbf{y}_\alpha^T\mathbf{Py}_\beta=(y_\alpha, y_\beta)\,. \tag{3.6}$$

This is a purely algebraic point of view and we may switch to a geometric one by setting

$$\mathbf{I}_m=[\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_m], \qquad \mathbf{I}_n=[\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2\cdots\boldsymbol{\varepsilon}_n], \tag{3.7}$$

where $\mathbf{I}_m$ and $\mathbf{I}_n$ are the $m\times m$ and $n\times n$ identity matrices, respectively. We say in this case that the $m\times 1$ vectors $\mathbf{e}_i$ having all elements zero except the i$^{\text{th}}$ one, which has the value 1, constitute a "natural" basis for the m-dimensional space X. A similar statement holds for the vectors $\boldsymbol{\varepsilon}_i$ in the n-dimensional space Y. We may set

$$x\equiv\sum_{i=1}^{m}x_i\mathbf{e}_i=[\mathbf{e}_1\mathbf{e}_2\cdots\mathbf{e}_m]\begin{bmatrix}x^1\\ \vdots\\ x^m\end{bmatrix}=\mathbf{I}_m\mathbf{x}=\mathbf{x}, \tag{3.8}$$

$$y\equiv\sum_{i=1}^{n}y_i\boldsymbol{\varepsilon}_i=[\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2\cdots\boldsymbol{\varepsilon}_n]\begin{bmatrix}y^1\\ \vdots\\ y^n\end{bmatrix}=\mathbf{I}_n\mathbf{y}=\mathbf{y}. \tag{3.9}$$

We may now view $\mathbf{x}$ and $\mathbf{y}$ as representations of the abstract elements x and y, respectively, with respect to the above choice of "natural" bases. We may change both bases and obtain different representations of x and y. Thus x stands for all the equivalent sets of parameter choices and y for all the equivalent sets of observables. The same can be said about the observations where $b\equiv\sum_i b_i\boldsymbol{\varepsilon}_i$ stands for all the equivalent sets of (synthetic) observations.

Remark:

It is usual when dealing with finite dimensional models to start with a particular representation $\mathbf{y}=\mathbf{A}\mathbf{x}$, where parameters $\mathbf{x}$, observables $\mathbf{y}$ and observations $\mathbf{b}$ have concrete physical meanings. In this case the corresponding abstract counterparts x, y and b simply stand for the whole set of alternative parametrizations (some of them with a physical meaning, most without any) and of alternative choices of synthetic observations. On the contrary when dealing with infinite dimensional problems the situation is usually the opposite, as far as parameterization is concerned. For example, when analyzing observations related to the study of the gravity field, we start with a more general unknown x, the gravity potential function. The (more or less arbitrary) choice of a specific basis e_i, i=1,2,..., (e.g. spherical harmonic functions) turns this function into a corresponding parameter vector $\mathbf{x}=(x_1,x_2,\ldots)$ (e.g. spherical harmonic coefficients), so that $x=x_1e_1+x_2e_2+\ldots$.

We consider a change of bases

$$\mathbf{e}'_k=\sum_{i=1}^{m}V_{ik}\mathbf{e}_k=[\mathbf{e}_1\cdots\mathbf{e}_m]\begin{bmatrix}S_{1k}\\ \vdots\\ S_{mk}\end{bmatrix}\quad\Leftrightarrow\quad[\mathbf{e}'_1\cdots\mathbf{e}'_m]=[\mathbf{e}_1\cdots\mathbf{e}_m]\mathbf{S}\tag{3.10}$$

$$\boldsymbol{\varepsilon}'_k=\sum_{i=1}^{n}U_{ik}\boldsymbol{\varepsilon}_k=[\boldsymbol{\varepsilon}_1\cdots\boldsymbol{\varepsilon}_n]\begin{bmatrix}T_{1k}\\ \vdots\\ T_{nk}\end{bmatrix}\quad\Leftrightarrow\quad[\boldsymbol{\varepsilon}'_1\cdots\boldsymbol{\varepsilon}'_n]=[\boldsymbol{\varepsilon}_1\cdots\boldsymbol{\varepsilon}_m]\mathbf{T}\tag{3.11}$$

with respect to which we have the representations

$$x=\sum_{i=1}^{m}x'_i\mathbf{e}'_i=[\mathbf{e}'_1\cdots\mathbf{e}'_m]\mathbf{x}'=[\mathbf{e}_1\cdots\mathbf{e}_m]\mathbf{S}\mathbf{x}'=\mathbf{S}\mathbf{x}'=\mathbf{x}\,,\tag{3.12}$$

$$y=\sum_{i=1}^{n}y'_i\boldsymbol{\varepsilon}'_i=[\boldsymbol{\varepsilon}'_1\cdots\boldsymbol{\varepsilon}'_n]\mathbf{y}'=[\boldsymbol{\varepsilon}_1\cdots\boldsymbol{\varepsilon}_n]\mathbf{T}\mathbf{y}'=\mathbf{T}\mathbf{y}'=\mathbf{y}\,,\tag{3.13}$$

where the above relations $\mathbf{S}\mathbf{x}'=\mathbf{x}$ and $\mathbf{T}\mathbf{y}'=\mathbf{y}$, expressing the change in coordinates due to change in bases, are equivalent to the formerly introduced respective algebraic transformations $\mathbf{x}'=\mathbf{S}^{-1}\mathbf{x}$ and $\mathbf{y}'=\mathbf{T}^{-1}\mathbf{y}$.

Since any basis is as good as any other, the questions arise whether there exists an optimal choice of bases $\{\mathbf{e}'_i\}$ and $\{\boldsymbol{\varepsilon}'_i\}$, such that the mapping A represented by the matrix $\mathbf{A}$ in the original bases, is represented in the new ones by a matrix $\mathbf{A}'$, which is of such a simple form that makes the properties of the mapping A very transparent and furthermore allows an easy construction of the inverse matrix $\mathbf{G}'$ which represents a mapping G "inverse" to the mapping A.

In the case of a bijective mapping ($n=m=r$) we may resort to the eigenvalues and eigenvectors of the matrix $\mathbf{A}$, defined by

$$\mathbf{A}\mathbf{u}_i=\lambda_i\mathbf{u}_i\,,\qquad i=1,2,\ldots,n\,.\tag{3.14}$$

Setting

$$\mathbf{U}=[\mathbf{u}_1\cdots\mathbf{u}_n]\,,\qquad\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&\lambda_n\end{bmatrix},\tag{3.15}$$

these equations can be combined into a single one

$$\mathbf{AU}=\mathbf{A}[\mathbf{u}_1\cdots\mathbf{u}_n]=[\lambda_1\mathbf{u}_1\cdots\lambda_n\mathbf{u}_n]=[\mathbf{u}_1\cdots\mathbf{u}_n]\begin{bmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{bmatrix}=\mathbf{U}\boldsymbol{\Lambda} \tag{3.16}$$

and

$$\mathbf{A}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}, \qquad \boldsymbol{\Lambda}=\mathbf{U}^{-1}\mathbf{A}\mathbf{U}. \tag{3.17}$$

The problem is that in general the entries of the matrices $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are complex numbers, except for the special case where $\mathbf{A}$ is a symmetric matrix, in which case eigenvalues and eigenvectors are real. The eigenvalues form an orthogonal and by proper scaling an orthonormal system, so that $\mathbf{U}$ is orthogonal and $\mathbf{A}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$. Applying the change of bases (3.10-3.11) with $\mathbf{T}=\mathbf{S}=\mathbf{U}$, we have

$$[\mathbf{e}'_1\cdots\mathbf{e}'_n]=[\mathbf{e}_1\cdots\mathbf{e}_n]\mathbf{U}, \qquad [\boldsymbol{\varepsilon}'_1\cdots\boldsymbol{\varepsilon}'_n]=[\boldsymbol{\varepsilon}_1\cdots\boldsymbol{\varepsilon}_n]\mathbf{U}, \tag{3.18}$$

with new representations ($\mathbf{U}^{-1}=\mathbf{U}^T$)

$$\mathbf{x}'=\mathbf{U}^T\mathbf{x}, \quad \mathbf{y}'=\mathbf{U}^T\mathbf{y}, \quad \mathbf{b}'=\mathbf{U}^T\mathbf{b}, \qquad \mathbf{A}'=\mathbf{U}^T\mathbf{A}\,\mathbf{U}=\mathbf{U}^T(\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T)\mathbf{U}=\boldsymbol{\Lambda} \tag{3.19}$$

and the system becomes in the new bases

$$\mathbf{b}'=\mathbf{A}'\mathbf{x}'=\boldsymbol{\Lambda}\mathbf{x}', \qquad b'_i=\lambda_i x'_i \tag{3.20}$$

with a simple solution

$$\hat{\mathbf{x}}'=\boldsymbol{\Lambda}^{-1}\mathbf{b}', \qquad \hat{x}'_i=\frac{1}{\lambda_i}b'_i, \tag{3.21}$$

that always exists and is unique. We have managed to simplify and solve the original equation $\mathbf{b}=\mathbf{A}\mathbf{x}$ by employing the *eigenvalue decomposition (EDV)* $\mathbf{A}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ of the symmetric matrix $\mathbf{A}$. The required inverse in this case is $\mathbf{G}'=\boldsymbol{\Lambda}^{-1}=(\mathbf{A}')^{-1}$ in the new bases and $\mathbf{G}=\mathbf{U}^T\boldsymbol{\Lambda}^{-1}\mathbf{U}=\mathbf{A}^{-1}$ in the original ones, since $\mathbf{GA}=\mathbf{AG}=\mathbf{I}$ as it can be easily verified.

A choice of the proper transformation matrices $\mathbf{T}$ and $\mathbf{S}$, leading to a very simple form of the matrix $\mathbf{A}'$ representing the same operator A (which is represented by the matrix $\mathbf{A}$ in the original bases), is based on the so called *singular value decomposition* (Lanczos, 1961, Schaffrin et al, 1977), presented in Appendix A.

With tools presented therein we attack now the estimation or inversion problem from a deterministic point of view, where we first treat the most general case and then specialize to simpler special cases.

3.1 The general case : Overdetermined and underdetermined system without full rank ($r<\min(n,m)$)

As shown in Appendix A, the operator A represented in the original bases by the matrix

$$\underset{n\times m}{\mathbf{A}}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T}\ \underset{m\times m}{\mathbf{Q}}=\begin{bmatrix}\underset{n\times r}{\mathbf{U}_1} & \underset{n\times f}{\mathbf{U}_2}\end{bmatrix}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{m\times r}{\mathbf{V}_1} & \underset{m\times d}{\mathbf{V}_2}\end{bmatrix}^T\underset{m\times m}{\mathbf{Q}}=\mathbf{U}_1\mathbf{\Lambda}\,\mathbf{V}_1^T\mathbf{Q} \tag{3.22}$$

where $f=n-r$ and $d=m-r$. The singular value decomposition (SVD) is defined by means of the transformations

$$\mathbf{x}'=\mathbf{V}^{-1}\mathbf{x}=\mathbf{V}^T\mathbf{Q}\mathbf{x}\,,\quad \mathbf{y}'=\mathbf{U}^{-1}\mathbf{y}=\mathbf{U}^T\mathbf{P}\mathbf{y}\,. \tag{3.23}$$

The operator A is represented with respect to the new (SVD) bases by the matrix

$$\underset{n\times m}{\mathbf{A}'}=\mathbf{U}^{-1}\mathbf{A}\mathbf{V}=\mathbf{U}^T\mathbf{P}\mathbf{A}\mathbf{V}=\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}, \tag{3.24}$$

where the matrices $\mathbf{U}$ and $\mathbf{V}$ have as columns the eigenvectors of the matrices $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$, respectively, ordered by descending eigenvalue magnitude, where $\mathbf{A}^*$ is the adjoint matrix of $\mathbf{A}$, representing the adjoint operator A^* defined by

$$(y,Ax)=(x,A^*y)\qquad\Leftrightarrow\qquad \mathbf{x}^T\mathbf{A}^T\mathbf{P}\mathbf{y}=\mathbf{x}^T\mathbf{Q}\mathbf{A}^*\mathbf{y} \tag{3.25}$$

and consequently

$$\mathbf{A}^*=\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P}\,. \tag{3.26}$$

The diagonal elements $\Lambda_{ii}=\lambda_i$ of the diagonal matrix $\mathbf{\Lambda}$ are the square roots of the non-vanishing common eigenvalues λ_i^2 of $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$. The complete SVD (Singular Value Decomposition) relations, as derived in Appendix A, are

$$\mathbf{A}\mathbf{V}=\mathbf{U}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\qquad \mathbf{A}^*\mathbf{U}=\mathbf{V}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{3.27}$$

$$(\mathbf{A}\mathbf{A}^*)\mathbf{U}=\mathbf{U}\begin{bmatrix}\mathbf{\Lambda}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\qquad (\mathbf{A}^*\mathbf{A})\,\mathbf{V}=\mathbf{V}\begin{bmatrix}\mathbf{\Lambda}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{3.28}$$

accompanied by the orthogonality relations

$$\mathbf{U}^T\mathbf{P}\mathbf{U}=\mathbf{I}, \qquad \mathbf{U}\mathbf{U}^T=\mathbf{P}^{-1}, \qquad \mathbf{V}^T\mathbf{Q}\mathbf{V}=\mathbf{I}, \qquad \mathbf{V}\mathbf{V}^T=\mathbf{Q}^{-1}. \tag{3.29}$$

which yield the inverse transformations from the SVD to the original bases

$$\mathbf{x}=\mathbf{V}\mathbf{x}', \qquad \mathbf{y}=\mathbf{U}\mathbf{y}'. \tag{3.30}$$

Note that as a consequence of (3.29) and (3.5), where now $\mathbf{S}=\mathbf{V}$ and $\mathbf{T}=(\mathbf{U}^T\mathbf{P})^{-1}=\mathbf{U}$, the weight matrices in the SVD system are $\mathbf{Q}'=\mathbf{V}^T\mathbf{Q}\mathbf{V}=\mathbf{I}$ and $\mathbf{P}'=\mathbf{U}^T\mathbf{P}\mathbf{U}=\mathbf{I}$.

For any element x the corresponding image $y=Ax$ has the SVD representation

$$\underset{n\times 1}{\mathbf{y}'}=\begin{bmatrix}\mathbf{y}'_1\\ r\times 1\\ \mathbf{y}'_2\\ f\times 1\end{bmatrix}=\mathbf{A}'\mathbf{x}'=\begin{bmatrix}\boldsymbol{\Lambda} & \mathbf{0}\\ r\times r & r\times d\\ \mathbf{0} & \mathbf{0}\\ f\times r & f\times d\end{bmatrix}\begin{bmatrix}\mathbf{x}'_1\\ r\times 1\\ \mathbf{x}'_2\\ d\times 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Lambda}\mathbf{x}'_1\\ \mathbf{0}\end{bmatrix}, \qquad \mathbf{y}'_1=\boldsymbol{\Lambda}\mathbf{x}'_1, \qquad \mathbf{y}'_2=\mathbf{0}. \tag{3.31}$$

Two conclusions can be drawn from the above relations:

(a) an arbitrary element $y\in Y$ is the image of some $x\in X$ (in other words the equation $y=Ax$ is consistent) only when $\mathbf{y}'_2=\mathbf{0}$ (consistency condition).

(b) When the vector x has an SVD representation with $\mathbf{x}'_1=\mathbf{0}$, its image vanishes, $Ax=0$ ($\mathbf{y}'=\mathbf{0}$).

The set of all $y\in Y$ which are the image of some $x\in X$ having SVD representations of the form $\mathbf{y}'=\begin{bmatrix}\mathbf{y}'_1\\ \mathbf{0}\end{bmatrix}$ constitute a linear subspace $R(A)\subset Y$ of dimension r (equal to the number of elements in $\mathbf{y}'_1$), which is called the *range* of the operator A.

Any element $z\in Y$ with SVD representation of the form $\mathbf{z}'=\begin{bmatrix}\mathbf{0}\\ \mathbf{z}'_2\end{bmatrix}$, is orthogonal to any element $y\in R(A)$, since

$$(z,y)=(\mathbf{z}')^T\mathbf{y}'=\begin{bmatrix}\mathbf{0}\\ \mathbf{z}'_2\end{bmatrix}^T\begin{bmatrix}\mathbf{y}'_1\\ \mathbf{0}\end{bmatrix}=\mathbf{0}\mathbf{y}'_1+(\mathbf{z}'_2)^T\mathbf{0}=0 \qquad \Rightarrow \qquad z\perp y. \tag{3.32}$$

The set of all such z orthogonal to $R(A)$ constitute a linear subspace $R(A)^\perp\subset Y$ of dimension f (equal to the number of elements in $\mathbf{z}'_2$) which is called the *orthogonal complement* of $R(A)$ with respect to Y.

Any vector $y\in Y$ can be uniquely decomposed into the sum of two elements

$$y=y_{R(A)}+y_{R(A)^{\perp}}, \qquad \mathbf{y}'=\mathbf{y}'_{R(A)}+\mathbf{y}'_{R(A)^{\perp}}=\begin{bmatrix}\mathbf{y}'_1\\ \mathbf{0}\end{bmatrix}+\begin{bmatrix}\mathbf{0}\\ \mathbf{y}'_2\end{bmatrix}, \tag{3.33}$$

where $y_{R(A)}\in R(A)$ and $y_{R(A)^{\perp}}\in R(A)^{\perp}$. We use in this respect symbolism

$$Y=R(A)\oplus R(A)^{\perp}. \tag{3.34}$$

The set of all $x\in X$ with vanishing images $Ax=0$, having SVD representations of the form $\mathbf{x}'=\begin{bmatrix}\mathbf{0}\\ \mathbf{x}'_2\end{bmatrix}$ constitute a linear subspace $N(A)\subset X$ of dimension d (equal to the number of elements in $\mathbf{x}'_2$) called the *null space* of A.

Any element $w\in X$ with SVD representation of the form $\mathbf{w}'=\begin{bmatrix}\mathbf{w}'_1\\ \mathbf{0}\end{bmatrix}$, is orthogonal to any element $x\in R(A)$, since

$$(w,x)=(\mathbf{w}')^T\mathbf{x}'=\begin{bmatrix}\mathbf{w}'_1\\ \mathbf{0}\end{bmatrix}^T\begin{bmatrix}\mathbf{0}\\ \mathbf{x}'_2\end{bmatrix}=(\mathbf{w}'_1)^T\mathbf{0}+\mathbf{0}\mathbf{x}'_2=0 \qquad \Rightarrow \qquad w\perp x. \tag{3.35}$$

The set of all such w orthogonal to $N(A)$ constitute a linear subspace $N(A)^{\perp}\subset X$ of dimension r (equal to the number of elements in $\mathbf{w}'_1$), which is called the *orthogonal complement* of $N(A)$ with respect to X.

Any vector $x\in X$ can be decomposed into the sum of two elements

$$x=x_{N(A)}+x_{N(A)^{\perp}}, \qquad \mathbf{x}'=\mathbf{x}'_{N(A)}+\mathbf{x}'_{N(A)^{\perp}}=\begin{bmatrix}\mathbf{0}\\ \mathbf{x}'_2\end{bmatrix}+\begin{bmatrix}\mathbf{x}'_1\\ \mathbf{0}\end{bmatrix}, \tag{3.36}$$

where $x_{N(A)}\in N(A)$ and $x_{N(A)^{\perp}}\in N(A)^{\perp}$. We use in this respect symbolism

$$X=N(A)\oplus N(A)^{\perp}. \tag{3.37}$$

We will construct a solution $\hat{x}=Gb$ to the estimation problem with $b\notin R(A)$ in two steps. In the first step we shall construct a consistent equation $\hat{y}=Ax$ by applying the *least squares* principle

$$\|b-\hat{y}\|=\min_{y\in R(A)}\|b-y\|, \tag{3.38}$$

i.e. by choosing the element $\hat{y} \in R(A)$ which is closest to the observed b among all the elements of $R(A)$. In the second step we will apply the *minimum norm* principle

$$\|\hat{x}\| = \min_{Ax=\hat{y}} \|x\| \tag{3.39}$$

i.e. by choosing among all (least-squares) solutions of the consistent equation $\hat{y}=Ax$ the one with minimum norm.

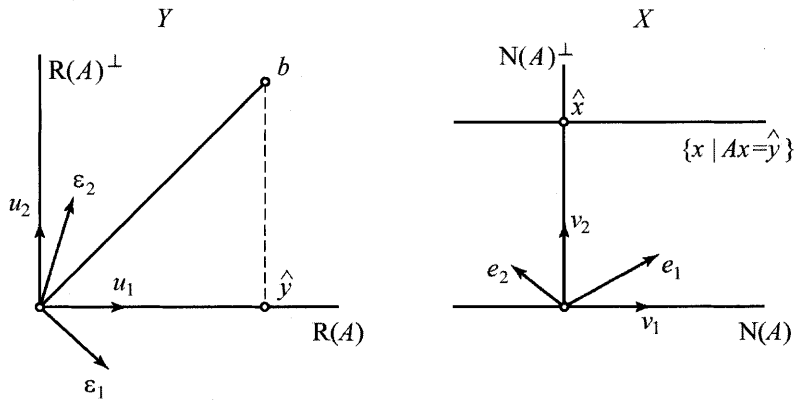


Fig. 2: An illustration of the two stage solution: least squares followed by minimum norm.

The application of the least squares principle in the SVD representation gives

$$\|\mathbf{b}'-\mathbf{y}'\|^2=(\mathbf{b}'-\mathbf{y}')^T(\mathbf{b}'-\mathbf{y}')=\begin{bmatrix}\mathbf{b}'_1-\mathbf{y}'_1\\ \mathbf{b}'_2\end{bmatrix}^T\begin{bmatrix}\mathbf{b}'_1-\mathbf{y}'_1\\ \mathbf{b}'_2\end{bmatrix}=$$

$$=(\mathbf{b}'_1-\mathbf{y}'_1)^T(\mathbf{b}'_1-\mathbf{y}'_1)+(\mathbf{b}'_2)^T\mathbf{b}'_2=\min \tag{3.40}$$

with obvious solution $\hat{\mathbf{y}}'_1=\mathbf{b}'_1$ or, in combination with the consistency condition $\hat{\mathbf{y}}'_2=\mathbf{0}$,

$$\hat{\mathbf{y}}'=\begin{bmatrix}\mathbf{b}'_1\\ \mathbf{0}\end{bmatrix}. \tag{3.41}$$

The consistent system $\hat{y}=Ax$ with SVD representation

$$\hat{\mathbf{y}}'=\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\mathbf{A}'\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\end{bmatrix}=\begin{bmatrix}\mathbf{\Lambda}\mathbf{x}_1'\\ \mathbf{0}\end{bmatrix} \tag{3.42}$$

has an infinite number of *least squares* solutions $\mathbf{x}'$ with $\mathbf{x}_1'=\mathbf{\Lambda}^{-1}\mathbf{b}_1'$, and arbitrary $\mathbf{x}_2'$. Any least squares solution

$$\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{x}_2'\end{bmatrix}=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}+\begin{bmatrix}\mathbf{0}\\ \mathbf{x}_2'\end{bmatrix}, \tag{3.43}$$

consists of a fixed component $\mathbf{x}_0'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}\in N(A)^{\perp}$, and an arbitrary component $\mathbf{x}_{N(A)}'=\begin{bmatrix}\mathbf{0}\\ \mathbf{x}_2'\end{bmatrix}\in N(A)$. The set of all least squares solutions constitute a linear variety $S_{\hat{y}}=x_0+N(A)$ of X, which is a parallel transport of the space $N(A)$ by the element represented by $\mathbf{x}_0'$ (see fig. 2). In a more general set up, for any $y\in R(A)$ we call $S_y=\{x\in X\,|\,Ax=y\}$ the *solution space* of the consistent equation $y=Ax$.

Equation (3.42) is the SVD representation of the *normal equations*, which every least squares solution must satisfy. They can be transformed into the original bases using $\mathbf{x}'=\mathbf{V}^T\mathbf{Q}\mathbf{x}$, $\mathbf{b}'=\mathbf{U}^T\mathbf{P}\mathbf{b}$ to yield

$$\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\end{bmatrix}=\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix} \quad\Rightarrow\quad \begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T\mathbf{Q}\mathbf{x}\\ \mathbf{V}_2^T\mathbf{Q}\mathbf{x}\end{bmatrix}=\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^T\mathbf{P}\mathbf{b}\\ \mathbf{U}_2^T\mathbf{P}\mathbf{b}\end{bmatrix} \quad\Rightarrow$$

$$\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q}\mathbf{x}=\mathbf{U}_1^T\mathbf{P}\mathbf{b}. \tag{3.44}$$

These are the normal equations, with respect to the original bases, to be satisfied by any least squares solution $\mathbf{x}$. Computing

$$\mathbf{A}^T\mathbf{P}\mathbf{A}=(\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q})^T\mathbf{P}\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q}=\mathbf{Q}\mathbf{V}_1\mathbf{\Lambda}^2\mathbf{V}_1^T\mathbf{Q} \tag{3.45}$$

$$\mathbf{A}^T\mathbf{P}\mathbf{b}=(\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q})^T\mathbf{P}\mathbf{b}=\mathbf{Q}\mathbf{V}_1\mathbf{\Lambda}\mathbf{U}_1^T\mathbf{P}\mathbf{b} \tag{3.46}$$

and multiplying (3.44) from the left with $\mathbf{Q}\mathbf{V}_1\mathbf{\Lambda}$ the normal equations obtain the more familiar form

$$(\mathbf{A}^T\mathbf{P}\mathbf{A})\mathbf{x}=\mathbf{A}^T\mathbf{P}\mathbf{b}. \tag{3.47}$$

Among all possible (least-squares) solutions (3.43) we shall select the one with minimum norm, which satisfies

$$\|\mathbf{x}'\|^2=(\mathbf{x}')^T\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{x}_2'\end{bmatrix}^T\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{x}_2'\end{bmatrix}=(\mathbf{\Lambda}^{-1}\mathbf{b}_1')^T(\mathbf{\Lambda}^{-1}\mathbf{b}_1')+(\mathbf{x}_2')^T\mathbf{x}_2'=\min. \tag{3.48}$$

Since the first summand is constant the obvious solution is $\hat{\mathbf{x}}_2'=\mathbf{0}$ and the *least-squares solution of minimum norm* becomes

$$\hat{\mathbf{x}}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix}=\mathbf{G}'\mathbf{b}' \tag{3.49}$$

where

$$\mathbf{G}'=\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{3.50}$$

Note that $\hat{\mathbf{x}}'$ is no other that the common $N(A)^{\perp}$-component x_0' of any least squares solution and the least-squares solution space can also be expressed as $S_{\hat{y}}=$ $=S_{P_{R(A)}b}=\hat{x}+N(A)$. In fact $S_{\hat{y}}\cap N(A)^{\perp}=\{\hat{x}\}$.

If we compute the matrices

$$\mathbf{P}'_{R(A)}=\mathbf{A}'\mathbf{G}=\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\qquad \mathbf{P}'_{N(A)^{\perp}}=\mathbf{G}'\mathbf{A}'=\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \tag{3.51}$$

we can easily establish the following four properties of the matrix $\mathbf{G}'$ (Rao and Mitra, 1971)

(G1) $$\mathbf{A}'\mathbf{G}'\mathbf{A}'=\mathbf{A}' \tag{3.52}$$

(G2) $$\mathbf{G}'\mathbf{A}'\mathbf{G}'=\mathbf{G}' \tag{3.53}$$

(G3) $$(\mathbf{A}'\mathbf{G}')^T=\mathbf{A}'\mathbf{G}' \tag{3.54}$$

(G4) $$(\mathbf{G}'\mathbf{A}')^T=\mathbf{G}'\mathbf{A}'. \tag{3.55}$$

Property (G1) is the defining property which characterizes $\mathbf{G}'$ as a *generalized inverse* of $\mathbf{A}'$. The additional property (G2) means that $\mathbf{A}'$ is in turn a generalized inverse of $\mathbf{G}'$. It characterizes $\mathbf{G}'$ as a *reflexive* generalized inverse.

Property (G3) in combination with (G1) characterizes $\mathbf{G}'$ as a *least squares* generalized inverse of $\mathbf{A}'$. Any matrix $\mathbf{G}'^{(1,3)}$ satisfying (G1) and (G3) yields a least squares solution $\mathbf{x}'=\mathbf{G}'^{(1,3)}\mathbf{b}'$. This becomes obvious from the fact that

$$\mathbf{y}'=\mathbf{A}'\mathbf{x}'=\mathbf{A}'\mathbf{G}'^{(1,3)}\mathbf{b}'=\mathbf{P}'_{R(A)}\mathbf{b}'=\begin{bmatrix}\mathbf{b}'_1\\ \mathbf{0}\end{bmatrix}=\hat{\mathbf{y}}'\in R(A)\,. \tag{3.56}$$

Property (G4) in combination with (G1) characterizes $\mathbf{G}'$ as a *minimum norm* generalized inverse of $\mathbf{A}'$. This means that any matrix $\mathbf{G}'^{(1,4)}$ satisfying (G1) and (G4) yields a minimum norm solution $\tilde{\mathbf{x}}'=\mathbf{G}'^{(1,4)}\mathbf{y}'$ only when it is applied to $\mathbf{y}'\in R(A)$. In other words the matrices $\mathbf{G}'^{(1,4)}$ provide minimum norm solutions for consistent equations only. Indeed if $\mathbf{y}'\in R(A)$ then $\mathbf{y}'=\mathbf{A}'\mathbf{x}'$ for some $\mathbf{x}'\in X$ and

$$\tilde{\mathbf{x}}'=\mathbf{G}'^{(1,4)}\mathbf{y}'=\mathbf{G}'^{(1,4)}\mathbf{A}'\mathbf{x}'=\mathbf{P}'_{N(A)^\perp}\mathbf{x}'=\begin{bmatrix}\mathbf{x}'_1\\ \mathbf{0}\end{bmatrix}\in N(A)^\perp\,. \tag{3.57}$$

The combination of all four properties characterizes $\mathbf{G}'$ as the unique *pseudoinverse* of $\mathbf{A}'$.

The $n\times n$ matrix $\mathbf{P}'_{R(A)}=\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$ is the SVD representation the operator $P_{R(A)}$ of orthogonal projection from Y to $R(A)$.

The $n\times n$ matrix $\mathbf{P}'_{R(A)^\perp}=\mathbf{I}-\mathbf{P}'_{R(A)}=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_f\end{bmatrix}$ is the SVD representation the operator $P_{R(A)^\perp}$ of orthogonal projection from Y to $R(A)^\perp$.

The $m\times m$ matrix $\mathbf{P}'_{N(A)^\perp}=\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$ is the SVD representation the operator $P_{N(A)^\perp}$ of orthogonal projection from Y to $N(A)^\perp$.

The $m\times m$ matrix $\mathbf{P}'_{N(A)}=\mathbf{I}-\mathbf{P}'_{N(A)^\perp}=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_d\end{bmatrix}$ is the SVD representation the operator $P_{N(A)}$ of orthogonal projection from Y to $N(A)$.

What remains is to translate all the above results from the SVD to the original bases, using the transformation relations $\mathbf{x}'=\mathbf{V}^T\mathbf{Q}\mathbf{x}$, $\mathbf{y}'=\mathbf{U}^T\mathbf{P}\mathbf{y}$, $\mathbf{b}'=\mathbf{U}^T\mathbf{P}\mathbf{b}$, $\mathbf{A}'=\mathbf{U}^T\mathbf{P}\mathbf{A}(\mathbf{V}^T\mathbf{Q})^{-1}=\mathbf{U}^T\mathbf{P}\mathbf{A}\mathbf{V}$, $\mathbf{G}'=\mathbf{V}^T\mathbf{Q}\mathbf{G}(\mathbf{U}^T\mathbf{P})^{-1}=\mathbf{V}^T\mathbf{Q}\mathbf{G}\mathbf{U}$, as well as their inverses. Using the introduced partitions we have

$$\mathbf{x}'=\mathbf{V}^T\mathbf{Q}\mathbf{x}\qquad\Rightarrow\qquad\mathbf{x}'_1=\mathbf{V}_1^T\mathbf{Q}\mathbf{x}\,,\qquad\mathbf{x}'_2=\mathbf{V}_2^T\mathbf{Q}\mathbf{x}\,, \tag{3.58}$$

$$\mathbf{y}'=\mathbf{U}^T\mathbf{P}\mathbf{y} \quad \Rightarrow \quad \mathbf{y}'_1=\mathbf{U}_1^T\mathbf{P}\mathbf{y}\,, \qquad \mathbf{y}'_2=\mathbf{U}_2^T\mathbf{P}\mathbf{y}\,, \tag{3.59}$$

$$\mathbf{b}'=\mathbf{U}^T\mathbf{P}\mathbf{b} \quad \Rightarrow \quad \mathbf{b}'_1=\mathbf{U}_1^T\mathbf{P}\mathbf{b}\,, \qquad \mathbf{b}'_2=\mathbf{U}_2^T\mathbf{P}\mathbf{b}\,. \tag{3.60}$$

We can now find elements of X and Y which characterize the subspaces $N(A)$, $N(A)^\perp$, $R(A)$ and $R(A)^\perp$. We use the notation $span(\mathbf{M})$ to characterize the subspace consisting of all the linear combinations of the columns of the matrix $\mathbf{M}$. It is now easy to see that

$$y\in R(A) \quad \Leftrightarrow \quad \mathbf{0}=\mathbf{y}'_2=\mathbf{U}_2^T\mathbf{P}\mathbf{y} \quad \Leftrightarrow \quad \mathbf{y}^T\mathbf{P}\mathbf{U}_2=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{y}\perp span(\mathbf{U}_2) \quad \Rightarrow$$

$$span(\mathbf{U}_2)=R(A)^\perp. \tag{3.61}$$

$$y\in R(A)^\perp \quad \Leftrightarrow \quad \mathbf{0}=\mathbf{y}'_1=\mathbf{U}_1^T\mathbf{P}\mathbf{y} \quad \Leftrightarrow \quad \mathbf{y}^T\mathbf{P}\mathbf{U}_1=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{y}\perp span(\mathbf{U}_1) \quad \Rightarrow$$

$$span(\mathbf{U}_1)=R(A)\,. \tag{3.62}$$

$$x\in N(A) \quad \Leftrightarrow \quad \mathbf{0}=\mathbf{x}'_1=\mathbf{V}_1^T\mathbf{Q}\mathbf{x} \quad \Leftrightarrow \quad \mathbf{x}^T\mathbf{Q}\mathbf{V}_1=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{x}\perp span(\mathbf{V}_1) \quad \Rightarrow$$

$$span(\mathbf{V}_1)=N(A)^\perp. \tag{3.63}$$

$$x\in N(A)^\perp \quad \Leftrightarrow \quad \mathbf{0}=\mathbf{x}'_2=\mathbf{V}_2^T\mathbf{Q}\mathbf{x} \quad \Leftrightarrow \quad \mathbf{x}^T\mathbf{Q}\mathbf{V}_2=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{x}\perp span(\mathbf{V}_2) \quad \Rightarrow$$

$$span(\mathbf{V}_2)=N(A)\,. \tag{3.64}$$

Thus the eigenvectors of $\mathbf{A}\mathbf{A}^*$ corresponding to non-zero eigenvalues, i.e. the columns of $\mathbf{U}_1$ span $R(A)$, while the ones corresponding to zero eigenvalues, i.e. the columns of $\mathbf{U}_2$ span $R(A)^\perp$. The eigenvectors of $\mathbf{A}^*\mathbf{A}$ corresponding to non-zero eigenvalues, i.e. the columns of $\mathbf{V}_1$ span $N(A)^\perp$, while the ones corresponding to zero eigenvalues, i.e. the columns of $\mathbf{V}_2$ span $N(A)$.

The orthogonality of the subspaces $R(A)$ and $R(A)^\perp$ is reflected in the orthogonality condition $\mathbf{U}_1^T\mathbf{P}\mathbf{U}_2=\mathbf{0}$ which is a direct consequence of $\mathbf{U}^T\mathbf{P}\mathbf{U}=\mathbf{I}$. The orthogonality of the subspaces $N(A)$ and $N(A)^\perp$ is reflected in the orthogonality condition $\mathbf{V}_2^T\mathbf{Q}\mathbf{V}_1=\mathbf{0}$ which is a direct consequence of $\mathbf{V}^T\mathbf{Q}\mathbf{V}=\mathbf{I}$.

The consistency condition, with respect to the original bases, becomes

$$\mathbf{b}'_2=\mathbf{U}_2^T\mathbf{P}\mathbf{b}=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{b}^T\mathbf{P}\mathbf{U}_2=\mathbf{0} \quad \Leftrightarrow \quad \mathbf{b}\perp span(\mathbf{U}_2)=R(A)^\perp \quad \Leftrightarrow$$

$$\mathbf{b}\in R(A)=span(\mathbf{U}_1) \tag{3.65}$$

The least squares solution becomes

$$\hat{\mathbf{y}}=\mathbf{U}\hat{\mathbf{y}}'=[\mathbf{U}_1\ \mathbf{U}_2]\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\mathbf{U}_1\mathbf{b}_1'=\mathbf{U}_1\mathbf{U}_1^T\mathbf{P}\mathbf{b}\,. \tag{3.66}$$

The minimum norm solution becomes

$$\hat{\mathbf{x}}=\mathbf{V}\hat{\mathbf{x}}'=[\mathbf{V}_1\ \mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{b}_1'=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P}\mathbf{b}=\mathbf{G}\mathbf{b}\,, \tag{3.67}$$

$$\mathbf{G}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P} \tag{3.68}$$

or

$$\mathbf{G}=\mathbf{V}\mathbf{G}'\mathbf{U}^T\mathbf{P}=[\mathbf{V}_1\ \mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^T\mathbf{P}\\ \mathbf{U}_2^T\mathbf{P}\end{bmatrix}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P}\,. \tag{3.69}$$

Recalling that

$$\mathbf{A}=\mathbf{U}\mathbf{A}'\mathbf{V}^T\mathbf{Q}=[\mathbf{U}_1\ \mathbf{U}_2]\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^T\mathbf{Q}\\ \mathbf{V}_2^T\mathbf{Q}\end{bmatrix}=\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q} \tag{3.70}$$

we can compute

$$\mathbf{P}_{R(A)}=\mathbf{A}\mathbf{G}=\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q}\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P}=\mathbf{U}_1\mathbf{U}_1^T\mathbf{P}\,, \tag{3.71}$$

$$\hat{\mathbf{y}}=\mathbf{P}_{R(A)}\mathbf{b}=\mathbf{U}_1\mathbf{U}_1^T\mathbf{P}\mathbf{b}\,, \tag{3.72}$$

$$\mathbf{P}_{N(A)^\perp}=\mathbf{G}\mathbf{A}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P}\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q}=\mathbf{V}_1\mathbf{V}_1^T\mathbf{Q}\,, \tag{3.73}$$

$$\mathbf{P}_{R(A)^\perp}=\mathbf{I}-\mathbf{P}_{R(A)}=\mathbf{I}-\mathbf{U}_1\mathbf{U}_1^T\mathbf{P}=\mathbf{U}_2\mathbf{U}_2^T\mathbf{P}\,, \tag{3.74}$$

$$\mathbf{P}_{N(A)}=\mathbf{I}-\mathbf{P}_{N(A)^\perp}=\mathbf{I}-\mathbf{V}_1\mathbf{V}_1^T\mathbf{Q}=\mathbf{V}_2\mathbf{V}_2^T\mathbf{Q}\,. \tag{3.75}$$

We have used above the orthogonality relations

$$\mathbf{U}^T\mathbf{P}\mathbf{U}=\mathbf{I} \qquad \Rightarrow \qquad \mathbf{U}_1^T\mathbf{P}\mathbf{U}_1=\mathbf{I} \tag{3.76}$$

$$\mathbf{V}^T\mathbf{Q}\mathbf{V}=\mathbf{I} \qquad \Rightarrow \qquad \mathbf{V}_1^T\mathbf{Q}\mathbf{V}_1=\mathbf{I} \tag{3.77}$$

$$\mathbf{U}\mathbf{U}^T\mathbf{P}=\mathbf{U}_1\mathbf{U}_1^T\mathbf{P}+\mathbf{U}_2\mathbf{U}_2^T\mathbf{P}=\mathbf{I} \tag{3.78}$$

$$\mathbf{V}\mathbf{V}^T\mathbf{Q}=\mathbf{V}_1\mathbf{V}_1^T\mathbf{Q}+\mathbf{V}_2\mathbf{V}_2^T\mathbf{Q}=\mathbf{I}\,. \tag{3.79}$$

It is easy to see that the relations (G1), (G2) hold, while (G3), (G4) should be modified, taking into account (3.71) and (3.73), in order to obtain the new set

(G1) $$\mathbf{AGA}=\mathbf{A} \tag{3.80}$$

(G2) $$\mathbf{GAG}=\mathbf{G} \tag{3.81}$$

(G3) $$(\mathbf{AGP}^{-1})^T=\mathbf{AGP}^{-1} \tag{3.82}$$

(G4) $$(\mathbf{GAQ}^{-1})^T=\mathbf{GAQ}^{-1}\,. \tag{3.83}$$

If the adjoints of the operators $G:Y\to X$, $AG:X\to X$, $GA:Y\to Y$ are introduced it holds that

$$(x,Gy)=\mathbf{y}^T\mathbf{G}^T\mathbf{Q}\mathbf{x}=(y,G^*x)=\mathbf{y}^T\mathbf{P}\mathbf{G}^*\mathbf{x} \quad\Rightarrow\quad \mathbf{G}^*=\mathbf{P}^{-1}\mathbf{G}^T\mathbf{Q} \tag{3.84}$$

$$(y_\alpha, AGy_\beta)=\mathbf{y}_\beta^T\mathbf{G}^T\mathbf{A}^T\mathbf{P}\mathbf{y}_\alpha=((AG)^*y_\alpha, y_\beta)=\mathbf{y}_\beta^T\mathbf{P}(\mathbf{AG})^*\mathbf{y}_\alpha \quad\Rightarrow$$

$$(\mathbf{AG})^*=\mathbf{P}^{-1}\mathbf{G}^T\mathbf{A}^T\mathbf{P}=\mathbf{G}^*\mathbf{A}^*=(\mathbf{AGP}^{-1})^T\mathbf{P}=\mathbf{AGP}^{-1}\mathbf{P}=\mathbf{AG} \tag{3.85}$$

$$(x_\alpha, GAx_\beta)=\mathbf{x}_\beta^T\mathbf{A}^T\mathbf{G}^T\mathbf{Q}\mathbf{x}_\alpha=((GA)^*x_\alpha, x_\beta)=\mathbf{x}_\beta^T\mathbf{Q}(\mathbf{GA})^*\mathbf{x}_\alpha \quad\Rightarrow$$

$$(\mathbf{GA})^*=\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{G}^T\mathbf{Q}=\mathbf{A}^*\mathbf{G}^*=(\mathbf{GAQ}^{-1})^T\mathbf{Q}=\mathbf{GAQ}^{-1}\mathbf{Q}=\mathbf{GA} \tag{3.86}$$

where the relations (3.26), (G3) and (G4), have been implemented. Thus (G3) and (G4) can be replaced, respectively by $(\mathbf{AG})^*=\mathbf{AG}$ and $(\mathbf{GA})^*=\mathbf{GA}$, which are in turn representations of the relations $(AG)^*=AG$ and $(GA)^*=GA$. In conclusion the pseudoinverse operator G can be characterized in a "coordinate-free" way by

$$AGA=A\,, \qquad GAG=G\,, \qquad (AG)^*=AG\,, \qquad (GA)^*=GA\,. \tag{3.87}$$

In cases where the choice of $\mathbf{P}$ is based on probabilistic reasoning, which does not apply to the choice of $\mathbf{Q}$, one may well use any least squares solution, i.e. a solution of the normal equations (3.47). In this case the generalized inverse G needs only to satisfy $AGA=A$ and the least squares property $(AG)^*=AG$.

Remark:

In geodesy the reflectivity property $GAG=G$ is also included, in an implicit way by resorting to generalized inversion of the normal equations (3.47), which any least squares solution should satisfy. Such reflexive least-squares generalized inverses are not obtained directly, but in an implicit way based on the introduction of *minimal constraints*, i.e. linear constraints on the parameters which can be written in the form $\mathbf{C}^T\mathbf{Q}\mathbf{x}=\mathbf{0}$, where $\mathbf{C}$ is a $m\times d$ matrix with $rank(\mathbf{C})=m$ and such that $span(\mathbf{C})\cap N(A)=\{0\}$. A particular set of choices of $\mathbf{C}$ are the ones for which $span(\mathbf{C})=N(A)^{\perp}$, in which case the minimal constraints are called *inner constraints* and lead to the minimum norm (pseudoinverse) solution.

A complete investigation of the various types of generalized inverses based on the singular value decomposition is given in Dermanis (1998, ch. 6).

3. 2 The regular case ($r=m=n$)

In the case of a full-rank square $n\times n$ matrix $\mathbf{A}$ we have $rank(\mathbf{A}^T\mathbf{A})=$ $=rank(\mathbf{A}\mathbf{A}^T)=rank(\mathbf{A})=n$, $d=m-r=0$, $f=n-m=0$ and the singular value decomposition equations become

$$\mathbf{A}\mathbf{V}=\mathbf{U}\mathbf{L}, \qquad (\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})\mathbf{U}=\mathbf{V}\boldsymbol{\Lambda}, \tag{3.88}$$

$$[\mathbf{A}(\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})]\mathbf{U}=\mathbf{U}\boldsymbol{\Lambda}^2, \qquad [(\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})\mathbf{A}]\mathbf{V}=\mathbf{V}\boldsymbol{\Lambda}^2 \tag{3.89}$$

$$\underset{n\times n}{\mathbf{A}}=\underset{n\times n}{\mathbf{U}}\ \underset{n\times n}{\boldsymbol{\Lambda}}\ \underset{n\times n}{\mathbf{V}^T}\ \underset{n\times n}{\mathbf{Q}}, \qquad \mathbf{A}'=\boldsymbol{\Lambda}=\mathbf{U}^T\mathbf{P}\mathbf{A}\mathbf{V}. \tag{3.90}$$

The SVD representation of the equation $b=Ax$ is $\mathbf{b}'=\boldsymbol{\Lambda}\mathbf{x}'$ and has always a solution $\mathbf{x}'=\boldsymbol{\Lambda}^{-1}\mathbf{b}'$, which is also unique. In the original bases we obtain through the transformations $\mathbf{x}'=\mathbf{V}^T\mathbf{Q}\mathbf{x}$, $\mathbf{x}=\mathbf{V}\mathbf{x}'$, $\mathbf{b}'=\mathbf{U}^T\mathbf{P}\mathbf{b}$ the solution

$$\mathbf{x}=\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T\mathbf{P}\mathbf{b}=[(\mathbf{U}^T\mathbf{P})^{-1}\boldsymbol{\Lambda}\mathbf{V}^{-1}]^{-1}\mathbf{b}=(\mathbf{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\mathbf{Q})^{-1}\mathbf{b}=\mathbf{A}^{-1}\mathbf{b} \tag{3.91}$$

In this case there exist always a solution, no matter what $\mathbf{b}$ is, because

$$R(A)=span(\mathbf{U})=Y, \qquad R(A)^{\perp}=\{0\} \tag{3.92}$$

and the solution is unique because ($\mathbf{b}=\mathbf{0} \Rightarrow \mathbf{x}=\mathbf{A}^{-1}\mathbf{b}=\mathbf{0}$)

$$N(A)=\{0\}, \qquad N(A)^{\perp}=span(\mathbf{V})=X. \tag{3.93}$$

The inverse $\mathbf{G}=\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{P}$ satisfies all the four pseudoinverse conditions and it is furthermore both a left ($\mathbf{GA}=\mathbf{I}$) and a right ($\mathbf{AG}=\mathbf{I}$) inverse of $\mathbf{A}$ and thus a regular inverse $\mathbf{G}=\mathbf{A}^{-1}$.

3. 3 The full-rank overdetermined case ($r=m<n$)

In the case where $\mathbf{A}$ has full-column-rank, we have $rank(\mathbf{A}^T\mathbf{A})=m$, $d=m-r=0$, $f=n-m$ and the singular value decomposition becomes

$$\underset{n\times m}{\mathbf{A}} = \underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{m\times m}{\mathbf{\Lambda}}\\ \underset{f\times m}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T}\ \underset{m\times m}{\mathbf{Q}} = \begin{bmatrix}\underset{n\times m}{\mathbf{U}_1} & \underset{n\times f}{\mathbf{U}_2}\end{bmatrix}\begin{bmatrix}\underset{m\times m}{\mathbf{\Lambda}}\\ \underset{f\times m}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T}\ \underset{m\times m}{\mathbf{Q}} = \mathbf{U}_1\mathbf{\Lambda}\mathbf{V}^T\mathbf{Q}. \tag{3.94}$$

The SVD representation of the operator A is $\mathbf{A}'=\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}$ and for any vector x the corresponding image $y=Ax$ is represented by

$$\mathbf{y}'\equiv\begin{bmatrix}\underset{m\times 1}{\mathbf{y}_1'}\\ \underset{f\times 1}{\mathbf{y}_2'}\end{bmatrix}=\mathbf{A}'\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda}\mathbf{x}'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{y}_1'\\ \mathbf{0}\end{bmatrix}, \qquad \mathbf{y}_1'=\mathbf{\Lambda}\mathbf{x}', \quad \mathbf{y}_2'=\mathbf{0}. \tag{3.95}$$

We have again the same compatibility condition $\mathbf{y}_2'=\mathbf{0}$, which guarantees the existence of a solution, which in this case is unique $\mathbf{x}'=\mathbf{\Lambda}^{-1}\mathbf{y}_1'$.

The solution to the homogeneous consistent system $Ax=0$ is represented by the unique vector $\mathbf{x}'=\mathbf{\Lambda}^{-1}\mathbf{0}=\mathbf{0}$. In this case the null space has a single element $N(A)=\{0\}$ while its orthogonal complement takes up the whole space $N(A)^{\perp}=X$.

For the arbitrary observation vector $\mathbf{b}$, there is no solution and we must apply the least squares principle to obtain its orthogonal projection on $R(A)=span(\mathbf{U}_1)$

$$\hat{\mathbf{y}}'=\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix}=\mathbf{P}_{R(A)}'\mathbf{b}'. \tag{3.96}$$

The unique least squares solution follows from

$$\hat{\mathbf{x}}'=\mathbf{\Lambda}^{-1}\mathbf{b}_1'=[\mathbf{\Lambda}^{-1}\ \mathbf{0}]\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix}\equiv\mathbf{G}'\mathbf{b}', \qquad \mathbf{G}'=[\mathbf{\Lambda}^{-1}\ \mathbf{0}]. \tag{3.97}$$

It is very easy to show that the inverse $\mathbf{G}'$ satisfies the four properties (G1), (G2), (G3), (G4). Furthermore

$$\mathbf{G'A'}=[\mathbf{\Lambda}^{-1}\ \mathbf{0}]\begin{bmatrix}\mathbf{\Lambda}\\ \mathbf{0}\end{bmatrix}=\mathbf{\Lambda}^{-1}\mathbf{\Lambda}=\mathbf{I}. \tag{3.98}$$

This means that the operator G represented by $\mathbf{G'}$ is not only a pseudoinverse but also a *left inverse* of the operator A.

With respect to the original bases we have again the consistency condition $\mathbf{U}_2^T\mathbf{Pb}=\mathbf{0}$, the projection $\hat{y}=P_{R(A)}b$ is represented by

$$\hat{\mathbf{y}}=\mathbf{U}_1\mathbf{U}_1^T\mathbf{Pb} \tag{3.99}$$

and the unique least squares solution becomes

$$\hat{\mathbf{x}}=\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{Pb}\equiv\mathbf{Gb}, \qquad \mathbf{G}=\mathbf{V}[\mathbf{\Lambda}^{-1}\ \mathbf{0}]\mathbf{U}^T\mathbf{P}=\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{P}. \tag{3.100}$$

It is easy to show that $\mathbf{AGA}=\mathbf{A}$, $\mathbf{GAG}=\mathbf{G}$, $\mathbf{AG}=(\mathbf{AG})^T$, $\mathbf{GA}=\mathbf{I}=(\mathbf{GA})^T$ so that the matrix $\mathbf{G}$ is a pseudoinverse and also a left inverse of the matrix $\mathbf{A}$.

To express the solution $\hat{\mathbf{x}}$ in a more familiar form, we compute

$$(\mathbf{A}^T\mathbf{PA})^{-1}=(\mathbf{QV}\,\mathbf{\Lambda}\,\mathbf{U}_1^T\mathbf{PU}_1\mathbf{\Lambda}\,\mathbf{V}^T\mathbf{Q})^{-1}=(\mathbf{QV}\,\mathbf{\Lambda}^2\mathbf{V}^T\mathbf{Q})^{-1}=\mathbf{V}\,\mathbf{\Lambda}^{-2}\mathbf{V}^T, \tag{3.101}$$

$$\mathbf{A}^T\mathbf{Pb}=\mathbf{QV}\,\mathbf{\Lambda}\,\mathbf{U}_1^T\mathbf{Pb} \tag{3.102}$$

and the solution (3.100) can also be written as

$$\hat{\mathbf{x}}=\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}_1^T\mathbf{Pb}=\mathbf{V}\mathbf{\Lambda}^{-2}\,\mathbf{V}^T\mathbf{QV}\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Pb}=(\mathbf{A}^T\mathbf{PA})^{-1}\mathbf{A}^T\mathbf{Pb}. \tag{3.103}$$

3. 4 The full-rank underdetermined case ($r=n<m$)

In the case where $\mathbf{A}$ has full-row-rank, we have $rank(\mathbf{AA}^T)=n$, $f=n-r=0$, $d=m-n$ and the singular value decomposition becomes

$$\underset{n\times m}{\mathbf{A}}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{n\times n}{\mathbf{\Lambda}} & \underset{n\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T\mathbf{Q}}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{n\times n}{\mathbf{\Lambda}} & \underset{n\times d}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{n\times m}{\mathbf{V}_1^T\mathbf{Q}}\\ \underset{d\times m}{\mathbf{V}_2^T\mathbf{Q}}\end{bmatrix}=\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}_1^T\mathbf{Q}. \tag{3.104}$$

The SVD representation of the operator A is $\mathbf{A'}=[\mathbf{\Lambda}\ \mathbf{0}]$ and for any vector x the corresponding image $y=Ax$ is represented by

$$\mathbf{y}'=\mathbf{A}'\mathbf{x}'=[\mathbf{\Lambda}\ \mathbf{0}]\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\end{bmatrix}=\mathbf{\Lambda}\mathbf{x}_1'. \tag{3.105}$$

The equation $b=Ax$ is always consistent and has infinite solutions of the form $\mathbf{x}'=\mathbf{\Lambda}^{-1}\mathbf{b}'$, i.e.,

$$\mathbf{x}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}'\\ \mathbf{x}_2'\end{bmatrix} \tag{3.106}$$

where $\mathbf{x}_2'$ takes any value. The minimum norm solution is

$$\hat{\mathbf{x}}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix}\mathbf{b}'\equiv\mathbf{G}'\mathbf{b}',\qquad \mathbf{G}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix} \tag{3.107}$$

which is related to any other solution $\mathbf{x}'$ given by (3.106) through

$$\hat{\mathbf{x}}'=\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}'\\ \mathbf{x}_2'\end{bmatrix}=\mathbf{P}'_{N(A)^\perp}\mathbf{x}'. \tag{3.108}$$

The minimum norm solution is the orthogonal projection $\hat{x}=P_{N(A)^\perp}x$ of any solution x on the subspace $N(A)^\perp=span(\mathbf{V}_1)$, while $N(A)=span(\mathbf{V}_2)$.

It is easy to show that the matrix $\mathbf{G}'$ satisfies the relations (G1), (G2), (G3), (G4), while in addition

$$\mathbf{A}'\mathbf{G}'=[\mathbf{\Lambda}\ \mathbf{0}]\begin{bmatrix}\mathbf{\Lambda}^{-1}\\ \mathbf{0}\end{bmatrix}=\mathbf{\Lambda}\mathbf{\Lambda}^{-1}=\mathbf{I}. \tag{3.109}$$

This means that the operator G represented by $\mathbf{G}'$ is not only a pseudoinverse but also a *right inverse* of the operator A.

With respect to the original bases the minimum norm solution becomes

$$\hat{\mathbf{x}}=\mathbf{V}\hat{\mathbf{x}}'=[\mathbf{V}_1\ \mathbf{V}_2]\begin{bmatrix}\mathbf{\Lambda}^{-1}\mathbf{b}'\\ \mathbf{0}\end{bmatrix}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{b}'=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{P}\mathbf{b}=\mathbf{G}\mathbf{b},\quad \mathbf{G}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{P}. \tag{3.110}$$

It is easy to show that $\mathbf{AGA}=\mathbf{A}$, $\mathbf{GAG}=\mathbf{G}$, $\mathbf{AG}=(\mathbf{AG})^T$, $\mathbf{AG}=\mathbf{I}=(\mathbf{AG})^T$ so that the matrix $\mathbf{G}$ is a pseudoinverse and also a right inverse of the matrix $\mathbf{A}$.

To express the solution $\hat{\mathbf{x}}$ in a more familiar form, we compute

$$\mathbf{A}^T=\mathbf{Q}\mathbf{V}_1\ \mathbf{\Lambda}\ \mathbf{U}^T,\qquad \mathbf{Q}^{-1}\mathbf{A}^T=\mathbf{V}_1\mathbf{\Lambda}\ \mathbf{U}^T, \tag{3.111}$$

$$(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}=(\mathbf{U}\,\mathbf{\Lambda}\,\mathbf{V}_1^T\mathbf{V}_1\mathbf{\Lambda}\,\mathbf{U}^T)^{-1}=(\mathbf{U}\,\mathbf{\Lambda}^2\mathbf{U}^T)^{-1}=\mathbf{P}\mathbf{U}\,\mathbf{\Lambda}^{-2}\mathbf{U}^T\mathbf{P} \tag{3.112}$$

and the solution (3.110) can also be written as

$$\hat{\mathbf{x}}=\mathbf{V}_1\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{P}\mathbf{b}=\mathbf{V}_1\mathbf{\Lambda}\mathbf{U}^T\mathbf{P}\mathbf{U}\mathbf{\Lambda}^{-2}\mathbf{U}^T\mathbf{P}\mathbf{b}=\mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}\mathbf{b}\,. \tag{3.113}$$

3. 5 The hybrid solution (Tikhonov regularization)

Instead of the two step approach to the solution in the general case ($r<\min(n,m)$), where parameter norm minimization (uniqueness) follows the application of the least squares principle (existence), it is possible to seek a solution $\hat{x}$ satisfying

$$\|b-A\hat{x}\|^2+\alpha\|\hat{x}\|^2=\min_{x\in X}\{\|b-Ax\|^2+\alpha\|x\|^2\} \tag{3.114}$$

where $\alpha>0$ is a balancing "regularization" parameter (Tikhonov and Arsenin, 1977). In the SVD bases the minimization of the above "hybrid" norm takes the form

$$\phi\equiv\|\mathbf{b}'-\mathbf{A}'\mathbf{x}'\|^2+\alpha\|\mathbf{x}'\|^2=(\mathbf{b}'-\mathbf{A}'\mathbf{x}')^T(\mathbf{b}'-\mathbf{A}'\mathbf{x}')+\alpha\,(\mathbf{x}')^T\mathbf{x}'=\min \tag{3.115}$$

and since

$$\mathbf{b}'-\mathbf{A}^*\mathbf{x}'=\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix}-\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\end{bmatrix}=\begin{bmatrix}\mathbf{b}_1'-\mathbf{\Lambda}\mathbf{x}_1'\\ \mathbf{0}\end{bmatrix} \tag{3.116}$$

$$\phi=(\mathbf{b}_1'-\mathbf{\Lambda}\mathbf{x}_1')^T(\mathbf{b}_1'-\mathbf{\Lambda}\mathbf{x}_1')+\alpha[(\mathbf{x}_1')^T\mathbf{x}_1'+(\mathbf{x}_2')^T\mathbf{x}_2']=\min \tag{3.117}$$

with solution

$$\frac{\partial\phi}{\partial\mathbf{x}_1'}(\hat{\mathbf{x}})=-2(\mathbf{b}_1')^T\mathbf{\Lambda}+2(\hat{\mathbf{x}}_1')^T\mathbf{\Lambda}^2+2\alpha(\hat{\mathbf{x}}_1')^T=\mathbf{0}\quad\Rightarrow\quad\hat{\mathbf{x}}_1'=(\mathbf{\Lambda}^2+\alpha\mathbf{I})^{-1}\mathbf{\Lambda}\mathbf{b}_1' \tag{3.118}$$

$$\frac{\partial\phi}{\partial\mathbf{x}_2'}(\hat{\mathbf{x}})=2\alpha(\hat{\mathbf{x}}_2')^T=\mathbf{0}\qquad\Rightarrow\quad\hat{\mathbf{x}}_2'=\mathbf{0} \tag{3.119}$$

$$\hat{\mathbf{x}}'=\begin{bmatrix}(\mathbf{\Lambda}^2+\alpha\mathbf{I})^{-1}\mathbf{\Lambda}\mathbf{b}_1'\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}(\mathbf{\Lambda}^2+\alpha\mathbf{I})^{-1}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{b}_1'\\ \mathbf{b}_2'\end{bmatrix}. \tag{3.120}$$

In the SVD bases the hybrid solution becomes

$$\hat{\mathbf{x}}=\mathbf{V}\hat{\mathbf{x}}'=\mathbf{V}\begin{bmatrix}(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})^{-1}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{b}'=\mathbf{V}\begin{bmatrix}(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})^{-1}\boldsymbol{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{U}^T\mathbf{P}\mathbf{b} \tag{3.121}$$

or after the proper partitioning $\mathbf{V}=[\mathbf{V}_1\ \mathbf{V}_2]$, $\mathbf{U}=[\mathbf{U}_1\ \mathbf{U}_2]$

$$\hat{\mathbf{x}}=\mathbf{V}_1(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}\mathbf{U}_1^T\mathbf{P}\mathbf{b}=(\mathbf{V}_1\boldsymbol{\Lambda}^2\mathbf{V}_1^T\mathbf{Q}+\alpha\mathbf{I})^{-1}\mathbf{V}_1\boldsymbol{\Lambda}\mathbf{U}_1^T\mathbf{P}\mathbf{b}=$$

$$=\mathbf{V}_1\boldsymbol{\Lambda}\mathbf{U}_1^T(\mathbf{P}\mathbf{U}_1\boldsymbol{\Lambda}^2\mathbf{U}_1^T+\alpha\mathbf{I})^{-1}\mathbf{P}\mathbf{b}\,. \tag{3.122}$$

We have used above the matrix identities

$$\mathbf{V}_1(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})^{-1}=(\mathbf{V}_1\boldsymbol{\Lambda}^2\mathbf{V}_1^T\mathbf{Q}+\alpha\mathbf{I})^{-1}\mathbf{V}_1 \tag{3.123}$$

$$(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})^{-1}\boldsymbol{\Lambda}\mathbf{U}_1^T=\boldsymbol{\Lambda}\mathbf{U}_1^T(\mathbf{P}\mathbf{U}_1\boldsymbol{\Lambda}^2\mathbf{U}_1^T+\alpha\mathbf{I})^{-1}, \tag{3.124}$$

which can be easily proved starting from the obvious identities

$$\mathbf{V}_1\boldsymbol{\Lambda}^2+\alpha\mathbf{V}_1=\mathbf{V}_1\boldsymbol{\Lambda}^2\mathbf{V}_1^T\mathbf{Q}\mathbf{V}_1+\alpha\mathbf{V}_1 \quad\Rightarrow$$

$$\mathbf{V}_1(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})=(\mathbf{V}_1\boldsymbol{\Lambda}^2\mathbf{V}_1^T\mathbf{Q}+\alpha\mathbf{I})\mathbf{V}_1 \tag{3.125}$$

$$\boldsymbol{\Lambda}\mathbf{U}_1^T\mathbf{P}\mathbf{U}_1\boldsymbol{\Lambda}^2\mathbf{U}_1^T+\alpha\boldsymbol{\Lambda}\mathbf{U}_1^T=\boldsymbol{\Lambda}^3\mathbf{U}_1^T+\alpha\boldsymbol{\Lambda}\mathbf{U}_1^T \quad\Rightarrow$$

$$\boldsymbol{\Lambda}\mathbf{U}_1^T(\mathbf{P}\mathbf{U}_1\boldsymbol{\Lambda}^2\mathbf{U}_1^T+\alpha\mathbf{I})=(\boldsymbol{\Lambda}^2+\alpha\mathbf{I})\boldsymbol{\Lambda}\mathbf{U}_1^T\,. \tag{3.126}$$

If we use $\mathbf{A}=\mathbf{U}_1\boldsymbol{\Lambda}\mathbf{V}_1^T\mathbf{Q}$ to compute $\mathbf{A}^T=\mathbf{Q}\mathbf{V}_1\boldsymbol{\Lambda}\mathbf{U}_1^T$,

$$\mathbf{A}^T\mathbf{P}\mathbf{A}=\mathbf{Q}\mathbf{V}_1\boldsymbol{\Lambda}\mathbf{U}_1^T\mathbf{P}\mathbf{U}_1\boldsymbol{\Lambda}\mathbf{V}_1^T\mathbf{Q}=\mathbf{Q}\mathbf{V}_1\boldsymbol{\Lambda}^2\mathbf{V}_1^T\mathbf{Q}, \tag{3.127}$$

$$\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T=\mathbf{U}_1\boldsymbol{\Lambda}\mathbf{V}_1^T\mathbf{Q}\mathbf{V}_1\boldsymbol{\Lambda}\mathbf{U}_1^T=\mathbf{U}_1\boldsymbol{\Lambda}^2\mathbf{U}_1^T, \tag{3.128}$$

we can rewrite the hybrid solution (3.122) in the more familiar form

$$\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{P}\mathbf{A}+\alpha\mathbf{Q})^{-1}\mathbf{A}^T\mathbf{P}\mathbf{b}=\mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T+\alpha\mathbf{P}^{-1})^{-1}\mathbf{b}\,. \tag{3.129}$$

In the overdetermined full-rank case ($r=m<n$), where $\mathbf{V}_1=\mathbf{V}$ the hybrid solution degenerates into the least squares solution . In the underdetermined full-rank case ($r=n<m$), where $\mathbf{U}_1=\mathbf{U}$, the hybrid solution degenerates into the minimum norm solution.

When a probabilistic justification is given for the choice of both weight matrices $\mathbf{P}$ and $\mathbf{Q}$ then the hybrid solution is the reasonable to follow, since it treats the norms in both spaces X and Y simultaneously on an equal base (apart from the balancing factor α). The two step approach gives full priority to the norm minimization in Y and it is more appropriate when probabilistic justification is available for the choice of $\mathbf{P}$ only. In such a case the second step of norm minimization in X plays a minor role in securing the computation of a solution, which is as good as any other least squares solution, and its role becomes more clear by resorting to the so called full rank factorization of the mapping A.

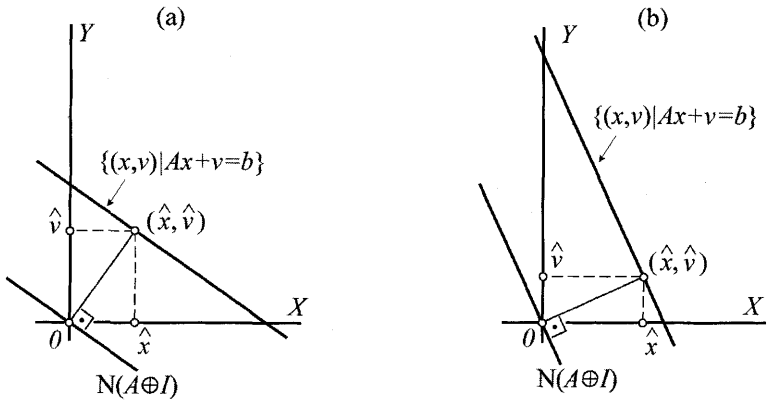


Fig. 3: The geometry of Tikhonov regularization, viewed as a full-rank underdetermined problem.
(a) large value of the regularization parameter α (small value of $\|\hat{x}\|$)
(b) small value of the regularization parameter α (large value of $\|\hat{x}\|$)

Remark:

Once we allow for the discrepancy $v=b-y=b-Ax$, we can also "restate" the model as a full rank overdetermined one for a new operator $A\oplus I: X\times Y\to Y: (x,v)\to Ax+v$, defined by $(A\oplus I)(x,v)\equiv Ax+v$. This is equivalent to setting $b=Ax+v=(A\oplus I)(x,v)$ with the particular representation $\mathbf{b}=[\mathbf{A}\ \mathbf{I}]\begin{bmatrix}\mathbf{x}\\ \mathbf{v}\end{bmatrix}$. The regularization solution corresponds to a minimum-norm solution, where the norm in $X\times Y$ is defined by

$$\|(x,v)\|^2=\begin{bmatrix}\mathbf{x}\\ \mathbf{v}\end{bmatrix}^T\begin{bmatrix}\alpha\mathbf{Q} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{v}\end{bmatrix}=\alpha\,\mathbf{x}^T\mathbf{Q}\mathbf{x}+\mathbf{v}^T\mathbf{P}\mathbf{v}\,.$$

Note that different values of the regularization parameter α correspond to different "geometries" for the space $X\times Y$ (see fig. 3).

3. 6 The full rank factorization

Looking into the singular value decomposition for the general case ($r<\min(n,m)$),

$$\underset{n\times m}{\mathbf{A}}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T}\underset{m\times m}{\mathbf{Q}}=\begin{bmatrix}\underset{n\times r}{\mathbf{U}_1} & \underset{n\times f}{\mathbf{U}_2}\end{bmatrix}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{m\times r}{\mathbf{V}_1} & \underset{m\times d}{\mathbf{V}_2}\end{bmatrix}^T\underset{m\times m}{\mathbf{Q}}=\mathbf{U}_1\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q} \tag{3.130}$$

it is easy to see that the matrix $\mathbf{A}$ as well as the operator A it represents can be expressed as a product $\mathbf{A}=\mathbf{BC}$ (respectively as a combination of operators $A=BC$)

$$\underset{n\times m}{\mathbf{A}}=\underset{n\times r}{\mathbf{B}}\,\underset{r\times m}{\mathbf{C}}=\left(\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}^{1/2}}\\ \underset{f\times r}{\mathbf{0}}\end{bmatrix}\right)\left(\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}^{1/2}} & \underset{r\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T}\underset{m\times m}{\mathbf{Q}}\right)=(\underset{n\times r}{\mathbf{U}_1}\,\underset{r\times r}{\mathbf{\Lambda}^{1/2}})(\underset{r\times r}{\mathbf{\Lambda}^{1/2}}\,\underset{r\times m}{\mathbf{V}_1^T}\,\underset{m\times m}{\mathbf{Q}}) \tag{3.131}$$

where

$$\mathbf{B}=\mathbf{U}_1\mathbf{\Lambda}^{1/2},\qquad \mathbf{C}=\mathbf{\Lambda}^{1/2}\mathbf{V}_1^T\mathbf{Q}. \tag{3.132}$$

We have chosen to "split" the matrix $\mathbf{\Lambda}=\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}$ equally between the two factors, for "symmetry" reasons, while any other choice such as $\mathbf{B}=\mathbf{U}_1$, $\mathbf{C}=\mathbf{\Lambda}\mathbf{V}_1^T\mathbf{Q}$ or $\mathbf{B}=\mathbf{U}_1\mathbf{\Lambda}$, $\mathbf{C}=\mathbf{V}_1^T\mathbf{Q}$, are equally acceptable. In fact there is an infinite number of such full rank factorizations of the form

$$\mathbf{B}=\mathbf{U}_1\mathbf{\Lambda}^{1/2}\mathbf{R},\qquad \mathbf{C}=\mathbf{R}^{-1}\mathbf{\Lambda}^{1/2}\mathbf{V}_1^T\mathbf{Q}, \tag{3.133}$$

where $\mathbf{R}$ runs over the set of all $r\times r$ regular (invertible) matrices.

With the help of a full rank factorization, the original problem is replaced by two problems, an overdetermined full-rank one $\mathbf{y}=\mathbf{Bz}$, with data $\mathbf{b}$ and an underdetermined full-rank one $\mathbf{z}=\mathbf{Cx}$. A unique least squares solution $\hat{\mathbf{z}}=(\mathbf{B}^T\mathbf{PB})^{-1}\mathbf{B}^T\mathbf{Pb}$ may be followed by a minimum norm solution $\hat{\mathbf{x}}=\mathbf{Q}^{-1}\mathbf{C}^T(\mathbf{CQ}^{-1}\mathbf{C}^T)^{-1}\hat{\mathbf{z}}$. Whenever there is no justification for selecting the minimum norm solution among all possible solutions of $\hat{\mathbf{z}}=\mathbf{Cx}$, while the application of the least squares principle $\mathbf{v}^T\mathbf{Pv}=\min$ ($\mathbf{v}=\mathbf{b}-\mathbf{y}$) is justifiable on probabilistic grounds, the parameters $\mathbf{z}$ are uniquely identifiable (determinable) quantities from the available data, and the same holds for any linear(ized) function $q=\mathbf{d}^T\mathbf{z}$ of $\mathbf{z}$. On the contrary a similar function $q=\mathbf{a}^T\mathbf{x}$, is identifiable only when it can be factored as a function $q=\mathbf{a}^T\mathbf{x}=\mathbf{d}^T\mathbf{Cx}=\mathbf{d}^T\mathbf{z}$, that is whenever there exists a vector $\mathbf{d}$ such that the given vector $\mathbf{a}$ can be factored according to $\mathbf{a}^T=\mathbf{d}^T\mathbf{C}$. This "identifiability" condition is usually introduced within a probabilistic framework as the "estimability" condition $\mathbf{C}^T\mathbf{d}=\mathbf{a}$ or $\mathbf{a}\in span(\mathbf{C}^T)$.

A standard problem in geodesy involves angular and distance observables, which relate to any set of parameters $\mathbf{z}$ defining the shape and size of a geodetic network, but they have no sufficient content to determine its position and orientation, which in addition to shape and size are described by the network point coordinates $\mathbf{x}$. This type of "improper modeling" where coordinates are used as parameters for the sake of convenience, although they are not identifiable, leads to a "false" character of underdetermination as opposed to that of uniquely underdetermined problems. Any least squares solution, i.e. any solution $\tilde{\mathbf{x}}$ satisfying $\mathbf{C}\tilde{\mathbf{x}}=\hat{\mathbf{z}}=(\mathbf{B}^T\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{P}\mathbf{b}$, leads to the same estimates $\hat{q}=\mathbf{a}^T\tilde{\mathbf{x}}=\mathbf{d}^T\mathbf{C}\tilde{\mathbf{x}}=\mathbf{d}^T\hat{\mathbf{z}}=\mathbf{d}^T(\mathbf{B}^T\mathbf{P}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{P}\mathbf{b}$ for identifiable quantities $q=\mathbf{a}^T\mathbf{x}$. In this sense the choice of solution $\tilde{\mathbf{x}}$ serves as an information depository for the computation of estimates of identifiable (estimable) quantities, while the computation of estimates $\hat{p}=\mathbf{h}^T\tilde{\mathbf{x}}$ of non-identifiable quantities $p=\mathbf{h}^T\mathbf{x}$, $\mathbf{h}\not\in span(\mathbf{C}^T)$, is possible but nevertheless meaningless.

Factorization is an important tool in modeling, especially when dealing with infinite dimensional models. As an example, consider observables y related to the relative positions and velocities of points on the earth and points, which are the instantaneous positions of satellites orbiting the earth. The orbits are governed by the attraction of the earth masses, fully determined by the density distribution function of the earth x. The original problem $y=Ax$ is factored as $y=Ax=BCx=Bz$ and $z=Cx$, where z is the gravitational potential function of the earth. The geodesist is concerned only with the solution of the problem $y=Bz$ leaving the remaining part $z=Cx$ to the geophysisist.

4 The statistical approach to parameter determination: Estimation and prediction.

The inverse problem approach followed in the previous chapter led to solutions which are not completely determined as a serious problem remains open: which values of the weight (metric) matrices $\mathbf{P}$ and $\mathbf{Q}$ should be used? In a more elaborate language: what metric properties should be given to the spaces X and Y, so that the solutions following from norm minimization are optimal in some sense that remains to be clarified. This problem is the inner product choice problem, or as it has been labeled in geodesy the *norm choice problem.*

The problem first arose in the case of full-rank overdetermined problems, where only the weight matrix $\mathbf{P}$ needs to be chosen. When observations of the same type were performed with the same instrument, the natural answer is to prescribe equal weights $p_i=p$ by setting $\mathbf{P}=p\mathbf{I}$, or $\mathbf{P}=\mathbf{I}$, since multiplication by a scalar does not alter the least-squares solution, e.g. $\hat{\mathbf{x}}=[\mathbf{A}^T(k\mathbf{P})\mathbf{A}]^{-1}\mathbf{A}^T(k\mathbf{P})\mathbf{b}=(\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}\mathbf{b}$.

The use of observations of different physical dimensions (e.g. angles and distances), or of the same type with different instruments of varying accuracy , makes the solu-

tion less easy to get, although a general guideline is obvious: prescribe larger weights to more accurate observations and vice-versa.

What is really needed in this case is a measure of the accuracy which comes from statistics. The accuracy is inverse to the variance of the observations performed by the instrument on the same observable over an infinite number of repetitions. Thus $p_i = 1/\sigma_i^2$, where σ_i^2 is the instrument error variance and $P_{ij} = \delta_{ij}\sigma_i^2$, where the zero non-diagonal elements reflect the statistical independence (no correlation) between different observational errors.

A formal justification of the above choice is given by the celebrated Gauss-Markov theorem. As already explained in section 2, a probabilistic or "stochastic" model is attached to the deterministic or "functional" model $\mathbf{y}=\mathbf{Ax}$, by setting $\mathbf{b}=\mathbf{Ax}+\mathbf{v}$, where $\mathbf{v}\sim(\mathbf{0},\mathbf{C}_v)$, meaning that $\mathbf{v}$ has zero mean $E\{\mathbf{v}\}=\mathbf{0}$ and (variance-)covariance matrix $E\{\mathbf{vv}^T\}=\mathbf{C}_v$, where $E\{\ \}$ is the expectation operator (mean over all possible realizations). It follows that $\mathbf{b}\sim(\mathbf{Ax},\mathbf{C}_v)$ is also a random vector.

We may attack directly the problem of parameter estimation $\hat{\mathbf{x}}=\mathbf{Gb}$ and try to find the optimal matrix $\mathbf{G}$, which minimizes an appropriate target function such as

$$E\{(\hat{\mathbf{x}}-\mathbf{x})^T(\hat{\mathbf{x}}-\mathbf{x})\}=trace[E\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^T\}] \tag{4.1}$$

(mean square error estimation) or a slightly more general one $E\{(\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{V}(\hat{\mathbf{x}}-\mathbf{x})\}$, where $\mathbf{V}$ is a positive definite matrix. The answer to such a minimization problem is obvious, trivial and useless: $\hat{\mathbf{x}}=\mathbf{x}$! In order to get rid of the dependence of the estimate on the unknown it seeks to estimate, a side criterion has to be introduced: the minimization is restricted to the class of uniformly unbiased estimates, i.e. the ones satisfying $E\{\hat{\mathbf{x}}\}=\mathbf{x}$, whatever the value of $\mathbf{x}$ may be. Since $E\{\hat{\mathbf{x}}\}=\mathbf{G}E\{\mathbf{b}\}=\mathbf{GAx}=\mathbf{x}$ the condition for uniform unbiasedness takes the algebraic form $\mathbf{GA}-\mathbf{I}=\mathbf{0}$. The solution of the minimization problem

$$\phi(\mathbf{G})=E\{(\hat{\mathbf{x}}-\mathbf{x})^T(\hat{\mathbf{x}}-\mathbf{x})\}=\mathbf{x}^T(\mathbf{GA}-\mathbf{I})^T(\mathbf{GA}-\mathbf{I})\mathbf{x}+\mathbf{GC}_v\mathbf{G}^T=\min_{\mathbf{GA}-\mathbf{I}=\mathbf{0}} \tag{4.2}$$

involves some complicated matrix differentiation and manipulation, which we prefer to avoid by following an algebraically simpler and yet more general approach, which applies also to the case where $r=rank(\mathbf{A})<m<n$. Recalling that the choice of parameters is not unique, we seek to estimate an arbitrary linear(ized) function of the parameters $q=\mathbf{a}^T\mathbf{x}$ by a linear function of the observations $\hat{q}=\mathbf{d}^T\mathbf{b}$, in such a way that the mean square estimation error $E\{e^2\}=E\{(\hat{q}-q)^2\}$ is minimized over all vectors $\mathbf{d}$ which yield uniformly unbiased estimates, i.e. estimates such that their bias $\beta\equiv E\{\hat{q}\}-q$ vanishes for any value of $\mathbf{x}$. Since $E\{\hat{q}\}=\mathbf{d}^T\mathbf{Ax}$ and $\beta=(\mathbf{A}^T\mathbf{d}-\mathbf{a})^T\mathbf{x}$, the condition for uniformly unbiased estimation (i.e. $\beta=0$ for any value of $\mathbf{x}$) is $\mathbf{A}^T\mathbf{d}-\mathbf{a}=\mathbf{0}$ and the minimization problem becomes

$$\phi(\mathbf{d})=E\{(\hat{q}-q)^2\}=[(\mathbf{A}^T\mathbf{d}-\mathbf{a})^T\mathbf{x}]^2+\mathbf{d}^T\mathbf{C}_v\mathbf{d}=\min_{\mathbf{A}^T\mathbf{d}-\mathbf{a}=\mathbf{0}}. \tag{4.3}$$

The minimization problem is easily solved by setting equal to zero the derivatives of the Lagrangian function $\Phi=\phi-2\mathbf{k}^T(\mathbf{A}^T\mathbf{d}-\mathbf{a})$ with respect to d and the vector of Lagrangean multiplies k. This leads to the system of equations

$$(\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{A})\mathbf{k}=\mathbf{a}, \qquad \mathbf{d}=\mathbf{C}_v^{-1}\mathbf{A}\mathbf{k} \tag{4.4}$$

and the estimate

$$\hat{q}=\mathbf{k}^T\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{b}. \tag{4.5}$$

For the overdetermined full-rank case ($r=m<n$) there exists a unique solution of (4.4) and the estimate becomes

$$\hat{q}=\mathbf{a}^T(\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{b}=\mathbf{a}^T\hat{\mathbf{x}}, \qquad \hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_v^{-1}\mathbf{b} \tag{4.6}$$

where the estimate $\hat{\mathbf{x}}$ follows by replacing $q=\mathbf{a}^T\mathbf{x}$ with each component $x_j=\mathbf{e}_j^T\mathbf{x}$, separately. Comparison with (3.103) shows that the optimal weight matrix choice problem has been solved by setting $\mathbf{P}=\sigma^2\mathbf{C}_v^{-1}$, where the scalar factor σ^2 does not affect the results. This means that if we set $\mathbf{C}_v=\sigma^2\mathbf{Q}_v$, then we have $\mathbf{P}=\mathbf{Q}_v^{-1}$, i.e. we need to know only $\mathbf{Q}_v$, while the "reference variance" σ^2 may be unknown. The Gauss-Markov theorem states that the best (=minimum mean square estimation error) uniformly (=whatever $\mathbf{x}$ might be) unbiased estimate of any linear function $q=\mathbf{a}^T\mathbf{x}$ is given by $\hat{q}=\mathbf{a}^T\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the least squares ($\mathbf{v}^T\mathbf{P}\mathbf{v}=\min$) solution of $\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{v}$, provided that the weight matrix $\mathbf{P}$ is related to the covariance matrix of the observations $\mathbf{C}_v=\sigma^2\mathbf{Q}_v$, where $\mathbf{Q}_v$ is known and σ^2 unknown, by $\mathbf{P}=\mathbf{Q}_v^{-1}$.

The generalized Gauss-Markov theorem for $r<\min(m,n)$.

The essential point here is that a uniformly unbiased estimate of $q=\mathbf{a}^T\mathbf{x}$ is possible only when the condition $\mathbf{A}^T\mathbf{d}=\mathbf{a}$ is a consistent equation, i.e. only when $\mathbf{a}\in R(\mathbf{A}^T)$. The unknown parameters $\mathbf{x}$ do not satisfy such a restriction and they are non-estimable quantities. However, if $\hat{\mathbf{x}}$ is any of the solutions of the normal equations

$$(\mathbf{A}^T\mathbf{Q}_v^{-1}\mathbf{A})\hat{\mathbf{x}}=\mathbf{A}^T\mathbf{Q}_v^{-1}\mathbf{b}, \tag{4.7}$$

i.e. any least-square solution with the choice $\mathbf{P}=\mathbf{Q}_v^{-1}$, then we can write the estimate (4.5) in the form

$$\hat{q} = \mathbf{k}^T \mathbf{A}^T \mathbf{Q}_v^{-1} \mathbf{b} = \mathbf{k}^T (\mathbf{A}^T \mathbf{Q}_v^{-1} \mathbf{A}) \hat{\mathbf{x}} = [(\mathbf{A}^T \mathbf{Q}_v^{-1} \mathbf{A})^T \mathbf{k}]^T \hat{\mathbf{x}} = [(\mathbf{A}^T \mathbf{Q}_v^{-1} \mathbf{A}) \mathbf{k}]^T \hat{\mathbf{x}} = \mathbf{a}^T \hat{\mathbf{x}} . \qquad (4.8)$$

This means that we may compute any one of the least squares solutions $\hat{\mathbf{x}}$ of (4.7), which is meaningless by itself since $\mathbf{x}$ is non-estimable, and use it as a "depository" for the computation of the best uniformly unbiased linear estimate $\hat{q}=\mathbf{a}^T \hat{\mathbf{x}}$ of any quantity $q=\mathbf{a}^T \mathbf{x}$, which is estimable, i.e. $\mathbf{a} \in R(\mathbf{A}^T)$.

The case of stochastic parameters

We can now turn to the case where we assign a stochastic character to the parameters also, i.e. we assume that $\mathbf{x}$ is a random vector $\mathbf{x} \sim (\mathbf{m}_x, \mathbf{C}_x)$ with mean $\mathbf{m}_x = E\{\mathbf{x}\}$ and covariance matrix $\mathbf{C}_x = E\{(\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T\}$. In this case we speak of a prediction rather than estimation of $\mathbf{x}$, since we seek to determine an estimate, not of the fixed value of a deterministic parameter, but of the outcome of the random variable in the specific realization (experiment) where the observations $\mathbf{b}$ have been determined as the outcomes of corresponding random variables. Some confusion may arise by the usual choice of denoting random variables and their outcomes with the same symbols: When we make computations the equations refer to the outcomes, while when, e.g., we apply the expectation operator $E\{\ \}$ we refer to the random variables.

The solution to the prediction error comes from a more general result which states that if $\mathbf{x} \sim (\mathbf{m}_x, \mathbf{C}_x)$ and $\mathbf{z} \sim (\mathbf{m}_z, \mathbf{C}_z)$ are stochastically dependent random vectors, in which case $\mathbf{C}_{zx} = \{(\mathbf{z}-\mathbf{m}_z)(\mathbf{x}-\mathbf{m}_x)^T\} \neq \mathbf{0}$ then the unbiased linear prediction $\hat{x}_i$ with minimum mean square error for any component x_i of $\mathbf{x}$ is provided by

$$\hat{\mathbf{x}} = \mathbf{m}_x + \mathbf{C}_{xz} \mathbf{C}_z^{-1} (\mathbf{z}-\mathbf{m}_z) . \qquad (4.9)$$

Here unbiased means that $E\{\hat{\mathbf{x}}\} = E\{\mathbf{x}\}$ while linear means of the form $\hat{x}_i = \mathbf{d}^T \mathbf{z} + \kappa$ (inhomogeneous linear). The restriction to the (homogeneous) linear class $\hat{x}_i = \mathbf{d}^T \mathbf{z}$ leads to a slightly different result which corresponds to what is called "Kriging" in geostatistics (see, e.g. Christakos, 1992).

In this case the model $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v}$, which is called a *random effects* model in the statistical classification of linear models, only serves for the determination of the relevant means and covariances (assuming $\mathbf{C}_{xv} = \mathbf{0}$):

$$\mathbf{m}_b = \mathbf{A}\mathbf{m}_x , \qquad \mathbf{C}_b = \mathbf{A}\mathbf{C}_x\mathbf{A}^T + \mathbf{C}_v , \qquad \mathbf{C}_{xb} = \mathbf{C}_x\mathbf{A}^T . \qquad (4.10)$$

Direct application of (4.9) with $\mathbf{z}=\mathbf{b}$ gives

$$\hat{\mathbf{x}} = \mathbf{m}_x + \mathbf{C}_{xb}\mathbf{C}_b^{-1}(\mathbf{b}-\mathbf{m}_b) = \mathbf{m}_x + \mathbf{C}_x\mathbf{A}^T(\mathbf{A}\mathbf{C}_x\mathbf{A}^T + \mathbf{C}_v)^{-1}(\mathbf{b}-\mathbf{A}\mathbf{m}_x) . \qquad (4.11)$$

It is possible to rewrite the model and the prediction in the forms

$$(\mathbf{b}-\mathbf{A}\mathbf{m}_{\mathbf{x}})=\mathbf{A}(\mathbf{x}-\mathbf{m}_{\mathbf{x}})+\mathbf{v}\,, \tag{4.12}$$

$$(\hat{\mathbf{x}}-\mathbf{m}_{\mathbf{x}})=\mathbf{C}_{\mathbf{xb}}\mathbf{C}_{\mathbf{b}}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{m}_{\mathbf{x}})=\mathbf{C}_{\mathbf{x}}\mathbf{A}^{T}(\mathbf{A}\mathbf{C}_{\mathbf{x}}\mathbf{A}^{T}+\mathbf{C}_{\mathbf{v}})^{-1}(\mathbf{b}-\mathbf{A}\mathbf{m}_{\mathbf{x}})\,, \tag{4.13}$$

which means that by setting $\mathbf{x}-\mathbf{m}_{\mathbf{x}}\rightarrow\mathbf{x}$, $\mathbf{b}-\mathbf{A}\mathbf{m}_{\mathbf{x}}\rightarrow\mathbf{b}$, we can restrict ourselves without any loss of generality to the model $\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{v}$ with $\mathbf{x}\sim(\mathbf{0},\mathbf{C}_{\mathbf{x}})$ and optimal prediction

$$\hat{\mathbf{x}}=\mathbf{C}_{\mathbf{xb}}\mathbf{C}_{\mathbf{b}}^{-1}\mathbf{b}=\mathbf{C}_{\mathbf{x}}\mathbf{A}^{T}(\mathbf{A}\mathbf{C}_{\mathbf{x}}\mathbf{A}^{T}+\mathbf{C}_{\mathbf{v}})^{-1}\mathbf{b}\,. \tag{4.14}$$

In fact since we are mostly dealing with linearized models we can use $\mathbf{m}_{\mathbf{x}}=E\{\mathbf{x}^{a}\}$ as approximate values $\mathbf{x}^{0}$ of the original unknowns $\mathbf{x}^{a}$, in which the resulting corrections $\mathbf{x}=\mathbf{x}^{a}-\mathbf{x}^{0}=\mathbf{x}^{a}-\mathbf{m}_{x}$, which appear as unknowns in the linearized model, will have zero mean $E\{\mathbf{x}\}=E\{\mathbf{x}^{a}\}-\mathbf{m}_{x}=\mathbf{0}$.

Remark:
In any case we impose the invariance of the model under translations (as well as under multiplication by non-singular matrices) as a requirement for a "reasonable" estimation or prediction method. This is a must because the choice of parameters, of approximate values in the linearization and the use of synthetic and/or reduced observations is a matter of fact in geodetic practice. Other estimation and prediction methods are possible which do not obey these invariance characteristics, but they have to offer a certain "robustness" against the use of e.g. wrong means $\mathbf{m}_{\mathbf{x}}$. We refer to Schaffrin (1983, 1985) for these methods, as well as to Dermanis (1988) for an alternative down-to-earth derivation of the algorithms used in some of them and to Dermanis (1991) for a study of their invariance properties.

Let us know assume that $\mathbf{C}_{\mathbf{x}}=\sigma_{x}^{2}\mathbf{Q}_{\mathbf{x}}$, $\mathbf{C}_{\mathbf{v}}=\sigma_{v}^{2}\mathbf{Q}_{\mathbf{v}}$, where $\mathbf{Q}_{\mathbf{x}}$, $\mathbf{Q}_{\mathbf{v}}$ are known and σ_{x}^{2}, σ_{v}^{2} unknown. In this case (4.14) can be written in the alternative form

$$\hat{\mathbf{x}}=\mathbf{Q}_{\mathbf{x}}\mathbf{A}^{T}(\mathbf{A}\mathbf{Q}_{\mathbf{x}}\mathbf{A}^{T}+\frac{\sigma_{v}^{2}}{\sigma_{x}^{2}}\mathbf{Q}_{\mathbf{v}})^{-1}\mathbf{b}\,. \tag{4.15}$$

This solution can be identified with the hybrid solution (3.129) of Tikhonov regularization, provided that we choose $\mathbf{P}=\mathbf{Q}_{\mathbf{v}}^{-1}$, $\mathbf{Q}=\mathbf{Q}_{\mathbf{x}}^{-1}$ and $\alpha=\frac{\sigma_{v}^{2}}{\sigma_{x}^{2}}$. Thus a probabilistic justification of the hybrid norm is provided.

The combination of deterministic and stochastic parameters

In some cases a probabilistic model can be justified only for a subset $\mathbf{x}_1$ of the parameters, while the remaining ones $\mathbf{x}_2$ retain their deterministic character. The linear model $\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{v}=\mathbf{A}_1\mathbf{x}_1+\mathbf{A}_2\mathbf{x}_2+\mathbf{v}$ can also be written in the form

$$\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{G}\mathbf{s}+\mathbf{v}, \quad \mathbf{s}\sim(\mathbf{0},\mathbf{C}_s), \quad \mathbf{v}\sim(\mathbf{0},\mathbf{C}_v), \quad \mathbf{C}_{sv}=\mathbf{0}, \tag{4.16}$$

where we have let $\mathbf{A}_1\to\mathbf{A}$, $\mathbf{x}_1\to\mathbf{x}$, $\mathbf{A}_2\to\mathbf{G}$, $\mathbf{x}_2\to\mathbf{s}$. This is the *mixed effects model* in the statistical terminology related to the linear model.

The problem now can be solved in two stages according to the separation

$$\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{e}, \qquad \mathbf{e}=\mathbf{G}\mathbf{s}+\mathbf{v}, \tag{4.17}$$

where we can directly compute the covariance matrices of $\mathbf{e}\sim(\mathbf{0},\mathbf{C}_e)$

$$\mathbf{C}_e=\mathbf{G}\mathbf{C}_s\mathbf{G}^T+\mathbf{C}_v, \qquad \mathbf{C}_{se}=\mathbf{C}_s\mathbf{G}^T, \qquad \mathbf{C}_{ve}=\mathbf{C}_v. \tag{4.18}$$

In the first stage we estimate

$$(\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A})\hat{\mathbf{x}}=\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{b}, \qquad \hat{\mathbf{e}}=\mathbf{b}-\mathbf{A}\hat{\mathbf{x}} \tag{4.19}$$

and in the second stage we predict

$$\hat{\mathbf{s}}=\mathbf{C}_{s\hat{e}}\mathbf{C}_{\hat{e}}^{-1}\hat{\mathbf{e}}, \qquad \hat{\mathbf{v}}=\mathbf{C}_{v\hat{e}}\mathbf{C}_{\hat{e}}^{-1}\hat{\mathbf{e}}. \tag{4.20}$$

Restricting ourselves to the regular case $r(\mathbf{A})=m<n$, where $|\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A}|\neq 0$, we have

$$\hat{\mathbf{x}}=(\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{b}, \tag{4.21}$$

$$\hat{\mathbf{e}}=\mathbf{b}-(\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{b}=[\mathbf{I}-(\mathbf{A}^T\mathbf{C}_e^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{C}_e^{-1}]\mathbf{e}\equiv\mathbf{H}\mathbf{e}. \tag{4.22}$$

The "projection" matrix $\mathbf{H}=\mathbf{P}_{R(A)^\perp}$ has the properties $\mathbf{H}^2=\mathbf{H}$ and $\mathbf{H}^T=\mathbf{C}_e^{-1}\mathbf{H}\mathbf{C}_e$, which can be implemented to prove that $\mathbf{C}_{s\hat{e}}\mathbf{C}_{\hat{e}}^{-1}=\mathbf{C}_{se}\mathbf{C}_e^{-1}$, $\mathbf{C}_{v\hat{e}}\mathbf{C}_{\hat{e}}^{-1}=\mathbf{C}_{ve}\mathbf{C}_e^{-1}$ and the predictions in the form

$$\hat{\mathbf{s}}=\mathbf{C}_{se}\mathbf{C}_e^{-1}\hat{\mathbf{e}}=\mathbf{C}_s\mathbf{G}^T(\mathbf{G}\mathbf{C}_s\mathbf{G}^T+\mathbf{C}_v)^{-1}(\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}), \tag{4.23}$$

$$\hat{\mathbf{v}}=\mathbf{C}_{ve}\mathbf{C}_e^{-1}\hat{\mathbf{e}}=\mathbf{C}_v(\mathbf{G}\mathbf{C}_s\mathbf{G}^T+\mathbf{C}_v)^{-1}(\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}). \tag{4.24}$$

If $\mathbf{C}_s=\sigma_s^2\mathbf{Q}_s$, $\mathbf{C}_v=\sigma_v^2\mathbf{Q}_v$, where $\mathbf{Q}_s$, $\mathbf{Q}_v$ are known and σ_s^2, σ_v^2 unknown, we can write the solution in the form

$$\hat{\mathbf{x}}=[\mathbf{A}^T(\mathbf{G}\mathbf{Q}_s\mathbf{G}^T+\alpha\mathbf{Q}_v)^{-1}\mathbf{A}]^{-1}\mathbf{A}^T(\mathbf{G}\mathbf{Q}_s\mathbf{G}^T+\alpha\mathbf{Q}_v)^{-1}\mathbf{b} \tag{4.25}$$

$$\hat{\mathbf{s}}=\mathbf{Q}_s\mathbf{G}^T(\mathbf{G}\mathbf{Q}_s\mathbf{G}^T+\alpha\mathbf{Q}_v)^{-1}(\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}), \tag{4.26}$$

$$\hat{\mathbf{v}}=\alpha\mathbf{Q}_v(\mathbf{G}\mathbf{Q}_s\mathbf{G}^T+\alpha\mathbf{Q}_v)^{-1}(\mathbf{b}-\mathbf{A}\hat{\mathbf{x}}), \tag{4.27}$$

where the ratio $\alpha=\dfrac{\sigma_v^2}{\sigma_s^2}$, should be known, or properly chosen.

5 From finite to infinite-dimensional models (or from discrete to continuous models)

The generalization of the model $y=Ax$, to infinite dimensional x and y, is by no means trivial and requires a satisfactory mathematical background on functional analysis and the theory of stochastic processes (random fields). We will take here a more modest approach, where we will try to point out the essential features of this generalization, sacrificing mathematical rigor and refer to Sansò (1986) for a more advanced approach. To be more precise the model consists of the following elements:

(a) a definition of the space X,
(b) a definition of the space Y,
(c) a definition of the mapping $A:X\rightarrow Y$
(d) a known element $b\in Y$.

In the finite dimensional case, the only choice associated with X and Y is that of their metric (weight) matrices in some particular representations. On the contrary, the characterization of the infinite dimensional spaces X and Y is a more delicate matter from both a physical and a mathematical point of view.

We should first examine whether there is a need to consider infinite dimensional spaces at all. In many cases the unknown x is, or includes, one or more unknown functions, e.g. the potential function of the gravity field of the earth. There are situations where a function may be completely, or at least efficiently described by a finite number of parameters. For example, in order to describe the attractive force of the sun on a satellite orbiting the earth, it is sufficient to know the components of the satellite-to-sun vector. However we need theoretically an infinite number of parameters to describe the corresponding attraction of the earth, due to the presence of unknown variations in the density of the earth masses, whose effect cannot be ignored. It is true, that in practice we can always find a finite dimensional representation of the unknown function which is more than sufficient to describe the corresponding physical effects in the framework of the analysis of a specific data set, but the number of necessary parameters cannot be determined a priori. The choice of a very large number of pa-

rameters leads to another problem: there might be no sufficient information in the observations to determine these parameters, so that the obtained estimates are meaningless, when seen by themselves. For these reasons, a representation of the unknown function with an infinite number of parameters is useful and necessary, if one is to have a control over computational procedures, which will finally implement only a finite number of them. (The computation of an infinite number of parameter estimates would require infinite computation time!).

Among the various types of function representations, series representations are the most popular ones. One reason for this preference is the linear or "vector-type" character of the representation, because the function ϕ is represented by a linear combination $\phi=a_1\phi_1+a_2\phi_2+\ldots$ of known base functions ϕ_k, in the same way that a vector is represented by a linear combination of base vectors. The existence of an infinite number of "components" requires a selection and ordering of the base functions in such a way that the components a_k decline fast enough for the series to converge.

Spectral representations, employing the use of spherical harmonic base functions, have played a dominant role, despite certain disadvantages in comparison to the use of "localized" functions (e.g. point masses, splines, wavelets, etc.) which are best adapted to the representation of local features.

The reason of the popularity of spectral representations lies in the fact that they use base functions which are the eigenfunctions, $L(\phi_k)=\lambda_k\phi_k$, of a wide class operators L, identified, or closely related, to the operator A present in a data analysis problem $y=Ax$.

The origin of spectral representations is Fourier analysis on the real line, where the trigonometric functions $\phi_\omega(t)=e^{i\omega t}$ ($i=\sqrt{-1}$) are eigenfunctions of any translation-invariant linear operator L (usually called a linear system in signal analysis), i.e. any linear operator which commutes ($U_\tau L=LU_\tau$) with the function-translation operators $(U_\tau f)(t)\equiv f(t+\tau)$. Such linear translation-invariant integral operators accept a convolution representation $(Lf)(t)=\int_{-\infty}^{+\infty}k(t,s)f(s)ds$, with a kernel $k(t,s)=k(|t-s|)$, which depends only on the absolute value of the time difference $|t-s|$.

In geodesy we are concerned with functions defined on a (unit) sphere and the linear operators which are rotation-invariant are having spherical harmonic functions as their eigenfunctions. More precisely let $R=R(\theta_1,\theta_2,\theta_3)$ be a rotation operator on the sphere, represented e.g. by the orthogonal matrix $\mathbf{R}(\boldsymbol{\theta})=\mathbf{R}_3(\theta_3)\mathbf{R}_2(\theta_2)\mathbf{R}_1(\theta_1)$ and L an operator acting on functions defined on a sphere, which is linear

$$L(a_1f_1+a_2f_2)=a_1L(f_1)+a_2L(f_2) \tag{5.1}$$

and commutes with rotations, i.e., it satisfies $RL=LR$, or explicitly

$$(RLf)(\boldsymbol{\eta})\equiv(Lf)(\mathbf{R}\boldsymbol{\eta})=(LRf)(\boldsymbol{\eta})=(Lf_R)(\boldsymbol{\eta}),\qquad f_R(\boldsymbol{\xi})=(Rf)(\boldsymbol{\xi})\equiv f(\mathbf{R}\boldsymbol{\xi}) \tag{5.2}$$

for a wide class of functions f ($\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are unit vectors). In this case

$$Le_{nm}=\lambda_n e_{nm} \tag{5.3}$$

where

$$e_{nm}(\lambda,\theta)=z_{nm}\cos(m\lambda)P_{nm}(\cos\theta)\,, \tag{5.4}$$

$$e_{n,-m}(\lambda,\theta)=z_{nm}\sin(m\lambda)P_{nm}(\cos\theta)\,, \qquad n=0,1,\ldots\,, \quad m=0,1,\ldots,n\,, \tag{5.5}$$

are the spherical harmonic functions, P_{nm} are the associated Legendre functions and λ,θ,r are the usual spherical coordinates (longitude, co-latitude, distance from origin), The constant normalization factor z_{nm} is introduced in order to make the "norm" $\|e_{nm}\|$ of the spherical harmonics unity, i.e.,

$$\|e_{nm}\|^2=\frac{1}{4\pi}\int_\sigma (e_{nm})^2\,d\sigma=1\,. \tag{5.6}$$

The linear rotation-invariant integral operators on the unit sphere accept a convolution representation

$$(Lf)(\xi)=\int_\sigma k(\xi,\boldsymbol{\eta})f(\boldsymbol{\eta})\,d\sigma(\boldsymbol{\eta})=\int_\sigma k(\xi\cdot\boldsymbol{\eta})f(\boldsymbol{\eta})\,d\sigma(\boldsymbol{\eta}) \tag{5.7}$$

where the kernel depends only on the inner product $\xi\cdot\boldsymbol{\eta}=\cos\psi_{\xi,\eta}$, i.e. on the spherical distance $\psi_{\xi,\eta}$ between the points defined by the unit vectors ξ and $\boldsymbol{\eta}$. A typical example is the Stokes integral operator, which maps gravity anomalies into geoid undulations within the framework of a spherical approximation (Heiskanen & Moritz, 1967).

The translation and rotation invariance are important for two reasons, corresponding to two different ways in which the transformations may arise. True physical transformations mean that we apply the action described by the operator at a different epoch or at a different part of the earth and we expect to have same the result, except that it comes out "transformed" in the same way that the input was transformed (delayed or rotated). On the other hand transformations may result without any physical change of time or place, but simply as a result of a change in the reference system used. In this case invariance means that the physical effect of the operator, as described by the particular mathematical representation, is independent of the coordinate system used either for time or for the representation of points on a sphere. An approach for the treatment of discrete data based on invariance requirements has been presented by Sansò (1978).

We may now return to the question of the choice of a proper function space X, where the unknown function f, namely the potential function of the gravity field of

the earth, belongs. Based on the physics of the problem we may start with $X=H_\Sigma$ as the set of all spatial functions which are harmonic outside the earth surface $\partial\Sigma$, i.e. functions satisfying $(\Delta f)(P)=0$, $P\in\Sigma^c$, where Σ is the part of space which is occupied by earth masses and $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$ is the Laplace operator. In addition the regularity condition $\lim_{r\to\infty} f=0$ is added in order to get rid of any additive constant.

The problem is that such a set of functions cannot be described by simple (or even modestly complex) mathematical tools, because of the complex and irregular shape of the earth surface $\partial\Sigma$. For this reasons we use a space $X=H_S$ of functions which are harmonic in the exterior S^c of a sphere ∂S instead of outside the earth Σ. The origin of the sphere $\partial S=\{P\sim(\lambda,\theta,r)|r=R\}$ of radius R coincides with the coordinate origin, which is the mass center of the earth. Such a sphere ∂S of radius R can be identified with the unit sphere σ, through $f(\lambda,\theta,R)=f(\lambda,\theta)$, where the left side is a function on ∂S and the right one a function defined on σ.

Two choices are possible, the Brillouin sphere ($S\supset\Sigma$) and the Bjerhammar sphere ($S\subset\Sigma$). In the case of the Brillouin sphere $H_\Sigma\subset H_S$ and the potential function f still belongs to H_S. However the representation by spherical harmonics is not guaranteed to converge also in $\Sigma\cap S^c$ (i.e., down to the earth surface) and for this reason we use instead the Bjerhammar sphere. The actual potential function is not harmonic down to the Bjerhammar sphere. It satisfies instead the Poisson equation $\Delta f=-4\pi G\rho$, in $\Sigma\cap S^c$, where ρ is the density function of the earth and G the gravitational constant, and thus $f\notin H_S$. To circumvent this difficulty we resort to the famous Runge-Krarup theorem, which states that although $f\notin H_S$ it can be approximated arbitrarily well (in a certain sense) by elements of H_S.

Even in this case H_S is too large a function space to lead to a tractable "estimation" model where some of the solution concepts developed for the finite-dimensional case may be adapted.

The use of a spherical boundary for the domain of harmonicity allows us to switch from the space of functions harmonic outside the sphere to the space of functions on the surface of the sphere, thanks to the Dirichlet principle. This principles states that a harmonic function can be uniquely defined by its values on the boundary of its harmonicity domain. To any function $x(\lambda,\theta)$ defined on the sphere surface, with spherical harmonic representation

$$x(\lambda,\theta)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}x_{nm}e_{nm}(\lambda,\theta) \tag{5.8}$$

corresponds a unique spatial function $f=Ux$

$$f(\lambda,\theta,r)=(Ux)(\lambda,\theta,r)=\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n}x_{nm}e_{nm}(\lambda,\theta) \tag{5.9}$$

harmonic outside the sphere. U is the "upward harmonic continuation" operator, while its inverse U^{-1} is the "restriction to the sphere" operator

$$x(\lambda,\theta)=(U^{-1}f)(\lambda,\theta)=f(\lambda,\theta,r)_{|r=R}=f(\lambda,\theta,R)\,. \tag{5.10}$$

The use of the operators U and U^{-1} plays an essential role in transforming a problem originally defined in the space outside the sphere, to an equivalent problem defined on the sphere, so that the eigenvalue decomposition can be applied in a simpler and more straightforward way.

To extend the results of the finite-dimensional inversion problem to the present infinite-dimensional situation we need a definition of the inner product and we may choose the one which makes the spherical harmonics orthonormal

$$<f,g>=\frac{1}{4\pi}\int_{\sigma}fg\,d\sigma\,, \tag{5.11}$$

$$<e_{nm},e_{pq}>=\begin{cases}1 & n=p,\ m=q\\ 0 & \text{otherwise}\end{cases}\quad\Rightarrow\quad <f,g>=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}f_{nm}g_{nm}\,. \tag{5.12}$$

The convergence of the last expansion is guaranteed (in view of the Cauchy-Schwarz inequality $<f,g>\leq\|f\|\|g\|$) if we restrict ourselves to functions such that

$$\|f\|^2=\int_{\sigma}f^2d\sigma=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(f_{nm})^2<\infty\,. \tag{5.13}$$

Thus we have arrived at a space $X=H_{L^2(\sigma)}$ of harmonic functions with square integrable restrictions on the surface of the unit sphere and our original unknown potential function f has been replaced by its restriction to the sphere $x=U^{-1}f$ which serves as a new unknown.

We now turn our attention to the space Y and the observables y. The case of infinite-dimensional Y and an observed function y corresponds to continuous observations over a certain domain, typically the surface of the earth $\partial\Sigma$. This is of course a theoretical abstraction, while true observations are always carried out at discrete points. The continuous observations can be seen as a limiting case when the density of the observation points becomes larger and larger. The case where we assume that the observed function b is identified with the observable function y is of great theoreti-

cal interest, since a meaningful solution to the problem $y=Ax$ is the first step to the determination of x from noisy data $b=y+v=Ax+v$.

A mathematical difficulty arises in passing from the discrete to the continuous case, because discrete observation errors v_i which are assumed to be independent random variables, will have to be replaces by a "white noise" random function v with erratic behavior. A treatment of this problem can be found in the approach developed by Sansò and Sona (1995) which introduces random functions defined over subsets of the relevant domain of definition (Wiener measure).

We will restrict ourselves here to the study of two cases:

(a) continuous observations $b=y=Ax$, which are assumed to be errorless.

(b) discrete observations $b_i+y_i+v_i$ where each observable $y_i=(Ax)(P_i)$ is the result of the evaluation at a particular point of the image Ax of the unknown function x under one (or more) linear(ized) operator A.

An example of such an operator A is the gravity anomaly operator

$$Ax=DUx=(-\tfrac{2}{r}-\tfrac{\partial}{\partial r})Ux\,, \tag{5.14}$$

where U is the upward harmonic continuation operator defined by equation (5.9) and $D=(-\frac{2}{r}-\frac{\partial}{\partial r})$ is a spatial partial differential operator.

5. 1 Continuous observations without errors

Within a spherical approximation, we may consider continuous coverage on a spherical earth with radius coinciding with the Bjerhammar radius R. Then the restriction to the sphere operator U^{-1} should be also included as a last (from the left) factor of A, which can now be seen as a mapping with domain and range functions defined on the unit sphere.

$$y=Ax=U^{-1}DUx=\left[-\frac{2}{r}(Ux)-\frac{\partial(Ux)}{\partial r}\right]_{r=R}=-\frac{2}{R}(Ux)_{r=R}-\left.\frac{\partial(Ux)}{\partial r}\right|_{r=R}=$$

$$=-\frac{2}{R}x-\left.\frac{\partial(Ux)}{\partial r}\right|_{r=R}. \tag{5.15}$$

Introducing the *spatial spherical harmonics* $\varepsilon_{nm}=Ue_{nm}$ or explicitly

$$\varepsilon_{nm}(\lambda,\theta,r)=\left(\frac{R}{r}\right)^{n+1}e_{nm}(\lambda,\theta)\,, \tag{5.16}$$

we have

$$D\varepsilon_{nm} = -\frac{2}{r}\varepsilon_{nm} - \frac{\partial \varepsilon_{nm}}{\partial r} = \frac{n-1}{r}\varepsilon_{nm} \tag{5.17}$$

and therefore

$$Ae_{nm} = U^{-1}DUe_{nm} = U^{-1}D\varepsilon_{nm} = U^{-1}\left(\frac{n-1}{r}\varepsilon_{nm}\right) = \frac{n-1}{R}e_{nm} \doteq \lambda_{nm}e_{nm}, \tag{5.18}$$

$$\lambda_{nm} = \lambda_n = \frac{n-1}{R}, \tag{5.19}$$

which means that the spherical harmonics constitute a set of eigenfunctions of A.

The gravity anomaly function $y = \Delta g$ on the unit sphere can similarly be expressed in an eigenvalue expansion $y = \Sigma_{nm} y_{nm} e_{nm}$. In this case, within a linearization procedure, the original potential function f is replaced by the disturbing potential function $T = f - f_0$, where the "normal potential" f_0 is a known approximation of f and $x = \Sigma_{nm} x_{nm} e_{nm} = UT$ is the restriction of T to the sphere.

Thus the original equation $y = Ax$, has obtained the eigenvalue decomposition (EVD) form

$$y_{nm} = \lambda_{nm} x_{nm} = \frac{n-1}{R} x_{nm}. \tag{5.20}$$

For $n=1$ we have $y_{1,-1} = y_{1,0} = y_{1,1} = 0$ and $x_{1,-1}$, $x_{1,0}$, $x_{1,1}$ may take any arbitrary value. The components $x_{1,-1}$, $x_{1,0}$, $x_{1,1}$ are in fact scalar multiples of the cartesian coordinates x_c, y_c, z_c, respectively, of the center of mass of the earth. The minimum norm solution with $x_{1,-1} = x_{1,0} = x_{1,1} = 0$, corresponds to choosing to use a reference frame with origin at the geocenter. For $n=0$ the component v_{00} is a scalar multiple of the difference $\delta M = M - M_0$, between the actual mass of the earth M and the mass M_0 implicit in the definition of the used normal potential f_0.

If we restrict ourselves to functions with vanishing zero and first degree terms and to operators A which have the spherical harmonics e_{nm} as eigenfunctions ($Ae_{nm} = \lambda_{nm} e_{nm}$) with eigenvalues $\lambda_{nm} = \lambda_n$ which depend only on the degree n and not on the order m, then the equation $y = Ax$ has a unique solution $x = A^{-1}y$, which takes the EVD (eigenvalue decomposition) form

$$x_{nm} = \frac{1}{\lambda_n} y_{nm}. \tag{5.21}$$

In order to get an explicit form for x we will use the relation

$$y_{nm} = <e_{nm}, y> = \frac{1}{4\pi} \int_{\sigma} e_{nm}(\boldsymbol{\eta})\, y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) \tag{5.22}$$

and the addition theorem

$$P_n(\boldsymbol{\xi} \cdot \boldsymbol{\eta}) = \frac{1}{2n+1} \sum_{m=-n}^{n} e_{nm}(\boldsymbol{\xi}) e_{nm}(\boldsymbol{\eta}) \tag{5.23}$$

to obtain

$$x(\boldsymbol{\xi}) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} x_{nm} e_{nm}(\boldsymbol{\xi}) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \frac{1}{\lambda_n} y_{nm} e_{nm}(\boldsymbol{\xi}) =$$

$$= \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \frac{1}{\lambda_n} \left[\frac{1}{4\pi} \int_{\sigma} e_{nm}(\boldsymbol{\eta})\, y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) \right] e_{nm}(\boldsymbol{\xi}) = \frac{1}{4\pi} \int_{\sigma} \left[\sum_{n=2}^{\infty} \frac{1}{\lambda_n} \sum_{m=-n}^{n} e_{nm}(\boldsymbol{\eta}) e_{nm}(\boldsymbol{\xi}) \right] y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) =$$

$$= \frac{1}{4\pi} \int_{\sigma} \left[\sum_{n=2}^{\infty} \frac{1}{\lambda_n} (2n+1) P_n(\boldsymbol{\xi} \cdot \boldsymbol{\eta}) \right] y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) \cdot \tag{5.24}$$

Thus the inverse operator $K \equiv A^{-1}$ is an integral rotational invariant operator

$$x(\boldsymbol{\xi}) = (A^{-1} y)(\boldsymbol{\xi}) = \frac{1}{4\pi} \int_{\sigma} k(\boldsymbol{\xi} \cdot \boldsymbol{\eta})\, y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) = <k(\boldsymbol{\xi} \cdot \boldsymbol{\eta}), y(\boldsymbol{\eta})>_{\boldsymbol{\eta}} \tag{5.25}$$

with kernel depending only on the spherical distance $\psi = \psi_{\boldsymbol{\xi}, \boldsymbol{\eta}} = \arccos(\boldsymbol{\xi} \cdot \boldsymbol{\eta})$

$$k(\psi) = k(\boldsymbol{\xi} \cdot \boldsymbol{\eta}) = \sum_{n=2}^{\infty} \frac{2n+1}{\lambda_n} P_n(\cos\psi) \cdot \tag{5.26}$$

A spatial extension $K' \equiv UA^{-1}$ of the inverse operator may be obtained by

$$x(\boldsymbol{\xi}) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \left(\frac{R}{r}\right)^{n+1} x_{nm} e_{nm}(\boldsymbol{\xi}) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \left(\frac{R}{r}\right)^{n+1} \frac{1}{\lambda_n} y_{nm} e_{nm}(\boldsymbol{\xi}) =$$

$$= \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \left(\frac{R}{r}\right)^{n+1} \frac{1}{\lambda_n} \left[\frac{1}{4\pi} \int_{\sigma} e_{nm}(\boldsymbol{\eta})\, y(\boldsymbol{\eta})\, d\sigma(\boldsymbol{\eta}) \right] e_{nm}(\boldsymbol{\xi}) =$$

$$= \frac{1}{4\pi}\int_\sigma \left[\sum_{n=2}^{\infty}\frac{1}{\lambda_n}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n} e_{nm}(\eta)e_{nm}(\xi)\right] y(\eta)\, d\sigma(\eta) =$$

$$= \frac{1}{4\pi}\int_\sigma \left[\sum_{n=2}^{\infty}\frac{1}{\lambda_n}\left(\frac{R}{r}\right)^{n+1}(2n+1)P_n(\xi\cdot\eta)\right] y(\eta)\, d\sigma(\eta) = \frac{1}{4\pi}\int_\sigma k'(\xi\cdot\eta)\, y(\eta)\, d\sigma(\eta)\,, \qquad (5.27)$$

with "spatial" kernel

$$k'(\psi,r,R) = \sum_{n=2}^{\infty}\frac{1}{\lambda_n}\left(\frac{R}{r}\right)^{n+1}(2n+1)P_n(\cos\psi)\,. \qquad (5.28)$$

We may now apply the above results to particular cases, starting with the gravity anomaly operator $A=U^{-1}DU$, where $\lambda_n=\frac{n-1}{R}$ and $S=A^{-1}$ is the *Stokes operator* with kernel

$$s(\psi)=R\sum_{n=2}^{\infty}\frac{2n+1}{n-1}P_n(\cos\psi)=R\left[\frac{1}{\sin\frac{1}{2}\psi}+1-6\sin\tfrac{1}{2}\psi-5\cos\psi-3\cos\psi\,\ln\left(\sin\tfrac{1}{2}\psi+\sin^2\tfrac{1}{2}\psi\right)\right] \qquad (5.29)$$

$S'=US$ is the *spatial Stokes operator* with kernel

$$s'(\psi,r,R)=R\sum_{n=2}^{\infty}\frac{2n+1}{n-1}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi)=R\left[\frac{2R}{l}+\frac{R}{r}-\frac{3Rl}{r^2}-\frac{R^2}{r^2}\cos\psi\left(5+3\ln\frac{r-R\cos\psi+l}{2r}\right)\right] \qquad (5.30)$$

where $l=\sqrt{R^2+r^2-2Rr\cos\psi}$, which solves the third boundary problem on the sphere:

$$\Delta f=0 \text{ in } S^c, \qquad \lim_{r\to\infty} f=0, \qquad -\frac{\partial f}{\partial r}-\frac{2}{r}f=y \text{ on } \partial S\,. \qquad (5.31)$$

The differential operator $D_r=-\frac{\partial}{\partial r}$ is associated with gravity disturbances δg which appear (instead of the gravity anomalies) when the classical free geodetic boundary value problem is replaced by a fixed one, where the shape of boundary surface to which the data refer is known. The operator $A=U^{-1}D_rU$ has eigenvalues $\lambda_n=\frac{n+1}{R}$, so that the inverse operator $H=A^{-1}$ has kernel

$$h(\psi)=R\sum_{n=0}^{\infty}\frac{2n+1}{n+1}P_n(\cos\psi)=R\left[\frac{1}{\sin\frac{1}{2}\psi}-\ln\frac{1+\sin\frac{1}{2}\psi}{\sin\frac{1}{2}\psi}\right], \qquad (5.32)$$

with spatial counterpart $H'=UH$ with kernel (*Hotine* or *Neumann kernel*)

$$h'(\psi,r,R)=R\sum_{n=0}^{\infty}\frac{2n+1}{n+1}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi)=R\left[\frac{2R}{l}-\ln\frac{l+R-r\cos\psi}{r(1-\cos\psi)}\right], \tag{5.33}$$

which solves the second (Neumann) boundary value problem for the sphere

$$\Delta f=0 \text{ in } S^c, \qquad \lim_{r\to\infty} f=0, \qquad -\frac{\partial f}{\partial r}=y \text{ on } \partial S. \tag{5.34}$$

The differential operator $D_{rr}=\frac{\partial^2}{\partial r^2}$, is associated with one of the gradiometry observations. Then $A=U^{-1}D_{rr}U$ has eigenvalues $\lambda_n=\frac{(n+1)(n+2)}{R^2}$ and the inverse operator $G=A^{-1}$ has kernel

$$g(\psi)=R^2\sum_{n=0}^{\infty}\frac{2n+1}{(n+1)(n+2)}P_n(\cos\psi)=R^2\sum_{n=0}^{\infty}\left(\frac{3}{n+2}-\frac{1}{n+1}\right)P_n(\cos\psi)=$$

$$=R^2\left[-3+6\sin\tfrac{1}{2}\psi+(3\cos\psi-1)\ln\frac{1+\sin\frac{1}{2}\psi}{\sin\frac{1}{2}\psi}\right]. \tag{5.35}$$

The spatial extension $G'=UG$ has kernel

$$g'(\psi,r,R)=R^2\sum_{n=0}^{\infty}\frac{2n+1}{(n+1)(n+2)}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi)=R^2\left[\frac{3(l-r)}{R}+(3\frac{r}{R}\cos\psi-1)\ln\frac{1+\sin\frac{1}{2}\psi}{\sin\frac{1}{2}\psi}\right]. \tag{5.36}$$

The upward continuation operator U can also expressed in an integral form

$$(Ux)(\xi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}x_{nm}e_{nm}(\xi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left(\frac{R}{r}\right)^{n+1}\left[\frac{1}{4\pi}\int_{\sigma}e_{nm}(\eta)x(\eta)\,d\sigma(\eta)\right]e_{nm}(\xi)=$$

$$=\frac{1}{4\pi}\int_{\sigma}\left[\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n}e_{nm}(\eta)e_{nm}(\xi)\right]x(\eta)\,d\sigma(\eta)=$$

$$=\frac{1}{4\pi}\int_{\sigma}\left[\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}(2n+1)P_n(\xi\cdot\eta)\right]x(\eta)\,d\sigma(\eta)=\frac{1}{4\pi}\int_{\sigma}u'(\xi\cdot\eta)x(\eta)\,d\sigma(\eta), \tag{5.37}$$

with kernel (Poisson kernel)

$$u'(\psi,r,R)=\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi)=R\frac{r^2-R^2}{l^3}, \tag{5.38}$$

which solves the first (Dirichlet) boundary problem on the sphere:

$$\Delta f=0 \text{ in } S^c, \qquad \lim_{r\to\infty} f=0, \qquad f=y \text{ on } \partial S. \tag{5.39}$$

To summarize, we have solved the problem of continuous data by resorting to the following:

(a) A convenient choice of the space X and the unknown function $x\in X$. We have replaced the actual space H_Σ of functions harmonic outside the earth, first with the space H_S of functions harmonic outside the Bjerhammar sphere and next with a space of functions $F(\sigma)$ on the unit sphere, by replacing each $f\in H_S$ by its restriction $x=U^{-1}f=f_{r=R}$ on the Bjerhammar sphere ∂S with radius R, which we identified with the unit sphere σ. Finally, we have added an inner product (and norm) definition to $F(\sigma)$ to obtain the space of square integrable functions $X=L^2(\sigma)\subset F(\sigma)$, which has the advantage that it accepts the spherical harmonics as a set of appropriate base functions (to be precise: as a complete orthonormal sequence of the separable Hilbert space $L^2(\sigma)$).

(b) A convenient choice of the space Y and a consequent interpretation of the observed function $y\in Y$. The actual space of functions $F(\partial\Sigma)$ defined on the earth surface $\partial\Sigma$ is replaced by a set of functions on a sphere ∂S (such as the Bjerhammar sphere, but not necessarily), which in turn was identified with the unit sphere. This is a convenient choice, but not an ingenious one, since the approximation $\partial S\approx\partial\Sigma$ introduces large errors, which will eventually affect the obtained solution. Thus $F(\partial\Sigma)$ is replaced with $Y=L^2(\sigma)$, or rather with a subspace $Y\subset L^2(\sigma)$, depending on the properties of the relevant mapping.

(c) The original mapping $A_0=R_{\partial\Sigma}D:H_\Sigma\to F(\partial\Sigma)$, where D is a given differential operator defined on H_Σ while $R_{\partial\Sigma}$ is the mapping of restriction on the earth surface, is replaced by an operator $A=U^{-1}DU:X=L^2(\sigma)\to Y\subset L^2(\sigma)$, where D is the same operator as before, U is the upward harmonic continuation from $\sigma\sim\partial S$ to H_S and U^{-1} the restriction to the sphere ∂S.

With the above choices and approximations we have succeeded in ending up with a mapping A which is invertible (we may have to remove the first two degree terms to make it so) and furthermore it has the spherical harmonics as eigenfunctions. Thus the spherical harmonics $e_k=e_{k(n,m)}=e_{nm}$ play the same role as the eigenvectors $u_i\in X$ and $v_i\in Y$, of the finite dimensional case. The corresponding eigenvalues

$\lambda_k = \lambda_{k(n,m)} = \lambda_{nm}$, are the diagonal elements of an infinite dimensional square matrix $\underset{\infty\times\infty}{\mathbf{\Lambda}}$ which is the representation of A with respect to the spherical harmonics basis for both X and Y. The functions x and y are represented by their spherical harmonic coefficients $x_k = x_{k(n,m)} = x_{nm}$ and $y_k = y_{k(n,m)} = y_{nm}$, which are the elements of corresponding infinite-dimensional vectors $\underset{\infty\times 1}{\mathbf{x}}$ and $\underset{\infty\times 1}{\mathbf{y}}$, respectively.

The original model $y = Ax$ is represented by the "matrix" equation $\underset{\infty\times 1}{\mathbf{y}} = \underset{\infty\times\infty}{\mathbf{\Lambda}}\ \underset{\infty\times 1}{\mathbf{x}}$, with obvious solution $\underset{\infty\times 1}{\mathbf{x}} = \underset{\infty\times\infty}{\mathbf{\Lambda}^{-1}}\ \underset{\infty\times 1}{\mathbf{x}}$, where the matrix $\mathbf{\Lambda}^{-1}$ is the corresponding representation of A^{-1}.

Setting $\underset{1\times\infty}{\mathbf{e}} = [e_1 \ldots e_k \ldots]$ for the infinite-dimensional row-matrix of the spherical harmonics, we have $x = \mathbf{ex}$, $y = \mathbf{ey}$, $\mathbf{x} = [<e_1, x> \ldots <e_k, x> \ldots]^T \equiv <\mathbf{e}, x>$, $\mathbf{y} = [<e_1, y> \ldots <e_k, y> \ldots]^T \equiv <\mathbf{e}^T, y>$ so that the solution takes the form

$$x(\xi) = \mathbf{e}(\xi)\mathbf{x} = \mathbf{e}(\xi)\mathbf{\Lambda}^{-1}\mathbf{y} = \mathbf{e}(\xi)\mathbf{\Lambda}^{-1} <\mathbf{e}^T(\eta), y(\eta)>_\eta = <\mathbf{e}(\xi)\mathbf{\Lambda}^{-1}\mathbf{e}(\eta)^T, y(\eta)>_\eta \equiv$$

$$\equiv <k(\xi,\eta), y(\eta)>_\eta = \frac{1}{4\pi}\int_\sigma k(\xi,\eta) y(\eta)\, d\sigma(\eta) \tag{5.40}$$

where ξ, η are unit vectors (i.e. points on the unit sphere) and

$$k(\xi,\eta) = \mathbf{e}(\xi)\mathbf{\Lambda}^{-1}\mathbf{e}(\eta) = [\ldots e_{nm}(\xi)\ldots] \begin{bmatrix} \ddots & & 0 \\ & 1/\lambda_{nm} & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ e_{nm}(\eta) \\ \vdots \end{bmatrix} = \sum_{n,m} \frac{1}{\lambda_{nm}} e_{nm}(\xi) e_{nm}(\eta). \tag{5.41}$$

The fact that $\lambda_{nm} = \lambda_n$, (which is a consequence of the fact that D and therefore $A = U^{-1}DU$ is an isotropic, i.e. rotationally invariant operator) allows us to use the addition theorem and obtain an isotropic kernel $k(\xi,\eta) = k(\xi \bullet \eta)$ and therefore an isotropic integral operator A^{-1}.

Equation (5.40) is in fact analogous to the equation

$$\mathbf{x} = \mathbf{U}\mathbf{x}^* = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{y}^* = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{y} = \mathbf{A}^{-1}\mathbf{y}, \tag{5.42}$$

which solves the finite dimensional problem $\mathbf{y} = \mathbf{Ax}$ via the eigenvalue decomposition of $\mathbf{A}$, with the following correspondences:
$\mathbf{x} \to x$, $\mathbf{y} \to y$, $\mathbf{x}^* \to \mathbf{x}$, $\mathbf{y}^* \to \mathbf{y}$, $\mathbf{U} = [\ldots \mathbf{u}_k \ldots] \to \mathbf{e} = [\ldots e_k \ldots]$, $\mathbf{A} \to A$, $\mathbf{A}^{-1} \to A^{-1}$.

5. 2 Discrete observations affected by noise

When instead of the function $y=U^{-1}DUx$ only its values $y_i=y(P_i)$, at a finite number $n<\infty$ of discrete points are given, we obviously are facing an underdetermined problem, which we may solve by a minimum norm principle $\|x\|=\min$, or by regularization. As a matter of fact we can remove in this case the spherical approximation for the earth surface and even the points P_i may be of arbitrary position. In this way we consider instead point observations $y_i=E_i(y)\equiv y(P_i)$ where $y=DUx$, or even different differential operators D_k at the various points, giving rise to different types of observed functions $y^{(k)}=D_kUx$ (gravity anomalies, geoid heights, gravity gradients, etc.). If $\mathbf{y}$ is the vector of the observables y_i we can set

$$y_i=E_i(y^{(k)})=y^{(k)}(P_i)=E_i(D_kUx)=(D_kUx)(P_i) \tag{5.43}$$

where E_i is the "evaluation functional" associated with the point P_i, which maps the corresponding function into a real number. In the special case when $D_k=I$ (identity operator) we have a typical *interpolation* problem: Determine a function $y=Ux$ from its point values $y_i=y(P_i)=(Ux)(P_i)$ at discrete points P_i, $i=1,\ldots,n$, of its domain of definition. In this sense eq. (5.43) represents a "generalized" interpolation problem.

A formulation of the model (5.43), which is more convenient for its conversion into a "spectral" form, is based on the factorization

$$y_i=E_iD_kUx=E_iUU^{-1}D_kUx=E_iUL_kx\,, \qquad L_k=U^{-1}D_kU\,, \tag{5.44}$$

where L_k is an operator "confined" to functions on the unit sphere, corresponding to a particular case of the operator $A=U^{-1}DU$ of the continuous case (eq. 5.15). For the sake of notational convenience we will denote the operator associated with an observation at a point P_i by L_i rather than L_k, where the point-index i is replacing the operator-type-index k. Of course identical operators may appear at different points and we may even have more than one operators (and observations) at the same point.

Introducing the evaluation operator E and the "combined" differential operator L by means of

$$E=\begin{bmatrix}E_1\\E_2\\\vdots\\E_n\end{bmatrix}, \qquad L=\begin{bmatrix}L_1&0&\cdots&0\\0&L_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&L_n\end{bmatrix} \tag{5.45}$$

the model can be written in the form

$$\mathbf{y}=EULx=Ax\,. \tag{5.46}$$

Using the spherical harmonic representation (5.8) in the "matrix" form

$$x=\sum_{p,q}x_{pq}e_{pq}=[\cdots e_{pq}\cdots]\begin{bmatrix}\vdots\\ x_{pq}\\ \vdots\end{bmatrix}\equiv \underset{1\times\infty}{\mathbf{e}}\ \underset{\infty\times 1}{\mathbf{x}} \tag{5.47}$$

we can transform the model into a "spectral" form by taking into account that

$$L_i e_{pq}=\lambda_p^{(i)}e_{pq} \tag{5.48}$$

$$UL_i e_{pq}=U\left(\lambda_p^{(i)}e_{pq}\right)=\lambda_p^{(i)}Ue_{pq}=\lambda_p^{(i)}\left(\frac{R}{r}\right)^{n+1}e_{pq}\,. \tag{5.49}$$

Thus the matrix-spectral form of the model becomes

$$\underset{n\times 1}{\mathbf{y}}=EULx=EUL\mathbf{ex}=\begin{bmatrix}E_1UL_1\mathbf{e}\\ \vdots\\ E_nUL_n\mathbf{e}\end{bmatrix}\mathbf{x}=\begin{bmatrix}\cdots & E_1UL_1(e_{pq}) & \cdots\\ \cdots & \vdots & \cdots\\ \cdots & E_nUL_n(e_{pq}) & \cdots\end{bmatrix}\mathbf{x}=$$

$$=\begin{bmatrix}\cdots & E_1\left[\lambda_p^{(1)}\left(\frac{R}{r}\right)^{p+1}e_{pq}\right] & \cdots\\ \vdots & \vdots & \vdots\\ \cdots & E_n\left[\lambda_p^{(n)}\left(\frac{R}{r}\right)^{p+1}e_{pq}\right] & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}=\begin{bmatrix}\cdots & \lambda_p^{(1)}\left(\frac{R}{r_1}\right)^{p+1}e_{pq}(\lambda_1,\theta_1) & \cdots\\ \vdots & \vdots & \vdots\\ \cdots & \lambda_p^{(n)}\left(\frac{R}{r_n}\right)^{p+1}e_{pq}(\lambda_n,\theta_n) & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}=$$

$$=\begin{bmatrix}\cdots & \lambda_p^{(1)}\varepsilon_{pq}(P_1) & \cdots\\ \vdots & \vdots & \vdots\\ \cdots & \lambda_p^{(n)}\varepsilon_{pq}(P_n) & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}\equiv \underset{n\times\infty}{\mathbf{A}}\ \underset{\infty\times 1}{\mathbf{x}} \tag{5.50}$$

This is an underdetermined problem and we may consider a minimum norm solution, satisfying

$$\|x\|_Q^2=<x,x>_Q=<\mathbf{ex},\mathbf{ex}>_Q=\underset{1\times\infty}{\mathbf{x}^T}\ \underset{\infty\times\infty}{\mathbf{Q}}\ \underset{\infty\times 1}{\mathbf{x}}=\min\,,\qquad Q_{ik}\equiv<e_i,e_k>_Q\,, \tag{5.51}$$

which can be determined by applying the solution (3.113)

$$\mathbf{x}=\mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}\mathbf{y}\,, \tag{5.52}$$

provided that all the sums resulting from the presence of infinite dimensional matrices converge.

We shall examine the possibility of applying the inner product (5.11) for the domain space $X=L^2(\sigma)$, with respect to which the spherical harmonics are orthonormal yielding $<e_i,e_k>=\delta_{ik}$ and $\mathbf{Q}=\mathbf{I}$. We leave to the reader the general case where the matrix $\mathbf{A}$ is defined as in eq. (5.50). Instead we will consider only the simple case where $D_i=I \;\Rightarrow\; L_i=I$, so that $y=Ux$ and the observations are carried on an (approximately) spherical earth, identified with the Bjerhammar sphere. For such a spherical approximation we may omit the upward continuation ($y=x$) and write the model in the form $\mathbf{y}=Ex$, or explicitly recalling (5.47)

$$\underset{n\times 1}{\mathbf{y}}=Ex=E\mathbf{e}\mathbf{x}=\begin{bmatrix}E_1\mathbf{e}\\ \vdots\\ E_n\mathbf{e}\end{bmatrix}\mathbf{x}=\begin{bmatrix}\cdots & E_1(e_{pq}) & \cdots\\ \cdots & \vdots & \cdots\\ \cdots & E_n(e_{pq}) & \cdots\end{bmatrix}\mathbf{x}=\begin{bmatrix}\cdots & e_{pq}(P_1) & \cdots\\ \cdots & \vdots & \cdots\\ \cdots & e_{pq}(P_n) & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}=$$

$$=\begin{bmatrix}\mathbf{e}(P_1)\\ \vdots\\ \mathbf{e}(P_n)\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}=\underset{n\times\infty}{\mathbf{A}}\;\underset{\infty\times 1}{\mathbf{x}}\,. \tag{5.53}$$

A minimum norm solution will have the form $\hat{x}=\sum_{n,m}e_{nm}\hat{x}_{nm}=\mathbf{e}\hat{\mathbf{x}}$, where ($\mathbf{Q}=\mathbf{I}$)

$$\hat{\mathbf{z}}=\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}\,. \tag{5.54}$$

The matrix to be inverted has elements

$$(\mathbf{A}\mathbf{A}^T)_{ik}=\sum_j A_{ij}A_{kj}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}e_{nm}(P_i)e_{nm}(P_k)=\sum_{n=0}^{\infty}(2n+1)P_n(\cos\psi_{ik})=\infty\,! \tag{5.55}$$

and we have a serious problem! Even without the spherical approximation, with $P_i\sim(\lambda_i,\theta_i,r_i)$, $P_k\sim(\lambda_k,\theta_k,r_k)$, we arrive at

$$(\mathbf{A}\mathbf{A}^T)_{ik}=\sum_{n=0}^{\infty}\left(\frac{R}{r_i}\right)^{n+1}\left(\frac{R}{r_k}\right)^{n+1}\sum_{m=-n}^{n}e_{nm}(\lambda_i,\theta_i)e_{nm}(\lambda_k,\theta_k)=$$

$$=\sum_{n=0}^{\infty}\left(\frac{R^2}{r_ir_k}\right)^{n+1}(2n+1)P_n(\cos\psi_{ik})=\infty\,. \tag{5.56}$$

To give a mathematical description of this problem (within the spherical approximation) we consider a single observation, which we write as

$$y_i = x(P_i) = \sum_{n,m} e_{nm}(P_i) x_{nm} = [\cdots e_{nm}(P_i) \cdots] \begin{bmatrix} \vdots \\ x_{nm} \\ \vdots \end{bmatrix} \equiv \mathbf{e}_{P_i}^T \mathbf{x}, \tag{5.57}$$

where, in view of (5.47), the $\infty \times 1$ vector $\mathbf{e}_{P_i}$ consists of the components of a new function

$$e_{P_i} = \mathbf{e}\,\mathbf{e}_{P_i} = [\cdots e_{nm} \cdots] \begin{bmatrix} \vdots \\ e_{nm}(P_i) \\ \vdots \end{bmatrix} = \sum_{n,m} e_{nm}(P_i) e_{nm}, \tag{5.58}$$

which (if it exists) has the property

$$\langle e_{p_i}, x \rangle = \langle \sum_{n,m} e_{nm}(P_i) e_{nm}, \sum_{p,q} x_{pq} e_{pq} \rangle = \sum_{n,m} \sum_{p,q} e_{nm}(P_i) x_{pq} \langle e_{nm}, e_{pq} \rangle =$$

$$= \sum_{n,m} \sum_{p,q} e_{nm}(P_i) x_{pq} \delta_{np} \delta_{mq} = \sum_{nm} x_{nm} e_{nm}(P_i) = x(P_i) = E_i(x). \tag{5.59}$$

The function e_{P_i} is called the *representer* of the functional E_i (see e.g. Taylor and Lay, 1980) and its norm in $L^2(\sigma)$ is given by

$$\| e_{P_i} \|^2 = \mathbf{e}_{P_i}^T \mathbf{e}_{P_i} = \sum_{n,m} e_{nm}(P_i) e_{nm}(P_i) = \sum_{n=0}^{\infty} (2n+1) P_n(\cos 0) = \sum_{n=0}^{\infty} (2n+1) = \infty. \tag{5.60}$$

This means that there exists no element $e_{P_i} \in L^2(\sigma)$ such that $E_i(x) = \langle e_{P_i}, x \rangle$. In fact this possibility is restricted to the class H^* of functionals F on a Hilbert space H, such as $L^2(\sigma)$, which are bounded, i.e., $|F(x)| < C\|x\|$ for some constant C. The bounded functionals are also continuous and vice-versa. H^* is called the *dual* of the Hilbert spaces H.

Since the functionals E_i at hand cannot be changed, the only way out of the problem is to abandon $L^2(\sigma)$ for the sake of another Hilbert space H_k, large enough to accommodate the representers of all E_i, i.e. such that $E_i \in H_k^*$.

Since the inner product $\langle \, , \, \rangle_k$ of H_k is different from that of $L^2(\sigma)$, the spherical harmonics are not any more orthonormal, in general, but instead

$$<e_{nm},e_{pq}>_k = Q_{nm,pq}\,, \tag{5.61}$$

so that

$$<f,g>_k = <\sum_{nm} f_{nm}e_{nm}, \sum_{pq} g_{pq}e_{pq}>_k = \sum_{nm}\sum_{pq} f_{nm}g_{pq}<e_{nm},e_{pq}>_k =$$

$$= \sum_{nm}\sum_{pq} f_{nm}g_{pq}Q_{nm,pq} = [\cdots f_{nm} \cdots\cdots] \begin{bmatrix} & & \vdots & \\ \cdots & \cdots & Q_{nm,pq} & \cdots \\ & & \vdots & \end{bmatrix} \begin{bmatrix} \vdots \\ \vdots \\ g_{pq} \\ \vdots \end{bmatrix} =$$

$$= \underset{1\times\infty}{\mathbf{f}^T}\ \underset{\infty\times\infty}{\mathbf{Q}}\ \underset{\infty\times 1}{\mathbf{g}}\,. \tag{5.62}$$

Comparing $<f,g>_k = \mathbf{f}^T\mathbf{Q}\mathbf{g}$ with $<f,g> = \mathbf{f}^T\mathbf{g}$, we see that we have switched to a "weighted" inner product, where the weight matrix has to be selected in such a way that the series implicit in $\mathbf{f}^T\mathbf{Q}\mathbf{g}$ converge.

On the other hand, in order to make $\|e_{P_i}\|_k = \sqrt{<e_{P_i},e_{P_i}>_k}$ finite, it is sufficient to include in (5.60) a factor $k_n^2 \sim O(n^{-1})$ so that $\|e_{P_i}\|_k^2 = \sum_{n=0}^{\infty} k_n^2(2n+1) < \infty$. Both requirements $<f,g>_k = \mathbf{f}^T\mathbf{Q}\mathbf{g} < \infty$ and $\|e_{P_i}\|_k = < \infty$ can be fulfilled by using a diagonal weight matrix, with diagonal elements which depend only on the degree n

$$Q_{nm,pq} = \delta_{np}\delta_{mq}k_n^2\,. \tag{5.63}$$

With this choice $<f,g>_k = \sum_{nm} k_n^2 f_{nm}g_{nm}$, $<e_{nm},e_{pq}>_k = \delta_{nm}\delta_{nm}k_n^2$ and if we introduce

$$\tilde{e}_{nm} = \frac{1}{k_n}e_{nm}, \qquad \tilde{f}_{nm} = k_n f_{nm} \tag{5.64}$$

we have for $f,g \in L^2(\sigma)$ the representations

$$f = \sum_{nm} f_{nm}e_{nm} = \sum_{nm}(k_n f_{nm})(\frac{1}{k_n}e_{nm}) = \sum_{nm}\tilde{f}_{nm}\tilde{e}_{nm}\,, \tag{5.65}$$

$$\tilde{f}_{nm} = \sum_{nm} <f,\tilde{e}_{nm}>_k\,, \tag{5.66}$$

$$<f,g>_k=\sum_{nm}\tilde{f}_{nm}\tilde{g}_{nm}\ ,\qquad \|f\|_k^2=\sum_{nm}\tilde{f}_{nm}^2=\sum_{nm}k_n^2 f_{nm}^2\ , \tag{5.67}$$

where the "weighted" spherical harmonics $\tilde{e}_{nm}=\frac{1}{k_n}e_{nm}$ form an orthonormal system

$$<\tilde{e}_{nm},\tilde{e}_{pq}>_k=\delta_{nm}\delta_{pq}\ . \tag{5.68}$$

We take H_k to be the Hilbert space where the functions $\tilde{e}_{nm}$ form a complete orthonormal system, which means that $f\in H_k$ whenever

$$\|f\|_k^2=\sum_{nm}k_n^2 f_{nm}^2=\sum_{nm}k_n^2<f,e_{nm}>^2<\infty\ . \tag{5.69}$$

With a choice such that $k_n^2\to 0$ $(n\to\infty)$ it holds that

$$\|f\|_k^2=\sum_{nm}k_n^2 f_{nm}^2<\sum_{nm}f_{nm}^2=\|f\|^2\ , \tag{5.70}$$

which means that $\|f\|^2<\infty \ \Rightarrow\ \|f\|_k^2<\infty$, i.e., $f\in L^2(\sigma) \ \Rightarrow\ f\in H_k$, so that $L^2(\sigma)\subset H_k$. We have thus enlarged the space X so that it can accommodate the representers of the evaluation functionals. Indeed

$$E_i(x)=x(P_i)=\sum_{nm}x_{nm}e_{nm}(P_i)=\sum_{nm}(k_n x_{nm})[\frac{1}{k_n}e_{nm}(P_i)]=\sum_{nm}\tilde{x}_{nm}\tilde{e}_{nm}(P_i)=$$

$$=\sum_{nm}<x,\tilde{e}_{nm}>_k\tilde{e}_{nm}(P_i)=<x,\sum_{nm}\tilde{e}_{nm}(P_i)\tilde{e}_{nm}>_k=<x,e_{P_i}>_k\ , \tag{5.71}$$

where the representer of E_i is $e_{P_i}=\sum_{nm}\tilde{e}_{nm}(P_i)\tilde{e}_{nm}$ with norm

$$\|e_{P_i}\|^2=\sum_{nm}[\tilde{e}_{nm}(P_i)]^2=\sum_{n=0}^{\infty}k_n^2\sum_{m=-n}^{n}e_{nm}(P_i)e_{nm}(P_i)=\sum_{n=0}^{\infty}k_n^2(2n+1)<\infty\ . \tag{5.72}$$

This means that $e_{P_i}\in H_k$ and consequently $E_i\in H_k^*$, i.e. the evaluation functionals, when considered to act on elements of H_k are bounded (continuous) functionals.

We may now return to our norm minimization problem, which with $X=H_k$ can be restated as

$$\mathbf{y}=Ex=\begin{bmatrix}\cdots & \tilde{e}_{nm}(P_1) & \cdots\\ \cdots & \vdots & \cdots\\ \cdots & \tilde{e}_{nm}(P_n) & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ \tilde{x}_{nm}\\ \vdots\end{bmatrix}=\tilde{\mathbf{A}}\,\tilde{\mathbf{x}}\,. \tag{5.73}$$

The minimum norm solution will have the form

$$\hat{\tilde{\mathbf{x}}}=\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\mathbf{y} \tag{5.74}$$

and thus

$$\hat{x}(P)=\sum_{n,n}\hat{\tilde{x}}_{nm}\tilde{e}_{nm}(P)=\tilde{\mathbf{e}}(P)\,\hat{\tilde{\mathbf{x}}}=\tilde{\mathbf{e}}(P)\,\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)^{-1}\mathbf{y}\equiv\mathbf{k}_P^T\mathbf{K}^{-1}\mathbf{y}\,, \tag{5.75}$$

where we have introduced the notation

$$\mathbf{K}=\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\,,\qquad \mathbf{k}_P=\tilde{\mathbf{A}}\,\tilde{\mathbf{e}}(P)^T\,, \tag{5.76}$$

and we should recall that $\tilde{\mathbf{e}}=[\cdots\tilde{e}_{nm}\cdots]$. The two matrices required for the solution have elements

$$\mathbf{K}_{ik}=(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)_{ik}=\sum_j\tilde{A}_{ij}\tilde{A}_{kj}=\sum_{nm}\tilde{e}_{nm}(P_i)\tilde{e}_{nm}(P_k)=$$

$$=\sum_{n=0}^{\infty}k_n^2\sum_{m=-n}^{n}e_{nm}(P_i)e_{nm}(P_k)=\sum_{n=0}^{\infty}k_n^2(2n+1)P_n(\cos\psi_{P_iP_k})\,, \tag{5.77}$$

$$(\mathbf{k}_P)_i=\left(\tilde{\mathbf{A}}\,\tilde{\mathbf{e}}(P)^T\right)_i=\sum_j\tilde{A}_{ij}\left(\tilde{\mathbf{e}}(P)\right)_j=\sum_{nm}\tilde{e}_{nm}(P_i)\tilde{e}_{nm}(P)=\sum_{n=0}^{\infty}k_n^2(2n+1)P_n(\cos\psi_{P,P_i})\,. \tag{5.78}$$

If we introduce the two-point function

$$k(P,Q)=\sum_{nm}\tilde{e}_{nm}(P)\tilde{e}_{nm}(Q)=\sum_{n=0}^{\infty}k_n^2(2n+1)P_n(\cos\psi_{PQ}) \tag{5.79}$$

the elements of the matrices $\mathbf{K}$ and $\mathbf{k}_P$ are simply given by

$$K_{ij}=k(P_i,P_j)\,,\qquad(\mathbf{k}_P)_i=k(P,P_i)\,. \tag{5.80}$$

The function $k(P,Q)$ has the "reproducing" property

$$<k(P,Q),f(Q)>_Q=<\sum_{nm}\tilde{e}_{nm}(P)\tilde{e}_{nm}(Q),\sum_{pq}\tilde{f}_{pq}\tilde{e}_{pq}(Q)>_Q=$$

$$=\sum_{nm}\sum_{pq}\tilde{f}_{pq}\tilde{e}_{nm}(P)<\tilde{e}_{nm}(Q),\tilde{e}_{pq}(Q)>_Q=\sum_{nm}\sum_{pq}\tilde{f}_{pq}\tilde{e}_{nm}(P)\delta_{np}\delta_{mq}=$$

$$=\sum_{nm}\tilde{f}_{pq}\tilde{e}_{nm}(P)=f(P) \tag{5.81}$$

and it is therefore called the *reproducing kernel* of H_k, which becomes a "reproducing kernel Hilbert space" (RKHS).

The generalization to a non-spherical approximation, with observations carried on the surface of the earth, or even at any point outside the earth, leads to exactly the same results, the only difference being that the reproducing kernel takes the form

$$k(P,Q)=\sum_{nm}\left(\frac{R^2}{r_P r_Q}\right)^{n+1}\tilde{e}_{nm}(P)\tilde{e}_{nm}(Q)=\sum_{n=0}^{\infty}\left(\frac{R^2}{r_P r_Q}\right)^{n+1}k_n^2(2n+1)P_n(\cos\psi_{PQ})\,. \tag{5.82}$$

The generalization to observables $y_i=(D_kUx)(P_i)=(UL_kx)(P_i)$, with even different operators D_k and $L_k=U^{-1}D_kU$ for different functionals E_i, i.e. a combination of different types of observables, is also easy. The only difference is that the elements of the relevant matrices will be instead

$$K_{ij}=F_iF_jk(P_i,P_j)\,,\qquad(\mathbf{k}_P)_i=F_ik(P,P_i)\,. \tag{5.83}$$

where $F_i=E_iD_k$ are functionals resulting from the application of an operator D_k on $k(P,Q)$ of eq. (5.82), viewed as a function of one of the points only while the other is held fixed, and next evaluating the resulting function at a particular point P_i.

The functions k^P and k^Q, defined by $k^P(Q)\equiv k(P,Q)$ and $k^Q(P)\equiv k(P,Q)$, are elements of H_k ($k^P\in H_k$, $k^Q\in H_k$). Conveniently identifying operators by point rather than type indices, we can rewrite (5.83) in the more rigorous form

$$F_jk(P,P_j)=(D_jk^P)(P_j)\equiv\phi_j(P)\,,\qquad K_{ij}=F_iF_jk(P_i,P_j)=(D_j\phi_j)(P_i) \tag{5.84a}$$

$$(\mathbf{k}_P)_i=F_ik(P,P_i)=(D_ik^P)(P_i)\,. \tag{5.84b}$$

Usually the values of the observables $\mathbf{y}$ are not available and we have instead observations $\mathbf{b}=\mathbf{y}+\mathbf{v}$ affected by errors. Then a two-stage solution is equivalent to replacing $\mathbf{y}$ by $\mathbf{b}$ in equation (5.74). This follows from the fact that the operator A in $\mathbf{y}=Ax$ is surjective, i.e. $R(E)=R(A)=R^m$ ($\mathbf{y},\mathbf{b}\in R^m$) and the application of the

least-squares principle $\mathbf{v}^T\mathbf{P}\mathbf{v}=\min$ in the first stage simply gives $\hat{\mathbf{v}}=0$ and $\hat{\mathbf{y}}=\mathbf{b}$. Of the resulting estimate $\hat{z}(P)=\mathbf{k}_P^T\mathbf{K}^{-1}\mathbf{b}$ is affected by the true values of the errors present in $\mathbf{b}$.

To avoid absorbing the errors, a hybrid norm minimization, or regularization approach may be used, where the estimate $\hat{x}$ of the model $\mathbf{b}=Ax+\mathbf{v}=\tilde{\mathbf{A}}\tilde{\mathbf{x}}+\mathbf{v}$ should satisfy

$$\mathbf{v}^T\mathbf{P}\mathbf{v}+\alpha\|x\|_k^2=\mathbf{v}^T\mathbf{P}\mathbf{v}+\alpha\,\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}=\min\,. \tag{5.85}$$

The solution is given by

$$\hat{\tilde{\mathbf{x}}}=\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T+\alpha\,\mathbf{P}^{-1})^{-1}\mathbf{b}\,,\qquad \hat{x}(P)=\mathbf{e}(P)\hat{\tilde{\mathbf{x}}}=\mathbf{e}(P)\tilde{\mathbf{A}}^T(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T+\alpha\,\mathbf{P}^{-1})^{-1}\mathbf{b} \tag{5.86}$$

or in terms of the notation introduced above

$$\hat{x}(P)=\mathbf{k}_P^T(\mathbf{K}+\alpha\,\mathbf{P}^{-1})^{-1}\mathbf{b}\,. \tag{5.87}$$

5. 3 The stochastic approach

The deterministic approach followed above has solved the problem formally, but gives no hint to what is the proper choice of the weight coefficients k_n^{-2}. The unknown function contribution to the minimized quantity is

$$\|x\|^2=\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}=\sum_{nm}\tilde{x}_{nm}^2=\sum_{nm}(\frac{1}{k_n}x_{nm})^2=\sum_{nm}\frac{1}{k_n^2}x_{nm}^2=$$

$$=[\cdots x_{nm}\cdots]\begin{bmatrix}\ddots & \ddots & 0\\ \ddots & \frac{1}{k_n^2} & \ddots\\ 0 & \ddots & \ddots\end{bmatrix}\begin{bmatrix}\vdots\\ x_{nm}\\ \vdots\end{bmatrix}=\mathbf{x}^T\mathbf{Q}\mathbf{x} \tag{5.88}$$

and thus k_n^{-2} are indeed the "weights" in the norm definition. The choice of k_n^{-2} is equivalent to the choice of the norm of the inner product and essentially of the space H_k itself. As in the finite-dimensional case, resorting to a stochastic model for the unknown function will solve the problem of the proper choice of the weights (norm choice problem) x. Thus we assume that x a random function, usually called a stochastic process or a random field.

A complete characterization of the random field x requires the knowledge of the joint probability distribution of the random variables $x(P_1)$, $x(P_2)$, ..., $x(P_q)$, for any finite number of points in the domain of definition of x. For the purpose of esti-

mation (or rather prediction) we need only to know the first and second order moments

$$E\{x(P)\}=m(P)=0\,, \qquad E\{x(P)x(Q)\}=\sigma(P,Q)\,. \tag{5.89}$$

We have set the mean function $m(P)=0$, because any non-zero mean can be included in the approximate (normal) function used for the linearization of the model.

Remark:

The above characterization of a stochastic process by the joint distribution of any set of point values, due to Kolmogorov, does not extend directly to the case of a random function x with realizations in a Hilbert space H. Instead we must know the joint distribution of the random variables $u_i^*(x)$, where u_1^*, u_2^*, ..., u_q^* is any arbitrary set of bounded (continuous) linear functionals ($u_i^* \in H^*$). The reflexivity property allows us to identify H with H^*, since every bounded linear functional $u_i^* \in H^*$ corresponds to a unique element of the Hilbert space $u_i \in H$ and vice-versa, through $u_i^*(z)=<u_i,z>$ for every $z \in H$ (Riesz representation theorem). Therefore the "Hilbert space valued random variable" or simply the random field x is fully characterized by the joint probability distributions of the random variables $<u_1,x>$, $<u_2,x>$, ..., $<u_q,x>$, for any arbitrary set u_1, u_2, ..., u_q of elements of H.

The optimal (minimum mean square error unbiased linear) prediction of the random variable $y(P)=E_P(L_P x)$ utilizing the observed random variables $y_i=E_i(L_i x)=$ $=(L_i x)(P_i)$ is given, as already seen, by

$$\hat{y}(P)=\mathbf{c}_P^T\mathbf{C}^{-1}\mathbf{y}\,, \tag{5.90}$$

$$C_{ik}=\sigma\big((L_i x)(P_i),(L_k x)(P_k)\big)=L_i L_k \sigma\big(x(P_i),x(P_k)\big)=L_i L_k \sigma(P_i,P_k)\,, \tag{5.91}$$

$$(\mathbf{c}_P)_i=\sigma\big((L_P x)(P),(L_i x)(P_i)\big)=L_P L_i \sigma\big(x(P),x(P_i)\big)=L_P L_i \sigma(P,P_i)\,. \tag{5.92}$$

If the observations are affected by errors $\mathbf{b}=\mathbf{y}+\mathbf{v}$, with $\mathbf{v}\sim(\mathbf{0},\mathbf{C}_\mathbf{v})$ we have instead

$$\hat{y}(P)=\mathbf{c}_P^T(\mathbf{C}+\mathbf{C}_\mathbf{v})^{-1}\mathbf{y}\,. \tag{5.93}$$

Comparison with the deterministic results shows that the two approaches are identical when the reproducing kernel $k(P,Q)$ of the deterministic approach is taken to coincide with the covariance function $\sigma(P,Q)$ of the stochastic approach.

The regularization factor α may arise when $\sigma(P,Q)=\sigma^2 q(P,Q)$ so that $\mathbf{C}=\sigma^2\mathbf{Q}$, $\mathbf{c}_P=\sigma^2\mathbf{q}_P$ and also $\mathbf{C}_v=\sigma_v^2\mathbf{Q}_v$, with $\mathbf{Q}$, $\mathbf{q}_P$ $\mathbf{Q}_v$ known and σ^2, σ_v^2 unknown. In this case (5.93) becomes

$$\hat{y}(P)=\mathbf{q}_P^T(\mathbf{Q}+\alpha\mathbf{Q}_v)^{-1}\mathbf{y}\,, \tag{5.94}$$

where the ratio $\alpha=\frac{\sigma_v^2}{\sigma^2}$ must be known or properly chosen.

6 Beyond the standard formulation: Two examples from satellite geodesy

The standard formulation with a straightforward model $y=Ax$, which we have so far used in order to discover the fundamental features of geodetic data analysis, covers a wide class of applications but not all. In fact the main techniques used for the analysis of observations involving satellite tracking, aiming at the determination of the gravity field and station position, will force us to deviate to a certain degree from the previous formulation. One could say that practical applications pose particularities which do not allow them to fit directly into elegant theoretical schemes of data analysis. However, despite the required modifications, the basic principles remain unchanged.

6. 1 Determination of gravity potential coefficients

Observational data used for the determination of the gravity field of the earth, as expressed by its potential function V, do not relate directly to the unknown function $x=V$, but only to the orbits of a number of satellites, which are represented by their state vector functions

$$\mathbf{z}(t)=\begin{bmatrix}\mathbf{x}(t)\\ \frac{d\mathbf{x}}{dt}(t)\end{bmatrix}, \tag{6.1}$$

where $\mathbf{x}(t)$ are the (inertial) coordinates of the satellite at epoch t. Thus we have an intermediate unknown, say z, so that $y=g(z)$ which in turn depends on the original unknown $z=h(x)$. If the last expression was explicitly available we would have to deal with a usual factorization of the original model $y=f(x)$ through

$$y=g(z)=g\big(h(x)\big)=(g\circ h)(x)=f(x) \quad\Rightarrow\quad f=g\circ h\,. \tag{6.2}$$

The problem here is that the second factor $z=h(x)$ is not given explicitly but only implicitly, through the equations of satellite motion

$$\frac{d\mathbf{z}}{dt}(t)=\begin{bmatrix}\frac{d\mathbf{x}}{dt}(t)\\ \frac{d^2\mathbf{x}}{dt^2}(t)\end{bmatrix}=\begin{bmatrix}\frac{d\mathbf{x}}{dt}(t)\\ gradV(\mathbf{x}(t),t)\end{bmatrix}\equiv\mathbf{a}\big(t,\mathbf{z}(t),V(t)\big)=\mathbf{a}\big(t,\mathbf{z}(t),\mathbf{c}\big), \tag{6.3}$$

where we have ignored all other forces acting on the satellite for the sake of simplicity (see Schneider, 1988, ch.22, for a more general exposition) and $\mathbf{c}$ is a vector of a finite number of parameters which are chosen to represent the potential $V\big(\mathbf{x}(t),t\big)=$ $=V\big((\mathbf{x}(t),t,\mathbf{c}\big)$, typically the coefficients of its expansion in spherical harmonics.

In principle, but only in principle, we can write

$$\mathbf{z}(t)=\mathbf{z}(t_0)+\int_{t_0}^{t}\mathbf{a}\big(\mathbf{z}(\tau),V(\tau)\big)\,d\tau\equiv\mathbf{z}(t,\mathbf{z}_0,V)=\mathbf{z}(t,\mathbf{z}_0,\mathbf{c}),\qquad \mathbf{z}_0\equiv\mathbf{z}(t_0), \tag{6.4}$$

which is a concrete version of its abstract counterpart $z=z(x)$. The integration in (6.4) can be carried out analytically using perturbation techniques, on the basis of certain approximations which are not always acceptable (Colombo, 1986, p. 267). In general analytical methods are more appropriate for qualitative studies, while data processing relies on a numerical approach. Numerical techniques are used not for carrying out the integration in (6.4) but rather for the computational of the partials which will appear in the linearized version of the model, namely $\frac{\partial\mathbf{z}(t)}{\partial\mathbf{z}_0}$ and $\frac{\partial\mathbf{z}(t)}{\partial\mathbf{c}}$.

Any observable y is typically a function of the satellite position and velocity, i.e. of its state vector $\mathbf{z}(t)$, as well as of the inertial coordinates $\mathbf{x}_i(t)=\mathbf{R}\big(\boldsymbol{\theta}(t)\big)\mathbf{u}_i$ of the tracking station i, where $\mathbf{u}_i$ are the corresponding terrestrial coordinates, $\mathbf{R}$ is an orthogonal rotation matrix and $\boldsymbol{\theta}(t)$ a vector of functions describing earth rotation. For simplicity we will assume known tracking station coordinates $\mathbf{u}_i$ and earth rotation parameters $\boldsymbol{\theta}(t)$, in which case the model for the observable becomes

$$y=y\big(\mathbf{z}(t)\big)=y\big(\mathbf{z}(t,\mathbf{z}_0,\mathbf{c})\big)=y\big(\mathbf{z}(t,\mathbf{p})\big),\qquad \mathbf{p}=\begin{bmatrix}\mathbf{z}_0\\ \mathbf{c}\end{bmatrix}. \tag{6.5}$$

The linearized version is

$$y-y\big(\mathbf{z}(t,\mathbf{p}^0)\big)=y-y\big(\mathbf{z}(t,\mathbf{z}_0^0,\mathbf{c}^0)\big)=\left.\frac{\partial y}{\partial\mathbf{p}}\right|_0(\mathbf{p}-\mathbf{p}^0)=\left.\frac{\partial y}{\partial\mathbf{z}_0}\right|_0(\mathbf{z}_0-\mathbf{z}_0^0)+\left.\frac{\partial y}{\partial\mathbf{c}}\right|_0(\mathbf{c}-\mathbf{c}^0). \tag{6.6}$$

The chain rule $\frac{\partial y}{\partial\mathbf{p}}=\frac{\partial y}{\partial\mathbf{z}}\frac{\partial\mathbf{z}}{\partial\mathbf{p}}$ cannot be implemented analytically because no explicit expression is available for the symbolic relation $\mathbf{z}(t)=\mathbf{z}(t,\mathbf{z}_0,\mathbf{c})=\mathbf{z}(t,\mathbf{p})$, in order to

compute the terms $\frac{\partial \mathbf{z}}{\partial \mathbf{z}_0}$ and $\frac{\partial \mathbf{z}}{\partial \mathbf{c}}$ present in $\frac{\partial \mathbf{z}}{\partial \mathbf{p}}$. Thus another way must be found for the computation of the derivatives $\frac{\partial \mathbf{z}}{\partial \mathbf{p}}$.

The trick is to differentiate (6.3) of the form $\frac{d\mathbf{z}}{dt} = \mathbf{a}(t, \mathbf{z}, \mathbf{c}) = \mathbf{a}(t, \mathbf{z}, \mathbf{p})$, with respect to $\mathbf{p}$ in order to obtain (all partial derivatives been the explicit ones)

$$\frac{\partial}{\partial \mathbf{p}}\left(\frac{d\mathbf{z}}{dt}\right) = \frac{d}{dt}\left(\frac{\partial \mathbf{z}}{\partial \mathbf{p}}\right) = \frac{\partial \mathbf{a}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{p}} + \frac{\partial \mathbf{a}}{\partial \mathbf{p}} = \begin{bmatrix} \frac{\partial \mathbf{a}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{z}_0} + \frac{\partial \mathbf{a}}{\partial \mathbf{z}_0} \\ \frac{\partial \mathbf{a}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{c}} + \frac{\partial \mathbf{a}}{\partial \mathbf{c}} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{a}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{z}_0} \\ \frac{\partial \mathbf{a}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{c}} + \frac{\partial \mathbf{a}}{\partial \mathbf{c}} \end{bmatrix} \tag{6.7}$$

Setting

$$\frac{\partial \mathbf{z}}{\partial \mathbf{p}}(t) \equiv \mathbf{Z}(t), \qquad \frac{\partial \mathbf{a}}{\partial \mathbf{z}}(t) \equiv \mathbf{A}(t), \qquad \frac{\partial \mathbf{a}}{\partial \mathbf{p}}(t) = \begin{bmatrix} \frac{\partial \mathbf{a}}{\partial \mathbf{z}_0}(t) \\ \frac{\partial \mathbf{a}}{\partial \mathbf{c}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \frac{\partial \mathbf{a}}{\partial \mathbf{c}}(t) \end{bmatrix} \equiv \mathbf{C}(t), \tag{6.8}$$

the differential equations (6.7) take the form of the so called *variational equations* (see e.g. Reigber, 1989)

$$\frac{d\mathbf{Z}}{dt}(t) = \mathbf{A}(t)\mathbf{Z}(t) + \mathbf{C}(t), \tag{6.9}$$

where $\mathbf{A}(t)$ and $\mathbf{C}(t)$ are known matrices, resulting from the partial differentiation of the analytically known function $\mathbf{a}(t, \mathbf{z}, \mathbf{c})$. The variational equations can be solved by numerical integration

$$\mathbf{Z}(t) = \mathbf{Z}(t_0) + \int_{t_0}^{t} [\mathbf{A}(\tau)\mathbf{Z}(\tau) + \mathbf{C}(\tau)]\, d\tau, \tag{6.10}$$

in order to compute the factor $\mathbf{Z} = \frac{\partial \mathbf{z}}{\partial \mathbf{p}} = \begin{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{z}_0} \\ \frac{\partial \mathbf{z}}{\partial \mathbf{c}} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{Z}_{\mathbf{z}_0} \\ \mathbf{Z}_{\mathbf{c}} \end{bmatrix}$ in the linearized model

$$y - y\left(\mathbf{z}(t, \mathbf{z}_0^0, \mathbf{c}^0)\right) = \left.\frac{\partial y}{\partial \mathbf{z}}\right|_0 \mathbf{Z}\,(\mathbf{p} - \mathbf{p}^0) = \left.\frac{\partial y}{\partial \mathbf{z}}\right|_0 \mathbf{Z}_{\mathbf{z}_0}(\mathbf{z}_0 - \mathbf{z}_0^0) + \left.\frac{\partial y}{\partial \mathbf{z}}\right|_0 \mathbf{Z}_{\mathbf{c}}\,(\mathbf{c} - \mathbf{c}^0). \tag{6.11}$$

The analysis based on the linearized model follows the already explained procedure. Note that only a finite number of spherical harmonic coefficients $\mathbf{c}$ is retained, with a much larger number of observations, so that the original underdetermined problem (the potential function V is an infinite-dimensional object) is converted into an over-

determined problem which admits a unique least-squares solution. However the residuals $\mathbf{v}=\mathbf{b}-\mathbf{y}$ do not only contain the observational errors, as well as other modeling errors, but also the error committed in replacing V not with the infinite set of all the spherical harmonic coefficients, but with a finite "truncated" set of only the ones up to a maximal degree N.

6. 2 GPS Observables and integer unknowns

Observables of the Global Positioning System are related to a nominal time delay $\tau_R^S=t_R-t^S$ between receiver clock reading $t_R=t_{R,true}+\delta t_R$, at the epoch of signal arrival and satellite clock reading $t^S=t_{true}^S+\delta t^S$, at the epoch of signal departure. The observed delay can be related to true delay and satellite-to-receiver distance ρ_R^S through the known velocity of light c, since

$$\tau_R^S=t_R-t^S=t_{R,true}-t_{true}^S+\delta t_R-\delta t^S=\frac{1}{c}\rho_R^S+\delta t_R-\delta t^S=$$

$$=\frac{1}{c}\sqrt{(\mathbf{x}_R-\mathbf{x}^S)^T(\mathbf{x}_R-\mathbf{x}^S)}+\delta t_R-\delta t^S, \tag{6.12}$$

where $\mathbf{x}^S=\mathbf{x}^S(t_{true}^S)$ is the satellite and $\mathbf{x}_R=\mathbf{x}_R(t_{R,true})$ the receiver position at the corresponding epochs of signal departure and arrival.

In code observations the delay τ_R^S is related to pseudo-distances $p_R^S=c\tau_R^S$, and the model for the "observables" becomes

$$p_R^S=\sqrt{(\mathbf{x}_R-\mathbf{x}^S)^T(\mathbf{x}_R-\mathbf{x}^S)}+c\delta t_R-c\delta t^S=\sqrt{(\mathbf{u}_R-\mathbf{u}^S)^T(\mathbf{u}_R-\mathbf{u}^S)}+c\delta t_R-c\delta t^S \tag{6.13}$$

where inertial satellite coordinates $\mathbf{x}^S$ are usually assumed to be known and converted to terrestrial ones $\mathbf{u}^S=\mathbf{R}\mathbf{x}^S$ through a known earth rotation matrix $\mathbf{R}$. Thus the model unknowns are the receiver terrestrial coordinates $\mathbf{u}_R$ and the satellite and receiver clock errors δt^S and δt_R, respectively.

When phases are observed, however, the observable is not related to the delay τ_R^S, but to the remainder Φ_R^S of its division with the period $T=\frac{1}{f}$, where f is the frequency of the signal. Thus $\tau_R^S=N_R^S T+\Phi_R^S=\frac{N_R^S}{f}+\Phi_R^S$ where N_R^S is an unknown integer, and the model becomes

$$\Phi_R^S=\tau_R^S-\frac{1}{f}N_R^S=\frac{1}{c}\sqrt{(\mathbf{u}_R-\mathbf{u}^S)^T(\mathbf{u}_R-\mathbf{u}^S)}+\delta t_R-\delta t^S-\frac{1}{f}N_R^S. \tag{6.14}$$

Multiplication with the velocity of light c and introduction of the signal wavelength $\lambda=cT=\frac{c}{f}$, leads to the model for phase pseudo-distances $L_R^S=c\Phi_R^S$

$$L_R^S=\sqrt{(\mathbf{u}_R-\mathbf{u}^S)^T(\mathbf{u}_R-\mathbf{u}^S)}+c\delta t_R-c\delta t^S-\lambda N_R^S+I_R^S+T_R^S, \tag{6.15}$$

where the ionospheric I_R^S and the tropospheric T_R^S influence of atmospheric refraction has been included. The only unknowns of interest are the receiver coordinates $\mathbf{u}_R$, the rest being "nuisance parameters". Clock errors are eliminated by taking single and double differences

$$L_{AB}^S\equiv L_B^S-L_A^S=\rho_B^S-\rho_A^S+c\delta t_B-c\delta t_A-\lambda N_B^S+\lambda N_A^S+I_B^S-I_A^S+T_B^S-T_A^S\equiv$$

$$\equiv\rho_B^S-\rho_A^S+c\delta t_B-c\delta t_A-\lambda N_{AB}^S+I_{AB}^S+T_{AB}^S, \tag{6.16}$$

$$L_{AB}^{jk}\equiv L_{AB}^k-L_{AB}^j=\rho_B^k-\rho_A^k-\rho_B^j+\rho_A^j-\lambda N_{AB}^k+\lambda N_{AB}^j+I_{AB}^k-I_{AB}^j+T_{AB}^k-T_{AB}^j\equiv$$

$$\equiv\rho_B^k-\rho_A^k-\rho_B^j+\rho_A^j-\lambda N_{AB}^{jk}+I_{AB}^{jk}+T_{AB}^{jk}. \tag{6.17}$$

The atmospheric influences are computed or eliminated, and the remaining unknowns are either station coordinates, which have real values and the integer number of cycles N_R^S or their linear combinations, e.g.

$$N_{AB}^{jk}=N_{AB}^k-N_{AB}^j=N_B^k-N_A^k+N_B^j-N_A^j. \tag{6.18}$$

The presence of integer unknowns is the new feature, which makes the classical least squares solution not applicable, because it is designed specifically for real unknowns. The solution of the least squares problem $\phi=(\mathbf{b}-\mathbf{A}\mathbf{x})^T\mathbf{P}(\mathbf{b}-\mathbf{A}\mathbf{x})=\min$ by $\frac{\partial\phi}{\partial\mathbf{x}}=0$, is possible only for a real-valued variable $\mathbf{x}$, since the derivative with respect to an integer variable makes no sense. We are therefore faced with a new least squares problem of the form

$$\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{v}=[\mathbf{A}_\mathbf{z}\ \ \mathbf{A}_n]\begin{bmatrix}\mathbf{z}\\ \mathbf{n}\end{bmatrix}+\mathbf{v}=\mathbf{A}_\mathbf{z}\mathbf{z}+\mathbf{A}_n\mathbf{n}+\mathbf{v},\qquad \phi=\mathbf{v}^T\mathbf{Q}^{-1}\mathbf{v}=\min, \tag{6.19}$$

with real unknowns $\mathbf{z}$ and integer ones $\mathbf{n}$. Due to the discrete nature of $\mathbf{n}$, the problem must be solved in two steps. In the first step, we take separately any value of $\mathbf{n}$, we keep it fixed and we find the corresponding optimal $\tilde{\mathbf{z}}=\tilde{\mathbf{z}}(\mathbf{n})$ which minimizes ϕ,

$$\phi(\tilde{\mathbf{z}}(\mathbf{n}))=\min_{\mathbf{z}}\phi(\mathbf{z},\mathbf{n}). \tag{6.20}$$

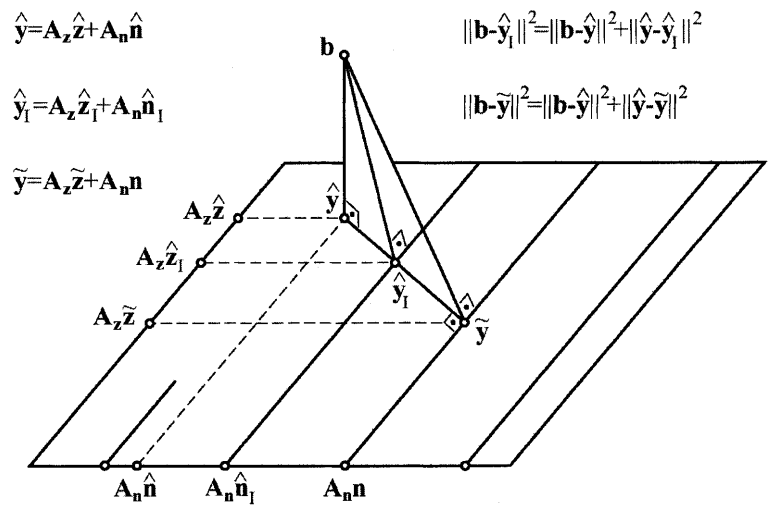


Fig. 4: The geometry of the least-squares optimal solution when the model $\mathbf{b}=\mathbf{y}+\mathbf{v}$, $\mathbf{y}=\mathbf{A}_z\mathbf{z}+\mathbf{A}_n\mathbf{n}$ includes both real ($\mathbf{z}$) and integer unknowns ($\mathbf{n}$).

To any fixed integer-valued $\mathbf{n}$ corresponds an optimal value $\tilde{\mathbf{y}}=\tilde{\mathbf{y}}(\mathbf{n})=\mathbf{A}_z\tilde{\mathbf{z}}+\mathbf{A}_n\mathbf{n}$ of the observables, which is closest to the data $\mathbf{b}$ among all values $\mathbf{y}=\mathbf{A}_z\mathbf{z}+\mathbf{A}_n\mathbf{n}$ and a corresponding optimal value of the real unknowns $\tilde{\mathbf{z}}=\tilde{\mathbf{z}}(\mathbf{n})$. As $\mathbf{n}$ varies among all integer-valued vectors, there exists one, denoted by $\hat{\mathbf{n}}_I$, giving the observable value $\hat{\mathbf{y}}_I=\tilde{\mathbf{y}}(\hat{\mathbf{n}}_I)$ closest to $\mathbf{b}$ among all $\tilde{\mathbf{y}}(\mathbf{n})$, which together with the corresponding real values $\hat{\mathbf{z}}_I=\tilde{\mathbf{z}}(\hat{\mathbf{n}}_I)$ constitutes the desired optimal solution. The least-squares solution obtained by pretending that $\mathbf{n}$ is real-valued (float solution) corresponds to observables $\hat{\mathbf{y}}=\mathbf{A}_z\hat{\mathbf{z}}+\mathbf{A}_n\hat{\mathbf{n}}$ closest to the data $\mathbf{b}$, with corresponding "optimal" real values of the unknowns $\hat{\mathbf{n}}$ and $\hat{\mathbf{z}}$. The Pythagorean theorem $\|\mathbf{b}-\tilde{\mathbf{y}}\|^2=\|\mathbf{b}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\tilde{\mathbf{y}}\|^2$, where $\|\mathbf{b}-\hat{\mathbf{y}}\|=\text{const.}$, allows to replace the minimization of $\|\mathbf{b}-\tilde{\mathbf{y}}\|^2$ with the minimization of $\|\hat{\mathbf{y}}-\tilde{\mathbf{y}}\|^2=(\hat{\mathbf{n}}-\mathbf{n})^T\mathbf{Q}_{\hat{\mathbf{n}}}^{-1}(\hat{\mathbf{n}}-\mathbf{n})$, by varying $\mathbf{n}$ over all integer-valued vectors to obtain the optimal solution $\hat{\mathbf{n}}_I$, $\hat{\mathbf{z}}_I$ and $\hat{\mathbf{y}}_I=\mathbf{A}_z\hat{\mathbf{z}}_I+\mathbf{A}_n\hat{\mathbf{n}}_I$.

In the second step we let $\mathbf{n}$ vary (in a discrete way!) until $\phi(\tilde{\mathbf{z}}(\mathbf{n}))$ is minimized in order to obtain the final optimal values $\hat{\mathbf{z}}_I$ and $\hat{\mathbf{n}}_I$, where the subscript I emphasizes the fact that $\hat{\mathbf{n}}_I$ has integer values. Thus the second step is

$$\phi(\hat{\mathbf{z}}_I,\hat{\mathbf{n}}_I)=\phi(\tilde{\mathbf{z}}(\hat{\mathbf{n}}_I),\hat{\mathbf{n}}_I)=\min_{\mathbf{n}}\phi(\tilde{\mathbf{z}}(\mathbf{n}),\mathbf{n})=\min_{\mathbf{n}}\left[\min_{\mathbf{z}}\phi(\mathbf{z},\mathbf{n})\right]. \tag{6.21}$$

The solution to the first step is a classical least-squares solution for the modified model $\mathbf{b}-\mathbf{A}_{\mathbf{n}}\mathbf{n}=\mathbf{A}_{\mathbf{z}}\mathbf{z}+\mathbf{v}$ ($\mathbf{n}$ is known and fixed), i.e.

$$\tilde{\mathbf{z}}(\mathbf{n})=(\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}})^{-1}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{A}_{\mathbf{n}}\mathbf{n})\equiv\mathbf{N}_{\mathbf{z}}^{-1}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{A}_{\mathbf{n}}\mathbf{n}). \tag{6.22}$$

To this "estimate" corresponds a value of the observables

$$\tilde{\mathbf{y}}(\mathbf{n})=\mathbf{A}\tilde{\mathbf{x}}(\mathbf{n})=\mathbf{A}_{\mathbf{z}}\tilde{\mathbf{z}}(\mathbf{n})+\mathbf{A}_{\mathbf{n}}\mathbf{n}=\mathbf{A}_{\mathbf{z}}\mathbf{N}_{\mathbf{z}}^{-1}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}(\mathbf{b}-\mathbf{A}_{\mathbf{n}}\mathbf{n})+\mathbf{A}_{\mathbf{n}}\mathbf{n} \tag{6.23}$$

and of the minimized quantity

$$\tilde{\phi}(\mathbf{n})=\|\mathbf{b}-\tilde{\mathbf{y}}(\mathbf{n})\|^2=(\mathbf{b}-\tilde{\mathbf{y}}(\mathbf{n}))^T\mathbf{Q}^{-1}(\mathbf{b}-\tilde{\mathbf{y}}(\mathbf{n})). \tag{6.24}$$

What remains is to minimize $\tilde{\phi}(\mathbf{n})$ by examining all possible integer values of $\mathbf{n}$. This seems to be an impossible task without some external guidance which will limit the search to a smaller finite set of values of $\mathbf{n}$. The required help comes from the usual least-squares solution $\hat{\mathbf{z}}$, $\hat{\mathbf{n}}$, obtained by pretending that $\mathbf{n}$ is real valued (floating solution)

$$\begin{bmatrix}\hat{\mathbf{z}}\\ \hat{\mathbf{n}}\end{bmatrix}=\begin{bmatrix}\mathbf{Q}_{\hat{\mathbf{z}}} & \mathbf{Q}_{\hat{\mathbf{z}}\hat{\mathbf{n}}}\\ \mathbf{Q}_{\hat{\mathbf{z}}\hat{\mathbf{n}}}^T & \mathbf{Q}_{\hat{\mathbf{n}}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{b}\\ \mathbf{A}_{\mathbf{n}}^T\mathbf{Q}^{-1}\mathbf{b}\end{bmatrix}, \tag{6.25}$$

where

$$\mathbf{Q}_{\hat{\mathbf{n}}}=[\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}}-\mathbf{A}_{\mathbf{n}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}}(\mathbf{A}_{\mathbf{n}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{n}})^{-1}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{n}}]^{-1}, \tag{6.26}$$

$$\mathbf{Q}_{\hat{\mathbf{z}}\hat{\mathbf{n}}}=-(\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}})^{-1}\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{n}}\mathbf{Q}_{\hat{\mathbf{n}}}, \tag{6.27}$$

$$\mathbf{Q}_{\hat{\mathbf{z}}}=(\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}})^{-1}-\mathbf{Q}_{\hat{\mathbf{z}}\hat{\mathbf{n}}}\mathbf{A}_{\mathbf{n}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}}(\mathbf{A}_{\mathbf{z}}^T\mathbf{Q}^{-1}\mathbf{A}_{\mathbf{z}})^{-1}. \tag{6.28}$$

It can be shown that the corresponding estimates of the observables $\hat{\mathbf{y}}=\mathbf{A}\hat{\mathbf{x}}=$ $=\mathbf{A}_{\mathbf{z}}\hat{\mathbf{z}}+\mathbf{A}_{\mathbf{n}}\hat{\mathbf{n}}$ satisfy the orthogonality relation

$$(\mathbf{b}-\hat{\mathbf{y}})^T \mathbf{Q}^{-1}\left(\hat{\mathbf{y}}-\tilde{\mathbf{y}}(n)\right)=0\,, \tag{6.29}$$

which allows us to set the quantity to be minimized in the form

$$\hat{\phi}(\mathbf{n})=\|\mathbf{b}-\tilde{\mathbf{y}}(\mathbf{n})\|^2=\|\mathbf{b}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\tilde{\mathbf{y}}(\mathbf{n})\|^2=\min_{\mathbf{n}}\,. \tag{6.30}$$

Since $\|\mathbf{b}-\hat{\mathbf{y}}\|^2=\text{const.}$, we can minimize instead

$$\phi'(\mathbf{n})=\|\hat{\mathbf{y}}-\tilde{\mathbf{y}}(\mathbf{n})\|^2=\min_{\mathbf{n}}\,. \tag{6.31}$$

It can be shown (see e.g. Dermanis, 1999, for details) that

$$\phi'(\mathbf{n})=(\mathbf{n}-\hat{\mathbf{n}})^T \mathbf{Q}_{\hat{\mathbf{n}}}^{-1}(\mathbf{n}-\hat{\mathbf{n}})\,. \tag{6.32}$$

In this way one of different search strategies (Leick, 1995, Hofmann-Wellenhof et al, 1979, Teunissen and Kleusberg, 1998) can be incorporated for locating, in a neighborhood of the real-valued (floating) solution $\hat{\mathbf{n}}$, the optimal value $\hat{\mathbf{n}}_I$ satisfying

$$\phi'(\hat{\mathbf{n}}_I)=\min_{\mathbf{n}}\phi'(\mathbf{n})=\min_{\mathbf{n}}\left[(\mathbf{n}-\hat{\mathbf{n}})^T \mathbf{Q}_{\hat{\mathbf{n}}}^{-1}(\mathbf{n}-\hat{\mathbf{n}})\right]. \tag{6.33}$$

We conclude by bringing attention to another critical difference between real- and integer-valued unknown estimation, which is in fact the "heart" of GPS data analysis strategies.

In the case of real-valued unknowns assume that an initial set of observations $\mathbf{b}_1$ is used in association with the model

$$\mathbf{b}_1=\mathbf{A}_1\mathbf{x}+\mathbf{A}_{11}\mathbf{x}_1+\mathbf{v}_1\,, \tag{6.34}$$

for the determination of estimates $\hat{\mathbf{x}}_{|\mathbf{b}_1}$ and $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$. Let us also assume that the data $\mathbf{b}_1$ are sufficient to determine $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$ with very good accuracy. Next a second set of observations

$$\mathbf{b}_2=\mathbf{A}_2\mathbf{x}+\mathbf{A}_{22}\mathbf{x}_2+\mathbf{v}_2 \tag{6.35}$$

becomes available which is used (in combination to the previous ones $\mathbf{b}_1$) to produce updated estimates $\hat{\mathbf{x}}_{|\mathbf{b}_1,\mathbf{b}_2}$ and $\hat{\mathbf{x}}_{1|\mathbf{b}_1,\mathbf{b}_2}$. Due to the original accuracy of the estimates $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$, the "improvement" $\hat{\mathbf{x}}_{1|\mathbf{b}_1,\mathbf{b}_2}-\hat{\mathbf{x}}_{1|\mathbf{b}_1}$ will be very small but not zero. The estimates $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$ will "slide" to a nearby best value $\hat{\mathbf{x}}_{|\mathbf{b}_1,\mathbf{b}_2}$.

On the contrary, in the case where the parameters $\mathbf{x}_1$ are integer-valued, the integer estimates $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$ are not allowed to "slide" but only to "jump" to a nearby integer estimate $\hat{\mathbf{x}}_{|\mathbf{b}_1,\mathbf{b}_2}$. As demonstrated in eq. (6.33), "closeness" is associated with estimate accuracy (covariance factor matrix $\mathbf{Q}_{\hat{\mathbf{n}}}$). If $\hat{\mathbf{x}}_{1|\mathbf{b}_1}$ has been determined with sufficient accuracy, all integer values around it will be too far away and no "jump" to a better estimate is possible and thus $\hat{\mathbf{x}}_{1|\mathbf{b}_1,\mathbf{b}_2} = \hat{\mathbf{x}}_{1|\mathbf{b}_1}$. This means that once integer unknowns are efficiently determined from a subset of relevant observations, they cannot be further improved by "small" variations due to the "influence" of the other observations, as in the case of real-valued unknowns.

Therefore, the following strategy is followed for the analysis of GPS data:

Every data set $\mathbf{b}_i = \mathbf{A}_{\mathbf{z}}^i \mathbf{z} + \mathbf{A}_{\mathbf{n}}^i \mathbf{n}_i + \mathbf{v}_i$, $(i=1,2,\ldots)$ is separately analyzed for the estimation $\hat{\mathbf{n}}_i$ of its own particular integer unknowns. These values are then held fixed and the resulting models $\mathbf{b}_i' \equiv \mathbf{b}_i - \mathbf{A}_{\mathbf{n}}^i \hat{\mathbf{n}}_i = \mathbf{A}_{\mathbf{z}}^i \mathbf{z} + \mathbf{v}_i$ are combined for the optimal estimation $\hat{\mathbf{z}}$ of the remaining real-valued unknowns, i.e. the coordinates of points occupied by receivers for observations.

References

Bruns, H. (1878): *Die Figur der Erde.* Publication des Königl. Preussischen Geodätischen Instituts, Berlin.

Baarda, W. (1973): *S-transformations and Criterion Matrices.* Publications on Geodesy, Vol. 5, Nr. 5. Netherlands Geodetic Commission, Delft.

Christakos, G. (1992): *Random Field Models in Earth Sciences.* Academic Press.

Colombo, O.L. (1986): Notes on the Mapping of the Gravity Field Using Satellite Data. In: H. Sünkel (ed.): *Mathematical and Numerical Techniques in Physical Geodesy*, Lecture Notes in Earth Sciences, vol. 7, 49-155, Springer.

Dermanis, A. (1978): Adjustment of geodetic observations in the presence of signals. International School of Advanced Geodesy, Erice, Sicily, May-June 1978. *Bollettino di Geodesia e Scienze Affini*, 38 (1979), 4, 513-539.

Dermanis, A. (1988): *Geodetic Applications of Interpolation and Prediction.* Inter. School of Geodesy "A. Marussi", 4th Course: *Applied and Basic Geodesy: Present and Future Trends*, Ettore Majorana Centre for Scientific Culture, Erice-Sicily, 15-25 June 1987. *Eratosthenes*, 22, 229-262.

Dermanis, A. (1991): *A Unified Approach to Linear Estimation and Prediction.* Presented at the 20th General Assembly of the IUGG, Vienna, August 1990.

Dermanis, A. (1998): Generalized inverses of nonlinear mappings and the nonlinear geodetic datum problem. *Journal of Geodesy*, 72, 2, 71-100.

Dermanis A. (1999): *Space Techniques in Geodesy and Geodynamics – GPS*. Ziti Editions, Thessaloniki (in Greek).

Heiskanen, W.A. and H. Moritz (1967): *Physical Geodesy*. Freeman, San Francisco.

Helmert, F. R. (1880): *Die mathematischen und physikalischen Theorien der höheren Geodäsie*. B.G. Teubner, Leipzig.

Hofmann-Wellenhof, B., H. Lichtenegger and J. Collins (1997): *Global Positioning System. Theory and Practice*. Fourth Revised Edition. Springer.

Hotine, M. (1969): *Mathematical Geodesy*. U.S. Department of Commerce, Washington, D.C.

Koch, K.-R. (1999): *Parameter Estimation and Hypothesis Testing in Linear Models*. Second Edition. Springer, Berlin.

Krarup, T. (1969): *A contribution to the mathematical foundation of physical geodesy*. Danish Geodetic Institute, Meddelelse no. 44, Copenhagen.

Lanczos, C. (1961,1967): *Linear Differential Operators*. Van Nostrand, London. (Dover Reprint, 1967.)

Lauritzen, S. (1973): *The Probabilistic Background of Some Statistical Methods in Physical Geodesy*. Danish Geodetic Institute, Meddelelse no. 48, Copenhagen.

Leick, A. (1995): *GPS Satellite Surveying*. 2nd Edition. Wiley.

Rao, C.R. and S.K. Mitra (1971): *Generalized Inverse of Matrices and its Applications*. Wiley, New York.

Reigber, Ch. (1989): Gravity field recovery from satellite data. In: F. Sansò & R. Rummel, Eds., 'Theory of Satellite Geodesy and Gravity Field Determination', Lecture Notes in Earth Sciences 25, 197-234, Springer-Verlag.

Rummel, R. (1997): Spherical spectral properties of the earth's gravitational potential and its first and second order derivatives. In: F. Sansò & R. Rummel, Eds., *'Geodetic Boundary Value Problems in View of the One Centimeter Geoid'*, Lecture Notes in Earth Sciences 65, 359-404. Springer-Verlag.

Rummel, R. and M. van Gelderen (1995): Meissl scheme - Spectral characteristics of Physical Geodesy. *Manuscripta Geodaetica*, 20, 379-385.

Sansò, F. (1978): The minimum mean square estimation principle in physical geodesy (stochastic and non-stochastic interpretation). 7th Hotine Symposium on Mathematical Geodesy, Assissi, June 1978. *Bollettino di Geodesia e Scienze Affini*, 39 (1980), 2, 111-129.

Sansò, F. (1986): Statistical Methods in Physical Geodesy. In: H. Sünkel (ed.): *Mathematical and Numerical Techniques in Physical Geodesy*, Lecture Notes in Earth Sciences, vol. 7, 49-155, Springer.

Sansò F. and G. Sona (1995): The theory of optimal linear estimation for continuous fields of measurements. *Manuscripta Geodaetica*, 20, 204-230.

Schaffrin, B. (1983): *Model Choice and Adjustment Techniques in the Presence of Prior Information*. Ohio State University Department of Geodetic Science and Surveying, Report No. 351, Columbus.

Schaffrin, B. (1985): *Das geodätische Datum mit stochastischer Vorinformation*. Habilitationsschrift, Universität Stuttgart. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 313, München.

Schaffrin, B., E. Grafarend, G. Sschmitt (1977): Kaninisches Design Geodätischer Netze I, *Manuscripta Geodaetica*, 2, 263-306.

Schneider, M. (1988): *Satellitengeodäsie*. Bibliographisches Institut, Mannheim.

Taylor, A.E. and D.C. Lay (1980): *Introduction to Functional Analysis*. Second Edition. Wiley, New York.

Teunissen, P.J.G. and A. Kleusberg, Eds. (1998): *GPS for Geodesy*. 2nd Edition. Springer.

Tikhonov, A. and V. Arsenin (1977): *Solutions of Ill-Posed Problems*. Wiley, New York.

Acknowledgment

This work has been completed during a visit of the first author to Germany. The financial support of the Alexander-von-Humboldt Foundation is gratefully acknowledged.

Appendix A: The Singular Value Decomposition

Let $\mathbf{A}$ be an $n\times m$ matrix with $rank(\mathbf{A})=r$, representing a mapping $A:X\rightarrow Y$, where X and Y are Euclidean spaces with inner products

$$(x_\alpha,x_\beta)=\mathbf{x}_\alpha^T\mathbf{Q}\,\mathbf{x}_\beta,\qquad (y_\alpha,y_\beta)=\mathbf{y}_\alpha^T\mathbf{P}\,\mathbf{y}_\beta. \tag{A1}$$

We shall examine first the simpler case where $\mathbf{Q}=\mathbf{I}$, $\mathbf{P}=\mathbf{I}$ and then we will generalize the results to any weight matrices $\mathbf{P}$ and $\mathbf{Q}$.

We will show that in the general case, with $r\leq\min(n,m)$, the matrix $\mathbf{A}$ can be decomposed according to

$$\underset{n\times m}{\mathbf{A}}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times d}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T} \tag{A2}$$

where the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, $\mathbf{\Lambda}$ is diagonal, $d=m-r$ is the injectivity defect and $f=n-r$ is the surjectivity defect. If such a representation is possible, it must also hold, in view of the orthogonality of $\mathbf{U}$ and $\mathbf{V}$ that

$$\underset{m\times m}{\mathbf{A}^T\mathbf{A}}=\underset{m\times m}{\mathbf{V}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}^2} & \underset{r\times d}{\mathbf{0}}\\ \underset{d\times r}{\mathbf{0}} & \underset{d\times d}{\mathbf{0}}\end{bmatrix}\underset{m\times m}{\mathbf{V}^T},\qquad \underset{n\times n}{\mathbf{A}\mathbf{A}^T}=\underset{n\times n}{\mathbf{U}}\begin{bmatrix}\underset{r\times r}{\mathbf{\Lambda}^2} & \underset{r\times f}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}}\end{bmatrix}\underset{n\times n}{\mathbf{U}^T}. \tag{A3}$$

We may consider the eigenvector-eigenvalue relations of the symmetric composite matrix

$$\begin{bmatrix}\underset{m\times m}{\mathbf{0}} & \underset{m\times n}{\mathbf{A}^T}\\ \underset{n\times m}{\mathbf{A}} & \underset{n\times n}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{m\times(n+m)}{\mathbf{W}}\\ \underset{n\times(n+m)}{\mathbf{Z}}\end{bmatrix}=\begin{bmatrix}\underset{m\times(n+m)}{\mathbf{W}}\\ \underset{n\times(n+m)}{\mathbf{Z}}\end{bmatrix}\begin{bmatrix}\underset{r\times r}{\mathbf{M}} & \underset{r\times s}{\mathbf{0}}\\ \underset{s\times r}{\mathbf{0}} & \underset{s\times s}{\mathbf{0}}\end{bmatrix} \tag{A4}$$

where $s=n+m-r=n+d=m+f$. We have used the fact that a symmetric matrix has orthonormal eigenvectors and the number of its positive eigenvalues is equal to its rank, the rest being zero. Thus $\mathbf{M}$ is the diagonal matrix having as diagonal elements the positive eigenvalues of the composite matrix. Carrying our the multiplication we obtain

$$\mathbf{A}^T\mathbf{Z}=\mathbf{W}\begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\qquad\&\qquad\mathbf{A}\mathbf{W}=\mathbf{Z}\begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\qquad\Rightarrow \tag{A5}$$

$$\mathbf{A}\mathbf{A}^T\mathbf{Z}=\mathbf{A}\mathbf{W}\begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}=\mathbf{Z}\begin{bmatrix}\mathbf{M}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \qquad \mathbf{A}^T\mathbf{A}\mathbf{W}=\mathbf{A}^T\mathbf{Z}\begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}=\mathbf{W}\begin{bmatrix}\mathbf{M}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \quad \text{(A6)}$$

We may now split

$$\underset{m\times(n+m)}{\mathbf{W}}=\begin{bmatrix}\underset{m\times m}{\mathbf{W}_1} & \underset{m\times n}{\mathbf{W}_2}\end{bmatrix}, \qquad \underset{n\times(n+m)}{\mathbf{Z}}=\begin{bmatrix}\underset{n\times n}{\mathbf{Z}_1} & \underset{n\times m}{\mathbf{Z}_2}\end{bmatrix} \quad \text{(A7)}$$

and rewrite (A5) as

$$\underset{m\times n}{\mathbf{A}^T}\underset{n\times n}{\mathbf{Z}_1}=\underset{m\times m}{\mathbf{W}_1}\begin{bmatrix}\underset{r\times r}{\mathbf{M}} & \underset{r\times f}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}}\end{bmatrix}, \qquad \underset{m\times n}{\mathbf{A}^T}\underset{n\times m}{\mathbf{Z}_2}=\underset{m\times m}{\mathbf{0}}, \quad \text{(A8)}$$

$$\underset{n\times m}{\mathbf{A}}\underset{m\times m}{\mathbf{W}_1}=\underset{n\times n}{\mathbf{Z}_1}\begin{bmatrix}\underset{r\times r}{\mathbf{M}} & \underset{r\times f}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}}\end{bmatrix}, \qquad \underset{n\times m}{\mathbf{A}}\underset{m\times n}{\mathbf{W}_2}=\underset{n\times n}{\mathbf{0}}, \quad \text{(A9)}$$

$$\underset{n\times n}{(\mathbf{A}\mathbf{A}^T)}\begin{bmatrix}\underset{n\times n}{\mathbf{Z}_1} & \underset{n\times m}{\mathbf{Z}_2}\end{bmatrix}=\begin{bmatrix}\underset{n\times n}{\mathbf{Z}_1} & \underset{n\times m}{\mathbf{Z}_2}\end{bmatrix}\begin{bmatrix}\underset{r\times r}{\mathbf{M}^2} & \underset{r\times f}{\mathbf{0}} & \underset{r\times m}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}} & \underset{f\times m}{\mathbf{0}}\\ \underset{m\times r}{\mathbf{0}} & \underset{m\times f}{\mathbf{0}} & \underset{m\times m}{\mathbf{0}}\end{bmatrix}, \quad \text{(A10)}$$

$$\underset{m\times m}{(\mathbf{A}^T\mathbf{A})}\begin{bmatrix}\underset{m\times m}{\mathbf{W}_1} & \underset{m\times n}{\mathbf{W}_2}\end{bmatrix}=\begin{bmatrix}\underset{m\times m}{\mathbf{W}_1} & \underset{m\times n}{\mathbf{W}_2}\end{bmatrix}\begin{bmatrix}\underset{r\times r}{\mathbf{M}^2} & \underset{r\times d}{\mathbf{0}} & \underset{r\times n}{\mathbf{0}}\\ \underset{d\times r}{\mathbf{0}} & \underset{d\times d}{\mathbf{0}} & \underset{d\times n}{\mathbf{0}}\\ \underset{n\times r}{\mathbf{0}} & \underset{n\times d}{\mathbf{0}} & \underset{n\times n}{\mathbf{0}}\end{bmatrix}. \quad \text{(A11)}$$

The last two relations can be split into

$$\underset{n\times n}{(\mathbf{A}\mathbf{A}^T)}\underset{n\times n}{\mathbf{Z}_1}=\underset{n\times n}{\mathbf{Z}_1}\begin{bmatrix}\underset{r\times r}{\mathbf{M}^2} & \underset{r\times f}{\mathbf{0}}\\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}}\end{bmatrix}, \qquad \underset{n\times n}{(\mathbf{A}\mathbf{A}^T)}\underset{n\times m}{\mathbf{Z}_2}=\underset{n\times m}{\mathbf{0}}, \quad \text{(A12)}$$

$$\underset{m\times m}{(\mathbf{A}^T\mathbf{A})}\underset{m\times m}{\mathbf{W}_1}=\underset{m\times m}{\mathbf{W}_1}\begin{bmatrix}\underset{r\times r}{\mathbf{M}^2} & \underset{r\times d}{\mathbf{0}}\\ \underset{d\times r}{\mathbf{0}} & \underset{d\times d}{\mathbf{0}}\end{bmatrix}, \qquad \underset{m\times m}{(\mathbf{A}^T\mathbf{A})}\underset{m\times n}{\mathbf{W}_2}=\underset{m\times n}{\mathbf{0}}. \quad \text{(A13)}$$

Comparison with (A3) shows that the obvious choice is $\mathbf{V}=\mathbf{W}_1$, $\mathbf{U}=\mathbf{Z}_1$, and $\mathbf{\Lambda}=\mathbf{M}$, provided that $\mathbf{Z}_1$ and $\mathbf{W}_1$ are orthogonal matrices. To show that the columns

$\mathbf{z}_i$ of $\mathbf{Z}_1$ form an orthonormal system and the same holds for the columns $\mathbf{w}_i$ of $\mathbf{W}_1$, we consider an eigenvector of the composite symmetric matrix corresponding to an eigenvalue μ_i

$$\begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix} = \mu_i \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix} \quad \Rightarrow \quad \begin{Bmatrix} \mathbf{A}^T \mathbf{z}_i = \mu_i \mathbf{w}_i \\ \mathbf{A}\mathbf{w}_i = \mu_i \mathbf{z}_i \end{Bmatrix} \Rightarrow \quad \begin{Bmatrix} \mathbf{A}^T (-\mathbf{z}_i) = (-\mu_i)\mathbf{w}_i \\ \mathbf{A}\mathbf{w}_i = (-\mu_i)(-\mathbf{z}_i) \end{Bmatrix}$$

$$\Rightarrow \quad \begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_i \\ -\mathbf{z}_i \end{bmatrix} = -\mu_i \begin{bmatrix} \mathbf{w}_i \\ -\mathbf{z}_i \end{bmatrix}. \tag{A14}$$

The eigenvectors corresponding to different eigenvalues are orthogonal to each other, the orthogonality relations for the pairs $\mu_i \neq \mu_k$ and $\mu_i \neq -\mu_k$ being

$$\begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_k \\ \mathbf{z}_k \end{bmatrix} = \mathbf{w}_i^T \mathbf{w}_k + \mathbf{z}_i^T \mathbf{z}_k = 0, \qquad \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_k \\ -\mathbf{z}_k \end{bmatrix} = \mathbf{w}_i^T \mathbf{w}_k - \mathbf{z}_i^T \mathbf{z}_k = 0. \tag{A15}$$

Adding and subtracting the above two relations we obtain the desired orthogonality relations

$$\mathbf{w}_i^T \mathbf{w}_k = 0, \qquad \mathbf{z}_i^T \mathbf{z}_k = 0 \qquad (i \neq k). \tag{A16}$$

The required normality relations are derived from

$$\begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix} = \mathbf{w}_i^T \mathbf{w}_i + \mathbf{z}_i^T \mathbf{z}_i = 1, \qquad \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix}^T \begin{bmatrix} \mathbf{w}_i \\ -\mathbf{z}_i \end{bmatrix} = \mathbf{w}_i^T \mathbf{w}_i - \mathbf{z}_i^T \mathbf{z}_i = 0, \tag{A17}$$

which added and subtracted yield

$$\mathbf{w}_i^T \mathbf{w}_i = 1, \qquad \mathbf{z}_i^T \mathbf{z}_i = 1. \tag{A18}$$

With the above identifications $\mathbf{V}=\mathbf{W}_1$, $\mathbf{U}=\mathbf{Z}_1$, $\mathbf{\Lambda}=\mathbf{M}$, we can write the basic relations (A8), (A9), (A10), (A11) in the final form

$$\underset{m\times n}{\mathbf{A}^T}\ \underset{n\times n}{\mathbf{U}} = \underset{m\times m}{\mathbf{V}} \begin{bmatrix} \underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times f}{\mathbf{0}} \\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}} \end{bmatrix}, \qquad \underset{n\times m}{\mathbf{A}}\ \underset{m\times m}{\mathbf{V}} = \underset{n\times n}{\mathbf{U}} \begin{bmatrix} \underset{r\times r}{\mathbf{\Lambda}} & \underset{r\times f}{\mathbf{0}} \\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}} \end{bmatrix}, \tag{A19}$$

$$(\mathbf{AA}^T)\underset{n\times n}{\mathbf{U}} = \underset{n\times n}{\mathbf{U}}\begin{bmatrix} \underset{r\times r}{\mathbf{\Lambda}^2} & \underset{r\times f}{\mathbf{0}} \\ \underset{f\times r}{\mathbf{0}} & \underset{f\times f}{\mathbf{0}} \end{bmatrix}, \qquad (\mathbf{A}^T\mathbf{A})\underset{m\times m}{\mathbf{V}} = \underset{m\times m}{\mathbf{V}}\begin{bmatrix} \underset{r\times r}{\mathbf{\Lambda}^2} & \underset{r\times d}{\mathbf{0}} \\ \underset{d\times r}{\mathbf{0}} & \underset{d\times d}{\mathbf{0}} \end{bmatrix}. \tag{A20}$$

In conclusion, the matrix $\mathbf{A}$ assumed the *singular value decomposition* (A2) where the orthogonal matrix $\mathbf{U}$ has columns the eigenvectors of $\mathbf{AA}^T$, the orthogonal matrix $\mathbf{V}$ has columns the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and the diagonal matrix $\mathbf{\Lambda}$ has diagonal elements equal to the common non vanishing eigenvalues of $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$. Note that eigenvectors and eigenvalues have been ordered in such a way that the first r columns of $\mathbf{U}$ and the first r columns of $\mathbf{V}$ correspond to non-vanishing eigenvalues.

We shall first examine the solution and properties of the two simpler cases where $\mathbf{A}$ has either full column rank or full row rank.

Generalization to any weight matrices

We have assumed so far that the finite-dimensional spaces X and Y of the unknown parameters and the observations, respectively, have metric properties defined by the simple Euclidean inner product. In the more general cases we have weight (metric) matrices different from the identity matrix and the inner products are instead

$$(x_\alpha, x_\beta) = \mathbf{x}_\alpha^T\mathbf{Q}\,\mathbf{x}_\beta, \qquad (y_\alpha, y_\beta) = \mathbf{y}_\alpha^T\mathbf{P}\,\mathbf{y}_\beta. \tag{A21}$$

The natural bases $\{\mathbf{e}_i\}$ of X and $\{\boldsymbol{\varepsilon}_i\}$ of Y are not any more orthonormal, since $\mathbf{e}_i^T\mathbf{Q}\mathbf{e}_k = Q_{ik} \neq \delta_{ik}$ and $\boldsymbol{\varepsilon}_i^T\mathbf{P}\boldsymbol{\varepsilon}_k = P_{ik} \neq \delta_{ik}$.
The positive definite symmetric matrices $\mathbf{Q}$ and $\mathbf{P}$ can be analyzed as

$$\mathbf{Q} = \mathbf{Q}_0^T\mathbf{Q}_0, \qquad \mathbf{P} = \mathbf{P}_0^T\mathbf{P}_0, \tag{A22}$$

where $\mathbf{Q}_0$ and $\mathbf{P}_0$ are invertible square matrices, which follow easily from the corresponding eigenvalue decompositions. We can formulate orthonormal bases

$$[\tilde{\mathbf{e}}] = [\tilde{\mathbf{e}}_1 \ldots \tilde{\mathbf{e}}_m] = [\mathbf{e}_1 \ldots \mathbf{e}_m]\mathbf{Q}_0^{-1} = [\mathbf{e}]\mathbf{Q}_0^{-1}, \tag{A23}$$

$$[\tilde{\boldsymbol{\varepsilon}}] = [\tilde{\boldsymbol{\varepsilon}}_1 \ldots \tilde{\boldsymbol{\varepsilon}}_n] = [\boldsymbol{\varepsilon}_1 \ldots \boldsymbol{\varepsilon}_n]\mathbf{P}_0^{-1} = [\boldsymbol{\varepsilon}]\mathbf{P}_0^{-1}, \tag{A24}$$

with corresponding component transformations $\tilde{\mathbf{x}} = \mathbf{Q}_0\mathbf{x}$, $\tilde{\mathbf{y}} = \mathbf{P}_0\mathbf{y}$ so that the inner products of vectors $x = [\mathbf{e}]\mathbf{x} = [\tilde{\mathbf{e}}]\tilde{\mathbf{x}}$ in X and $y = [\boldsymbol{\varepsilon}]\mathbf{y} = [\tilde{\boldsymbol{\varepsilon}}]\tilde{\mathbf{y}}$ in Y, obtain the simple "Euclidean" form

$$(x_\alpha, x_\beta) = \tilde{\mathbf{x}}_\alpha^T\tilde{\mathbf{x}}_\beta, \qquad (y_\alpha, y_\beta) = \tilde{\mathbf{y}}_\alpha^T\tilde{\mathbf{y}}_\beta. \tag{A25}$$

The model operator A is now represented by a matrix

$$\tilde{\mathbf{A}}=\mathbf{P}_0\mathbf{A}\mathbf{Q}_0^{-1} \tag{A26}$$

as follows from the fact that when $\mathbf{y}=\mathbf{A}\mathbf{x}$ then $\tilde{\mathbf{y}}=\mathbf{P}_0\mathbf{y}=\mathbf{P}_0\mathbf{A}\mathbf{x}=\mathbf{P}_0\mathbf{A}\mathbf{Q}_0^{-1}\tilde{\mathbf{x}}\equiv\tilde{\mathbf{A}}\tilde{\mathbf{x}}$. We can now write the relations (A19), (A20) as

$$\tilde{\mathbf{A}}^T\tilde{\mathbf{U}}=\tilde{\mathbf{V}}\begin{bmatrix}\tilde{\mathbf{\Lambda}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \qquad \tilde{\mathbf{A}}\tilde{\mathbf{V}}=\tilde{\mathbf{V}}\begin{bmatrix}\tilde{\mathbf{\Lambda}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{A27}$$

$$(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T)\tilde{\mathbf{U}}=\tilde{\mathbf{U}}\begin{bmatrix}\tilde{\mathbf{\Lambda}}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \qquad (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})\tilde{\mathbf{V}}=\tilde{\mathbf{V}}\begin{bmatrix}\tilde{\mathbf{\Lambda}}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{A28}$$

The columns of $\mathbf{V}$ and $\mathbf{U}$ transform according to $\tilde{\mathbf{v}}_i=\mathbf{Q}_0\mathbf{v}_i$, $\tilde{\mathbf{u}}_i=\mathbf{P}_0\mathbf{u}_i$, implying that $\tilde{\mathbf{V}}=\mathbf{Q}_0\mathbf{V}$, $\tilde{\mathbf{U}}=\mathbf{P}_0\mathbf{U}$, which substituted above together with $\tilde{\mathbf{A}}=\mathbf{P}_0\mathbf{A}\mathbf{Q}_0^{-1}$ lead after identifying $\tilde{\mathbf{\Lambda}}=\mathbf{\Lambda}$ to the relations

$$(\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})\mathbf{U}=\mathbf{V}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \qquad \mathbf{A}\mathbf{V}=\mathbf{U}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \tag{A29}$$

$$[\mathbf{A}(\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})]\mathbf{U}=\mathbf{U}\begin{bmatrix}\mathbf{\Lambda}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}, \qquad [(\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P})\mathbf{A}]\,\mathbf{V}=\mathbf{V}\begin{bmatrix}\mathbf{\Lambda}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{A30}$$

One should notice that the matrix $\mathbf{A}^*=\mathbf{Q}^{-1}\mathbf{A}^T\mathbf{P}$ represents the adjoint operator A^* of A defined by $(y,Ax)=(A^*y,x)$ for any x and y.
The SVD (A2) takes the form

$$\tilde{\mathbf{A}}=\tilde{\mathbf{U}}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\tilde{\mathbf{V}}^T, \qquad \mathbf{A}'=\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}=\tilde{\mathbf{U}}^T\tilde{\mathbf{A}}\tilde{\mathbf{V}}, \tag{A31}$$

which transforms into

$$\mathbf{A}=\mathbf{U}\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{V}^T\mathbf{Q}, \qquad \mathbf{A}'=\begin{bmatrix}\mathbf{\Lambda} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}=\mathbf{U}^T\mathbf{P}\mathbf{A}\mathbf{V}. \tag{A32}$$

The columns of $\mathbf{U}$ are the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^*$ and the columns of $\mathbf{V}$ the eigenvectors of $\mathbf{A}^*\mathbf{A}$. Since the matrices $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$ are not symmetric the

"eigenvector" matrices $\mathbf{U}$ and $\mathbf{V}$ are not orthogonal. However they satisfy "orthogonality" relations reflecting the orthogonality of the elements $u_i \in Y$, $v_i \in X$ that their respective columns $\mathbf{u}_i$, $\mathbf{v}_i$ represent. Replacing $\tilde{\mathbf{V}}=\mathbf{Q}_0\mathbf{V}$ and $\tilde{\mathbf{U}}=\mathbf{P}_0\mathbf{U}$ in the original orthogonality relations $\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}=\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T=\mathbf{I}$, $\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}=\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T=\mathbf{I}$ we obtain

$$\mathbf{U}^T\mathbf{P}\mathbf{U}=\mathbf{I}, \qquad \mathbf{U}\mathbf{U}^T=\mathbf{P}^{-1}, \qquad \mathbf{V}^T\mathbf{Q}\mathbf{V}=\mathbf{I}, \qquad \mathbf{V}\mathbf{V}^T=\mathbf{Q}^{-1}. \tag{A33}$$

The transformation equations can be written as $\mathbf{x}'=\tilde{\mathbf{V}}^T\tilde{\mathbf{x}}=\mathbf{V}^T\mathbf{Q}_0^T$, $\mathbf{y}'=\tilde{\mathbf{U}}^T\tilde{\mathbf{y}}=$ $=\mathbf{U}^T\mathbf{P}_0^T\tilde{\mathbf{y}}$, which combined with $\tilde{\mathbf{x}}=\mathbf{Q}_0\mathbf{x}$, $\tilde{\mathbf{y}}=\mathbf{P}_0\mathbf{y}$ yield the SVD transformation

$$\mathbf{x}'=\mathbf{V}^T\mathbf{Q}\mathbf{x}, \qquad \mathbf{y}'=\mathbf{U}^T\mathbf{P}\mathbf{y}. \tag{A34}$$

It is easy to see that in the SVD system the inner products of X and Y have the "Euclidean" forms $(x_\alpha, x_\beta)=\mathbf{x}'^T_\alpha\mathbf{x}'_\beta$, $(y_\alpha, y_\beta)=\mathbf{y}'^T_\alpha\mathbf{y}'_\beta$.

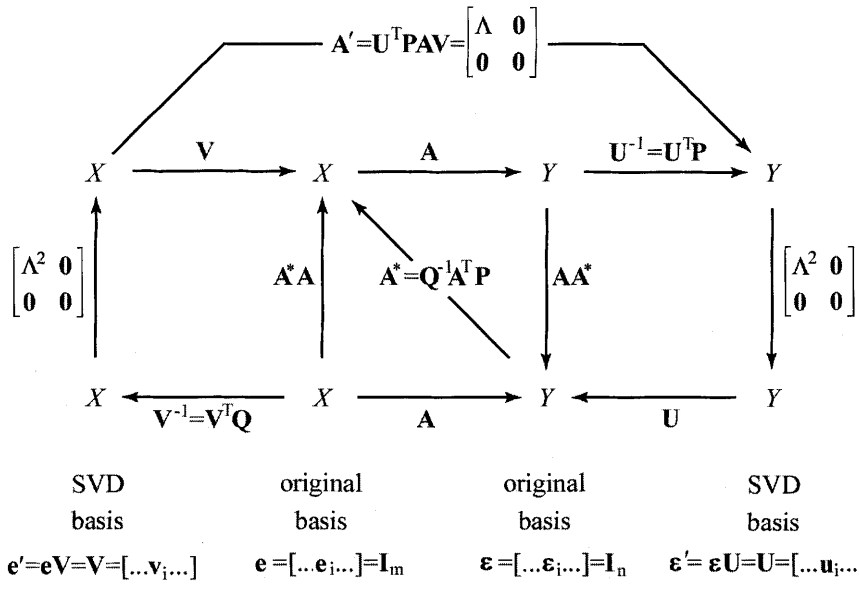


Fig. A1: An illustration of the Singular Value Decomposition (SVD).

The metric properties in the original bases are determined by the matrices $\mathbf{Q}$ in X with $Q_{ik}=<\mathbf{e}_i,\mathbf{e}_k>=\mathbf{e}_i^T\mathbf{Q}\mathbf{e}_k$ and $\mathbf{P}$ in Y with $P_{ik}=<\boldsymbol{\varepsilon}_i,\boldsymbol{\varepsilon}_k>=\boldsymbol{\varepsilon}_i^T\mathbf{P}\boldsymbol{\varepsilon}_k$.

In the SVD bases the corresponding metric properties are

$<\mathbf{v}_i,\mathbf{v}_k>=\mathbf{v}_i^T\mathbf{Q}\mathbf{v}_k=\delta_{ik}$ in X and $<\mathbf{u}_i,\mathbf{u}_k>=\mathbf{u}_i^T\mathbf{P}\mathbf{u}_k=\delta_{ik}$ in Y.

We may also have a coordinate-free view of the singular value decomposition: The columns $\mathbf{v}_i$, $\mathbf{u}_i$ of the matrices $\mathbf{V}$ and $\mathbf{U}$, respectively, represent elements $v_i\in X$, $i=1,\ldots,m$ and $u_i\in Y$, $i=1,\ldots,n$ which are the orthonormal eigen-elements of the symmetric operators

$$(A^*A)v_i=\lambda_i^2 v_i, \quad (AA^*)u_i=\lambda_i^2 u_i, \quad \lambda_i^2=0 \text{ for } i>r=rank(A) \tag{A35}$$

and they can serve as bases for the spaces X and Y. With respect to these SVD bases we have the representations

$$x=\sum_{i=1}^{m} x_i v_i, \quad y=\sum_{i=1}^{n} y_i u_i \tag{A36}$$

and the equation $y=Ax$ takes, with respect to the above SVD coordinates, the form

$$y_i=\lambda_i x_i, \quad i=1,2,\ldots,r, \quad y_i=0, \quad i=r+1,2,\ldots,n. \tag{A37}$$

LINEAR AND NONLINEAR INVERSE PROBLEMS

ROEL SNIEDER AND JEANNOT TRAMPERT
Dept. of Geophysics
Utrecht University
P.O. Box 80.021
3508 TA Utrecht
The Netherlands
email snieder@geo.uu.nl

1. Introduction

An important aspect of the physical sciences is to make inferences about physical parameters from data. In general, the laws of physics provide the means for computing the data values given a model. This is called the "forward problem", see figure 1. In the inverse problem, the aim is to reconstruct the model from a set of measurements. In the ideal case, an exact theory exists that prescribes how the data should be transformed in order to reproduce the model. For some selected examples such a theory exists assuming that the required infinite and noise-free data sets would be available. A quantum mechanical potential in one spatial dimension can be reconstructed when the reflection coefficient is known for all energies [*Marchenko*, 1955; *Burridge,* 1980]. This technique can be generalized for the reconstruction of a quantum mechanical potential in three dimensions [*Newton,* 1989], but in that case a redundant data set is required for reasons that are not well understood. The mass-density in a one-dimensional string can be constructed from the measurements of all eigenfrequencies of that string [*Borg,* 1946], but due to the symmetry of this problem only the even part of the mass-density can be determined. If the seismic velocity in the earth depends only on depth, the velocity can be constructed exactly from the measurement of the arrival time as a function of distance of seismic waves using an Abel transform [*Herglotz,* 1907; *Wiechert,* 1907]. Mathematically this problem is identical to the construction of a spherically symmetric quantum mechanical potential in three dimensions [*Keller et al.,* 1956]. However, the construction method

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 93–164, 2000.

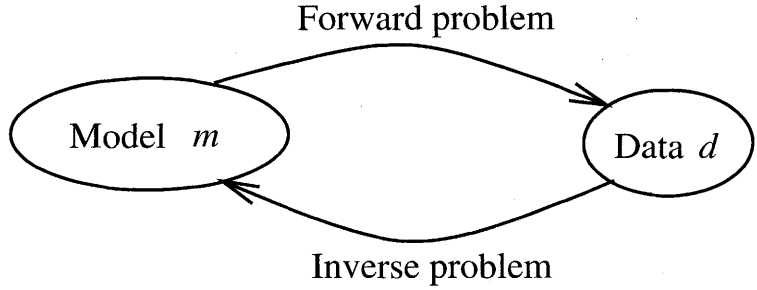


Figure 1. The traditional definition of the forward and inverse problems.

of Herglotz-Wiechert only gives an unique result when the velocity increases monotonically with depth [*Gerver and Markushevitch,* 1966]. This situation is similar in quantum mechanics where a radially symmetric potential can only be constructed uniquely when the potential does not have local minima [*Sabatier*, 1973].

Despite the mathematical elegance of the exact nonlinear inversion schemes, they are of limited applicability. There are a number of reasons for this. First, the exact inversion techniques are usually only applicable for idealistic situations that may not hold in practice. For example, the Herglotz-Wiechert inversion presupposes that the velocity in the earth depends only on depth and that the velocity increases monotonically with depth. Seismic tomography has shown that both requirements are not met in the earth's mantle [*Nolet et al.,* 1994]. Second, the exact inversion techniques often are very unstable. The presence of this instability in the solution of the Marchenko equation has been shown explicitly by *Dorren et al.* [1994]. However, the third reason is the most fundamental. In many inverse problems the model that one aims to determine is a continuous function of the space variables. This means that the model has infinitely many degrees of freedom. However, in a realistic experiment the amount of data that can be used for the determination of the model is usually finite. A simple count of variables shows that the data cannot carry sufficient information to determine the model uniquely. In the context of linear inverse problems this point has been raised by *Backus and Gilbert* [1967, 1968] and more recently by *Parker* [1994]. This issue is equally relevant for nonlinear inverse problems.

The fact that in realistic experiments a finite amount of data is available to reconstruct a model with infinitely many degrees of freedom necessarily means that the inverse problem is not unique in the sense that there are many models that explain the data equally well. The model obtained from the inversion of the data is therefore not necessarily equal to the

true model that one seeks. This implies that the view of inverse problems as shown in figure 1 is too simplistic. For realistic problems, inversion really consists of two steps. Let the true model be denoted by m and the data by d. From the data d one reconstructs an estimated model $\tilde{m}$, this is called the *estimation problem*, see figure 2. Apart from estimating a model $\tilde{m}$ that is consistent with the data, one also needs to investigate what relation the estimated model $\tilde{m}$ bears to the true model m. In the *appraisal problem* one determines what properties of the true model are recovered by the estimated model and what errors are attached to it. The essence of this discussion is that *inversion* = *estimation* + *appraisal*. It does not make much sense to make a physical interpretation of a model without acknowledging the fact of errors and limited resolution in the model [*Trampert*, 1998].

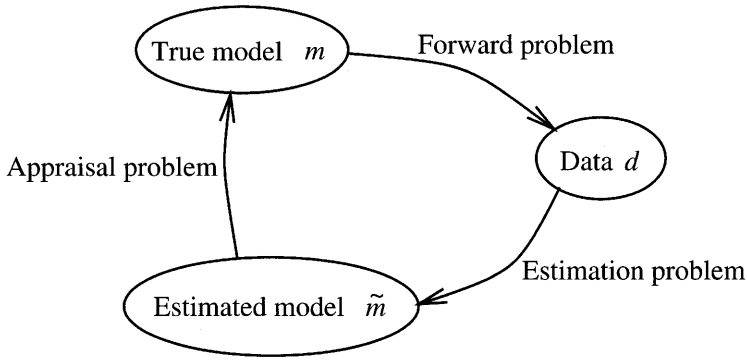


Figure 2. The inverse problem viewed as a combination of an estimation problem plus an appraisal problem.

In general there are two reasons why the estimated model differs from the true model. The first reason is the non-uniqueness of the inverse problem that causes several (usually infinitely many) models to fit the data. Technically, this model null-space exits due to inadequate sampling of the model space. The second reason is that real data (and physical theories more often than we would like) are always contaminated with errors and the estimated model is therefore affected by these errors as well. Therefore model appraisal has two aspects, non-uniqueness and error propagation.

Model estimation and model appraisal are fundamentally different for discrete models with a finite number of degrees of freedom and for continuous models with infinitely many degrees of freedom. Also, the problem of model appraisal is only well-solved for linear inverse problems. For this reason the inversion of discrete models and continuous models is treated separately, and the case of linear inversion and nonlinear inversion is also treated independently. In section 2 linear inversion for a finite number of

model parameters is discussed. This is generalized in section 3 to deal with linear inverse problems for continuous models with infinitely many degrees of freedom. In reality many inverse problems are not really linear, but often these problems can be linearized by making a suitable approximation. In section 4 the single-scattering approximation is derived. This technique forms the basis of imaging tools used in reflection seismology. Rayleigh's principle, as treated in section 5, is the linearization that forms the basis for the inversion for the Earth's structure using normal-mode frequencies. The linearization technique of seismic travel time tomography is based on Fermat's principle, which is treated in section 6. Nonlinear inverse problems are significantly more difficult than linear inverse problems. It is shown in section 7 that non-linearity can be a source of ill-posedness. Presently, there is no satisfactory theory for the appraisal problem for nonlinear inverse problems. In section 8 three methods are presented that can be used for the nonlinear appraisal problem. However, neither of these methods is quite satisfactory, which indicates that nonlinear inverse problem theory is a field with important research challenges.

2. Solving finite linear systems of equations

As argued in the previous section, the inverse problem maps a finite number of data onto a model. In most practical applications in geophysics the model is a continuous function of the space coordinates and therefore has infinitely many degrees of freedom. For the moment we will ignore this and will assume that the model can be characterized by a finite number of parameters. We will return to the important case of models that are infinitely dimensional in section 3.

2.1. LINEAR MODEL ESTIMATION

For a finite-dimensional model, the model parameters can be ordered in a vector $\boldsymbol{m}$, and similarly the data can be ordered in a vector $\boldsymbol{d}$. The matrix $\boldsymbol{A}$ relates the data to the model through the product $\boldsymbol{Am}$. This matrix is often referred to as the theory operator. Indeed, it contains all the information on physics and mathematics we have chosen to model in the given problem. In practice, the data are contaminated with errors $\boldsymbol{e}$, so that the recorded data and the model are related by:

$$\boldsymbol{d} = \boldsymbol{Am} + \boldsymbol{e} \tag{1}$$

It should be noted there often there is an certain arbitrariness in the choice of the model parameters that are contained in the model vector $\boldsymbol{m}$. For example, if one wants to describe the density in the earth one could choose

a model where the Earth's mantle and the core have a uniform density, in that case there are two model parameters. Alternatively, one could expand the density in the Earth in a large amount of eigenfunctions defined on the sphere such as spherical harmonics for lateral variations and polynomials for depth variations, in that case there are much more model parameters. These two different parameterizations of the same model correspond to different model parameters $\boldsymbol{m}$ and to a different matrix $\boldsymbol{A}$. This example illustrates that the model $\boldsymbol{m}$ is not necessarily the true model,[1] but that the choice of the model parameters usually contains a restriction on the class of models that can be constructed. Below we will refer to $\boldsymbol{m}$ as the true model regardless of the difficulties in its definition.

From the recorded data one makes an estimate of the model. Since this estimate in practice will be different from the true model the estimated model is denoted by $\tilde{\boldsymbol{m}}$. There are many ways for designing an inverse operator that maps the data on the estimated model [e.g. *Menke*, 1984; *Tarantola*, 1987; *Parker*, 1994]. Whatever estimator one may choose, the most general linear mapping from data to the estimated model can be written as:

$$\tilde{\boldsymbol{m}} = \boldsymbol{A}^{-g}\boldsymbol{d} \tag{2}$$

The operator $\boldsymbol{A}^{-g}$ is called the *generalized inverse* of the matrix $\boldsymbol{A}$. In general, the number of data is different from the number of model parameters. For this reason $\boldsymbol{A}$ is usually a non-square matrix, and hence its formal inverse does not exist. Later we will show how the generalized inverse $\boldsymbol{A}^{-g}$ may be chosen, but for the moment $\boldsymbol{A}^{-g}$ does not need to be specified. The relation between the estimated model $\tilde{\boldsymbol{m}}$ and the true model $\boldsymbol{m}$ follows by inserting (1) in expression (2):

$$\tilde{\boldsymbol{m}} = \boldsymbol{A}^{-g}\boldsymbol{A}\boldsymbol{m} + \boldsymbol{A}^{-g}\boldsymbol{e} \tag{3}$$

The matrix $\boldsymbol{A}^{-g}\boldsymbol{A}$ is called the *resolution kernel*, this operator is given by:

$$\boldsymbol{R} \equiv \boldsymbol{A}^{-g}\boldsymbol{A} \tag{4}$$

Expression (3) can be interpreted by rewriting it in the following form:

[1]We urge the reader to formulate a definition of the concept "true model." It is not so difficult to formulate a vague definition such as "the true model is the model that corresponds to reality and which is only known to the gods." However, we are not aware of any definition that is operational in the sense that it provides us with a set of actions that could potentially tell us what the true model really is.

$$\tilde{\boldsymbol{m}} = \boldsymbol{m} + \underbrace{\left(\boldsymbol{A}^{-g}\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{m}}_{\text{Limited Resolution}} + \underbrace{\boldsymbol{A}^{-g}\boldsymbol{e}}_{\text{Error propagation}} \tag{5}$$

In the ideal case, the estimated model equals the true model vector: $\tilde{\boldsymbol{m}} = \boldsymbol{m}$ meaning that our chosen parameters, ordered in vector $\boldsymbol{m}$, may be estimated independently from each other. The last two terms in equation (5) account for *blurring* and *artifacts* in the estimated model. The term $\left(\boldsymbol{A}^{-g}\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{m}$ describes the fact that components of the estimated model vector are linear combinations of different components of the true model vector. We only retrieve averages of our parameters and "blurring" occurs in the model estimation as we are not able to map out the finest details. In the ideal case this term vanishes; this happens when $\boldsymbol{A}^{-g}\boldsymbol{A}$ is equal to the identity matrix. With (4) this means that for perfectly resolved model parameters the resolution matrix is the identity matrix:

$$\text{Perfect resolution: } \boldsymbol{R} = \boldsymbol{I} \tag{6}$$

As noted earlier, usually there is a certain ambiguity in the definition of the model parameters that define the vector $\boldsymbol{m}$. The resolution operator tells us to what extend we can retrieve the model parameters independently from the estimation process. However, the resolution matrix does not tell us completely what the relation between the estimated model and the real underlying physical model is, because it does not take into account to what extent the choice of the model parameters has restricted the model that can be obtained from the estimation process.

The last term in (5) describes how the errors $\boldsymbol{e}$ are mapped onto the estimated model.[2] These errors are not known deterministically, otherwise they could be subtracted from the data. A statistical analysis is needed to describe the errors in the estimated model due to the errors in the data. When the data d_j are uncorrelated and have standard deviation σ_{d_j}, the standard deviation σ_{m_i} in the model estimate $\tilde{m}_i$, resulting from the propagation of data errors only, is given by:

$$\sigma_{m_i}^2 = \sum_j \left(A_{ij}^{-g}\sigma_{d_j}\right)^2 \tag{7}$$

Ideally, one would like to obtain both: a perfect resolution and no errors in the estimated model. Unfortunately this cannot be achieved in practice. The error propagation is, for instance, completely suppressed by using the generalized inverse $\boldsymbol{A}^{-g} = 0$. This leads to the (absurd) estimated model

[2] As shown by *Scales and Snieder* [1998] the concept of errors in inverse problems is not as simple as it appears.

$\tilde{\boldsymbol{m}} = 0$ which is indeed not affected by errors. However, this particular generalized inverse has a resolution matrix given by $\boldsymbol{R} = 0$, which is far from the ideal resolution matrix given in (6). Hence in practice, one has to find an acceptable trade-off between error-propagation and limitations in the resolution.

2.2. LEAST-SQUARES ESTIMATION

Let us for the moment consider the case where the number of independent data is larger than the number of unknowns. In that case, the system $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{m}$ cannot always be satisfied for any given model $\boldsymbol{m}$ because of possible errors contained in the data vector making the equations inconsistent. As an example, let us consider the following problem. We have two masses with weight m_1 and m_2. The weighing of the first mass yields a weight of 1 (kilo). Then one measures the second mass to find a weight of 2. Next, one weighs the masses together to find a combined weight of 2. The system of equations that corresponds to these measurements is given by:

$$\begin{aligned} m_1 &= d_1 &&= 1 \\ m_2 &= d_2 &&= 2 \\ m_1 + m_2 &= d_3 &&= 2 \end{aligned} \tag{8}$$

The matrix $\boldsymbol{A}$ for this problem is given by:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{9}$$

It is clear that this system of equations cannot be satisfied. It cannot be true that the first mass has a weight $m_1 = 1$, and the second mass has a weight $m_2 = 2$ while the combined mass is equal to $m_1 + m_2 = 2$. Clearly errors have been made during the measurements, but there is no reason to discard one of the three equations in favor of the other two. This problem is illustrated graphically in figure 3. The three equations (8) correspond to the three solid lines in the (m_1, m_2)-plane. The fact that the three lines do not intersect in a single point signifies that the linear equations are inconsistent. The inverse problem of determining the two masses thus consists in reconciling these equations in a meaningful way.

A common way to estimate a model is to seek the model $\tilde{\boldsymbol{m}}$ that gives the best fit to the data in the sense that the difference, measured by the L_2-norm, between the data vector $\boldsymbol{d}$ and the recalculated data $\boldsymbol{A}\tilde{\boldsymbol{m}}$ is made as small as possible. This means that the least-squares solution is given by the model that minimizes the following cost function:

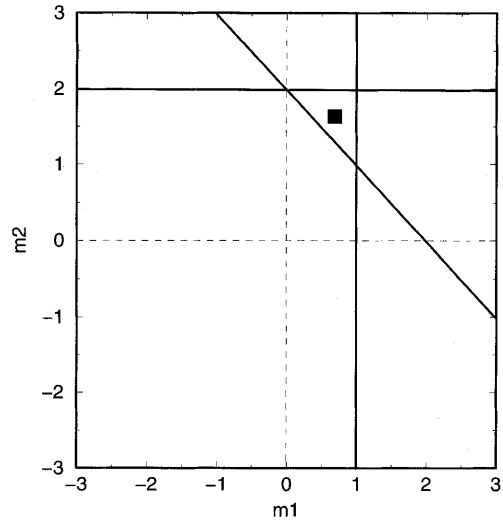


Figure 3. Geometrical interpretation of the linear equations (8).

$$S = \|\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}\|^2 \tag{10}$$

As shown in detail by *Strang* [1988] this quantity is minimized by the following model estimate:

$$\tilde{\boldsymbol{m}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d} \tag{11}$$

In the example of figure 3 the least-squares solution is the point in the (m_1, m_2)-plane that has the smallest distance to the three lines in that figure, this point is indicated by a black square. Using the matrix (9) one readily finds that the least-squares estimator of the problem (8) is given by:

$$\tilde{\boldsymbol{m}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{pmatrix}\boldsymbol{d} \tag{12}$$

For the used data vector this means that the estimated model is given by:

$$\begin{aligned} \tilde{m}_1 &= 2/3 \\ \tilde{m}_2 &= 5/3 \end{aligned} \tag{13}$$

2.3. MINIMUM NORM ESTIMATION

In some problems the number of unknowns is less than the number of parameters. Consider for example the situation where there are two masses

m_1 and m_2 and one has measured only the combined weight of these masses:

$$m_1 + m_2 = d = 2 \tag{14}$$

The matrix that corresponds to this system of one equation is given by:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \end{pmatrix} \tag{15}$$

Graphically this problem is depicted in figure 4. Clearly any model vector lying on the solid line fits the equation (14) exactly. There are thus infinitely many solutions, provided the masses are positive, that exactly fit the data. A model estimate can be defined by choosing a model that fits the data exactly and that has the smallest L_2-norm, this model is indicated by in figure 4 by the black square.

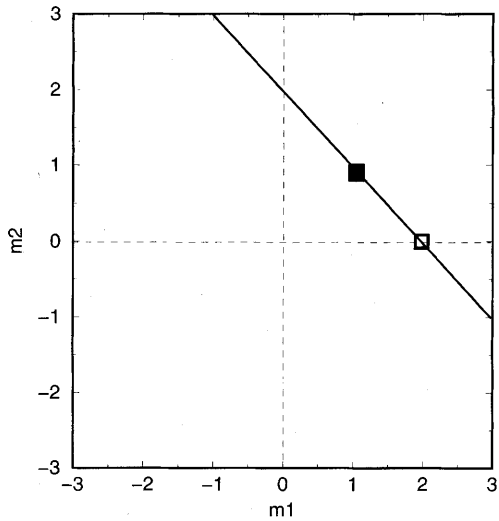


Figure 4. Geometrical interpretation of the linear equation (14) with two unknowns.

For a general underdetermined system of equations the minimum norm solution is defined as the model that fits the data exactly, $\boldsymbol{Am} = \boldsymbol{d}$, and that minimizes $\|\boldsymbol{m}\|^2$. Using Lagrange multipliers one can show that the minimum-norm solution is given by:

$$\tilde{\boldsymbol{m}} = \boldsymbol{A}^{\boldsymbol{T}} \left(\boldsymbol{A}\boldsymbol{A}^T \right)^{-1} \boldsymbol{d} \,, \tag{16}$$

a detailed derivation is given by *Menke* [1984]. One readily finds that the minimum norm solution of "system" (14) is given by

$$m_1 = m_2 = 1 \,. \tag{17}$$

In the least-squares estimation, we assumed that we had enough information to evaluate all model parameters, even though contradictions occurred due to measurement errors. The problem is then purely over-determined and as a consequence $\boldsymbol{A}^T\boldsymbol{A}$ is regular. In the minimum norm solution, we assumed no contradictions in the available information, but we don't have enough equations to evaluate all model parameters. This is the case of a purely under-determined problem and here $\boldsymbol{A}\boldsymbol{A}^T$ is regular. The most common case, however, is that we have contradictory information on some model parameters, while others cannot be assessed due to a lack of information. Then neither $\boldsymbol{A}^T\boldsymbol{A}$ nor $\boldsymbol{A}\boldsymbol{A}^T$ can be inverted and the problem is ill-posed. Even if the inverse matrices formally exist, they are often ill-conditioned meaning that small changes in the data vector lead to large changes in the model estimation. This means that errors in the data will be magnified in the model estimation. Clearly a trick is needed to find a model that is not too sensitive on small changes in the data. To this effect, *Levenberg* [1944] introduced a damped least-squares solution. From a mathematical point of view, ill-posedness and ill-conditioning result from zero or close to zero singular values of $\boldsymbol{A}$.

Suppose one has a matrix $\boldsymbol{M}$ with eigenvalues λ_n and eigenvectors $\hat{\boldsymbol{v}}_n$:

$$\boldsymbol{M}\hat{\boldsymbol{v}}_n = \lambda_n \hat{\boldsymbol{v}}_n \tag{18}$$

One readily finds that the matrix $(\boldsymbol{M}+\gamma\boldsymbol{I})$ has eigenvalues $(\lambda_n + \gamma)$:

$$(\boldsymbol{M}+\gamma\boldsymbol{I})\,\hat{\boldsymbol{v}}_n = (\lambda_n + \gamma)\,\hat{\boldsymbol{v}}_n \tag{19}$$

This means that the eigenvalues of a matrix can be raised by adding the scaled identity matrix to the original matrix. This property can be used to define the *damped least-squares solution*:

$$\tilde{\boldsymbol{m}} = \left(\boldsymbol{A}^T\boldsymbol{A}+\gamma\boldsymbol{I}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d} \tag{20}$$

Since the matrix $\boldsymbol{A}^T\boldsymbol{A}$ has positive eigenvalues[3] its eigenvalues are moved away from zero when the constant γ is positive. Alternatively, the solution (20) can be found by minimizing the following cost function:

[3]That the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ are positive follows from the following identity: $\left(\boldsymbol{x}\cdot\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}\right) = \left(\boldsymbol{A}^{TT}\boldsymbol{x}\cdot\boldsymbol{A}\boldsymbol{x}\right) = (\boldsymbol{A}\boldsymbol{x}\cdot\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|^2 \geq 0$. When $\boldsymbol{x}$ is the eigenvector $\hat{\boldsymbol{u}}^{(n)}$ of $\boldsymbol{A}^T\boldsymbol{A}$ with eigenvalue μ_n, this expression can be used to show that $\mu_n \left\|\hat{\boldsymbol{u}}^{(n)}\right\|^2 = \mu_n\left(\hat{\boldsymbol{u}}^{(\boldsymbol{n})}\cdot\hat{\boldsymbol{u}}^{(n)}\right) = \left(\hat{\boldsymbol{u}}^{(\boldsymbol{n})}\cdot\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{u}}^{(n)}\right) \geq 0$, hence the eigenvalues μ_n are positive.

$$S = \|\boldsymbol{d} - \boldsymbol{Am}\|^2 + \gamma \|\boldsymbol{m}\|^2 \tag{21}$$

This expression clearly shows what the effect of the damping is. Minimizing the first term of (21) amounts to finding the model that gives the best fit to the data. Minimizing the last term of (21) amounts to finding the model with the smallest norm. In general we cannot minimize both terms simultaneously, but in minimizing (21) we comprise in finding a model that both fits the data reasonably well and whose model size is not too large. The parameter γ controls the emphasis we put on these conflicting requirements and for this reason it is called the *trade-off parameter.*

For a number of applications the following matrix identity is extremely useful:[4]

$$\boxed{\left(\boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{A} + \boldsymbol{D}^{-1}\right)^{-1}\boldsymbol{A}^T\boldsymbol{B}^{-1} = \boldsymbol{D}\boldsymbol{A}^T\left(\boldsymbol{B} + \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T\right)^{-1}.} \tag{22}$$

In this expression $\boldsymbol{B}$ and $\boldsymbol{D}$ are regular square matrices, whereas $\boldsymbol{A}$ needs not to be square. This expression can be used to show that when damping or regularization is used, the least-squares solution and the minimum-norm solution (both supplied with a damping term) are identical. To see this use (22) with $\boldsymbol{B}^{-1} = \boldsymbol{I}$ and $\boldsymbol{D}^{-1} = \gamma\boldsymbol{I}$. It then follows that

$$\left(\boldsymbol{A}^T\boldsymbol{A} + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d} = \frac{1}{\gamma}\boldsymbol{A}^T\left(\boldsymbol{I} + \boldsymbol{A}\frac{1}{\gamma}\boldsymbol{A}^T\right)^{-1}\boldsymbol{d} = \boldsymbol{A}^T\left(\boldsymbol{A}\boldsymbol{A}^T + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{d}\,. \tag{23}$$

The left hand side corresponds to the damped least-squares solution (20) while the right hand side is the damped version of the minimum-norm solution (16). This implies that when damping is applied the least-squares solution and the minimum-norm solution are identical.

2.5. THE CONSISTENCY PROBLEM FOR THE LEAST-SQUARES SOLUTION

The least-squares solution appears to provide an objective method for finding solutions of overdetermined problems. However, there is trouble ahead. To see this, let us consider the overdetermined system of equations (8). Mathematically, this system of equations does not change when we

[4]This identity follows from the identity $\boldsymbol{A}^T + \boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T = \boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T + \boldsymbol{A}^T$. Write the first term on the left hand side as $\boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{B}$ and the last term on the right hand side as $\boldsymbol{D}^{-1}\boldsymbol{D}\boldsymbol{A}^T$. The resulting expression can then be written as $\boldsymbol{A}^T\boldsymbol{B}^{-1}\left(\boldsymbol{B} + \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T\right) = \left(\boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{A} + \boldsymbol{D}^{-1}\right)\boldsymbol{D}\boldsymbol{A}^T$. The expression (22) then follows by multiplying on the left with $\left(\boldsymbol{B} + \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^T\right)^{-1}$ and by multyplying on the right with $\left(\boldsymbol{A}^T\boldsymbol{B}^{-1}\boldsymbol{A} + \boldsymbol{D}^{-1}\right)^{-1}$.

multiply the last equation with a factor two. The following two systems of equations thus are completely equivalent:

$$\left.\begin{array}{rrr} m_1 = & d_1 & = 1 \\ m_2 = & d_2 & = 2 \\ m_1 + m_2 = & d_3 & = 2 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{rrrr} m_1 = & d_1 = & d_1' = & 1 \\ m_2 = & d_2 = & d_2' = & 2 \\ 2m_1 + 2m_2 = & 2d_3 = & d_3' = & 4 \end{array}\right. \tag{24}$$

The matrices of the original system and the new equivalent system are given by

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{A}' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{pmatrix} \tag{25}$$

In this section the unprimed quantities denote the original system of equations while the primed quantities refer to the transformed system of equations. One readily finds that the least-squares solution (11) of the original system and the transformed system are given by

$$\tilde{\boldsymbol{m}} = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \end{pmatrix} \boldsymbol{d} \quad \text{and} \quad \tilde{\boldsymbol{m}} = \frac{1}{9}\begin{pmatrix} 5 & -4 & 2 \\ -4 & 5 & 2 \end{pmatrix} \boldsymbol{d}' \,, \tag{26}$$

Using the numerical values of the original data vector $\boldsymbol{d}$ and the transformed data vector $\boldsymbol{d}'$ this leads to the following estimates of the model:

$$\tilde{\boldsymbol{m}} = \begin{pmatrix} 2/3 \\ 5/3 \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{m}} = \begin{pmatrix} 5/9 \\ 14/9 \end{pmatrix} \tag{27}$$

The problem is that these two estimators of the same model are *different*! This is surprising because the original system of equations and the transformed system of equations in (24) are mathematically equivalent. The reason that the two solutions are different is that the metric in the original data space and in the transformed data space has been changed by the transformation. This is a different way of saying that distances are measured in different ways in the least-squares criteria for solving the two systems of equations. Since the least-squares solution minimizes distances it makes sense that the least-squares solution changes when the metric (or measuring unit) of the data space is changed. This implies that the least-squares solution is not as objective as it appeared at first sight, because arbitrary transformations of the system of equations lead to different least-squares solutions!

For the least-squares solution the generalized inverse is given by $\boldsymbol{A}^{-g} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$. One readily deduces that both for the original system and

the transformed system the resolution kernel is equal to the identity matrix: $\boldsymbol{R} = \boldsymbol{A}^{-g}\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{R}' = \boldsymbol{A}^{-g'}\boldsymbol{A}' = \boldsymbol{I}$. Hence both systems have perfect resolution! The reader may be interested to pause and explain how this can be reconciled with the fact that the estimated models in (27) are different.

The reason for this discrepancy lies in the error propagation term $\boldsymbol{A}^{-g}\boldsymbol{e}$ in (5). We know that errors must be present in the data used in the systems defined in expression (24) because the equations are inconsistent. After scaling the equations, data and errors are reconciled in different ways in the two systems of equations, so that different model estimators are obtained. *It is thus the presence of inconsistencies in the system of equations caused by errors that creates a dependence of the least-squares solution to arbitrary scaling operations.*

Let us now consider the properties of the least-squares solution under transformations of the data vector and the model vector in a more general way. The initial system of equations is given by

$$\boldsymbol{A}\boldsymbol{m} = \boldsymbol{d} \tag{28}$$

This expression is not quite correct because we ignored the errors $\boldsymbol{e}$ which will always be present. This is the reason why the above expression can not exactly be satisfied, and we will have to seek the least-squares solution to this system of equations. Let us consider a transformation of the model parameters through a transformation matrix $\boldsymbol{S}$:

$$\boldsymbol{m}' = \boldsymbol{S}\boldsymbol{m}\ , \tag{29}$$

and a transformation of the data vector with a transformation matrix $\boldsymbol{Q}$:

$$\boldsymbol{d}' = \boldsymbol{Q}\boldsymbol{d}\ . \tag{30}$$

Assume that $\boldsymbol{S}$ has an inverse, the transformed system of equations then is given by

$$\boldsymbol{Q}\boldsymbol{A}\boldsymbol{S}^{-1}\boldsymbol{m}' = \boldsymbol{Q}\boldsymbol{d} = \boldsymbol{d}'. \tag{31}$$

The original system of equations (28) has the least-squares solution

$$\boxed{\tilde{\boldsymbol{m}}^{(1)} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d}} \tag{32}$$

The solution of the transformed system of equations (31) follows from the same expression, setting $\boldsymbol{A}' = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{S}^{-1}$ and replacing $\boldsymbol{d}$ by $\boldsymbol{Q}\boldsymbol{d}$. This, however, gives the solution to the transformed model vector $\boldsymbol{m}'$. In order to compare this solution with the model estimate (32) we need to transform back to the original model space, using the relation $\boldsymbol{m} = \boldsymbol{S}^{-1}\boldsymbol{m}'$. The

least-squares solution $\tilde{\boldsymbol{m}}^{(2)}$ that follows from the transformed system is then given by:

$$\tilde{\boldsymbol{m}}^{(2)} = \boldsymbol{S}^{-1}\left(\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{d} \tag{33}$$

Assuming again that the appropriate inverses exist, this expression can be simplified by repeatedly applying the matrix identity $(\boldsymbol{N}\boldsymbol{M})^{-1} = \boldsymbol{M}^{-1}\boldsymbol{N}^{-1}$ to the term $\left(\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\boldsymbol{S}^{-1}\right)^{-1}$, giving $\left(\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\boldsymbol{S}^{-1}\right)^{-1} \left(\boldsymbol{S}^{-1}\right)^{-1}\left(\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\right)^{-1}\left(\boldsymbol{S}^{T-1}\right)^{-1} = \boldsymbol{S}\left(\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\right)^{-1}\boldsymbol{S}^T$. The least-squares solution of the transformed system can then be written as:

$$\boxed{\tilde{\boldsymbol{m}}^{(2)} = \left(\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{d}} \tag{34}$$

Comparing this with the least-squares solution $\tilde{\boldsymbol{m}}^{(1)}$ in expression (32) of the original system one finds that the least-squares solution is invariant for:

- Transformations of the model vector altogether.
- Transformations of the data vector if $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$.

The first property can be understood if we recall that the cost function in the least-squares problem does not minimize the model length but only the data misfit. The last property can be understood by comparing the quantities that are minimized in the original system and for the transformed system. For the original system one minimizes:

$$S = \|\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}\|^2 = ((\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}) \cdot (\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m})) \ , \tag{35}$$

while in the transformed system one minimizes

$$\begin{aligned} S' = \|\boldsymbol{Q}\boldsymbol{d} - \boldsymbol{Q}\boldsymbol{A}\boldsymbol{m}\|^2 &= (\boldsymbol{Q}(\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}) \cdot \boldsymbol{Q}(\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m})) \\ &= (\boldsymbol{Q}^T\boldsymbol{Q}(\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}) \cdot (\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m})) \end{aligned} \tag{36}$$

These two quantities are identical when the transformation $\boldsymbol{Q}$ is unitary, i.e. when $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$. This result stems from the property that unitary matrices do not affect the norm of a vector.

2.6. THE CONSISTENCY PROBLEM FOR THE MINIMUM-NORM SOLUTION

Consistency problems do not only arise for the least-squares solution, the minimum-norm solution suffers from the same problem. As an example

let us return to the underdetermined "system" of equations (14). The minimum-norm solution of this problem is given by

$$\boxed{\tilde{m}_1 = 1 \quad , \quad \tilde{m}_2 = 1} \tag{37}$$

Now carry out a transformation from the model vector $\boldsymbol{m}$ to a new model vector $\boldsymbol{m}'$:

$$\begin{aligned} m_1' &= m_1 + m_2 \\ m_2' &= m_2 \end{aligned} \tag{38}$$

For this new model vector the "system" of equations is given by:

$$\tilde{m}_1' = d = 2 \tag{39}$$

Note that this transformed model vector brings out the fact that the system is undetermined much more clearly than the original system (14) because the new system imposes no constraint whatsoever on the model parameter m_2'. The minimum-norm solution of the transformed equation (39) is given by $\tilde{m}_1' = 2$, $\tilde{m}_2' = 0$. With the transformation (38) this solution in the transformed model space corresponds to the following solution in the original model space:

$$\boxed{\tilde{m}_1 = 2 \quad , \quad \tilde{m}_2 = 0} \tag{40}$$

This solution is shown by the open square in figure 4. Note that this solution differs from the minimum-norm solution (37) for the original system of equations. The reason for this discrepancy is similar to the consistency problem for the least-squares problem in section 2.5; the transformation (38) has changed the metric of model space, so that distances in the original model space and the transformed model space are measured in different ways. For this reason the minimum norm solution of the original problem and transformed problem are different.

We could carry out a similar general analysis for the transformation properties of the minimum norm solution under general transformations of the model vector and data vector as we carried out for the least-squares solution in section 2.5. However, in practice one applies regularization to the equations. As shown in equation (23) the damped least-squares solution and the damped minimum-norm solution are identical. For this reason the general transformation properties are treated in the next section for the damped least-squares solution.

2.7. THE NEED FOR A MORE GENERAL REGULARIZATION

The analysis of the transformation properties of the damped least-squares is completely analogous to the analysis of the undamped least-squares solution of section 2.5. Ignoring errors for the moment, the linear system of equations is given by (28): $\boldsymbol{Am} = \boldsymbol{d}$, and the transformation of the model vector and data vector is given by (29) and (30) respectively: $\boldsymbol{m}' = \boldsymbol{Sm}$ and $\boldsymbol{d}' = \boldsymbol{Qd}$. Assuming again that $\boldsymbol{S}^{-1}$ exists, the transformed system of equations is given by (31): $\boldsymbol{QAS}^{-1}\boldsymbol{m}' = \boldsymbol{Qd}$.

The damped least squares solution of the original system is given by:

$$\boxed{\tilde{\boldsymbol{m}}^{(1)} = \left(\boldsymbol{A}^T\boldsymbol{A}+\gamma\boldsymbol{I}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d}} \tag{41}$$

Analogously to (34) the damped least-squares solution of the transformed equations is given by:

$$\tilde{\boldsymbol{m}}^{(2)} = \boldsymbol{S}^{-1}\left(\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{QAS}^{-1} + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{S}^{T-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Qd} \tag{42}$$

The damping parameter here is not necessarily equal to the damping parameter in the original damped least-squares solution, but for our purpose we do not need to make the distinction. Expression (42) can be simplified using the same steps as in the derivation of (34). Writing the term $\gamma\boldsymbol{I}$ as $\gamma\boldsymbol{I} = \gamma\boldsymbol{S}^{T-1}\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{S}^{-1}$, it follows that

$$\boxed{\tilde{\boldsymbol{m}}^{(2)} = \left(\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{QA}+\gamma\boldsymbol{S}^T\boldsymbol{S}\right)^{-1}\boldsymbol{A}^T\boldsymbol{Q}^T\boldsymbol{Qd}} \tag{43}$$

This expression points to a fundamental problem: the damping term in the model space $\boldsymbol{m}'$ is given by the identity matrix $\gamma\boldsymbol{I}$ (see (42)) and the damping term is $\gamma\boldsymbol{S}^T\boldsymbol{S}$ when expressed in terms of the original model vector $\boldsymbol{m}$ (see (43)). This implies that the damping $\gamma\boldsymbol{I}$ is not invariant for transformations of the model parameters. The terms $\boldsymbol{Q}^T\boldsymbol{Q}$ appear when a transformation of the data vector is carried out. This implies that the damped least-squares solution is in general not invariant under transformations of the data vector or model vector.

There is therefore a need for a more general regularization which allows to change model and data space in a consistent manner so that the solution is coordinate independent. Such a general regularization can be found from (43) by setting $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{W}_d$ and by defining $\boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{W}_m$. The general least-squares solution is then given by:

$$\tilde{\boldsymbol{m}} = \left(\boldsymbol{A}^T\boldsymbol{W}_d\boldsymbol{A}+\gamma\boldsymbol{W}_m\right)^{-1}\boldsymbol{A}^T\boldsymbol{W}_d\boldsymbol{d}\,. \tag{44}$$

This solution minimizes the following cost function:

$$S = (\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m})^T \boldsymbol{W}_d (\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}) + \gamma \boldsymbol{m}^T \boldsymbol{W}_m \boldsymbol{m} \tag{45}$$

This expression shows that in general the weight matrices $\boldsymbol{W}_d$ and $\boldsymbol{W}_m$ can be anything (as long as they are positive definite to ensure that S has minima). Written in this way, γ may be seen as a trade-off parameter which compromises between two characteristics of the model: its size and its disagreement with the data. Both independent properties of the model cannot be arbitrary small together, hence there is a need for a balance. The choice of an optimum γ, however, is not an easy question. We have shown explicitly that when you start with a simple damped least-squares solution you can transform that problem into a more generally regularized least-squares solution in a different coordinate system and vice versa.

This implies that there is no reason to favor the damped least-squares solution over the more general least-squares solution (44). In fact, most inverse problems are ill-posed (partly underdetermined and partly overdetermined) and ill-conditioned (small errors in the data causes large variations in the model) which goes hand in hand with large null-spaces and hence non-unique solutions. Regularization is thus needed, but there is a large ambiguity in its choice [*Scales and Snieder*, 1997]. This reflects the fundamental difficulty that one faces in solving inverse problems: solving the system of equations is a minor problem, compared to choosing the regularization.

One approach is to use Bayesian statistics where one treats the inverse problem from a statistical point of view combining a-priori information about the data and the model with the data that are actually measured [*Tarantola and Valette,* 1982a; *Tarantola and Valette,* 1982b]. The weight matrices reflect true physical a-priori information (in a statistical sense) that one has of the data and the model, *independent* of the measured data. This includes for example the statistical noise characteristics of the instrument that produced the data, as well as information of the model and data that follow from other arguments. (For example, the mass-density in the Earth must be positive.) In such a Bayesian approach the weight matrices are given by

$$\boldsymbol{W}_d = \boldsymbol{C}_d^{-1} \quad , \quad \gamma \boldsymbol{W}_m = \boldsymbol{C}_m^{-1} \, , \tag{46}$$

where $\boldsymbol{C}_d^{-1}$ and $\boldsymbol{C}_m^{-1}$ are the a-priori covariance matrices for the data and model respectively:

$$\boldsymbol{C}_d = \left\langle (\boldsymbol{d} - \langle \boldsymbol{d} \rangle) (\boldsymbol{d} - \langle \boldsymbol{d} \rangle)^T \right\rangle \, , \tag{47}$$

$$\boldsymbol{C}_m = \left\langle (\boldsymbol{m} - \langle \boldsymbol{m} \rangle) (\boldsymbol{m} - \langle \boldsymbol{m} \rangle)^T \right\rangle \, . \tag{48}$$

In these expressions the brackets $\langle \cdots \rangle$ denote the expectation value. In this interpretation the estimator (44) corresponds to the most likely a-posteriori model when the error distribution is Gaussian. The statistical basis of Bayesian inversion leads to an objective solution if one respects the rule that the a-priori information has a true physical meaning. In practice however, one should realize that the choice of the a-priori distribution of the data and model is very often subjective as well. The reader can find further details in the column "To Bayes or not to Bayes" of *Scales and Snieder* [1997].

A different approach is to define the misfit function in such a way that it favours models with given properties (small, smooth, ...) [*Parker*, 1994]. Choosing a-priori information then amounts to defining an appropriate norm in which the data misfit and any given property of the model are measured. In our case, the weight matrices would then define a particular metric for the L_2-norm. As an example of choosing the weight matrices $\boldsymbol{W}_m$ the use of Occam's inversion is quite common [*Constable et al.*, 1987] where one seeks the smoothest model that is consistent with the data. Instead of putting a constraint on the model length, one seeks the square of its gradient to be as small as possible, i.e. the last term in (45) is a discretization of $\|\nabla m\|^2 = \int (\nabla m \cdot \nabla m)\, dV = -\int m\, \nabla^2 m\, dV$,[5] and hence $\boldsymbol{W}_m$ corresponds to a discretized form of the Laplacian $-\nabla^2$.

2.8. THE TRANSFORMATION RULES FOR THE WEIGHT MATRICES

One of the fundamental requirements of an inverse solution should be that the results of the inversion are independent of arbitrary scalings applied to the model vector or data vector. Alas, this requirement is often ignored which can render comparisons of different models quite meaningless. For practical implications see *Trampert and Lévêque* [1990] and *Trampert et al.* [1992]. Here we derive how the weight matrices $\boldsymbol{W}_m$ and $\boldsymbol{W}_d$ should scale under such transformations for the least-squares solution to remain invariant.

Let us first consider the scaling (29) of the model vector: $\boldsymbol{m}' = \boldsymbol{S}\boldsymbol{m}$. Under this transformation the model term in the least-squares quantity (45) transforms as

$$\boldsymbol{m}^T \boldsymbol{W}_m \boldsymbol{m} = \boldsymbol{m}'^T \boldsymbol{S}^{T-1} \boldsymbol{W}_m \boldsymbol{S}^{-1} \boldsymbol{m}' = \boldsymbol{m}'^T \boldsymbol{W}'_m \boldsymbol{m}' \,, \tag{49}$$

with

[5]Note that we tacitly assumed in the last identity that there are no nonzero terms arising from the boundary conditions. A formal treatment based on Green's theorem allows for the incorporation of nonzero boundary terms.

$$\boldsymbol{W}_m' = \boldsymbol{S}^{T-1}\boldsymbol{W}_m\boldsymbol{S}^{-1} \,. \tag{50}$$

Under this transformation rule for the model weight matrix the least-squares criterion is unchanged, hence the least-squares solution is not changed when the model weight matrix $\boldsymbol{W}_m$ is transformed. It is of interest to note that this rule implies that for Bayesian inversions, where the weight matrix is the inverse of the a-priori model covariance matrix ($\gamma\boldsymbol{W}_m = \boldsymbol{C}_m^{-1}$), the covariance matrix should transform as

$$\boldsymbol{C}_m' = \boldsymbol{S}\boldsymbol{C}_m\boldsymbol{S}^T \tag{51}$$

One easily verifies from definition (48) that this is indeed the transformation rule for covariance operators.

Next let us consider how the transformation (30) for the data vector $\boldsymbol{d}' = \boldsymbol{Q}\boldsymbol{d}$ affects the transformation of the data weight matrix $\boldsymbol{W}_d$. The matrix $\boldsymbol{A}$ scales under this transformation as $\boldsymbol{A}' = \boldsymbol{Q}\boldsymbol{A}$. Under this transformation the data term in the least-squares quantity (45) transforms as

$$\begin{aligned} &(\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m})^T \boldsymbol{W}_d (\boldsymbol{d} - \boldsymbol{A}\boldsymbol{m}) \\ &= \left(\boldsymbol{d}' - \boldsymbol{A}'\boldsymbol{m}\right)^T \boldsymbol{Q}^{T-1}\boldsymbol{W}_d\boldsymbol{Q}^{-1} \left(\boldsymbol{d}' - \boldsymbol{A}'\boldsymbol{m}\right) \\ &= \left(\boldsymbol{d}' - \boldsymbol{A}'\boldsymbol{m}\right)^T \boldsymbol{W}_d' \left(\boldsymbol{d}' - \boldsymbol{A}'\boldsymbol{m}\right) \end{aligned} \tag{52}$$

with

$$\boldsymbol{W}_d' = \boldsymbol{Q}^{T-1}\boldsymbol{W}_d\boldsymbol{Q}^{-1} \,. \tag{53}$$

For a Bayesian inversion the data weight matrix is the inverse of the data covariance matrix ($\boldsymbol{W}_d = \boldsymbol{C}_d^{-1}$), so that for a Bayesian inversion $\boldsymbol{C}_d$ should transform as

$$\boldsymbol{C}_d' = \boldsymbol{Q}\boldsymbol{C}_d\boldsymbol{Q}^T \,. \tag{54}$$

Note again that this is the correct transformation rule for a covariance matrix defined in (47). This implies that the Bayesian viewpoint, where $\boldsymbol{W}_m$ and $\boldsymbol{W}_d$ are the inverses of the model and data covariance matrices, ensures that the solution is invariant under transformations of the model vector and/or the data vector.

Although we have derived in which way the weight matrices $\boldsymbol{W}_m$ and $\boldsymbol{W}_d$ should transform under transformations of model and data vectors, this does by no means imply that these matrices can be defined in a unambiguous way. An ill-posed and/or ill-conditioned inverse problem can only be solved if one is willing to control the solution by imposing a regularization term. In general, there is no unique recipe for choosing the

weight matrices $\boldsymbol{W}_m$ and $\boldsymbol{W}_d$. It is the subjective input of the user that determines the choice of these matrices.

2.9. SOLVING THE SYSTEM OF LINEAR EQUATIONS

It should be noted that the least-squares solution always requires solving a set of linear algebraic equations. For instance, equation (44) may be written as

$$\left(\boldsymbol{A}^T\boldsymbol{W}_d\boldsymbol{A}+\gamma\boldsymbol{W}_m\right)\tilde{\boldsymbol{m}} = \boldsymbol{A}^T\boldsymbol{W}_d\boldsymbol{d}\ . \tag{55}$$

This represents a square system of linear equations, the so-called normal equations, of the form $\boldsymbol{Bx} = \boldsymbol{y}$. If we are merely interested in the estimation part of the problem, $\boldsymbol{B}$ doesn't need to be inverted. If we are also interested in the appraisal part of the problem (and we always should), it must be realized that $\boldsymbol{B}$ needs to be inverted at the cost of additional computer time. Many standard subroutine packages are available and *Press et al.* [1989] give a good and practical introduction to the subject. The reader should realize, however, that the system $\boldsymbol{Bx} = \boldsymbol{y}$ may become quite large for realistic geophysical problems which makes it worthwhile to consider a specialized routine best suited to the nature of $\boldsymbol{B}$ (symmetric, banded, sparse, ...). The dimension of the set of normal equations is also worth considering. Remember that the matrix $\boldsymbol{A}$ has the dimension $(N \times M)$, where N is the number of data and M the number of model parameters. System (55) has the dimension of the model space, but using (22) we may obtain a strictly equivalent system of the dimension of the data space. Choosing the smallest dimension to write the normal equation can save quite some computer time. Most techniques for solving the set of algebraic equations directly work with the matrix $\boldsymbol{B}$ as a whole requiring sufficient memory space to hold the matrix. But in global travel time tomography, for instance, these dimensions may become extremely large ($N > 10^6$ and $M > 10^5$) so that iterative methods need to employed which only work on parts of $\boldsymbol{B}$ at a time. Another problem which frequently occurs is that even though regularization is included in $\boldsymbol{B}$, it is singular or numerically very close to singular. A powerful technique, called Singular Value Decomposition (SVD), can diagnose precisely what the problem is and will give a useful numerical answer. SVD is the most effective tool in inverse theory to understand why a certain result has been obtained.

Iterative methods or SVD need not to work on square systems and may thus directly use the matrix $\boldsymbol{A}$. In this context it is useful to realize that the generalized least squares solution (44) is equivalent to the simple least squares solution of the system

$$\begin{pmatrix} \boldsymbol{W}_d^{1/2}\boldsymbol{A} \\ \cdots \\ \sqrt{\gamma}\ \boldsymbol{W}_m^{1/2} \end{pmatrix} \boldsymbol{m} = \begin{pmatrix} \boldsymbol{W}_d^{1/2}\boldsymbol{d} \\ \cdots \\ 0 \end{pmatrix} . \tag{56}$$

For a discussion on the meaning of a square root of a positive definite matrix the reader is referred to *Tarantola* [1987]. Keeping also in mind a certain freedom in choosing weighting matrices (see 2.7), the user might want to define directly $\boldsymbol{W}^{1/2}$ rather than $\boldsymbol{W}$. Expression (56) shows that regularization has the effect of adding extra rows to the system of linear equations, but the enlarged system is still of the form $\boldsymbol{Bx} = \boldsymbol{y}$, where the matrix $\boldsymbol{A}$ and the data vector $\boldsymbol{d}$ are augmented by extra rows that account for the regularization. $\boldsymbol{B}$ is not square anymore as in the case of normal equations. We will now illustrate in more detail the essence of singular value decomposition and iterative techniques as applied to the system $\boldsymbol{Bx} = \boldsymbol{y}$.

2.9.1. *Singular value decomposition*

Singular value decomposition was developed by *Lanczos* [1961], this technique is a generalization of the eigenvector decomposition of matrices to non-square matrices. Let us first consider a real symmetric $N \times N$-square matrix $\boldsymbol{B}$ with eigenvectors $\hat{\boldsymbol{v}}^{(n)}$ and eigenvalues λ_n. For such a matrix the eigenvectors form an orthonormal set, hence any vector $\boldsymbol{x}$ can be projected on these eigenvectors: $\boldsymbol{x} = \sum_{n=1}^{N} \hat{\boldsymbol{v}}^{(n)} \left(\hat{\boldsymbol{v}}^{(n)} \cdot \boldsymbol{x}\right)$. When $\boldsymbol{B}$ acts on this expression the result can be written as:

$$\boldsymbol{Bx} = \sum_{n=1}^{N} \lambda_n \hat{\boldsymbol{v}}^{(n)} \left(\hat{\boldsymbol{v}}^{(n)} \cdot \boldsymbol{x}\right) = \boldsymbol{y} . \tag{57}$$

Decomposing the vector $\boldsymbol{y}$ using the same eigenvectors $\hat{\boldsymbol{v}}^{(n)}$ gives $\boldsymbol{y} = \sum_{n=1}^{N} \hat{\boldsymbol{v}}^{(n)} \left(\hat{\boldsymbol{v}}^{(n)} \cdot \boldsymbol{y}\right)$, and inserting this into expression (57) yields the following expansion for the solution vector $\boldsymbol{x}$:

$$\boldsymbol{x} = \sum_{n=1}^{N} \frac{1}{\lambda_n} \hat{\boldsymbol{v}}^{(n)} \left(\hat{\boldsymbol{v}}^{(n)} \cdot \boldsymbol{y}\right) . \tag{58}$$

It can be seen that small eigenvectors can lead to instabilities in the solution $\boldsymbol{x}$. Singular value decomposition generalizes this expansion to non-square matrices. Details of this technique are given by *Lanczos* [1961] and by *Aki and Richards* [1980].

Now consider the following non-square system of equations:

$$\underbrace{\boldsymbol{B}}_{\substack{N \times M \\ matrix}} \quad \underbrace{\boldsymbol{x}}_{\substack{M \\ rows}} = \underbrace{\boldsymbol{y}}_{\substack{N \\ rows}} \tag{59}$$

Singular value decomposition is based on an expansion of $\boldsymbol{x}$ in an orthonormal set of eigenvectors $\hat{\boldsymbol{v}}^{(n)}$ and of $\boldsymbol{y}$ in an orthonormal set $\hat{\boldsymbol{u}}^{(n)}$. These vectors cannot be the eigenvectors of $\boldsymbol{B}$ because this matrix is not square, hence it does not have any eigenvectors. Instead, these vectors are related by the following relation:

$$\boldsymbol{B}\hat{\boldsymbol{v}}^{(n)} = \lambda_n \hat{\boldsymbol{u}}^{(n)} \quad , \quad \boldsymbol{B}^T \hat{\boldsymbol{u}}^{(n)} = \lambda_n \hat{\boldsymbol{v}}^{(n)} \tag{60}$$

It can easily be seen that the vectors $\hat{\boldsymbol{v}}^{(n)}$are the eigenvectors of $\boldsymbol{B}^T\boldsymbol{B}$ while the vectors $\hat{\boldsymbol{u}}^{(n)}$ are the eigenvectors of $\boldsymbol{B}\boldsymbol{B}^T$, hence these vectors can readily be determined. $\boldsymbol{B}^T\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{B}^T$ share the same nonzero eigenvectors λ_n^2. The λ_n are called the singular values of $\boldsymbol{B}$. When $\boldsymbol{B}$ acts on the vector $\boldsymbol{x}$ the result can be written as

$$\boldsymbol{B}\boldsymbol{x} = \sum_{n=1}^{P} \lambda_n \hat{\boldsymbol{u}}^{(n)} \left(\hat{\boldsymbol{v}}^{(n)} \cdot \boldsymbol{x} \right) . \tag{61}$$

The upper limit of the summation is determined by the number of eigenvalues that are nonzero because the vanishing eigenvalues do not contribute to the sum. This number P can be significantly less than the dimension of the problem: $P \leq N$ and $P \leq M$.

It is convenient to arrange the vectors $\hat{\boldsymbol{u}}^{(n)}$ and $\hat{\boldsymbol{v}}^{(n)}$ as the columns of matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, the eigenvectors from index P onwards correspond to zero eigenvalues and need to be included to make $\boldsymbol{U}$ and $\boldsymbol{V}$ complete:

$$\boldsymbol{U} = \left(\underbrace{\begin{array}{cccccccc} \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \hat{\boldsymbol{u}}^{(1)} & \hat{\boldsymbol{u}}^{(2)} & \cdots & \hat{\boldsymbol{u}}^{(P)} & \hat{\boldsymbol{u}}^{(P+1)} & \cdots & \hat{\boldsymbol{u}}^{(N)} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \end{array}}_{\boldsymbol{U}_p \qquad\qquad\qquad \boldsymbol{U}_0} \right) , \tag{62}$$

$$\boldsymbol{V} = \left(\underbrace{\begin{array}{cccccc} \vdots & \vdots & & \vdots & \vdots & \\ \hat{\boldsymbol{v}}^{(1)} & \hat{\boldsymbol{v}}^{(2)} & \cdots & \hat{\boldsymbol{v}}^{(P)} & \hat{\boldsymbol{v}}^{(P+1)} & \cdots \; \hat{\boldsymbol{v}}^{(M)} \\ \vdots & \vdots & & \vdots & \vdots & \end{array}}_{\boldsymbol{V}_p \qquad\qquad\qquad \boldsymbol{V}_0} \right), \tag{63}$$

The *orthogonality* of the eigenvectors implies that $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$. The *completeness* of the eigenvectors implies that $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$ and $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{I}$. Since the orthogonality of the eigenvectors also holds in the subspaces spanned by $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ we have $\boldsymbol{U}_p^T\boldsymbol{U}_p = \boldsymbol{I}$ and $\boldsymbol{V}_p^T\boldsymbol{V}_p = \boldsymbol{I}$. However, the vectors in these subspaces do in general not form a complete set, so that in general $\boldsymbol{U}\boldsymbol{U}_p^T \neq \boldsymbol{I}$ and $\boldsymbol{V}\boldsymbol{V}_p^T \neq \boldsymbol{I}$

The generalization of (61) to non-square systems can be written as

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{U}_p & \boldsymbol{U}_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_p^T \\ \boldsymbol{V}_0^T \end{pmatrix}, \tag{64}$$

where the matrix $\boldsymbol{\Sigma}$ is given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \lambda_p \end{pmatrix} \tag{65}$$

It follows from (61) that the eigenvectors $\hat{\boldsymbol{u}}^{(n)}$ that correspond to a vanishing eigenvalue do not contribute when $\boldsymbol{B}$ acts on a vector. These eigenvectors are ordered in the sub-matrix $\boldsymbol{U}_0$.This is equivalent to the statement that according to the representation (64) the matrix $\boldsymbol{B}$ can be constructed from $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ alone. $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$ are dark spots of the space not illuminated by operator $\boldsymbol{B}$. Since $\boldsymbol{U}_0^T\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$ the predicted data $\boldsymbol{B}\boldsymbol{x}$ are orthogonal to the subspace spanned by $\boldsymbol{U}_0$, see figure 5. This means that any components in the data vector that lie in $\boldsymbol{U}_0$ cannot be explained by *any* model. These components of the data vector necessarily correspond to errors in the data or errors in the operator $\boldsymbol{B}$ as a description of the physical problem. It is for this reason that $\boldsymbol{U}_0$ is called the *data-null-space.* In a least squares inversion one aims at minimizing the data misfit. Minimizing the data misfit then amounts to finding a model that produces a data vector in the subspace $\boldsymbol{U}_p$ that is closest to the true data. It follows from figure 5 that this is achieved by simply projecting the components of $\boldsymbol{U}_0$ contained in the data out of the problem. This is exactly what is done

by limiting the sum over eigenvalues in (64) to the nonzero eigenvalues only. Of course, when $\boldsymbol{U}_0$ is empty, one can always find $\boldsymbol{x}$ which explains the data $\boldsymbol{y}$ exactly because $\boldsymbol{U}_p$ spans the complete data space.

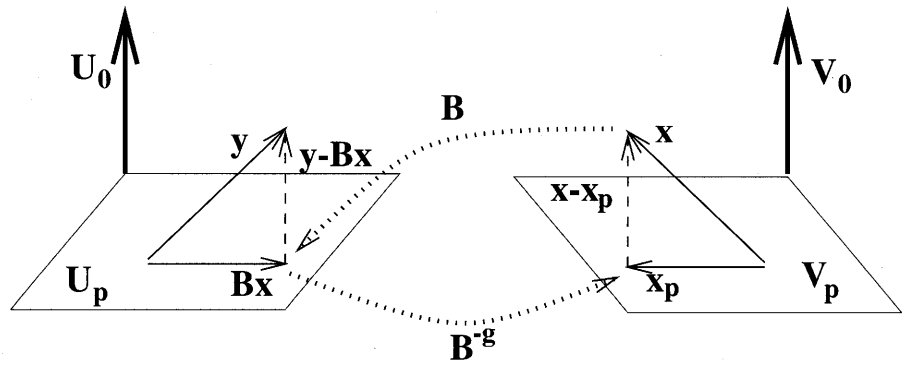


Figure 5. Geometrical interpretation of singular value decomposition. Note that starting from a model vector $\boldsymbol{x}$ and inverting the corresponding data $\boldsymbol{Bx}$ one obtains an estimated model vector $\boldsymbol{x}_p$ that in general differs from $\boldsymbol{x}$ because the resolution operator $\boldsymbol{R} = \boldsymbol{B}^{-g}\boldsymbol{B}$ in general is not equal to the identity matrix.

In a similar way, the restriction of the summation over eigenvalues to the nonzero eigenvalues has the effect that the model estimate lies in the subspace spanned by $\boldsymbol{V}_p$, but that the estimated model has no component in $\boldsymbol{V}_0$. Any component of the model in $\boldsymbol{V}_0$ does not affect the data because $\boldsymbol{BV}_0 = \boldsymbol{0}$. This means that $\boldsymbol{V}_0$ defines the *model-null-space.* The data have no bearing on the components of the model vector that lie in $\boldsymbol{V}_0$. Setting the component of the model vector in the model-null-space equal to zero then implies that in the model estimation one only takes the nonzero eigenvalues into account. Expanding $\boldsymbol{x}$ in the vectors $\hat{\boldsymbol{v}}^{(n)}$ and $\boldsymbol{y}$ in the vectors $\hat{\boldsymbol{u}}^{(n)}$ and taking only the nonzero eigenvalues into account one can thus generalize the solution (58) in the following way to non-square systems:

$$\boldsymbol{x} = \sum_{n=1}^{P} \frac{1}{\lambda_n} \hat{\boldsymbol{v}}^{(n)} \left(\hat{\boldsymbol{u}}^{(n)} \cdot \boldsymbol{y} \right) . \tag{66}$$

Using the matrices $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ this result can also be written as:

$$\boldsymbol{x} = \boldsymbol{V}_p \boldsymbol{\Sigma}^{-1} \boldsymbol{U}_p^T \boldsymbol{y} , \tag{67}$$

with $\boldsymbol{\Sigma}^{-1}$ given by

$$\boldsymbol{\Sigma}^{-1}=\begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_p \end{pmatrix} \tag{68}$$

Similar to the forward problem, the inverse problem is not a function of $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$. If both of these subspaces are zero, the operator $\boldsymbol{B}$ has an exact inverse. If $\boldsymbol{U}_0$ exists, one can show that the residual $\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}$ is perpendicular to $\boldsymbol{B}\boldsymbol{x}$ and hence the residual is minimum as for the least-squares solution. If $\boldsymbol{V}_0$ exists, solution (67) has no component in $\boldsymbol{V}_0$ and is therefore of minimum norm.

Clearly, small errors in $\boldsymbol{y}$ can lead to large errors in $\boldsymbol{x}$ when multiplied with $1/\lambda_n$ and the singular value is small. This process of error magnification can be controlled by limiting the summation in (66) to eigenvalues that differ significantly from zero. Alternatively, one can replace $1/\lambda_n$ by $\lambda_n/\left(\lambda_n^2+\gamma\right)$ with γ a positive constant. One can show that this is equivalent to the damped least-squares solution (20). See for example *Matsu'ura and Hirata* [1982] for a discussion on these different strategies. It should be noted that cutting off or damping small eigenvalues leads to different results. This makes it virtually impossible to quantitatively compare solutions of the same problem obtained with these fundamentally different strategies. One of the main reasons for the popularity of singular value decomposition is the control that one can exert over the error propagation in the solution. The drawback is that one needs to determine the eigenvectors of a matrix. For realistic large-scale problems ($P > 10^4$) this may require a prohibitive amount of CPU time. On the other hand, once the eigenvectors are calculated, resolution and error propagation are obtained at virtually no cost, since they merely involve the matrix multiplication $\boldsymbol{V}_p\boldsymbol{V}_p^T$.

2.9.2. *Iterative least-squares*

The least-squares solution of the system $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{y}$ is, as shown in (11), given by $\boldsymbol{x} = \left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T\boldsymbol{y}$. Note that the matrix $\boldsymbol{B}$ may contain some form of regularization as seen in expression (56). In practice, given the large-scale of many inverse problems, the matrix $\boldsymbol{B}^T\boldsymbol{B}$ may not fit in the computer memory. It is for this reason that one has developed iterative techniques which improve existing estimates of the solution.

Suppose one has in the n-th iteration of an iterative process a model estimate $\boldsymbol{x}_n$ and that one seeks an update $\delta\boldsymbol{x}_n$ such that the new model estimate $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \delta\boldsymbol{x}_n$ is a better estimate of the model. Inserting this

expression into the relation $\boldsymbol{Bx} = \boldsymbol{y}$ gives an expression for the model update:

$$\boldsymbol{B}\delta\boldsymbol{x}_n = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}_n) \tag{69}$$

Note that the right hand side of this expression is the residual of the model estimate $\boldsymbol{x}_n$, i.e. the difference $\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}_n$, is a measure to what extend the model estimate $\boldsymbol{x}_n$ does not explain the data. Expression (69) prescribes how the model should be updated in order to reduce the data residual. The least squares-solution of this expression is given by

$$\delta\boldsymbol{x}_n = \left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}_n) \tag{70}$$

However, we have not gained anything yet because this expression is just as difficult to solve as the original equation and we still need to deal with the inverse of $\boldsymbol{B}^T\boldsymbol{B}$.

The advantage of solving the problem iteratively is that in such an approach one can replace the inverse $\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}$ by a suitably chosen estimate $\boldsymbol{P}$ of this inverse, i.e. one computes a model update using the following expression:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{P}\boldsymbol{B}^T(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}_n) \tag{71}$$

The operator $\boldsymbol{P}$ is called a preconditioning operator. If one sets $\boldsymbol{P} = \left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}$ one retrieves the full solution in one step, but in that case one needs to compute $\left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}$, which is what we wanted to avoid. Recognizing that $\boldsymbol{B}^T(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{x}_n)$ is the direction of descent of the cost function at $\boldsymbol{x}_n$, one may, on the other end of the spectrum, choose $\boldsymbol{P} = c\boldsymbol{I}$, where c is a constant derived from a least-squares criterion to ensure the steepest possible descent [e.g. *Tarantola*, 1984]. In practice one has to find a balance between using an advanced preconditioning operator (that may be difficult to compute but that leads to a solution with a few iterations), and a simple preconditioning operator (that is easy to compute but that may require many iterations). The most commonly used algorithms in geophysics are SIRT (Simultaneous Iterative Reconstruction Technique) and LSQR (Least-Squares Conjugate Gradient). SIRT has the drawback of introducing implicit regularization into the solution [*van der Sluis and van der Vorst*, 1987] and if corrected for significantly decreases the convergence speed [*Trampert and Lévêque*, 1990]. A more successful balance is achieved by the LSQR algorithm [*Paige and Saunders*, 1982a, 1982b; *van der Sluis and van der Vorst*, 1987]. An iterative scheme that mimics the properties of SVD is given by *Nolet and Snieder* [1990].

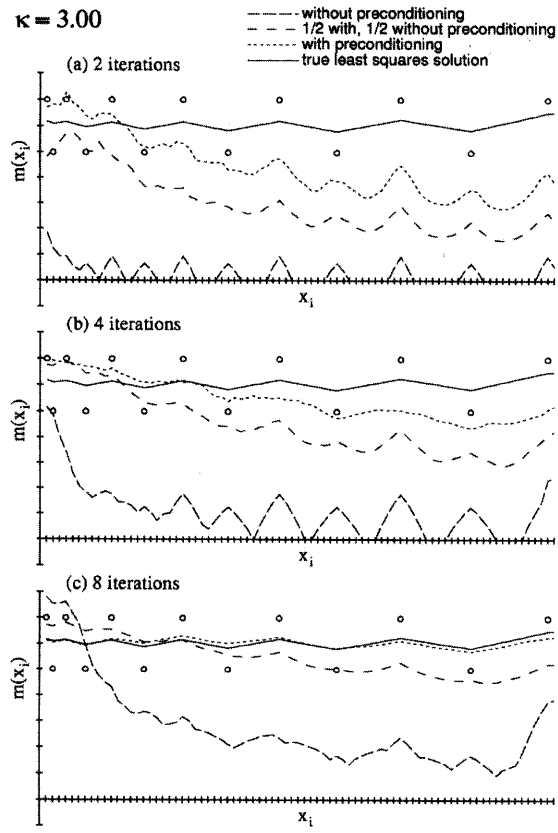


Figure 6. One-dimensional example of the convergence of a conjugate-gradient inversion at (a) iteration 2, (b) iteration 4 and (c) iteration 8. The solid line represents the true least-squares solution. The long-dashed lines are the interactive solutions without preconditioning while the dotted solutions are obtained using preconditioning.

Note that in the iterative least-squares algorithm (71) there is no need to invert a matrix, one only needs to multiply with the matrices $\boldsymbol{P}$ and $\boldsymbol{B}^T$, which can be done row by row. In many practical problems such as seismic tomography, the matrix $\boldsymbol{B}$ is very sparse which means that most of the matrix elements are zero, see section 6.1. For such a matrix the iterative least-squares algorithm is particularly efficient because the matrix multiplications only works on the nonzero matrix elements. The implementation of this idea in seismic tomography was developed by *Clayton and Comer*, [1983] and by *Nolet* [1985].

Despite the efficiency of the iterative least-squares problems one should

be aware that the convergence of this process may be very slow. An example from *VanDecar and Snieder* [1994] is shown in figure 6. In this example one fits a function whose values are given by the dots in figure 6 by a model defined as the samples of that function at the x-values indicated by the tick marks along the horizontal axis. This problem obviously is ill-posed because the function is not defined between the data points. In order to define a solution, Occam's method is used where $\boldsymbol{W}_m$ in (56) is a discretized version of the gradient operator. The true regularized least-squares solution of this problem is shown by the solid line. The iterative least-squares solution based on the conjugate-gradient algorithm after 2, 4 and 8 iterations is shown by the long-dashed line in the panels of figure 6.

It is clear that the convergence of the iterative least-squares algorithm is slow. The reason for this is easy to see. At every iteration, every point in the model interacts only with it's neighbouring points through the damping term $\boldsymbol{W}_m$ in the matrix. This means that a model element can only interact with a model element n positions further down the line after n iterations. In other words, the damping term imposes a global smoothness constraint on the solution, but in each iteration the model elements interact only locally. For this reason an inordinate number of iterations are required. *VanDecar and Snieder* [1994] developed a preconditioning operator that leads to a very rapid convergence, this is indicated by the dotted lines in figure 6. Note that this iterative solution has superior convergence properties.

A general drawback of all iterative techniques is that most of its advantages are lost if one is interested in the appraisal part of the problem since this involves the knowledge of $(\boldsymbol{B}^T\boldsymbol{B})^{-1}$, which an iterative technique doesn't explicitly compute. A way around this is to solve the system M times, each time replacing the data by a column of $\boldsymbol{B}$ for the resolution matrix for instance [*Trampert and Lévêque*, 1990].

3. Linear inverse problems with continuous models

Up to this point, the model vector was finite dimensional. In many inverse problems the model is a continuous function of the space coordinates; it therefore has infinitely many degrees of freedom. For example, in inverse problems using gravity data one wants to determine the mass density $\rho(\boldsymbol{r})$ in the earth. This is in general a continuous function of the space coordinates. In realistic inverse problems one has a finite amount of data. A simple variable count shows that it is impossible to determine a continuous model with infinitely many degrees of freedom from a finite amount of data in a unique way.

In order to make the problem manageable we will restrict ourselves in this section to linear inverse problems. For such problems the data and the model are related by

$$d_i = \int G_i(x) m(x) dx + e_i \ . \tag{72}$$

The notation in this section is one-dimensional, but the theory can be used without modification in higher dimensions as well. The model is a continuous model but the data vector is discrete and in practice has a finite dimension. The kernel $G_i(x)$ plays the same role as the matrix $\boldsymbol{A}$ in (1). Note that the data are contaminated with errors e_i.

Since the forward problem is linear, the estimated model is obtained by making a linear combination of the data (this is the most general description of a linear estimator):

$$\tilde{m}(x) = \sum_i a_i(x) d_i \ . \tag{73}$$

The coefficients $a_i(x)$ completely specify the linear inverse operator. By inserting (72) in (73) one arrives again at a relation between the true model $m(x)$ and the estimated model $\tilde{m}(x)$:

$$\tilde{m}(x) = \underbrace{\int R(x,x') m(x') dx'}_{\text{Finite resolution}} + \underbrace{\sum_i a_i(x) e_i}_{\text{Error propagation}} \ , \tag{74}$$

with the resolution kernel R given by

$$R(x,x') = \sum_i a_i(x) G_i(x') \ . \tag{75}$$

The first term in (74) accounts for averaging that takes places in the mapping from the true model to the estimated model. It specifies through the resolution kernel $R(x,x')$ what spatial resolution can be attained. In the ideal case the resolution or averaging kernel is a delta function: $R(x,x') = \delta(x-x')$. The resolution kernel, however, is a superposition of a finite amount of data kernels $G_i(x')$. These data kernels are in general continuous functions, and since a delta function cannot be constructed from the superposition of a finite number of continuous functions, the resolution kernel will differ from a delta function. In Backus-Gilbert theory [*Backus and Gilbert,* 1967; 1968] one seeks the coefficients $a_i(x)$ in such a way that the resolution kernel resembles a delta function as well as possible given a certain criterion which measures this resemblance.

The second term in (74) accounts for error propagation. A proper treatment of this term needs to be based on statistics. It is shown by *Backus and Gilbert* [1970] that one cannot simultaneously optimize the resolution and suppress the error propagation and that one has to seek a trade-off between finite resolution and error propagation. The above mentioned work tries to explain the data exactly, even if they contain errors. *Gilbert* [1971] extended the theory to explain data within their error bars only.

As mentionned above, the Backus-Gilbert strategy finds the coefficients $a_i(x)$ in equation (73) by imposing a condition on the averaging kernel. *Taratola and Valette* [1982] solve the problem in a Baysian framework. They introduce prior information on the data (Gaussian error distribution) and prior assumptions (also Gaussian) on the unknown function $m(x)$ which also yields the coefficients $a_i(x)$. In this approach, the resolution or averaging kernel is a consequence of the *a priori* information, but generally different from a delta function. They further show that the Backus-Gilbert approach is contained in Tarantola-Valette solution.

In summary, in both strategies the infinite dimensional problem is transformed into a finite dimensional problem by seeking local averages of the true model.

3.1. CONTINUOUS MODELS AND BASIS FUNCTIONS

Another approach is to evoke basis functions which amounts to changing the parameterization of the model. In general, any continuous model can be written as a superposition of a complete set of basis functions:

$$m(x) = \sum_{j=1}^{\infty} m_j B_j(x) \; . \tag{76}$$

In many global geophysical applications spherical harmonics are used to represent the seismic velocity or the density inside the earth, since they form a natural basis to describe a function on the sphere. In that case the $B_j(x)$ are the spherical harmonics and the sum over j stands for a sum over degree l and angular order m. The advantage of such an expansion is that one now deals with a discrete vector m_j of expansion coefficients rather than with a continuous function. However, the basis functions $B_j(x)$ only form a complete set when the sums is over infinitely many functions, and hence the problem has shifted from dealing with a continuous function to dealing with a vector with infinitely many components.

Inserting the expansion (76) into the forward problem (72) we may write

$$d_i = \sum_{j=1}^{\infty} A_{ij} m_j + e_i \ , \tag{77}$$

where the matrix elements A_{ij} are the projection of the data kernels onto the basis functions:

$$A_{ij} = \int G_i(x) B_j(x) dx \tag{78}$$

In practice, one cannot deal with infinitely many coefficients, and it is customary to ignore the fact that a model vector has infinitely many dimensions by taking only the first L basis functions into account. The resulting L-dimensional model vector will be denoted by $\boldsymbol{m}_L$, and a similar notation $\boldsymbol{A}_L$ is used for the first L rows of the matrix $\boldsymbol{A}$. The solution of the resulting finite dimensional model vector can be found using any technique shown in section 2. Recall that a general least-squares solution may be found by minimizing

$$S_L = (\boldsymbol{d} - \boldsymbol{A}_L \boldsymbol{m}_L)^T \boldsymbol{C}_d^{-1} (\boldsymbol{d} - \boldsymbol{A}_L \boldsymbol{m}_L) + \boldsymbol{m}_L^T \boldsymbol{C}_{mL}^{-1} \boldsymbol{m}_L \tag{79}$$

as a function of the truncated model vector $\boldsymbol{m}_L$. The weighting operators $\boldsymbol{C}_d$ and $\boldsymbol{C}_m$ may or may not be given the interpretation of covariance operators (see section 2.7). The resulting model estimate is then given by

$$\tilde{\boldsymbol{m}}_L = \left(\boldsymbol{A}_L^T \boldsymbol{C}_d^{-1} \boldsymbol{A}_L + \boldsymbol{C}_{mL}^{-1} \right)^{-1} \boldsymbol{A}_L^T \boldsymbol{C}_d^{-1} \boldsymbol{d} \tag{80}$$

3.2. SPECTRAL LEAKAGE, THE PROBLEM

Truncating the model vector after the first L elements may appear to be a convenient way to reduce the problem to a finite dimensional one. However, there are problems associated with this truncation. Analogously to (2) any linear estimator of the first L coefficients of the infinite vector $\boldsymbol{m}$ can be written as:

$$\tilde{\boldsymbol{m}}_L = \boldsymbol{A}_L^{-g} \boldsymbol{d} \ . \tag{81}$$

Let us now divide the sum over model elements in (77) as a sum over the first L elements that we are interested in and the remaining model elements:

$$\boldsymbol{d} = \boldsymbol{A}_L \boldsymbol{m}_L + \boldsymbol{A}_\infty \boldsymbol{m}_\infty + \boldsymbol{e} \ . \tag{82}$$

In this expression $\boldsymbol{m}_\infty$ denotes the infinitely dimensional vector with elements $(m_{L+1}, m_{L+2}, \ldots)$. Inserting this expression in (81) yields the relation between the estimated L model coefficients $\tilde{\boldsymbol{m}}_L$ and the true model $\boldsymbol{m}$ (for a discussion on the objectiveness of the concept of true model, see section 2.5):

$$\tilde{\boldsymbol{m}}_L = \boldsymbol{m}_L + \underbrace{\left(\boldsymbol{A}_L^{-g}\boldsymbol{A}_L - \boldsymbol{I}\right)\boldsymbol{m}_L}_{\text{Limited Resolution}} + \underbrace{\boldsymbol{A}_L^{-g}\boldsymbol{A}_\infty\boldsymbol{m}_\infty}_{\text{Spectral leakage}} + \underbrace{\boldsymbol{A}_L^{-g}\boldsymbol{e}}_{\text{Error propagation}} \tag{83}$$

The last three terms in this expression account for deviations in the estimated model from the true model. The second term and the last term are identical to the corresponding terms in expression (5) for finite-dimensional problems, accounting for finite resolution within the components of the vector $\boldsymbol{m}_L$ and error propagation respectively. The term $\boldsymbol{A}_L^{-g}\boldsymbol{A}_\infty\boldsymbol{m}_\infty$ has no counterpart in (5) and is responsible for a spurious mapping of the model coefficients $(m_{L+1}, m_{L+2}, \ldots)$ onto the estimated model coefficients $(\tilde{m}_1, \ldots, \tilde{m}_L)$. Since the basis function expansion in many problems is that of a spectral expansion, the mapping from the coefficients $\boldsymbol{m}_\infty$ onto the first L coefficients $\tilde{\boldsymbol{m}}_L$ will be referred to as *spectral leakage* [*Snieder et al.*, 1991].

An example of spectral leakage is shown in the figures 7-10. Suppose a function on a line, defined in the interval $0 \leq x < 1$, is expanded in normalized sines and cosines which have at most three wavelengths in the interval. This means that in this example $L = 7$. The function is sampled at given locations and the sampling is twice as dense on the subintervals $0 \leq x < 0.25$ and $0.5 \leq x < 0.75$ than on the remaining part of the line. The inverse problem consists in determining the expansion coefficients in the finite set of basis functions on the line given the sampled data points. Details of this example may be found in *Snieder* [1993].

In figure 7 the sampled function is a sine wave with exactly three wavelengths in the given interval. The sampling points are indicated by circles. The reconstructed function is given by the solid line and is indistinguishable from the true function. In figure 8 the input function is a sine wave with six wavelengths on the interval, this input function is indicated by the dashed line. Because of the orthogonality of the trigonometric functions, one would expect this function to have no projection onto the seven basis functions that are used in the inversion. Nevertheless, the reconstructed function shown by the solid line in figure 8 differs significantly from zero; it has about 50% of the magnitude of the input functions. As shown in *Snieder* [1993] the expansion coefficients have errors of about 100%! This difference between the estimated model and zero is entirely due to spectral leakage because the first L model components $\boldsymbol{m}_L$ are equal to zero and

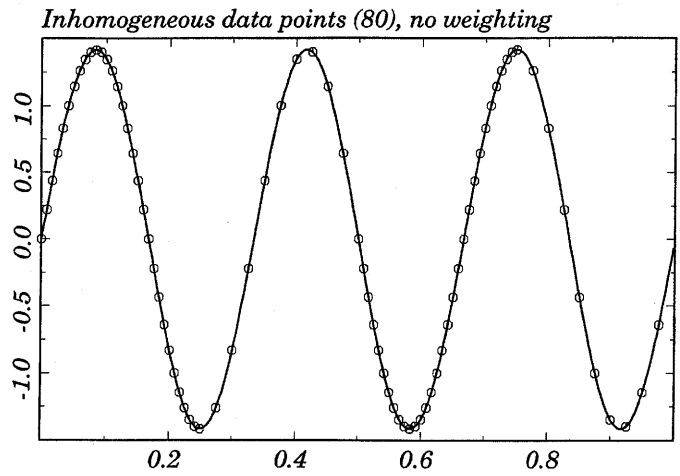


Figure 7. Unweighted least-squares fit of 80 data points (circles) sampling a sine wave with three wavelength along the interval. The estimated model is shown by the thick solid line, and is undistinguishable from the true projection on the basis functions.

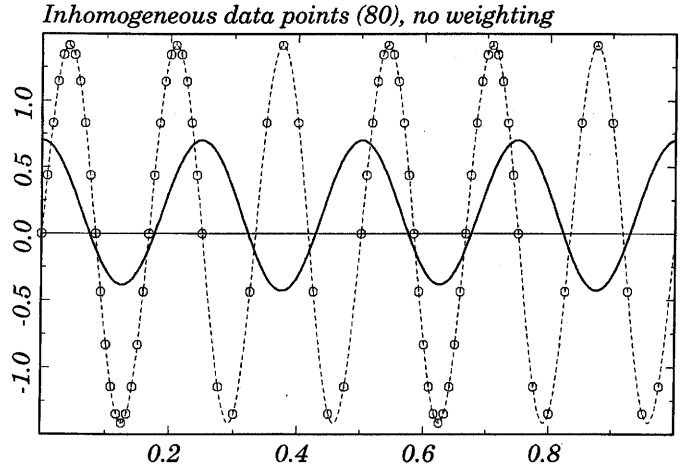


Figure 8. Unweighted least-squares fit of 80 data points (circles) sampling a sine wave with six periods (dashed line). The estimated model is shown by the thick solid line, the true projection on the basis functions is shown by the thin solid line.

there are no data errors in this example so that only the term $\boldsymbol{A}_L^{-g}\boldsymbol{A}_\infty\boldsymbol{m}_\infty$

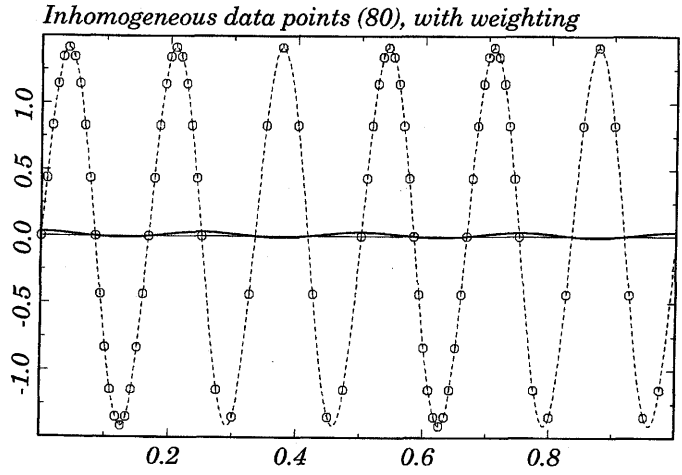


Figure 9. As the previous figure but for an inversion where each data point is with the inverse of the distance to neighbouring points.

in (83) can give rise to errors in the model reconstruction.

An interesting observation is that in the interval $0 \leq x < 0.25$ where the sampling is dense the input model and the reconstructed model are in phase whereas on the interval $0.25 \leq x < 0.5$ where the sampling is sparser the input model and reconstructed model are out off phase. This means that the least-squares criterion has selected a good fit in the densely sampled parts of the interval at the expense of a poorer fit in the sparsely sampled parts of the interval. This suggests a cure by down-weighting the contribution of the points in the densely sampled parts of the interval. Figure 9 shows the reconstructed function when the data points are weighted with weights that are inversely proportional to the sampling distance [*Snieder*, 1993], it can be seen that the reconstructed model indicated by the solid line is close to its true value (which is equal to zero). Weighting the data is the key to suppressing spectral leakage, we will return to this issue in section 3.3. The reason for the spectral leakage is that the orthogonality relation of the basis functions is weighted by the data kernels (i.e. $\sum_i \int G_i(x)B_j(x)dx \int G_i(y)B_k(y)dy$) and that the basis functions are not orthogonal for this inner product, i.e. this quantity is not equal to δ_{jk}.

Let us momentarily return to the unweighted inversion of figure 8. In that example 80 data points have been used. Note that the input function is not undersampled in any part of the interval, in other words there is

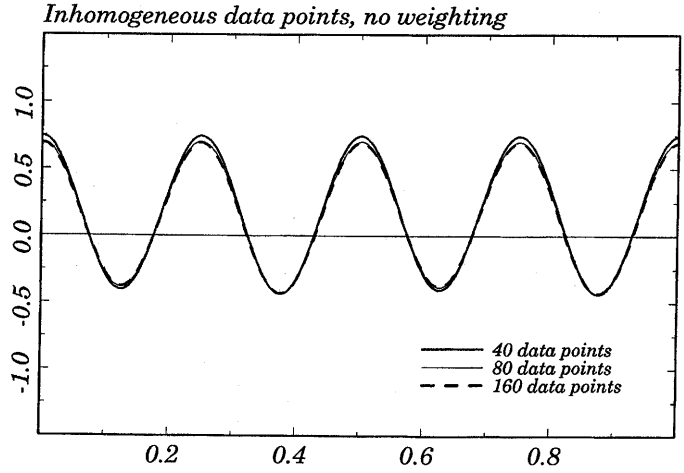


Figure 10. The estimated model when the true model is a sine wave with six wavelengths along the interval for inversions with different numbers of sampling points with the same sampling density along the line.

no aliasing [*Claerbout*, 1976] occurring in this problem. In figure 10 the reconstructed function is shown when 40, 80 or 160 data points have been used in the inversion. In all examples the density of data points was the same as in figure 9. Spectral leakage is not a problem due to the number of data points. Adding more data points while keeping the sampling density constant does not help to suppress spectral leakage.

Spectral leakage is a fundamentally different issue than aliasing!

In some studies, the reliability of estimated models has been studied by dividing the data set randomly in two parts, keeping thus the sampling density constant, and comparing the models reconstructed from the two halved data sets [e.g. *Woodhouse and Dziewonski*, 1984]. The fact that spectral leakage does not reduce when more data points are used implies that this test does not detect artifacts due to the variability in the data coverage, and therefore may give an overly optimistic impression of the reliability of the estimated model.

3.3. SPECTRAL LEAKAGE, THE CURE

The method presented in this section for suppressing spectral leakage was developed by *Trampert and Snieder* [1996]. A key element in the analysis is to acknowledge that more than the first L basisfunctions are needed to

describe the model. This is explicitly done by minimizing

$$S = (\boldsymbol{d} - \boldsymbol{A}_L\boldsymbol{m}_L - \boldsymbol{A}_\infty\boldsymbol{m}_\infty)^T \boldsymbol{C}_d^{-1} (\boldsymbol{d} - \boldsymbol{A}_L\boldsymbol{m}_L - \boldsymbol{A}_\infty\boldsymbol{m}_\infty) + \boldsymbol{m}_L^T\boldsymbol{C}_{mL}^{-1}\boldsymbol{m}_L + \boldsymbol{m}_\infty^T\boldsymbol{C}_{m\infty}^{-1}\boldsymbol{m}_\infty \tag{84}$$

The first term accounts for the minimization of the data misfit. Both, the first L basis functions contribute to the data misfit with $\boldsymbol{A}_L\boldsymbol{m}_L$, as well as the remaining basis functions through the term $\boldsymbol{A}_\infty\boldsymbol{m}_\infty$. The regularization terms $\boldsymbol{m}_L^T\boldsymbol{C}_{mL}^{-1}\boldsymbol{m}_L$ and $\boldsymbol{m}_\infty^T\boldsymbol{C}_{m\infty}^{-1}\boldsymbol{m}_\infty$ play a crucial role because they control to what extent the data misfit is distributed over the first L basisfunctions and over the remaining basis functions.

One has to minimize expression (84) both with respect to $\boldsymbol{m}_L$ and $\boldsymbol{m}_\infty$. Setting the derivatives $\partial S/\partial \boldsymbol{m}_L$ and $\partial S/\partial \boldsymbol{m}_\infty$ both equal to zero and eliminating $\boldsymbol{m}_\infty$ from these equations leads to the following estimate of the vector $\boldsymbol{m}_L$:

$$\tilde{\boldsymbol{m}}_L = \left(\boldsymbol{A}_L^T\boldsymbol{C}_{dL}^{-1}\boldsymbol{A}_L + \boldsymbol{C}_{mL}^{-1}\right)^{-1} \boldsymbol{A}_L^T\boldsymbol{C}_{dL}^{-1}\boldsymbol{d}\,, \tag{85}$$

where the new operator $\boldsymbol{C}_{dL}$ is given by

$$\boldsymbol{C}_{dL} = \boldsymbol{C}_d + \boldsymbol{A}_\infty\boldsymbol{C}_{m\infty}\boldsymbol{A}_\infty^T\,. \tag{86}$$

Expression (85) is very similar to the model estimate (80) obtained by simply ignoring all basis functions beyond degree L. The only difference is that the data weighting operator is given by (86) rather than by $\boldsymbol{C}_d$, but this difference is essential. The data weighting operator is positive definite. This implies that the new data weighting operator $\boldsymbol{C}_{dL}$ is always larger than the old operator $\boldsymbol{C}_d$. The new operator depends on the kernels $\boldsymbol{A}_\infty$ of the basisfunctions that are not inverted for, as well as the model weighting operator $\boldsymbol{C}_{m\infty}$ of these basis functions. From a Bayesian point of view these facts are all related by the observation that in this approach the data are partly explained by the infinite model vector $\boldsymbol{m}_\infty$. In the inversion for $\boldsymbol{m}_L$ only, the parts of the data which may be explained by $\boldsymbol{m}_\infty$ should be considered as noise, because they allow a variation in the data that is not cause by $\boldsymbol{m}_L$. For this reason this approach leads to an increase of the data variance $\boldsymbol{C}_{dL}$.

Although (86) looks simple, it is not trivial to evaluate because the product $\boldsymbol{A}_\infty\boldsymbol{C}_{m\infty}\boldsymbol{A}_\infty^T$ contains matrices of infinite dimensions. When the model weighting matrix $\boldsymbol{C}_{m\infty}$ is chosen to be proportional to the identity matrix ($\boldsymbol{C}_{m\infty} = \gamma\boldsymbol{I}$) there is a simple alternative because one only needs to evaluate the product $\boldsymbol{A}_\infty\boldsymbol{A}_\infty^T$. The ij-element of $\boldsymbol{A}_\infty\boldsymbol{A}_\infty^T$ can be written as $\left(\boldsymbol{A}_\infty\boldsymbol{A}_\infty^T\right)_{ij} = \sum_{k=L+1}^{\infty} A_{ik}A_{jk}$. Using expression (78) for the matrix

elements and interchanging the summation and the integration one can show that

$$\left(\boldsymbol{A}_\infty \boldsymbol{A}_\infty^T\right)_{ij} = \int dx \int dy\; G_i(x) G_j(y) \sum_{k=L+1}^{\infty} B_k(x) B_k(y)\;. \tag{87}$$

The sum over k can be written as a sum over all k-values minus the sum over the first L values:

$$\sum_{k=L+1}^{\infty} B_k(x) B_k(y) = \sum_{k=1}^{\infty} B_k(x) B_k(y) - \sum_{k=1}^{L} B_k(x) B_k(y)\;. \tag{88}$$

The basis functions form a complete set. Because of the closure relation $\sum_{k=1}^{\infty} B_k(x) B_k(y) = \delta\,(x-y)$ [e.g. p. 157 of *Merzbacher*, 1970] this can be written as:

$$\sum_{k=L+1}^{\infty} B_k(x) B_k(y) = \delta\,(x-y) - \sum_{k=1}^{L} B_k(x) B_k(y)\;. \tag{89}$$

Inserting this in (87) leaves after carrying out the y-integration in the first term and using (78) for the second term:

$$\left(\boldsymbol{A}_\infty \boldsymbol{A}_\infty^T\right)_{ij} = \boldsymbol{\Gamma}_{ij} - \left(\boldsymbol{A}_L \boldsymbol{A}_L^T\right)_{ij}\;, \tag{90}$$

where the Gram matrix $\boldsymbol{\Gamma}$ of our inverse problem is given by

$$\Gamma_{ij} = \int G_i(x) G_j(x) dx\;. \tag{91}$$

The first term in (90) can be computed by direct integration, whilst the second term now only involves multiplication of matrices of *finite* dimension. It should, however, be pointed out that not all problems in geophysics possess a well defined Gram matrix.

3.4. SPECTRAL LEAKAGE AND GLOBAL TOMOGRAPHY

The burning question is of course if spectral leakage is a serious problem in geophysics. We believe that the answer to this question is yes and illustrate this with a surface wave tomography problem. A surface wave of a given frequency sees an average structure of the outer-most layer of the Earth. The penetration of the wave depends on its frequency. The mapping problem concerns only a two-dimensional quantity (phase velocity as function of latitude and longitude) and is thus easier to represent than a

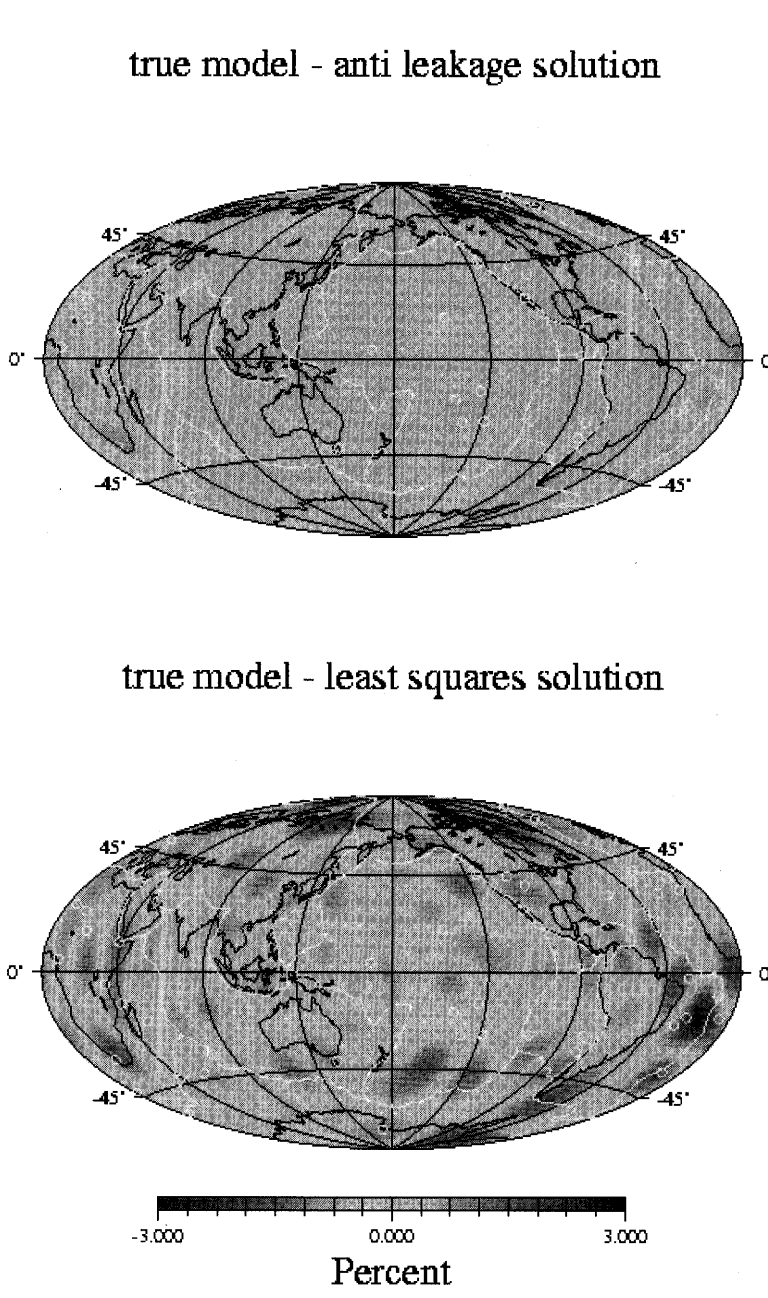


Figure 11. Example of phase velocity mapping with inhomogeneous sampling. The true model represents up to 6% perturbations with respect to a spherically average reference model. Plotted is the difference between true and estimated model: simple least-squares inversion (bottom) and anti-leakage inversion (top).

three-dimensional problem. To quantify the effect of leakage, we need to know the true answer to the problem.

We designed a synthetic experiment where we used a station-event distribution based on the global seismicity from 1989 together with the available global seismic stations. We have chosen arbitrarily 3000 paths giving a dense but uneven coverage (for details see *Trampert and Snieder*, [1996]). Next we took a Rayleigh wave phase velocity model [*Trampert and Woodhouse*, 1995] for a period of 80 seconds expressed in terms of spherical harmonics up to degree and order 40 and computed synthetic data for our chosen ray geometry. We added 10% random noise to the data and performed a classical least squares inversion up to degree and order 12. In figure 11 (bottom panel) we show the difference between the true degree 12 input model and the recovered model. This signal is far from zero and reaches in most places on the globe half the amplitude of the true model. If we applied the anti-leakage weighting defined in expression(86) the amplitudes of the artefacts are significantly reduced, see the top panel of figure 11. The anti-leakage operator does not completely prevent the bias because of the errors which we introduced in the synthetic data. To understand this, recall from equation (86) that the description of data errors and the anti-leakage operator add up linearly and thus any imperfect description of data errors will influence the anti-leakage. This suggests that our current tomographic models are likely to carry a bias from small-scale structures that are not accounted for by our current parameterizations.

4. The single scattering approximation and linearized waveform inversion

In the previous sections, the inverse problem was described when the forward problem was linear. Unfortunately, most wave propagation problems are nonlinear in the sense that the relation between the wave field and the medium is nonlinear. However, in many practical situations this relation can be linearized, notably with the single-scattering approximation (this section), with Fermat's theorem (section 6) and Rayleigh's principle (section 5).

4.1. THE BORN APPROXIMATION

Many wave propagation problems can symbolically be written in the form

$$Lu = F \tag{92}$$

In this expression u is the wave field, F accounts for the source that excites the waves and L is a differential operator that describes the wave

propagation. For example, for acoustic waves in a medium with constant density (92) is given by

$$\left(\nabla^2 + \frac{\omega^2}{c^2(\boldsymbol{r})}\right) u(\boldsymbol{r}) = F(\boldsymbol{r}) \ , \tag{93}$$

so that $L = \nabla^2 + \omega^2/c^2(\boldsymbol{r})$. Often, the operator L can be decomposed into an operator L_0 that correspond to a reference medium for which we can solve the problem easily plus a small perturbation εL_1 that accounts for a perturbation of the medium:

$$L = L_0 + \varepsilon L_1 \ . \tag{94}$$

The small parameter ε has been introduced to facilitate a systematic perturbation approach. As an example, the operator L in (93) can be decomposed using

$$\frac{1}{c^2(\boldsymbol{r})} = \frac{1}{c_0^2}\left(1 + \varepsilon n(\boldsymbol{r})\right) \ , \tag{95}$$

where c_0 is a constant velocity that described the mean properties of the medium well while $n(\boldsymbol{r})$ accounts for the velocity perturbations. Using this decomposition one obtains $L_1 = n(\boldsymbol{r})\omega^2/c_0^2$.

The wavefield is perturbed by the perturbation of the medium, this perturbation can be written as a regular perturbation series:

$$u = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots \tag{96}$$

A systematic perturbation approach is obtained by inserting (94) and (96) in (92) and by collection the terms that are of equal power in the perturbation strength ε. The terms proportional to ε^0 and ε^n respectively lead to the following equations:

$$L_0 u_0 = F \ , \tag{97}$$

$$L_0 u_n = -L_1 u_{n-1} \qquad for \qquad n \geq 1 \ . \tag{98}$$

Equations (97) and (98) are of the form $L_0 u =$ *forcing term* and hence can be solved using the Green's function G of the unperturbed problem, i.e. by using the Green's function that satisfies

$$L_0 G(\boldsymbol{r}, \boldsymbol{r}') = \delta(\boldsymbol{r} - \boldsymbol{r}') \ . \tag{99}$$

For vector waves the Green's function should be replaced by a second order tensor, but the principles are the same. Using this Green's function the solution of (97) and (98) is given by

$$u_0 = GF \qquad \text{and} \qquad u_n = -GL_1 u_{n-1} \qquad (n \geq 1) \ . \tag{100}$$

The solution can now be constructed by solving the above equation recursively and by inserting the result in (96):

$$u = \underbrace{u_0}_{\substack{\text{Unperturbed} \\ \text{wave}}} - \underbrace{\varepsilon G L_1 u_0}_{\substack{\text{Single} \\ \text{scattered} \\ \text{wave}}} + \underbrace{\varepsilon^2 G L_1 G L_1 u_0}_{\substack{\text{Double} \\ \text{scattered} \\ \text{wave}}} + \cdots \tag{101}$$

In this formulation the total wavefield is written as a sum over the unperturbed wave, the waves that are scattered once by the inhomogeneity L_1, the waves that are scattered twice by the perturbation L_1 and all the higher order scattered waves. The series (101) is called the *Neumann series* in scattering theory.

The Born approximation u^{Born} consists in truncating this multiple scattering series after the single scattered wave:

$$u^{Born} = u_0 - \varepsilon G L_1 u_0 \,. \tag{102}$$

The great advantage of this approximation is that the scattered waves are given by $-\varepsilon G L_1 u_0$, hence the scattered waves now depend linearly on the perturbation of the medium that is contained in the operator L_1. This means that the scattered waves are related in this approximation linearly to the perturbation of the medium. This makes it possible to use the theory of linear inverse problems as shown in the section 2 for the solution of this problem. In doing so, one must keep in mind that the Born approximation ignores multiple scattering effects. When such effects are present in the data one should be extremely cautious in using the Born approximation.

4.2. INVERSION AND MIGRATION

As one of the simplest examples let us return to the acoustic wave equation (93) with the perturbed velocity given in (95). The unperturbed problem is then given by $(\nabla^2 + k^2)\, u = F$ where the constant wavenumber is given by $k = \omega / c_0$. The Green's function for this differential equation is given by

$$G(\boldsymbol{r}, \boldsymbol{r}') = -\frac{e^{ik|\boldsymbol{r} - \boldsymbol{r}'|}}{4\pi\, |\boldsymbol{r} - \boldsymbol{r}'|} \,. \tag{103}$$

When a point source with spectrum $F(\omega)$ is located in $\boldsymbol{r}_s$ the unperturbed wave is given by

$$u_0(\boldsymbol{r}) = -\frac{e^{ik|\boldsymbol{r} - \boldsymbol{r}_s|}}{4\pi\, |\boldsymbol{r} - \boldsymbol{r}_s|} F(\omega) \,. \tag{104}$$

In the Born approximation, the scattered waves are scattered only once, it follows from (102) that the single-scattered waves at a receiver position $\boldsymbol{r}_r$ are given by:

$$u_s(\boldsymbol{r}_r) = -\frac{1}{(4\pi c_0)^2}\int \frac{e^{ik|\boldsymbol{r}_r-\boldsymbol{r}|}}{|\boldsymbol{r}_r-\boldsymbol{r}|} n(\boldsymbol{r}) \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}_s|}}{|\boldsymbol{r}-\boldsymbol{r}_s|} dV\, F(\omega)\ . \qquad (105)$$

Suppose one measures these single-scattered waves, as is done in a seismic reflection experiment. The inverse problem then consists in finding the model perturbation $n(\boldsymbol{r})$ given the recorded data $u_s(\boldsymbol{r}_r)$. The theory of section 2 can be used for this when one discretizes the volume integral in (105). After discretization the scattering integral by dividing the volume in small cells, this integral can be written in the form

$$u_i = \sum_j A_{ij} n_j\ . \qquad (106)$$

Since in a realistic imaging experiment one uses many source positions and records the wavefield at many receivers, u_i stands for the reflected wave for source-receiver pair and frequency component $\#i$. Because of the discretized volume integral, n_j denotes the perturbation of $1/c^2(\boldsymbol{r})$ in cell $\#j$. The matrix elements A_{ij} are given by

$$A_{ij} = -\frac{1}{(4\pi c_0)^2}\int_{cell\ j} \frac{e^{ik_i|\boldsymbol{r}_{ri}-\boldsymbol{r}|}}{|\boldsymbol{r}_{ri}-\boldsymbol{r}|} \frac{e^{ik_i|\boldsymbol{r}-\boldsymbol{r}_{si}|}}{|\boldsymbol{r}-\boldsymbol{r}_{si}|} dV\, F(\omega_i)\ . \qquad (107)$$

In principle, one can solve the linear system (106) by brute force. However, in many realistic problems the size of the system of equations is so large that this is practically speaking impossible. This is notably the case in seismic exploration where the number data and the number of model parameters are exceedingly large. For problems of this scale, iterative solutions of the linear system of equations seems the only realistic way of obtaining a solution. In fact, it is in practice only possible to carry out the first step of such an iterative process. Let us find an estimated model by using the first step of the iterative solution (71) of section 2.9.2 using the preconditioning operator $\boldsymbol{P} = const. \cdot \boldsymbol{I}$. In this approach the estimated model is given by:

$$\tilde{\boldsymbol{n}}(\boldsymbol{r}) \sim \boldsymbol{A}^\dagger \boldsymbol{d} \sim \sum_{\substack{sources \\ receivers \\ frequencies}} \frac{e^{-i\omega(|\boldsymbol{r}_r-\boldsymbol{r}|+|\boldsymbol{r}-\boldsymbol{r}_s|)/c}}{|\boldsymbol{r}_r-\boldsymbol{r}|\,|\boldsymbol{r}-\boldsymbol{r}_s|} d_{rs}(\omega)\ , \qquad (108)$$

where $d_{rs}(\omega)$ denotes the scattered wave for source s, recorded at receiver r with frequency ω. For simplicity it is assumed here that the source signal is a delta-pulse in the time domain, so that $F(\omega) = 1$, and that all cells have equal volume: $V_j = const.$ Note that the transpose $\boldsymbol{A}^T$ has been replaced by the Hermitian conjugate $\boldsymbol{A}^\dagger$.[6] The reason for this is that the analysis of section 2 was for real matrices. A similar analysis for complex matrices shows that the results are identical provided the transpose $\boldsymbol{A}^T$ is replaced by its complex conjugate. The summation in the matrix product effectively leads to a summation over all sources, receivers and frequency components because all these were labelled by the index i in (106).

It is instructive to consider the summation over frequencies in (108). At each frequency the data $d_{rs}(\omega)$ are multiplied with $\exp(-i\omega\tau)$ with $\tau = (|\boldsymbol{r}_r - \boldsymbol{r}| + |\boldsymbol{r} - \boldsymbol{r}_s|)/c$. This means that the frequency summation has the form of a Fourier transform so that up to a constant, the frequency summation gives the data in the time domain at time $t = (|\boldsymbol{r}_r - \boldsymbol{r}| + |\boldsymbol{r} - \boldsymbol{r}_s|)/c$:

$$\tilde{n}(\boldsymbol{r}) \sim \sum_{\substack{sources \\ receivers}} \frac{1}{|\boldsymbol{r}_r - \boldsymbol{r}|\,|\boldsymbol{r} - \boldsymbol{r}_s|}\, d_{rs}\left(t = \frac{(|\boldsymbol{r}_r - \boldsymbol{r}| + |\boldsymbol{r} - \boldsymbol{r}_s|)}{c}\right) \quad (109)$$

This expression implies that the image can be constructed by summing the data over all source-receiver pairs, and by considering for each target point the data at a time needed to travel from the source location $\boldsymbol{r}_s$ to the target point $\boldsymbol{r}$ to the receiver location $\boldsymbol{r}_r$.

This is the procedure followed in the imaging of seismic reflection data called "Kirchhoff migration" [*Claerbout, 1985*; *Yilmaz*, 1987]. Effectively one sums in such an approach over all the available data that have a travel time consistent with a scatterer at the target point $\boldsymbol{r}$. The only difference with the classical Kirchhoff migration is the presence of the geometrical spreading terms $1/(|\boldsymbol{r}_r - \boldsymbol{r}|\,|\boldsymbol{r} - \boldsymbol{r}_s|)$ that are not included in Kirchhoff migration. In non-destructive testing this approach is known as Synthetic Aperture Focussing Technique (SAFT) [*Mayer et al.*, 1990]. The main conclusion is that Kirchhoff migration (up to some constants) corresponds to the first step of an iterative solution of the linearized scattering equation. The derivation of this section is a simplified version of the derivation given by *Tarantola* [1984] who incorporates other source signals than a delta pulse.

[6]The complex conjugate of the transpose is by definition the Hermitian conjugate: $A^\dagger_{ij} = A^*_{ji}$.

There is a widespread belief that the Born approximation can only be used for truly scattered waves. However, the Born approximation can also be used to account for effects of medium perturbations on transmitted waves, but with a domain of applicability to transmission problems that is smaller than for reflection or true scattering problems. To see this, assume that the velocity has a small constant perturbation δc and that a wave propagates over a distance L. The wave will then experience a phase shift $\exp i\varphi$, given by $\varphi = -(\omega/c^2)\, L\delta c$. In the Born approximation this perturbation is replaced by $\exp i\varphi \approx 1+i\varphi$. This in only a good approximation when the phase shift φ is much less than a cycle. In practice, this limits the use of the Born approximation for the inversion of transmission data. Note that even when the velocity perturbation is small, the requirement that the phase shift is small breaks down for sufficiently large propagation distances L.

This does not imply that the Born approximation cannot be used for the inversion of transmitted waves. For surface waves propagating in the earth, the wavelength can be very long. (A Rayleigh wave at a period of $75s$ has a wavelength of about $300km$.) This means that these waves do not travel over very many wavelengths when they propagate for a few thousand kilometers. In addition, the heterogeneity in the Earth's mantle is not very large. This makes it possible to carry out inversions of transmitted surface wave data using the Born approximation.

The Born approximation for surface waves with the resulting scattering and mode-conversion coefficients is derived for a flat geometry by *Snieder* [1986a, 1986b], the generalization to a spherical geometry can be found in *Snieder and Nolet* [1987]. Although the Green's function and the scattering and mode-conversion coefficients are different for elastic surface waves than for acoustic waves, the Born approximation leads to a linearized relation between the perturbation δu of the waveform and the perturbation m of the Earth model:

$$\delta u_{rs}(\omega) = \iiint K_{rs}(\boldsymbol{r},\omega) m(\boldsymbol{r}) dV \;, \tag{110}$$

where $K_{rs}(\boldsymbol{r},\omega)$ contains the surface wave Green's function of the incoming and outgoing wave at the scatterer location $\boldsymbol{r}$ for source-receiver pair "rs" and frequency ω. The perturbation $m(\boldsymbol{r})$ stands for the perturbation in the elasticity tensor and/or the density. After discretization the volume integral can be written as a linear system of equations:

$$\delta u_i = \sum_j K_{ij} m_j \;, \tag{111}$$

where δu_i stands for the perturbation of the surface wave at source-receiver pair $\#i$ and frequency ω_i, while m_j stands for the perturbation of the

Earth model in cell $\#j$. This linear system can be solved numerically, and the specific implementation for the inversion of surface wave data is shown by *Snieder* [1988a].

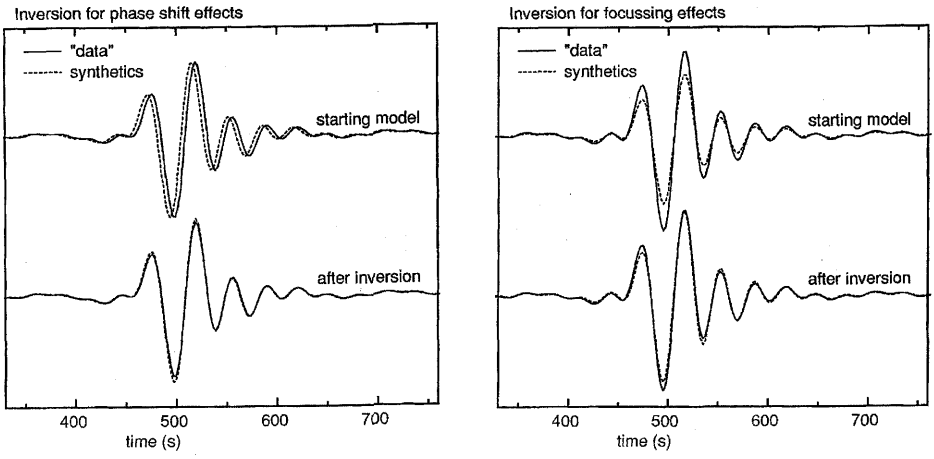


Figure 12. Phase shifted data (left panel) and data with an amplitude anomaly (right panel) before and after inversion. The synthetic data are shown with a solid line, the synthetics before and after inversion with a dashed line.

As an example how this algorithm operates consider the two upper seismograms shown in the two panels of figure 12. The seismogram of the left panel has been shifted in time with respect to the seismogram of a laterally homogeneous Earth model, whereas the amplitude of the seismogram of the right panel has been increased by 20%. Both seismograms are inverted simultaneously, and the corresponding model is shown in figure 13. (The map is shown only to display the scale of the problem but has otherwise no specific meaning.) The fit of the wave-forms after inversion is shown by the lower seismograms in figure 12, it can be seen that both the time shift and the amplitude change are accounted for well.

The triangle and square on the left are the source and receiver respectively of the time-shifted seismogram while the same symbols on the right are for the seismogram with the amplitude change. The velocity anomaly for the time-shifted seismogram is a negative velocity perturbation straddling the line joining the source and receiver on the left in figure 13. This negative velocity change gives the required time-shift. Note that this velocity perturbation is not confined to the source-receiver line; its finite extent is due to the fact that rays have no physical meaning and that a wave is influenced by the velocity perturbation averaged over the first Fresnel zone [*Snieder and Lomax*, 1996].

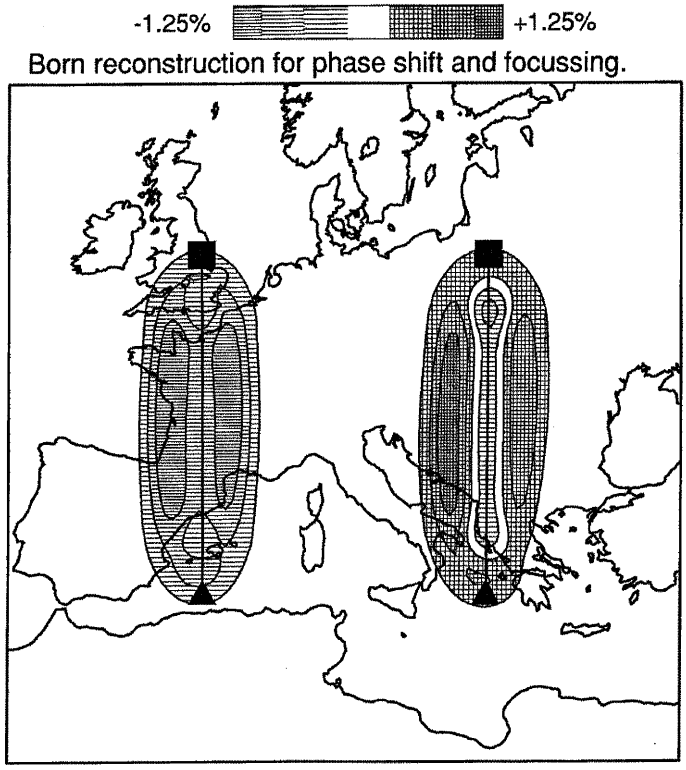


Figure 13. The velocity model obtained from the inversion of the seismograms in the previous figure. The source locations are marked with triangles, the receivers with squares. The source-receiver pair on the left is for the phase shifted seismogram, the pair on the right for the seismogram with an amplitude error. The map serves only to fix the scale.

The velocity structure on the right is for the seismogram with the perturbed amplitude. The perturbation is negative on the source-receiver line, but slightly further away from this line the velocity perturbation is positive. Since waves are deflected from areas of high velocity towards regions of low velocity, the wave energy is deflected towards the receiver. This means that the algorithm has realized a fit of the amplitude perturbation by creating a "surface wave lens" that produces just the right amount of focussing at the receiver to account for the amplitude perturbation.

Note that the model of figure 13 was constructed by numerically solving the linear system of equations (111). The input of the system of equations simple consisted of the real and imaginary components of the perturbation of the surface waves in the frequency domain. At no point in the inversion has the phase or amplitude of the waves been prescribed explicitly. However, the Born approximation contains all the relevant physics needed to

translate the perturbation of the real and imaginary components of the waves into the physics of focussing and phase retardation.

4.4. SURFACE WAVE INVERSION OF THE STRUCTURE UNDER NORTH-AMERICA

The surface waveform inversion has been applied to infer the shear-velocity structure under Europe and the Mediterranean by *Snieder* [1988b]. In this section a model for the shear-velocity under North-America is shown that was made by *Alsina et al.* [1996]. The shear-velocity perturbation in three layers with depths between 25 and 300 km is shown in figure 14. Red colours indicate slow velocities that are indicative of high temperature while green colours indicate high velocities correspond to low temperature.[7]

It is instructive to consider the west-coast of North America. Under the Gulf of California the velocity is very low. In this region, the East-Pacific rise meets the continent. (The East-Pacific rise is a spreading center in the ocean bottom comparable to the Mid-Atlantic ridge in the Atlantic Ocean.) Since in a spreading center hot material wells up from the mantle, one expects the temperature to be low, which is consistent with the slow velocity in that region.

Further north at the state of California the velocity perturbation is close to zero. In this area the Pacific plate slides horizontally along the North-American plate. This so called "strike-slip motion" gives rise to earthquakes in California. However, since the plate simply slide along each other, the temperature field is not perturbed very much, which is reflected by the neutral velocity perturbation.

However, northward of Cape Mendocino under the states of Oregon and Washington there is a distinct positive velocity perturbation. In this region oceanic plates slide eastward under the continents. This process is called "subduction." Since the subsiding plate is colder than the surrounding material this is reflected in the image as positive velocity anomalies.

The model shown in figure 14 and it's relation with the different tectonic regimes at the west coast of North-America show that with tomographic techniques as shown here it is possible to see plate-tectonics in action.

[7]The identification of velocity anomalies with temperature perturbations should be treated with care. Apart from temperature perturbations the seismic velocity is also affected by variations is compostion, the presence of volatiles such as water and possibly also pressure.

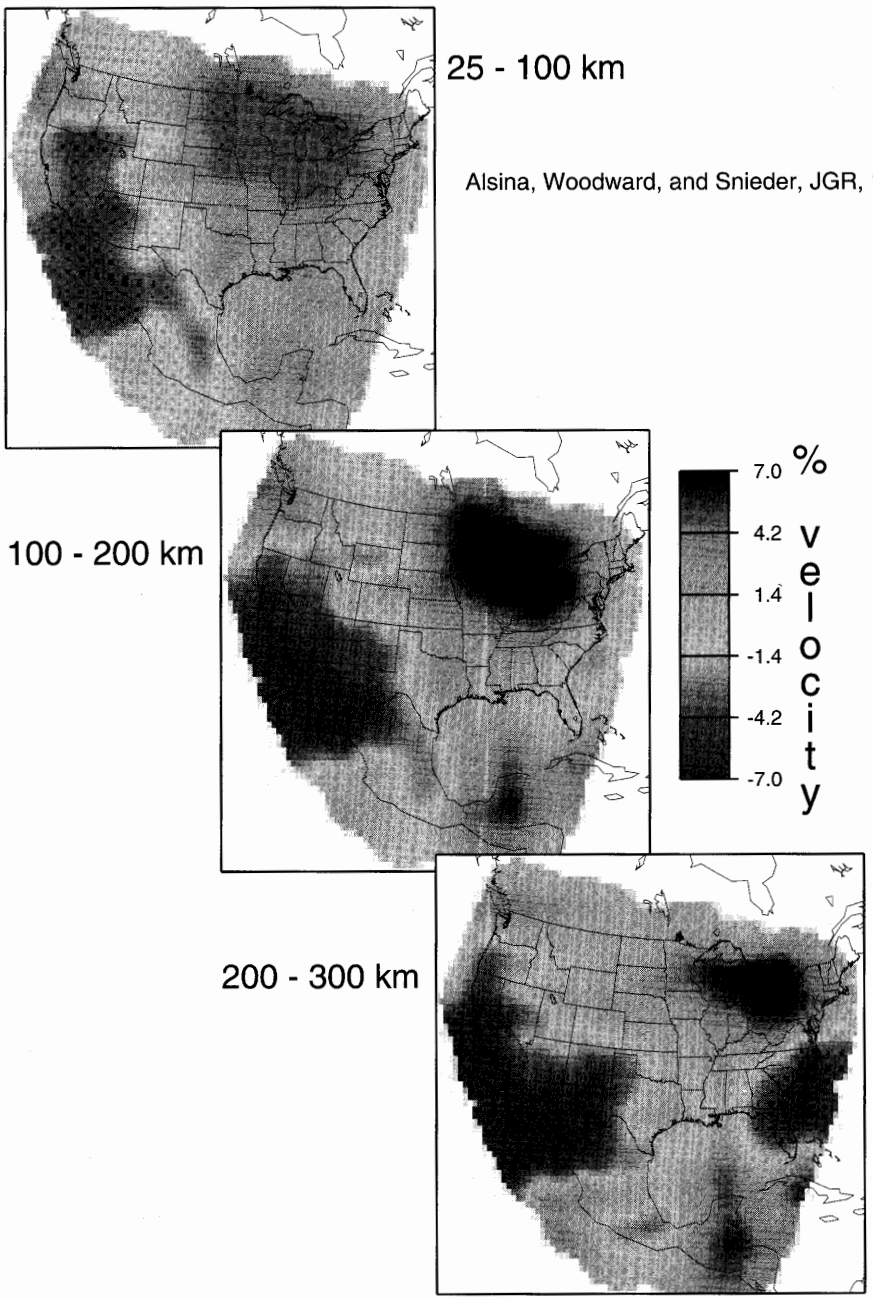


Figure 14. The relative shear velocity perturbation under North America in three layers at depths between 25 and 300 km as determined by [*Alsina et al.*, 1996].

5. Rayleigh's principle and perturbed eigenfrequencies

Important information about the structure of the Earth's interior follows form observations of perturbations of the frequencies of the free oscillations of the Earth as well as from measurements of the perturbation of the phase or group velocity of surface waves. This information can most easily be retrieved from observations when the relation between the frequency shift (or surface wave phase velocity) and the perturbation of the Earth model can be linearized. This can be achieved by applying Rayleigh-Schrödinger perturbation theory.

5.1. RAYLEIGH-SCHRÖDINGER PERTURBATION THEORY

Consider an eigenvalue problem of the following form:

$$Hu_n = \lambda_n r u_n \ . \tag{112}$$

In this expression H is an operator, for wave-propagation problems it usually is a differential operator. The eigenfunctions of this operator are denoted by u_n and the corresponding eigenvalues by λ_n. The function r denotes a positive weight function. For the Earth the equation of motion is given by

$$\partial_j \left(c_{ijkl} \partial_k u_l \right) = -\rho \omega^2 u_i \ , \tag{113}$$

where c_{ijkl} is the elasticity tensor. A comparison with (112) shows that $H\left(\bullet\right)_i = \partial_j \left(c_{ijkl} \partial_k \left(\bullet\right)_l \right)$, that the weight function r is given by the density ρ and that λ corresponds to $-\omega^2$. In this case, H is a Hermitian operator:

$$H^\dagger = H \ . \tag{114}$$

For the special case of elastic wave propagation, this property stems from the symmetry properties of the elasticity tensor and the stress-free boundary conditions at the Earth's surface [*Nolet*, 1981; *Dahlen and Tromp*, 1998]. For a Hermitian operator the eigenvalues are real and the eigenfunctions are orthonormal with respect to the following inner product

$$\langle u_n | r u_m \rangle = \int u_n^* r u_m dV = \delta_{nm} \ . \tag{115}$$

Let the operator H and the weight function r be decomposed in a reference operator H_0 and weight function r_0 for which we know the eigenfunctions $u_n^{(0)}$ and eigenvalues $\lambda_n^{(0)}$ and perturbations εH_1 and εr_1:

$$H = H_0 + \varepsilon H_1 \qquad , \qquad r = r_0 + \varepsilon r_1 \ . \tag{116}$$

Under this perturbation the eigenfunctions and eigenvalues are perturbed as well:

$$u_n = u_n^{(0)} + \varepsilon u_n^{(1)} + \varepsilon^2 u_n^{(2)} + \cdots \tag{117}$$

$$\lambda_n = \lambda_n^{(0)} + \varepsilon \lambda_n^{(1)} + \varepsilon^2 \lambda_n^{(2)} + \cdots \tag{118}$$

The goal of this analysis is to find the first order perturbation $\lambda_n^{(1)}$ of the eigenvalues. This can be achieved by inserting the expressions (116) through (118) in (112) and by collecting the terms of order ε^1, this gives:

$$H_1 u_n^{(0)} + H_0 u_n^{(1)} = \lambda_n^{(1)} r_0 u_n^{(0)} + \lambda_n^{(0)} r_1 u_n^{(0)} + \lambda_n^{(0)} r_0 u_n^{(1)} \tag{119}$$

The problem with this expression is that we are only interested in $\lambda_n^{(1)}$, but that in order to retrieve this term from (119) we need the first order perturbation $u_n^{(1)}$ of the eigenfunctions as well. The perturbation of the eigenvalue can be extracted by taking the inner product of (119) with the unperturbed eigenfunction $u_n^{(0)}$:

$$\begin{aligned} &\left\langle u_n^{(0)} | H_1 u_n^{(0)} \right\rangle + \underbrace{\left\langle u_n^{(0)} | H_0 u_n^{(1)} \right\rangle}_{(\heartsuit)} \\ &= \lambda_n^{(1)} \left\langle u_n^{(0)} | r_0 u_n^{(0)} \right\rangle + \lambda_n^{(0)} \left\langle u_n^{(0)} | r_1 u_n^{(0)} \right\rangle + \underbrace{\lambda_n^{(0)} \left\langle u_n^{(0)} | r_0 u_n^{(1)} \right\rangle}_{(\spadesuit)} \end{aligned} \tag{120}$$

Note that the perturbed eigenfunctions $u_n^{(1)}$ only appear in the terms marked ($\heartsuit$) and ($\spadesuit$). Using the fact that H_0 is Hermitian and that the eigenvalues $\lambda_n^{(0)}$ are real one finds that these terms are equal:

$$\begin{aligned} (\heartsuit) &= \left\langle u_n^{(0)} | H_0 u_n^{(1)} \right\rangle = \left\langle H_0^\dagger u_n^{(0)} | u_n^{(1)} \right\rangle = \left\langle H_0 u_n^{(0)} | u_n^{(1)} \right\rangle \\ &= \left\langle \lambda_n^{(0)} u_n^{(0)} | u_n^{(1)} \right\rangle = \lambda_n^{(0)} \left\langle u_n^{(0)} | u_n^{(1)} \right\rangle = (\spadesuit) \end{aligned} \tag{121}$$

This means that the terms containing the perturbed eigenfunctions cancel. Solving the remaining equation for the perturbation of the eigenvalue gives:

$$\lambda_n^{(1)} = \frac{\left\langle u_n^{(0)} | \left(H_1 - \lambda_n^{(0)} r_1 \right) u_n^{(0)} \right\rangle}{\left\langle u_n^{(0)} | r_0 u_n^{(0)} \right\rangle} . \tag{122}$$

The perturbation of the eigenvalue thus follows by evaluating the inner product of the perturbed operators sandwiched between the unperturbed eigenfunctions. This is one form of Rayleigh's principle: the eigenvalues are stationary to first order for perturbations of the eigenfunctions. The

crux is that according to 122 the first-order perturbation of the eigenvalues depends linearly on the perturbations of the operator H and weight r and hence linear inverse theory as exposed in the previous sections may be applied to infer the perturbations of H and r from the measured eigenvalues of the system.

5.2. THE PHASE VELOCITY PERTURBATION OF LOVE WAVES

Expression (122) can be used to evaluate the perturbation of the phase velocity of surface waves due to perturbations of the Earth model. This is shown here for the simplest example of Love waves in a plane geometry. The reader is referred to *Aki and Richards* [1980] for the extension to Rayleigh waves, while the analysis for a spherical geometry is treated by *Takeuchi and Saito* [1972]. As shown in *Aki and Richards* [1980] the Love wave eigenfunctions satisfy the following differential equation:

$$\partial_z(\mu\partial_z v) + \left(\rho\omega^2 - \mu k^2\right) v = 0 \,, \tag{123}$$

where μ is the shear modulus. The surface waves modes vanish at large depth ($z \to \infty$) and are stress-free at the surface. This gives the following boundary conditions:

$$\partial_z v(z=0) = 0 \qquad , \qquad v(z=\infty) = 0 \,. \tag{124}$$

In this analysis we assume that the frequency ω is a given constant. The operator H is given by $H = \partial_z\mu\partial_z + \rho\omega^2$, the weight function r is given by the shear modulus ($r(z) = \mu(z)$) while k^2 is the eigenvalue ($\lambda = k^2$). An integration by parts using the boundary conditions (124) can be used to show that the operator H is Hermitian. The theory of section 5.1 can then be used to find the perturbation δk of the wavenumber due to a perturbation μ_1 in the shear modulus and a perturbation ρ_1 in the density. Using the first order relation $\delta k^2 = 2k\delta k$ one obtains from (122) that:

$$\delta k = \frac{\int \left(\rho_1\omega^2 - \mu_1 k^2\right) v^2 dz - \int \mu_1 \left(\partial_z v\right)^2 dz}{2k \int \mu_0 v^2 dz} \,. \tag{125}$$

Since the phase velocity is given by $c = \omega/k$ the phase velocity perturbation follows to leading order from this expression by using the relation $\delta c/c = -\delta k/k$. This means that the first order effect of the perturbation of the Earth model on the phase velocity of Love waves can readily be computed once the unperturbed eigenfunctions $v(z)$ are known. A similar result can be derived for Rayleigh waves.

In general, the perturbation in the phase velocity due to perturbation in the density, P-wave velocity α and S-wave velocity β can be written as:

$$\frac{\delta c}{c} = \int K_\beta(z) \frac{\delta\beta(z)}{\beta(z)} dz + \int K_\alpha(z) \frac{\delta\alpha(z)}{\alpha(z)} dz + \int K_\rho(z) \frac{\delta\rho(z)}{\rho(z)} dz \,. \quad (126)$$

For Love waves the kernel $K_\alpha(z)$ vanishes because Love waves are independent of the bulk modulus. Examples of the kernels K_β, K_α and K_ρ are shown in figure 15 for the fundamental mode Rayleigh waves of periods of $100s$ (left panel) and $30s$ (right panel) respectively. It can be seen that the sensitivity of Rayleigh waves to the P-velocity α is much less than the sensitivity for changes in the shear velocity β and that the sensitivity kernel K_α decays more rapidly with depth than the other kernels. Both phenomena are a consequence of the fact that the compressive component of the motion become evanescent at much shallower depth than the shear component.

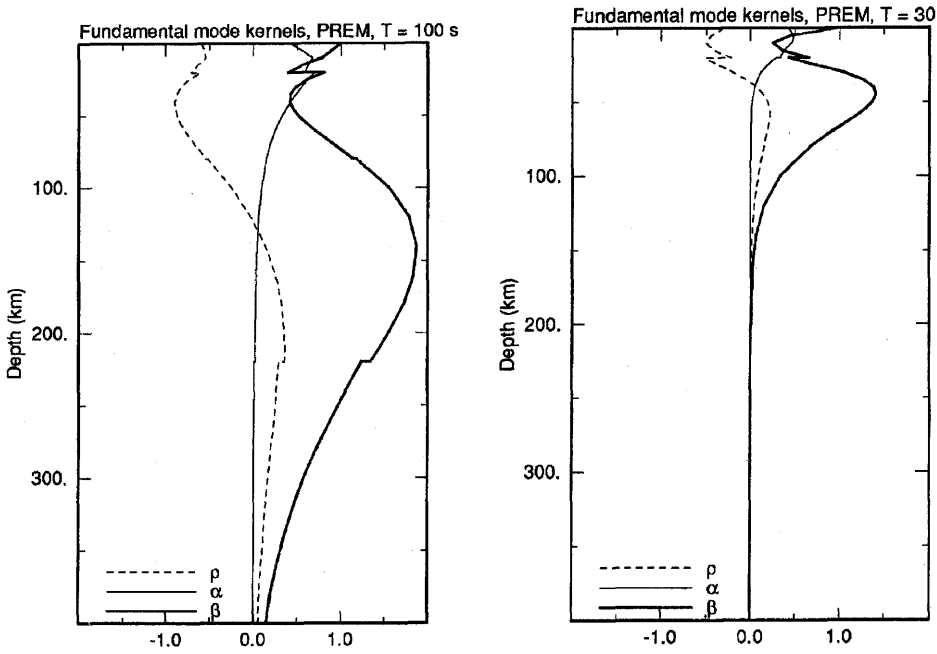


Figure 15. The sensitivity kernels K_α (thin solid line), K_β (thick solid line) and K_ρ (dashed line) for the fundamental mode Rayleigh wave for a period of $100s$ (left panel) and $30s$ (right panel). The kernels are computed for the PREM model [*Dziewonski and Anderson*, 1981].

For larger periods (left panel) the sensitivity kernels penetrate over a greater depth range than for shorter periods (right panel). This makes is it

possible to obtain a depth resolution in surface wave inversions. By measuring the phase velocity of surface waves at different periods, one obtains according to (126) and figure 15 the inner product of the perturbation of the Earth model with different weight functions $K(z)$. The theory of linear inversion in section 2 can then be used to determine the perturbation of the Earth model as a function of depth. Of course, only a finite depth resolution can be obtained, but using the expressions (4) or (75) one can determine the depth resolution that can be obtained. The depth resolution can be increased by using higher modes as well as the fundamental mode surface waves and by using both Love waves and Rayleigh waves [*Nolet,* 1977; *Cara*, 1978; *van Heijst and Woodhouse*, 1997].

The theory of this section can be extended to include anisotropic perturbations in the Earth as well [*Tanimoto*, 1986; *Montagner and Nataf*, 1986]. The depth-dependence of the sensitivity kernels of the Rayleigh wave phase velocity show a dependence of the anisotropic P-wave velocity that is fundamentally different than that of the isotropic P-wave velocity [*Muyzert and Snieder*, 1999].

6. Fermat's theorem and seismic tomography

Travel-time tomography is a technique where one aims to reconstruct the velocity structure of a body given the measurement of travel times of waves that have propagated through that body. The travel time along a ray is given by

$$T = \int_{\boldsymbol{r}[u]} u(\boldsymbol{r}) ds \; . \tag{127}$$

In this expression, u is the slowness which is defined as the reciprocal of the velocity: $u = 1/c$. The slowness is used rather than the velocity because now the integrand is linear to the quantity we aim to retrieve. It is tempting to conclude from (127) that the relation between the travel time and the slowness is linear. However, this is wrong! The reason for this is that the integration in (127) is along the path on which the waves travel. The rays are curves of stationary travel time, and hence the ray location depends on the slowness as well. Travel time tomography is thus a nonlinear inverse problem: the unknown slowness is present both in the integrand and it determines the ray position $\boldsymbol{r}[u]$ in the travel time integral (127). Linear inversion techniques can only be used when the relation between T and u is linearized.

Traditionally this is achieved by invoking Fermat's theorem which states that the travel along a ray does not change to first order when this ray is perturbed [e.g. *Nolet,* 1987; *Ben-Menahem and Singh*, 1981]. The proof of Fermat's theorem is based on variational calculus and invokes

the equations of kinematic ray tracing. However, the same result can be derived in a much simpler way when one starts from the eikonal equation.

6.1. FERMAT'S THEOREM, THE EIKONAL EQUATION AND SEISMIC TOMOGRAPHY

The derivation given in this section was formulated by *Aldridge* [1994] and is based on a perturbation analysis of the eikonal equation. Here, only the first order travel time perturbation will be derived, but *Snieder and Aldridge* [1995] have generalized the analysis to arbitrary order. Starting point is the eikonal equation which governs the propagation of wavefronts:

$$|\nabla T|^2 = u^2(\boldsymbol{r}) \ . \tag{128}$$

Consider a reference slowness $u_0(\boldsymbol{r})$ that is perturbed by a perturbation $\varepsilon u_1(\boldsymbol{r})$, where the parameter ε serves to facilitate a systematic perturbation approach:

$$u(\boldsymbol{r}) = u_0(\boldsymbol{r}) + \varepsilon u_1(\boldsymbol{r}) \ . \tag{129}$$

Under this perturbation the travel changes, and we assume that the travel time can be written as a regular perturbation series:[8]

$$T = T_0 + \varepsilon T_1 + \varepsilon^2 T_2 + \cdots \tag{130}$$

Inserting (129) and (130) in the eikonal equation (128) and collecting terms proportional to ε^0 and ε^1 gives:

$$|\nabla T_0|^2 = u_0^2(\boldsymbol{r}) \ , \tag{131}$$

$$(\nabla T_0 \cdot \nabla T_1) = u_0 u_1 \ . \tag{132}$$

The first equation is the eikonal equation for the reference travel time. Let the unit vector $\hat{\boldsymbol{t}}_0$ be directed along the gradient of T_0, then using (131) this gradient can be written as

$$\nabla T_0 = u_0 \hat{\boldsymbol{t}}_0 \ . \tag{133}$$

Taking the inner product of this expression with $\hat{\boldsymbol{t}}_0$ gives $\left(\hat{\boldsymbol{t}}_0 \cdot \nabla T_0\right) = u_0$, which can also be written as

$$\frac{dT_0}{ds_0} = u_0 \ . \tag{134}$$

[8]The assumption that the travel time perturbation is regular is not valid when caustics are present and the relation between the slownes perturbation and the travel time surface is not analytic.

In this expression $d/ds_0 = \hat{\boldsymbol{t}}_0 \cdot \nabla$ is the derivative along the unit vector $\hat{\boldsymbol{t}}_0$. Expression (134) can be integrated to give

$$T_0 = \int_{\boldsymbol{r}_0[u_0]} u_0(\boldsymbol{r}_0) ds_0 \ , \tag{135}$$

where $\boldsymbol{r}_0$ is the position of the ray in the reference slowness field.

Using (133), expression (132) for the travel time perturbation can be written as $\left(\hat{\boldsymbol{t}}_0 \cdot \nabla T_1\right) = u_1$, but since $\hat{\boldsymbol{t}}_0 \cdot \nabla = d/ds_0$ this can also be written as

$$\frac{dT_1}{ds_0} = u_1 \ . \tag{136}$$

This expression can be integrated to give

$$T_1 = \int_{\boldsymbol{r}_0[u_0]} u_1(\boldsymbol{r}_0) ds_0 \ . \tag{137}$$

The main point of this derivation is that the integration in (137) is along the *reference ray* $\boldsymbol{r}_0$ rather than along the true ray in the perturbed medium. Expression (137) constitutes a linearized relation between the travel time perturbation T_1 and the slowness perturbation u_1. When one divides the model in cells where one assumes the slowness perturbation is constant, then the discretized form of (137) can be written as

$$\delta T_i = \sum_j L_{ij} u_j \ . \tag{138}$$

In this expression, the subscript i labels the different travel times that are used in the inversion while j is the cell index. It follows from figure 16 that L_{ij} is the length of ray $\#i$ through cell $\#j$. Equation (138) forms a linear system of equations as discussed in section 2.

The matrix L_{ij} is in general very sparse for tomographic problems because every ray intersects only a small fraction of the cells, see figure 16. This is in particular the case in three dimensions where the relative number of intersected cells is much smaller than in two dimensions. This makes it particularly attractive to use iterative solutions for the linear system of equations as presented in section 2.9.2. It follows from expression (71) that in this iterative approach the solution is constructed by multiplication with the matrices $\boldsymbol{L}$ and $\boldsymbol{L}^T$, but that the inverse $\left(\boldsymbol{L}^T\boldsymbol{L}\right)^{-1}$ (which is not sparse) is not needed. In such an approach there not even a need to store the matrix $\boldsymbol{L}$, it is much efficient to store only the relatively few nonzero matrix elements and keep track of the indices of these elements. This approach has been developed by *Nolet* [1985] and has been used

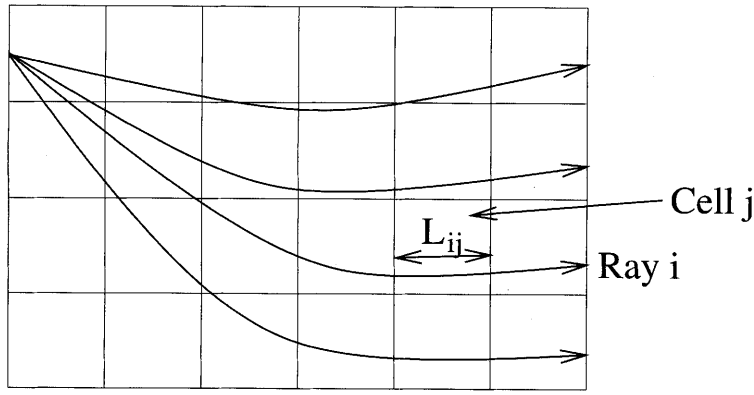


Figure 16. Diagram of a tomographic experiment.

successfully for extremely large-scale tomographic inversion for the interior of the Earth's mantle [e.g. *Spakman et al.*, 1993; *van der Hilst et al.*, 1997].

It should be remarked that there is no strict need to use cells to discretize the slowness (and the same holds for the treatment of linearized waveform inversion in the sections 4.2 and 4.3). As an alternative the slowness perturbation can be expanded in a finite set of basis functions $B_j(\boldsymbol{r})$:

$$u_1(\boldsymbol{r}) = \sum_{j=1}^{L} m_j B_j(\boldsymbol{r}) \ . \tag{139}$$

Inserting this in (137) one again arrives at a linear system of the form (138), but the matrix elements are now given by

$$L_{ij} = \int_{ref\ ray\ j} B_j(\boldsymbol{r}) ds_0 \ , \tag{140}$$

where the integration now is along reference ray $\#j$. However, one should realize that when the basis functions have global character, such as spherical harmonics, the matrix $\boldsymbol{L}$ is in general not sparse with this model parameterization. This means one cannot fully exploit the computational efficiency of iterative solutions of linear systems.

6.2. SURFACE WAVE TOMOGRAPHY

For most people Fermat's principle goes hand in hand with travel time tomography, but surface wave tomography actually relies on it since the early twenties when *Gutenberg* [1924], using data collected by *Tams* [1921], explained the dispersion differences between surface waves propagating

along continental and oceanic paths in terms of properties of the Earth's crust.

Surface wave tomography clearly means tomography based on surface waves. There are several ways of doing this. One may directly invert the waveform for structural parameters, provided the sensitivities are known. Generally, the structural parameters are non-linearly related to the waveforms and one way (among others) to linearize the problem has been outlined in section 4.3. A more classical approach, consists of the so-called two-step inversion where one first constructs models of phase or group velocity as a function of frequency. The information contained in these maps is then inverted for depth structure. This is possible because expression (126) gives a liner relation between the phase velocity as a function of frequency and the perturbation of the medium as a function a depth. The mapping from phase velocity to depth then reduces to a linear inverse problem to which the techniques of section 2 can be applied.

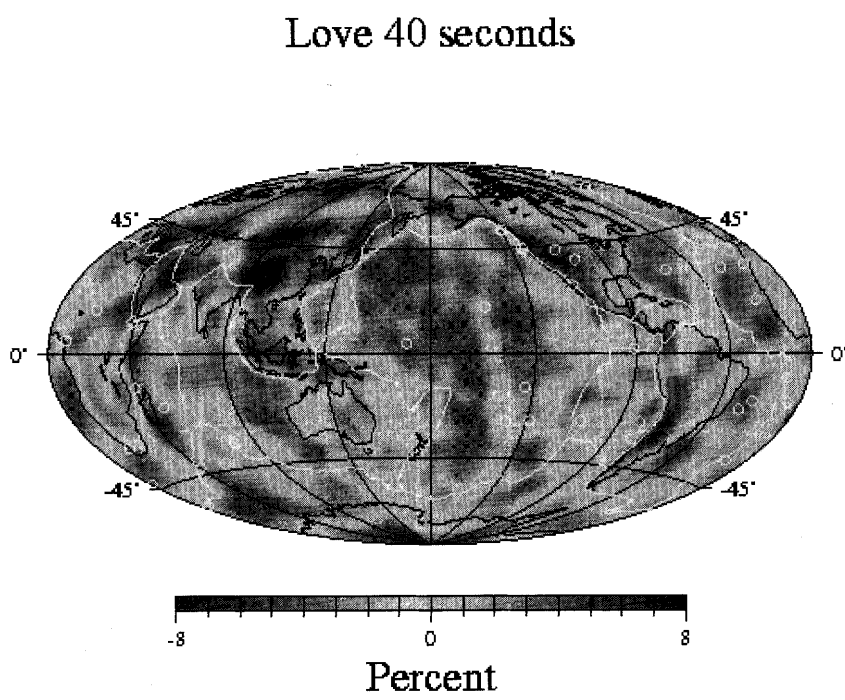


Figure 17. Love wave phase velocity perturbations at a period of 40 seconds. The variations are given in percent with respect to an average reference model. Yellow lines represent plate boundaries and yellow circles are hotspots.

It is this first step which commonly assumes Fermat's principle. If we assume that the Earth structure is sufficiently smooth compared to the

wavelength of a propagating wave, locally any wave of frequency ω may be approximated by a plane wave. We are then in the realm of ray theory and are able to measure the mean phase (or group) velocity along the ray path. If furthermore Fermat's principle holds, the measurements correspond to path integrals along the minor and major arc segments of the great circle connecting station and event. It is undoubtedly the case that there are many examples of off-great-circle propagation and non-theoretical effects. The extend to which such effects corrupt the models constructed under the simple assumptions of Fermat's principle can only be determined by numerical modelling, and is still an open question.

It is commonly accepted that Fermat's principle is valid for surface wave periods longer than 150 seconds roughly, and most work until the mid-nineties concentrated on these longer periods. *Trampert and Woodhouse* [1995, 1996] extend these classical methods in two ways. Firstly, they applied it to much shorter periods in the range of 40 to 150 seconds. Secondly, they exploited the by now tremendous wealth of assembled digital data in a systematic fashion by developing fully automatic methods for extracting path-averaged phase velocities of surface waves from recorded waveforms. The most original feature of these studies is the phase velocity maps at short periods which gives new global information on the uppermost structure of the Earth. As an example, the model for the phase velocity of Love waves at a period of 40 seconds is shown in figure 17. The Love-wave model shows a remarkable correlation with surface topography and bathymetry, and hence with crustal thickness. This is too be expected because these waves sample the surface with an exponentially decaying sensitivity. The reference model used in the inversion has a crustal thickness of 24.4 km. If the true Earth has a thicker crust, mantle material of the reference model is replaced with slower crustal material resulting in a slow velocity perturbation. This is the case for the continents. In oceans, the opposite takes place. The sensitivity of 40 second Love waves decays very rapidly with depth and hence tectonic signatures from the uppermost mantle are less pronounced. Nevertheless, a slow velocity anomaly is observed along most oceanic ridges, where hot, and hence slow, material is put into place.

7. Nonlinearity and ill-posedness

Nonlinearity complicates the estimation problem considerably. In many practical problems, nonlinear inversion is treated as a nonlinear optimization problem where a suitably chosen measure of the data misfit is reduced as a function of the model parameters. There is a widespread belief in the inverse problem community that the dominant effect of nonlinearity is the

creation of secondary minima in the misfit function that is minimized. This point of view is overly simplistic. Nonlinearity affects both the estimation problem and the appraisal problem. In sections 7.1 and 7.2 it is shown how non-linearity can be a source of ill-posedness of inverse problems. In section 8 the appraisal problem for nonlinear inverse problems is discussed.

7.1. EXAMPLE 1, NON-LINEARITY AND THE INVERSE PROBLEM FOR THE SCHRÖDINGER EQUATION

The inverse problem of the estimation of a quantum mechanical potential in one dimension from the measurement of the reflection coefficient of waves reflected by the potential is an interesting tool for studying nonlinear inversion because the inverse problem has a stunningly simple solution [*Marchenko,* 1955; *Burridge,* 1980]. This inverse problem is of direct relevance in the earth sciences; both the inverse problem of geomagnetic induction [*Weidelt,* 1972] as well as the seismic reflection problem [*Burridge,* 1980; *Newton,* 1981] can be reformulated similar to the problem treated in this section. For the Schrödinger equation the wavefield ψ satisfies the following differential equation:

$$\psi_{xx} + \Big(k^2 - V(x)\Big)\,\psi = 0\,. \tag{141}$$

Let the reflection coefficient after a Fourier transform to the time domain be denoted by $R(t)$. The potential $V(x)$ at a fixed location x follows from the reflection coefficient by solving the Marchenko equation:

$$K(x,t) + R(x+t) + \int_{-t}^{x} K(x,\tau)R(\tau+t)d\tau = 0\,, \tag{142}$$

for $K(x,t)$ and carrying out a differentiation:

$$V(x) = -2\frac{dK(x,x)}{dx}\,. \tag{143}$$

In figure 18 the results of a numerical solution of the Marchenko equation is shown. The potential is shown by a thick solid line. For this potential the reflection coefficient is computed and the potential is reconstructed by numerically solving the Marchenko equation (142) and carrying out the differentiation (143), details of the calculation can be found in [*Dorren et al.,* 1994]. The reconstructed potential of figure 18 (top panel) is on the scale of the figure indistinguishable from the true potential. In figure 18 (bottom panel) the same synthetic experiment is shown, the only difference is that the potential has been multiplied with a factor 2.5. If the

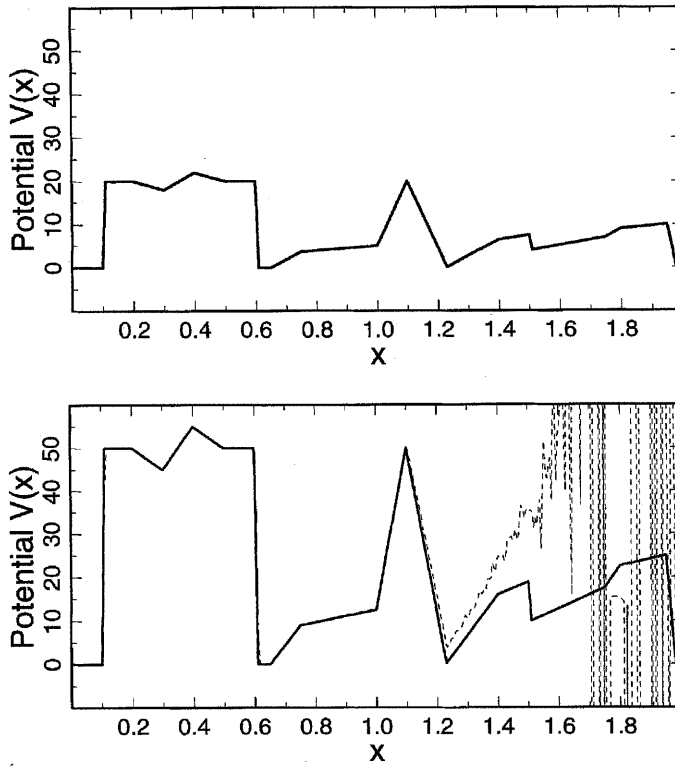


Figure 18. The original potential (solid lines) and the potential reconstructed with the Marchenko equation (dashed lines). The potential in the bottom panel is 2.5 times as strong as the potential in the top panel.

problem would be linear, the reflection coefficients would be 2.5 times as strong as for the potential in the top panel and the reconstructed potential would also be 2.5 times as strong. This would lead to a near-perfect reconstruction of the potential. However, the reconstructed potential is given by the dashed line. It can be seen that the left part of the potential (the side from which the waves are incident) the potential is reconstructed quite well, but that the part of the potential on the right is very poorly reconstructed. According to the reasoning above, the potential would have been reconstructed quite well if the problem had been linear. This implies that the instability in the reconstructed potential is due to the non-linearity of the problem.

The physical reason for this instability can relatively easy be understood. It follows from the Schrödinger equation (141) that the effective wavenumber is given by $\sqrt{k^2 - V(x)}$. When $k^2 < V(x)$ the wavenumber is complex which reflects the fact that the waves are evanescent when the

potential energy is larger than the total energy. In that case the wavefield decays exponentially within the potential. For a fixed energy k^2 the wavefield is more evanescent for the potential in the bottom panel of figure 18 than for the potential in the top panel of that figure, simply because the potential energy is 2.5 times as high. This has the result that the wave-field penetrates further in the potential in the top panel than in the potential in the bottom panel of figure 18. Obviously, the potential in a certain region is not constrained by the recorded wavefield if the wavefield does not sample the potential in that region. The strong evanescence of the waves in the potential in the bottom panel of figure 18 imlpies that that parts of that potential are effectively not sampled by the waves. In that case, the numerical details of the algorithm, including the numerical round-off error determine the reconstructed potential in that region. The essential point is that the values of the model parameters affect the way in which the wavefield interrogates the model.

Physically, the instability in the reconstructed potential in figure 18 can thus be understood. However, what does this imply for the inverse problem? What happens physically if the potential on the left side is high, is that the wavefield is prevented from sampling the potential on the right part. This means that for some values of the model parameters (that describe how high the potential is on the left), other model parameters (that describe the potential on the right) are unconstrained by the data. In terms of a misfit function this implies that the misfit does not depend on the model parameters that describe the right side of the potential (when the left side of the potential is high). In other words, the misfit function does not have a minimum, but is has a broad plateau. Note that as an additional complexity this only occurs for certain values of other model parameters (that describe how high the potential is on the left).

7.2. EXAMPLE 2, NON-LINEARITY AND SEISMIC TOMOGRAPHY

A second example of the ill-posedness introduced by the non-linearity in inverse problems is seismic tomography. Consider a cross-borehole tomographic experiment where rays travel from a source in a well to a string of receivers on another well. The case where the velocity is homogeneous is shown in the left panel of figure 19. In that case the rays are straight lines that travel from the source on the left to receivers $R1$ through $R5$ on the right. If the velocity is not homogeneous, the rays are curved. This implies that the value of the model parameters determine the way in which the rays interrogate the model.

Suppose that a low-velocity anomaly is present, see the middle panel of figure 19. Since the rays of first arrivals are curves of minimal travel

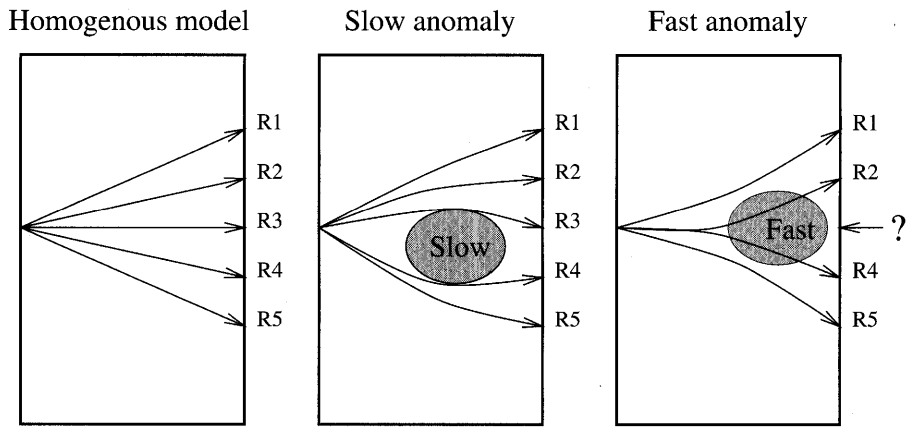


Figure 19. (a) Tomographic experiment where the velocity is homogeneous and the rays are straight. (b) Tomographic experiment where the rays curve around a low-velocity body. (c) Tomographic experiment where a high-velocity anomaly causes a shadow zone at the middle receiver.

time, the rays curve around the slow-velocity anomaly. If the velocity anomaly is sufficiently slow, all the rays may completely curve around the slow-velocity. In that situation the anomaly would not be sampled by any rays and the velocity within the anomaly cannot be determined because it is not sampled by any rays. At best one could derive an upper bound for the velocity within the anomaly. Just as in the previous section, the model parameters affect the way in which the probe (in this case the rays) samples the model. In terms of the misfit function this implies that for a certain range of model parameters, the misfit function does not depend on the model parameters at all, in other words: the misfit function in completely flat over an extended range of parameter values.

Let us now consider the opposite situation where a high-velocity anomaly is present, see the right panel of figure 19. Rays will be defocused by a high velocity body and a shadow-zone is formed behind the anomaly. This may mean that there is no ray that will hit the receiver $R3$, see the question mark in figure 19. This means that for some values of the model parameters it is impossible to compute the data because the travel time cannot be determined when there is no ray that hits the receiver. This means that for some values of the model parameters it is impossible to compute the corresponding data values given the theory that one uses. In this sense one can say that some values of the model parameters are "forbidden." It should be noted that this is not a complexity created in some exotic thought-experiment; the problem of the "missing rays" is a real problem in seismic travel time tomography [*Sambridge,* 1990], as well

as the tomographic imaging of the temperature field in ovens [*Natterer et al.*, 1997].

The critical reader might remark at this point that the fact that for certain values of the velocity model there are no rays hitting the receiver is due to the fact that ray theory is an approximation to a true theory (wave theory), and that wave theory predicts that some energy will diffract into the shadow-zones. Although this is correct, one should also note that the wavefield in the shadow zones is very weak (that is why they are called shadow zones) and that in practice this diffracted wavefield is usually not detectable.

8. Model appraisal for nonlinear inverse problems

In the previous section the effect of non-linearity on model estimation has been discussed. In this section an attempt is made to describe the effect of non-linearity on the appraisal problem where one wants to describe how the estimated model is related to the true model (see figure 2). However, it should be stressed that there is presently no general theory to deal with the appraisal problem for a truly nonlinear inverse problem with infinitely many degrees of freedom. In practice, one often linearizes the problem around the estimated model and then uses linear theory to make inferences about the resolution and reliability of the estimated model. The lack of a general theory for the appraisal problem should be seen as a challenge for theorists! In this section three attempts are described to carry out model assessment for a nonlinear inverse problem. These attempts follow the lines of formal theory (section 8.1), a numerical approach (section 8.2) and a pragmatic approach (section 8.3).

8.1. NONLINEAR BACKUS-GILBERT THEORY

Linear Backus-Gilbert theory is based on the equations (72)-(74) for linear inverse problems for continuous models. The model estimate $\tilde{m}(x)$ at location x is constructed by making the linear superposition (73) of the data. The resolution kernel in equation (74) specifies the relation between the estimated model and the true model. In the ideal case the resolution kernel is a delta function. *Backus and Gilbert* [1967, 1968] have shown that the criterion that the resolution kernel should resemble a delta function as much as possible can be used to determine the coefficients $a_i(x)$ in equation (73) which prescribes how a datum d_i affects the estimated model at location x.

Backus-Gilbert theory has been generalized by *Snieder* [1991] for the special case that the forward problem can be written as a perturbation

series:

$$d_i = \int G_i^{(1)}(x)m(x)dx + \iint G_i^{(2)}(x_1,x_2)m(x_1)m(x_2)dx_1dx_2 + \cdots \quad (144)$$

In a number of applications such a perturbation series arises naturally. Important examples are the Neumann series (101) in scattering theory where the scattering data are written as a sum of integrals that contain successively higher powers of the model perturbation, or ray perturbation theory where the travel time of rays is written as a sum of integrals with increasing powers of the slowness perturbation [e.g. *Snieder and Sambridge,* 1993; *Snieder and Aldridge,* 1995]. When the forward problem is nonlinear, the inverse problem is non-linear as well, this suggest that for a non-linear inverse problem the linear estimator (73) should be generalized to include terms that are nonlinear in the data as well:

$$\tilde{m}(x) = \sum_i a_i^{(1)}(x)d_i + \sum_{i,j} a_{ij}^{(2)}(x)d_id_j + \cdots \quad (145)$$

The key of non-linear Backus-Gilbert theory is to insert the expansion of the forward problem (144) in the estimator (145). The result can then be written as:

$$\tilde{m}(x) = \int R^{(1)}(x;x_1)m(x_1)dx_1 + \iint R^{(2)}(x;x_1,x_2)m(x_1)m(x_2)dx_1dx_2 + \cdots \quad (146)$$

This expression generalizes the linear resolution kernel of equation (74) to nonlinear inverse problems. The kernel $R^{(1)}(x;x_1)$ describes to what extent the estimated model is a blurred version of the true model. The higher order kernels such as $R^{(2)}(x;x_1,x_2)$ can be interpreted as nonlinear resolution kernels that describe to what extent there is a spurious nonlinear mapping from the estimated model onto the true model in the inversion process.

In the ideal case, the estimated model is equal to the true model: $\tilde{m}(x) = m(x)$. This is the case when the linear resolution kernel $R^{(1)}(x;x_1)$ is a delta function $\delta(x-x_1)$ and when the nonlinear resolution kernels are equal to zero: $R^{(n)}(x;x_1,\cdots,x_n) = 0$ for $n \geq 2$. However, as in equation (75) the linear resolution kernel $R^{(1)}(x;x_1)$ can be written as a sum of a finite amount of data kernels $G^{(1)}(x_1)$. Since a delta function cannot be obtained by summing a finite amount of smooth functions, the linear resolution kernel can never truly be a delta function. This reflects the fact that with a finite amount of data the estimated model will be a blurred version of the true model. *Snieder* [1991] treats the inverse problem of the determination of the mass-density of a vibrating string from the eigenfrequencies of the string. He shows that if only a finite amount of

eigenfrequencies are available, the nonlinear resolution kernels cannot be zero. This implies that the finiteness of the data set leads to a spurious nonlinear mapping from the true model to the estimated model. This can be related to the symmetries of the problem and to mode coupling "off the energy shell." The finite width of the linear resolution kernel and the fact that the nonlinear resolution kernels are nonzero imply not only that the estimated model is a blurred version of the true model, but also that the estimated model is biased. The reader is referred to *Snieder* [1991] for details. In that work it is also described how the coefficients $a^{(i)}$ in the estimator (145) can be determined.

Although nonlinear Backus-Gilbert theory is a new tool for dealing with the assessment problem for nonlinear inverse problems, one should realize that the theory can only be applied to weakly nonlinear problems where a (very) few orders are sufficient for an accurate description of both the forward and the inverse problem. In addition, the theory of *Snieder* [1991] is so complex that a reformulation is needed to make the theory applicable to the large-scale inverse problems that are being treated in practice.

It is important to realize that any regular (nonlinear) mapping from data d_i to an estimated model $\tilde{m}(x)$ can be written in the form of expression (145). The details of the employed algorithm then determine the coefficients $a^{(n)}_{i_1 \cdots i_n}$. The resolution analysis shown in equation (146) and the subsequent discussion therefore is applicable to the estimated model. The conclusions concerning the linear and nonlinear resolution kernels can thus be used for *any* regular mapping from the data to the estimated model.

8.2. GENERATION OF POPULATIONS OF MODELS THAT FIT THE DATA

Another approach to asses the reliability of estimated models is to generate not a single model that fit the data within a certain tolerance but to obtain an ensemble of models that fits the data within a certain tolerance [e.g. *Lomax and Snieder*, 1995]. An alternative approach is to compute the misfit for a very large class of models and to use the data fit, possibly in combination with Bayesian statistics to make inferences about the range of models that explain the data in a certain sense [*e.g. Mosegaard and Tarantola*, 1995; *Gouveia and Scales,* 1998, *Mosegaard,* 1998]. Obviously, this approach requires a numerical approach to create such ensembles, but with present day computers significance progress has been made.

An important concept in the generation of ensembles of models is the randomness in the search method that one employs. A descent method

contains no element of randomness whatsoever, whereas a Monte Carlo search where one randomly samples model space is completely random. In between are algorithms which have both, a random component as well as a mechanism to prefer models that fit the data well. Examples of such algorithms are simulated annealing [*Krikpatrick et al.,* 1983; *Rothman,* 1985] or genetic algorithms [*Sambridge and Drijkoningen,* 1992; *Sen and Stoffa,* 1992; *Lomax and Snieder,* 1995]. A promising technique is the adaptive search [*Mosegaard and Tarantola,* 1995] where in the process of carrying out a random search, information about the misfit function is being built up, and this misfit function is then used to drive the random search in an intelligent way.

The main merit of algorithms that determine an ensemble of models together with information on how well each model explains the data is that this ensemble can be used to make inferences about the model. These inferences may or may not be statistical. The possibility to analyse such ensembles of models in a meaningful way has not yet been fully exploited.

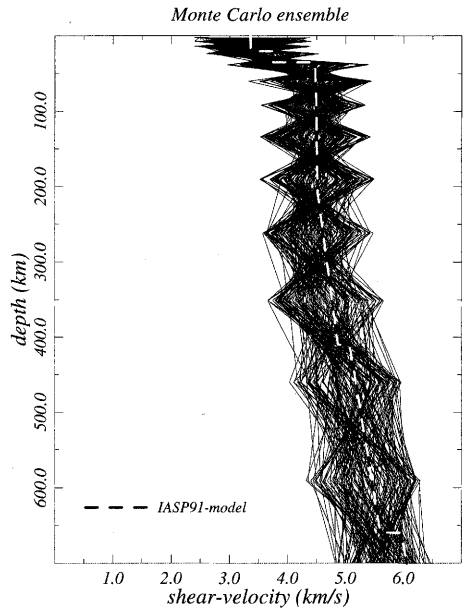


Figure 20. The true velocity model (broken curve) and an ensemble of models generated by a Monte Carlo search that fit surface wave group velocity data with a realistic tolerance (solid lines).

An example is shown from a study of *Douma et al.* [1996]. In their study synthetic group velocity data of the fundamental mode Rayleigh wave that propagates along the earth's surface were computed for periods between 10 and 300 s The true velocity model is shown as the dashed line in figure

20 as a function of depth. A Monte Carlo search was used to find models of the S-velocity that were consistent with the data within a realistic measurement error. The resulting population of models is shown in figure 20. In this study, the S-velocity was deliberately over-parameterized. As a result, the resulting models are highly oscillatory and contain strong trade-offs. The aim of the study of *Douma et al.* [1996] was to extract the robust features of the velocity model from the ensemble of models shown in figure 20. This was achieved by computing "Empirical Orthogonal Functions" (EOF's) from this ensemble. These functions give the patterns with the different degrees of variability within an ensemble.

The EOF's can be used to re-parameterize the model space in an intelligent way that reflects how well model perturbations are constrained by the data. Alternatively, the EOF's could be used to carry out a statistical analysis of the ensemble of models that explains the data. However, as noted by *Douma et al.* [1996] the EOF techniques is only useful for inverse problems that are weakly nonlinear.

8.3. USING DIFFERENT INVERSION METHODS

In the previous sections, a theoretical and a numerical method for model appraisal were presented. Apart from these more formal approaches, "common sense" is a powerful tool for carrying out model assessment. An important way to asses the reliability of a model is to determine the model in different ways. In the ideal case, different data sets are used by different research groups who use different theories to estimate the same properties. The agreement of disagreement between these models can be used as an indicator of the reliability of these models. It is admittedly difficult to quantify the sense of reliability that is thus obtained, but in the absence of an adequate theory to carry out model assessment for nonlinear inverse problems (and remember that we don't have such a theory) this may be the most practical approach. An example of this approach is given by *P*assier et al. [1997].

9. Epilogue

Linear inverse problem theory is an extremely powerful tool for solving inverse problems. Much of the information that we currently have on the Earth's interior is based on linear inverse problems. The success of modern-day oil exploration for an affordable price is to a large extent possible because of our ability to use imaging techniques that are based on single scattering theory to map oil reservoirs. One could argue that the modern world, which heavily relies on cheap access to hydrocarbons, might have

a drastically different form if single scattering would not explain seismic data so well.

Despite the success of linear inverse theory, one should be aware that for many practical problems our ability to solve inverse problems is largely confined to the estimation problem. This may surprise the reader, because for linear inverse problems the resolution kernel and the statistics of linear error propagation seem to be adequate tools for carrying out the model assessment. However, as noted in section 2, in order to compute the resolution kernel one needs to explicitly define the generalized inverse $\boldsymbol{A}^{-g}$. For many problems, this entails inverting the matrix $\boldsymbol{A}^T\boldsymbol{A}$ or an equivalent matrix that contains regularization terms as well. This task is for many important problem, such as imaging seismic reflection data, not feasible.

The results of section 8 show that for nonlinear inverse problems both the estimation problem and the appraisal problem are much more difficult. In the context of a data-fitting approach, the estimation problem amounts to a problem of nonlinear optimization. Researchers can use the results of this field of mathematics to solve the estimation problem. However, the reader should be aware of the fact that there is presently no theory to describe the appraisal problem for nonlinear inverse problems. Developing a theoretical framework for the appraisal of models obtained from nonlinear inversion of data is a task of considerable complexity which urgently needs the attention of the inverse problem community.

Acknowledgments: Discussions with John Scales and Jean-Jacques Lévêque have helped to clarify a number of issues, we appreciate their input very much.

References

1. Aldridge, D.F., Linearization of the eikonal equation, *Geophysics, 59*, 1631-1632, 1994.
2. Alsina, D., R.L. Woodward, and R.K. Snieder, Shear-Wave Velocity Structure in North America from Large-Scale Waveform Inversions of Surface Waves, *J. Geophys. Res., 101*, 15969-15986, 1996.
3. Aki, K., and P.G. Richards, *Quantitative Seismology (2 volumes)*, Freeman and Co. New York, 1980.
4. Backus, G., and J.F. Gilbert, Numerical applications of a formalism for geophysical inverse problems, *Geophys. J.R. Astron. Soc., 13*, 247-276, 1967.
5. Backus, G., and J.F. Gilbert, The resolving power of gross earth data, *Geophys. J.R. Astron. Soc., 16*, 169-205, 1968.
6. Backus, G. E. and F. Gilbert, Uniqueness in the inversion of inaccurate gross earth data, *Philos. Trans. R. Soc. London, Ser. A, 266*, 123-192, 1970.
7. Ben-Menahem, A. and S.J. Singh, *Seismic waves and sources*, Springer Verlag, New York, 1981.

8. Borg, G., Eine Umkehrung der Sturm-Liouvillischen Eigenwertaufgabe, Bestimmung der Differentialgleichung durch die Eigenwerte, *Acta Math., 78*, 1-96, 1946.
9. Burridge, R., The Gel'fand-Levitan, the Marchenko and the Gopinath-Sondi integral equations of inverse scattering theory, regarded in the context of the inverse impulse response problems, *Wave Motion, 2*, 305-323, 1980.
10. Cara, M., Regional variations of higher-mode phase velocities: A spatial filtering method, *Geophys. J.R. Astron. Soc.*, 54, 439-460, 1978.
11. Claerbout, J.F., *Fundamentals of Geophysical data processing*, McGraw-Hill, New York, 1976.
12. Claerbout, J.F., *Imaging the Earth's interior*, Blackwell, Oxford, 1985.
13. Clayton, R. W. and R. P. Comer, A tomographic analysis of mantle heterogeneities from body wave travel time data, *EOS, Trans. Am. Geophys. Un., 64*, 776, 1983.
14. Constable, S.C., R.L. Parker, and C.G. Constable, Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics, 52*, 289-300, 1987.
15. Dahlen, F.A., and J. Tromp, *Theoretical global seismology*, Princeton University Press, Princeton, 1998.
16. Dorren, H.J.S., E.J. Muyzert, and R.K. Snieder, The stability of one-dimensional inverse scattering, *Inverse Problems, 10*, 865-880, 1994.
17. Douma, H., R. Snieder, and A. Lomax, Ensemble inference in terms of Empirical Orthogonal Functions, *Geophys. J. Int., 127*, 363-378, 1996.
18. Dziewonski, A.M., and D.L. Anderson, Preliminary Reference Earth Model, *Phys. Earth. Plan. Int., 25*, 297-356, 1981.
19. Gerver, M.L. and V. Markushevitch, Determination of a seismic wave velocity from the travel time curve, *Geophys. J. Royal astro. Soc., 11* 165-173, 1966.
20. Gilbert, F., Ranking and winnowing gross Earth data for inversion and resolution, *Geophys. J. Royal astro. Soc., 23* 125-128, 1971.
21. Gouveia, W.P., and J.A. Scales, Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. Geophys. Res., 103*, 2759-2779, 1998.
22. Gutenberg, B., Dispersion und Extinktion von seismischen Oberflächenwellen und der Aufbau der obersten Erdschichten, *Physikalische Zeitschrift, 25*, 377-382, 1924.
23. Herglotz, G. Über das Benndorfsche Problem des Fortpflanzungsgeschwindigkeit der Erdbebenstrahlen, *Zeitschrift für Geophys., 8* 145-147, 1907.
24. Keller, J.B., I. Kay, and J. Shmoys, Determination of a potential from scattering data, *Phys. Rev., 102*, 557-559, 1956.
25. Kircpatrick, S., C. Gelatt, and M.P. Vechhis, Optimization by simulated annealing, *Science, 220*, 671-680, 1983.
26. Lanczos, C., *Linear Differential Operators,* Van Nostrand, London, 1961.
27. Levenberg, K., A method for the solution of certain nonlinear problems in least squares, *Quart. Appl. Math., 2*, 164-168, 1944.
28. Lomax, A., and R. Snieder, The contrast in upper-mantle shear-wave velocity between the East European Platform and tectonic Europe obtained with genetic algorithm inversion of Rayleigh-wave group dispersion, *Geophys. J. Int., 123*, 169-182, 1995.
29. Marchenko, V.A., The construction of the potential energy from the phases of scattered waves, *Dokl. Akad. Nauk, 104*, 695-698, 1955.
30. Matsu'ura M. and N. Hirata, Generalized least-squares solutions to quasi-linear inverse problems with a priori information, *J. Phys. Earth, 30*, 451-468, 1982.
31. Mayer, K., R. Marklein, K.J. Langenberg and T.Kreutter, Three-dimensional imaging system based on Fourier transform synthetic aperture focussing technique, *Ultrasonics*, 28, 241-255, 1990.
32. Menke, W., *Geophysical data analysis: discrete inverse theory*, Academic Press, San Diego, 1984.
33. Merzbacher, E., *Quantum mechanics (2nd ed.)*, Wiley, New York, 1970.

34. Montagner, J.P., and H.C. Nataf, On the inversion of the azimuthal anisotropy of surface waves, *J. Geophys. Res., 91*, 511-520, 1986.
35. Mosegaard, K., Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Problems, 14*, 405-426, 1998.
36. Mosegaard, K., and A. Tarantola, Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res., 100*, 12431-12447, 1995.
37. Muyzert, E., and R. Snieder, An alternative parameterization for surface waves in a transverse isotropic medium, *Phys. Earth Planet Int. (submitted)*, 1999.
38. Natterer, F., H. Sielschott, and W. Derichs, Schallpyrometrie, in *Mathematik - Sclüsseltechnologie für die Zukunft*, edited by K.H. Hoffmann, W. Jäger, T. Lochmann and H. Schunk, 435-446, Springer Verlag, Berlin, 1997.
39. Newton, R.G., Inversion of reflection data for layered media: A review of exact methods, *Geophys. J.R. Astron. Soc., 65*, 191-215, 1981.
40. Newton, R.G., *Inverse Schrödinger scattering in three dimensions*, Springer Verlag, Berlin, 1989.
41. Nolet, G., The upper mantle under Western-Europe inferred from the dispersion of Rayleigh wave modes, *J. Geophys., 43*, 265-285, 1977.
42. Nolet, G., Linearized inversion of (teleseismic) data, in *The Solution of the Inverse Problem in Geophysical Interpretation*, edited by R.Cassinis, Plenum Press, New York, 1981.
43. Nolet, G., Solving or resolving inadequate and noisy tomographic systems, *J. Comp. Phys., 61*, 463-482, 1985.
44. Nolet, G., Seismic wave propagation and seismic tomography, in *Seismic Tomography*, edited by G.Nolet, pp. 1-23, Reidel, Dordrecht, 1987.
45. Nolet, G., Partitioned waveform inversion and two-dimensional structure under the Network o f Autonomously Recording Seismographs, *J. Geophys. Res., 95*, 8499-8512, 1990.
46. Nolet, G., S.P. Grand, and B.L.N. Kennett, Seismic heterogeneity in the upper mantle, *J. Geophys. Res., 99*, 23753-23766, 1994.
47. Nolet, G, and R. Snieder, Solving large linear inverse problems by projection, *Geophys. J. Int., 103*, 565-568, 1990.
48. Paige, C.G., and M.A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least-squares, *ACM Trans. Math. Software, 8*, 43-71, 1982.
49. Paige, C.G., and M.A. Saunders, LSQR: Sparse linear equations and least-squares problems, *ACM Trans. Math. Software, 8*, 195-209, 1982.
50. Parker, R.L., *Geophysical Inverse Theory*, Princeton University Press, Princeton, New Jersey, 1994.
51. Passier, M.L., and R.K. Snieder, Using differential waveform data to retrieve local S-velocity structure or path-averaged S-velocity gradients, *J. Geophys. Res., 100*, 24061 - 24078, 1995.
52. Passier, M.L., and R.K. Snieder, Correlation between shear wave upper mantle structure and tectonic surface expressions: Application to central and southern Germany, *J. Geophys. Res., 101*, 25293-25304, 1996.
53. Passier, T.M., R.D. van der Hilst, and R.K. Snieder, Surface wave waveform inversions for local shear-wave velocities under eastern Australia, *Geophys. Res. Lett, 24*, 1291-1294, 1997.
54. Press, W.H., Flannery, B.P., Teukolsky, S.A. and W.T. Vetterling, *Numerical Recipies*, Cambridge University Press, Cambridge, 1989.
55. Rothman, D.H., Nonlinear inversion, statistical mechanics and residual statics estimation, *Geophysics, 50*, 2784-2796, 1985.
56. Sabatier, P.C., Discrete ambiguities and equivalent potentials, *Phys. Rev. A, 8*, 589-601, 1973.
57. Sambridge, M., Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D, *Geophys. J.R. Astron. Soc., 102*, 653-677, 1990.

58. Sambridge, M., and G. Drijkoningen, Genetic algorithms in seismic waveform inversion, *Geophys. J. Int., 109*, 323-342, 1992.
59. Scales, J., and R. Snieder, To Bayes or not to Bayes?, *Geophysics, 62*, 1045-1046, 1997.
60. Scales, J., and R. Snieder, What is noise?, *Geophysics, 63*, 1122-1124, 1998.
61. Sen, M.K., and P.L. Stoffa, Rapid sampling of model space using genetic algorithms: examples of seismic wave from inversion, *Geophys. J. Int., 198*, 281-292, 1992.
62. Sluis, A. van der, and H.A. van der Vorst, Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic tomography, with applications in global seismology and exploration geophysics*, edited by G. Nolet, Reidel, Dordrecht, 1987.
63. Snieder, R., 3D Linearized scattering of surface waves and a formalism for surface wave holography, *Geophys. J. R. astron. Soc., 84*, 581-605, 1986a.
64. Snieder, R., The influence of topography on the propagation and scattering of surface waves, *Phys. Earth Planet. Inter., 44*, 226-241, 1986b.
65. Snieder, R., Surface wave holography, in *Seismic tomography, with applications in global seismology and exploration geophysics*, edited by G. Nolet, pp. 323-337, Reidel, Dordrecht, 1987.
66. Snieder, R., Large-Scale Waveform Inversions of Surface Waves for Lateral Heterogeneity, 1, Theory and Numerical Examples, *J. Geophys. Res., 93*, 12055-12065, 1988.
67. Snieder, R., Large-Scale Waveform Inversions of Surface Waves for Lateral Heterogeneity, 2, Application to Surface Waves in Europe and the Mediterranean, *J. Geophys. Res., 93*, 12067-12080, 1988.
68. Snieder, R., A perturbative analysis of nonlinear inversion, *Geophys. J. Int., 101*, 545-556, 1990.
69. Snieder, R., The role of the Born-approximation in nonlinear inversion, *Inverse Problems, 6*, 247-266, 1990.
70. Snieder, R., An extension of Backus-Gilbert theory to nonlinear inverse problems, *Inverse Problems, 7*, 409-433, 1991.
71. Snieder, R., Global inversions using normal modes and long-period surface waves, in *Seismic tomography*, edited by H.M. Iyer and K. Hirahara, pp. 23-63, Prentice-Hall, London, 1993.
72. Snieder, R., and D.F. Aldridge, Perturbation theory for travel times, *J. Acoust. Soc. Am., 98*, 1565-1569, 1995.
73. Snieder, R.K., J. Beckers, and F. Neele, The effect of small-scale structure on normal mode frequencies and global inversions, *J. Geophys. Res., 96*, 501-515, 1991.
74. Snieder, R., and A. Lomax, Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes, *Geophys. J. Int., 125*, 796-812, 1996.
75. Snieder, R., and G. Nolet, Linearized scattering of surface waves on a spherical Earth, *J. Geophys., 61*, 55-63, 1987.
76. Snieder, R., and M. Sambridge, The ambiguity in ray perturbation theory, *J. Geophys. Res., 98*, 22021-22034, 1993.
77. Spakman, W., S. Van der Lee, and R.D. van der Hilst, Travel-time tomography of the European-Mediterranean mantle down to 1400 km, *Phys. Earth Planet. Int., 79*, 3-74, 1993.
78. Strang, *Linear algebra and its applications*, Harbourt Brace Jovanovisch Publishers, Fort Worth, 1988.
79. Takeuchi, H. and M. Saito, Seismic surface waves, in *Seismology: Surface waves and earth oscillations,* (Methods in computational physics, 11), Ed. B.A. Bolt, Academic Press, New York, 1972.

80. Tams, E., 1921. Über Fortplanzungsgeschwindigkeit der seismischen Oberflächenwellen längs kontinentaler und ozeanischer Wege, *Centralblatt für Mineralogie, Geologie und Paläontologie, 2-3*, 44-52, 1921.
81. Tanimoto, T., Free oscillations in a slighly anisotropic earth, *Geophys. J.R. Astron. Soc., 87*, 493-517, 1986.
82. Tarantola, A., Linearized inversion of seismic reflection data, *Geophys. Prosp., 32*, 998-1015, 1984.
83. Tarantola, A., *Inverse problem theory*, Elsevier, Amsterdam, 1987.
84. Tarantola, A. and B. Valette, Inverse problems = quest for information, *J. Geophys., 50*, 159-170, 1982a.
85. Tarantola, A., and B. Valette, Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys., 20*, 219-232, 1982b.
86. Trampert, J., Global seismic tomography: the inverse problem and beyond, *Inverse Problems, 14*, 371-385, 1998.
87. Trampert, J., and J.J. Lévêque, Simultaneous Iterative Reconstruction Technique: Physical interpretation based on the generalized least squares solution, *J. Geophys. Res.., 95*, 12553-12559, 1990.
88. Trampert, J., J.J. Lévêque, and M. Cara, Inverse problems in seismology, in *Inverse problems in scattering and imaging*, edited by M. Bertero and E.R. Pike, pp. 131-145, Adam Hilger, Bristol, 1992.
89. Trampert, J., and R. Snieder, Model estimations based on truncated expansions: Possible artifacts in seismic tomography, *Science, 271*, 1257-1260, 1996.
90. Trampert, J., and J.H. Woodhouse, Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int., 122*, 675-690, 1995.
91. Trampert, J., and J.H. Woodhouse, High resolution global phase velocity distributions, *Geophys. Res. Lett., 23*, 21-24, 1996.
92. VanDecar, J.C., and R. Snieder, Obtaining smooth solutions to large linear inverse problems, *Geophysics, 59*, 818-829, 1994.
93. van der Hilst, R.D., S. Widiyantoro, and E.R. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature, 386*, 578-584, 1997.
94. van der Hilst, R.D., and B.L.N. Kennett, Upper mantle structure beneath Australia from portable array deployments, *American Geophysical Union 'Geodynamics Series', 38*, 39-57, 1998.
95. van Heijst, H.J. and J.H. Woodhouse, Measuring surface wave overtone phase velocities using a mode-branch stripping technique, *Geophys. J. Int.,* 131, 209-230, 1997.
96. Weidelt, P., The inverse problem of geomagnetic induction, *J. Geophys., 38*, 257-289, 1972.
97. Wiechert, E., Über Erdbebenwellen. I. Theoretisches über die Ausbreitung der Erdbebenwellen, *Nachr. Ges. Wiss. Göttingen*, Math.-Phys. Klasse, 415-529 1907.
98. Woodhouse, J. H. and A. M. Dziewonski, Mapping the upper mantle: Three dimensional modelling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res, 89*, 5953-5986, 1984.
99. Yilmaz, O., Seismic data processing, *Investigations in geophysics, 2*, Society of Exploration Geophysicists, Tulsa, 1987.

Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images

Wolfgang Förstner

Institute for Photogrammetry, Bonn University
Nussallee 15, D-53115 Bonn, e-mail: wf@ipb.uni-bonn.de
http//:www.ipb.uni-bonn.de

Abstract. The paper discusses preprocessing for feature extraction in digital intensity, color and range images. Starting from a noise model, we develop estimates for a signal dependent noise variance function and a method to transform the image, to achieve an image with signal independent noise. Establishing significance tests and the fusion of different channels for extracting linear features is shown to be simplified .

1 Motivation

Signal analysis appears to be one of the most interesting problems common to Geodesy and Photogrammetry. Objects of interest primarily in Geodesy are the gravity field or the topography of the ocean. Objects of primary interest in Photogrammetry are photographic or digital images and all elements of topographic maps. Digital elevation models (DEM) and paths of sensor platforms of air or space vehicles are among the common interests to both fields.

Seen from the type of result Geodesists and Photogrammetrists, however, appear to show increasing differences, if one looks at both research and practical applications: Geodesists mainly are interested in the form, shape or position of geometric objects, especially point, scalar, vector or tensor fields. There was a long period of common interest in point determination, orientation, calibration and DEM derivation, all triggered by the strong tools from estimation and adjustment theory. Since about 10 years photogrammetric research has moved – away from Geodesy – to image interpretation, aiming at recovering not only the geometry of the objects but also their meaning (cf. the review by Mayer 1999).

This development did not really come over night, as the derivation of geometric structures from discretized continuous signals, especially structure lines in DEM appears to be a first step towards extracting symbolic information from sensor data. But this type of transition from an iconic, often raster type of description, to a symbolic, often vector type of description appears to be only slowly penetrate the research in Geodesy.

Interestingly enough, the signal processing part in image processing has only partly been realized by Geodesists, though the type of questions are very similar:

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 165–189, 2000.

predicting a possibly piecewise smooth continuum from sampled data. This task is very similar in both fields, thus also the type of approaches are similar, e. g. using Wiener prediction for restoring images. Of course there are differences in the underlying models. The main difference is the open world Photogrammetry has to cope with, i. e. the impossibility to impose very hard constraints on the observations, as this is the case e. g. in Physical Geodesy.

This is the reason why the author chose a topic for this paper in the overlapping area of Geodesy and Photogrammetry: the processing of two dimensional data prior to the derivation of geometric structures of these signals. Such two dimensional signals are either digital or digitized, black and white, color and multi spectral images or directly measured or derived DEM's. In both cases we assume the data to be given in a regular raster [1].

Starting with modeling the noise behavior of the given data, we develop methods for estimating the noise characteristics. As extracting geometric features starts with detecting signals in noise, which can be interpreted as hypothesis testing, we simplify processing by a noise variance equalization of the given signal. A locally adaptive Wiener filter can be used to smooth the signal depending on the local information content, aiming at smoothing homogeneous areas while preserving edges, lines, corners and isolated bright or dark points. In all cases we discuss digital images and DEM's, or so called range images, in parallel. The fusion of several channels in color images appears to be a challenge as there is no unique way to integrate the information.

The goal of this preprocessing is to exploit the statistical model of the signal as far as possible and thus reduce the number of control parameters for the algorithms. Actually we only need a significance level for distinguishing between signal and noise, as all other properties are estimated from the given data. The procedure for intensity images is partly described in an early version by Förstner 1994 and extended and thogoughly investigated by Fuchs 1998. First investigations into the performance of the feature extraction procedure are presented by Fuchs *et al.* 1994.

The paper collects research at the author's institute of about one decade. The references contain some review papers which point to further reading and alternative approaches.

Notation: We distinguish continuous signals with coordinates (x, y) and discrete signals with coordinates (r, c), r and c standing for rows and columns. Partial derivative operators are $\partial_x \doteq \partial/\partial x$ etc., discrete versions are $\partial_r \doteq \partial/\partial r$ etc. resp. Where necessary, stochastical variables are underscored, e. g. $\underline{g}$. The nor-

[1] This suggests to use Fourier techniques. But this is by no way reasonable: The objects shown in images do not show any periodicity nor is the statistical behavior homogeneous. Wavelet techniques seem to overcome some of the limitations, especially the assumption of homogeneity. But, at least in their Kronecker-version of basis-functions, they lack of providing rotation invariant image properties, and of including higher level knowledge. Only Gabor-wavelets have been widely used for more than two decades for describing texture in digital images for quite some time.

mal density is G_s with standard deviation s as parameter in both coordinate directions.

2 The Image Model

A digital image can be treated as a random process.

In a first approximation it therefore can be characterized by its mean and its variance-covariance structure, thus its first and second moments. In general both are inhomogeneous, thus location dependent, and anisotropic, thus orientation dependent. Moreover, we have to expect the noise statistics to be dependent on the signal.

The distribution of the random process can be expected to be quite complicated in case we start to model the generating process in detail, e. g. using the geometrical-physical model of the sensor. For simplicity we assume the stochastic process to be Gaussian. This allows to easily derive thresholds in the subsequent analysis steps. It is motivated by the central limit theorem and the experience, that deviations form the Gaussian distribution usually can be circumvented by a slight modification of the model.

For modeling reasons we distinguish three types of multi valued images:

1. the *true image* $\underline{\tilde{\boldsymbol{f}}}(x, y)$ is continuous. It is used for describing the image which we would have obtained with an ideal sensor with infinite resolution, no bias and no noise. The true image is modeled as a stochastic process. However, in our context we only impose very limited constraints on the true image: The image area $\mathcal{I}$ is assumed to be partitioned into regions $\mathcal{R}_i$ covering the complete image area. Thus:

$$\mathcal{I} = \bigcup_i \mathcal{R}_i, \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \quad \forall i, j \tag{1}$$

   Within each region the image function is assumed to be homogeneous, in this context smooth, with low gradient or low curvature in all channels. We will specify these properties later, when needed.
2. the *ideal image* $\underline{\boldsymbol{f}}(x, y)$ also is continuous but a blurred version of the true image. In a first instance it is used to explain the limitations of any feature extraction concerning the resolution. It also is used to cover the limited resolution of the sensor internal filtering processes. Technically it is necessary to fulfill the sampling theorem. The blurring also could cover small geometric errors, e. g. due to lens distortion.
3. the *real image* $\underline{\boldsymbol{g}}(r, c)$ is a sampled and noisy version of the ideal image. The sampling is indicated by the indices (r, c), representing *r*ows and *c*olumns of the digital image.

   The image $\underline{\boldsymbol{g}}$ therefore can be written as:

$$\underline{\boldsymbol{g}}(r, c) = \underline{\boldsymbol{f}}(r, c) + \underline{\boldsymbol{n}}(r, c) \tag{2}$$

As noise usually is small we can always assume it to be additive by assuming its characteristic to be signal dependent, i. e. dependent on $\boldsymbol{f}$.

In addition to the variance properties the correlations between neighboring pixels may be taken into account. In original images taken with a good camera or scanner the noise correlations between neighboring pixels are comparably small, i. e. usually less than 50 %. We therefore neglect spatial correlations and only discuss the variance of the noise.

We now discuss the different noise models in detail.

2.1 Intensity Images

Images of CCD-Cameras The noise in single channel intensity images $g(r, c)$ with CCD-Cameras contains three basic components:

1. The noise characteristics is dominated by the Poisson distribution of the photon flux (cf. Dainty and Shaw 1974). This holds, in case the intensity $\underline{g}$ is proportional to the number $\underline{N}$ of the photons. The Poisson distribution is characterized by $\mathsf{E}(\underline{N}) = \mathsf{D}(\underline{N}) = \mu_N = \sigma_N^2$. Thus this noise variance component increases linearly with the intensity.
2. The rounding errors, showing a variance of $1/12$.
3. Electronic noise, being independent on the intensity.

Therefore an adequate model for the noise is

$$\sigma_n^2(g) = a + b\, g \tag{3}$$

In the ideal case both parameters a and b are positive in order to guarantee a positive noise variance. The parameter a should be larger than $1/12$, as it covers the rounding errors. In case the image acquisition device performs a linear transformation on the intensities, the linear behavior of the noise variance still can be observed, however, possibly only in the interval $[g_{\min}, g_{\max}]$. Thus the following two conditions should hold

$$\sigma_n^2(g_{\min}) \geq 1/12, \quad b \geq 0 \tag{4}$$

In case the intensity is not proportional to the number N of photons, but e. g. proportional to $\log N$, the model eq. (3) needs to be modified.

Observe, the noise characteristics are dependent on the intensity which varies throughout the image, thus is totally inhomogeneous. However, for the same intensity it is invariant to position. This is one of the main reasons why it is not reasonable to use Fourier techniques.

In case the sensitivity of the individual pixels depends on the mean intensity in the neighborhood for increasing the radiometric resolution, the simple model eq. (3) does not hold anymore. Without knowing the type of adaptivity of the sensitivity to the intensity in the neighborhood there is no simple rule how to model the noise variance.

Digitized Photographs Digitized photographs, especially digitized aerial images show quite complex noise characteristics. In a first instance one would expect a similar increase of the noise variance with the signal as in CCD-images. Due to film characteristics, film development and scanner characteristics, especially also the scanner software, the noise characteristics will not follow a simple function of the intensity.

We now assume the noise characteristics only depends on the intensity, but in an arbitrary manner. Thus we have:

$$\sigma_n^2(g) = s(g), \quad s(g) \geq 1/12 \tag{5}$$

2.2 Color Images

Color images $\underline{g} = (\underline{g}_k)$ are multichannel images, where each of the K channels represents a specific spectral band. In principle the number of bands is not restricted to three, as in RGB-images, but may be any number, as in multi spectral images.

The sensitivity of the sensors or the filters used for separating the spectral bands may overlap. This results in correlations of the signals $\underline{g}_{k'}$ and $\underline{g}_{k''}$ in different channels k' and k''.

However, the *noise* of the different channels can certainly be assumed to be independent, unless the multi channel sensor does not transform originally independent channels, e. g. by a color transformation.

In the subsequent image analysis steps we do not refer to the colors in the way the visual system perceives them. We interpret the different channels as the result of independent physical sensing processes and therefore intensionally do *not* perform any color transformation.

The result of this argumentation is simple: Every channel has a statistically independent noise characteristic, either following eq. (3) leading to:

$$\sigma_{n_k}^2(g_k) = a_k + b_k\, g_k \quad k = 1, ..., K \tag{6}$$

with $2K$ parameters a_k and b_k specifying the noise variances or, following eq. (5), to

$$\sigma_{n_k}^2(g_k) = s_k(g_k) \quad k = 1, ..., K \tag{7}$$

with K different functions s_k.

2.3 Range Images

Range images in principle are single channel images as intensity images are. Depending on the type of sensor the noise can be assumed to depend on the distance, e. g. in case of time of flight sensors or image stereo sensors. The noise characteristics also may depend on the direction of the sensing ray and on the slope of the surface. The type of dependency not really has been investigated.

On the other hand sensor internal error sources or the resampling process, transferring irregular data into a regular grid, may dominate and lead to a noise

variance being constant over the full range image. We therefore assume the noise to be constant in range images. Making it dependent on position, orientation or other attributes can be taken into account if necessary.

As edge extraction in range images is based on the two channel image:

$$\underline{g}(r,c) = \nabla d(r,c) = \begin{pmatrix} d_r \\ d_c \end{pmatrix} = \begin{pmatrix} \dfrac{\partial d}{\partial r} \\ \dfrac{\partial d}{\partial c} \end{pmatrix} = \begin{pmatrix} \partial_r * d \\ \partial_c * d \end{pmatrix} \tag{8}$$

we want to discuss the noise characteristics of this two channel image.

The covariance matrix of the two values $\underline{g}_k(r,c), k = 1,2$ is diagonal:

$$\boldsymbol{\Sigma}_{gg} = \mathsf{D}(\nabla d) = \sigma_d^2 \, t\boldsymbol{I} \tag{9}$$

where the factor t depends on the filter kernel used for determining the gradient, and $\boldsymbol{I}$ is the unit matrix. This holds if the differentiation kernels $\partial_x(x,y)$ and $\partial_y(x,y)$ are orthogonal, thus

$$\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} \partial_x(x,y)\partial_y(x,y)\, dx\, dy = 0 \tag{10}$$

We often use Gaussian kernels

$$G_s(x,y) = \frac{1}{2\pi s^2} \mathrm{e}^{-\dfrac{(x^2+y^2)}{2s^2}} \tag{11}$$

or its derivatives, e. g.

$$G_{x;s}(x,y) = \frac{\partial}{\partial x} G_s(x,y) = -\frac{x}{s^2} G_s(x,y) \tag{12}$$

Then (10) obviously holds.

E. g. the Sobel kernels

$$\partial_{r,S} \doteq \left(\frac{\partial}{\partial r}\right)_S = \frac{1}{8} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \tag{13}$$

$$\partial_{c,S} \doteq \left(\frac{\partial}{\partial c}\right)_S = \frac{1}{8} \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \tag{14}$$

are discrete approximations of $G_{x;s}(x,y)$ and $G_{y;s}(x,y)$ with $s \approx 0.7$ which can be proved by comparing the energies of the two filter kernels:

$$\sum_{r,c} \partial_{r,S}^2 = \frac{3}{16}, \qquad \int_{x,y} G_{x;s}^2(x,y) dx\, dy = \frac{1}{8\pi s^4} \tag{15}$$

which are identical for $s = 0.6787\ldots$.

We now obtain the factor t_S in (9)

$$t_S = \left(\frac{1}{8}\right)^2 ((-1)^2+(-2)^2+(-1)^2+0^2+0^2+0^2+1^2+2^2+1^2) = \frac{12}{64} = \frac{3}{16} \quad (16)$$

which is due to the linear dependency of the gradient of the original distances:

$$d_r = \frac{1}{8}(- d(r-1,c-1) - 2d(r-1,c) - d(r-1,c+1) \quad (17)$$

$$+ d(r+1,c-1) + 2d(r+1,c) + d(r+1,c+1)) \quad (18)$$

$$d_c = \frac{1}{8}(- d(r-1,c-1) + d(r-1,c+1) - 2d(r,c-1) \quad (19)$$

$$+ 2d(r,c+1) - d(r+1,c-1) + d(r+1,c+1)) \quad (20)$$

and identical with the energy in (15a).

As the sum of products of the coefficients for d_r and d_c sum to 0, the two partial derivatives are uncorrelated, showing the Sobel to be a consistent approximation of the first Gaussian derivatives.

3 Noise Variance Estimation

The noise variance needs to be estimated from images. There are three possible methods to obtain such estimates:

1. *Repeated images*: Taking multiple images of the same scene without changing any parameters yields repeated images. This allows to estimate the noise variance for each individual pixel independently. This certainly is the optimal method in case no model for the noise characteristics is available and can be used as a reference.
   The method is the only one which can handle the case where there is no model for the noise characteristics.
   We used it for finding out the noise model for the scanner component of digitized aerial images (cf. fig. 1ff, taken from Waegli 1998).
   The disadvantage of this method is the need to have repeated images, which, e. g. in image sequences is difficult to achieve.
2. *Images of homogeneous regions:* Images of homogeneous regions, thus regions with piecewise constant or linear signal, allows to estimate the noise variance from one image alone.
   The disadvantage is the requirement for the segmentation of the images into homogeneous regions. Moreover, it is very difficult to guarantee the constancy or linearity of the true intensity image within the homogeneous regions. Small deviations from deficiencies in the illumination already jeopardize this method.
   The method is only applicable in case the noise only depends on the signal.
3. *Images with little texture:* Images with a small percentage of textured regions allow to derive the noise variance from the local gradients or curvature. For the larger part of the image they can be assumed to have approximately zero

mean. Thus presuming a small percentage of textured regions assumes the expectation of the gradient or the curvature in the homogeneous regions to be negligible compared to the noise.

Also this method is only applicable in case the noise characteristics is only depending on the signal.

We want to describe this method in more detail. We first discuss the method for intensity images. The generalization to range images is straight forward.

3.1 Estimation of the Noise Variance in Intensity Images

Case I: Constant Noise Variance The idea is to analyze the histogram of the gradient magnitude of the image in the area where there are no edges and no texture. The procedure given here is similar to that proposed in (Förstner 1991).

We now need to specify the model for the ideal image f. We assume that a significant portion $\mathcal{H}$ of the image area $\mathcal{I}$ is homogeneous, thus shows locally constant intensity, thus $\mu_f = const.$. Adopting notions from statistical testing $H_0 = (r,c) \in \mathcal{H}$ is the null-hypothesis, i. e. the hypothesis a pixel belongs to a homogeneous region. Thus

$$\mathsf{E}(\nabla \underline{f} | H_0) = \mathbf{0} \tag{21}$$

The other area $\mathcal{I} - \mathcal{H}$ covers edges and textured areas with significantly larger gradients.

Then, as to be shown, the histogram of the homogeneoity measure $h = |\nabla g|$ shows exponential behavior in its left part representing the noise in the image and arbitrary behavior in the right part representing the edges:

We assume the intensities to be Gaussian distributed with fixed mean and random noise. Assuming the simple gradient kernels

$$\left(\frac{\partial}{\partial r}\right)_0 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \qquad \left(\frac{\partial}{\partial c}\right)_0 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \tag{22}$$

neglecting the scaling factores $1/2$, we obtain the gradient

$$\nabla g = \begin{pmatrix} g_r \\ g_c \end{pmatrix} = \begin{pmatrix} g_{r+1,c} - g_{r-1,c} \\ g_{r,c+1} - g_{r,c-1} \end{pmatrix} \tag{23}$$

which is Gaussian distributed with covariance matrix

$$\mathsf{D}\left(\begin{matrix} \underline{g}_r \\ \underline{g}_c \end{matrix} \middle| H_0 \right) = \sigma_{n'}^2 \boldsymbol{I} \tag{24}$$

Here we use the convention

$$\sigma_{n'}^2 = \sigma_{n_r}^2 = \sigma_{n_c}^2 \tag{25}$$

which in general is given by (cf. eq. (15b), Fuchs 1998)

$$\sigma_{n'}^2 = \int_{x,y} G_{x;s}^2(x,y) dx\, dy = \frac{1}{8\pi s^4} \sigma_n^2\, \sigma_n^2, \quad \text{or} \quad \sigma_{n'}^2 = \sum_{r,c} \partial_r^2(r,c)\, \sigma_n^2 \tag{26}$$

In our case eq. (22) leads to

$$\sigma_{n'}^2 = 2\sigma_n^2 \tag{27}$$

The squared gradient magnitude measures the homogeneity h

$$h_\nabla(r,c) = |\nabla g(r,c)|^2 = g_r^2(r,c) + g_c^2(r,c) \tag{28}$$

It is the sum of two squares of Gaussian variables.

In case the mean $\mu_g = \mu_f$ of $\underline{g}$ is constant in a small region, thus the model eq. (21) holds, the squared gradient magnitude is χ_2^2 or exponentially distributed with density function (neglecting the index ∇ for simplicity)

$$p(h|H_0) = \frac{1}{\mu_h} \mathrm{e}^{-\frac{h}{\mu_h}} \tag{29}$$

and mean

$$\mathsf{E}(\underline{h}|H_0) = \mu_h = 4\sigma_n^2 \tag{30}$$

Therefore we are able to estimate the parameter μ_h from the empirical density function in the following way:

1. Set the iteration index $\nu = 0$. Specify an approximate value $\sigma_n^{(0)}$ for the noise standard deviation. Use $\mu_h^{(0)} = 4\sigma_n^{2(0)}$ as approximate value for the gradient magnitude.
2. Determine all $h(r,c)$
3. Take the mean $m^{(\nu)}$ of all values $h(r,c) < \mu_h^{(\nu)}$. Its expected value is given by

$$\mu_m^{(\nu)} = \frac{\int_{h=0}^{\mu_h^{(\nu)}} hp(h|H_0)dh}{\int_{h=0}^{\mu_h^{(\nu)}} p(h|H_0)dh} = \frac{\mathrm{e}-2}{\mathrm{e}-1}\mu_h^{(\nu)} \tag{31}$$

   in case the edges or textured areas do not significantly contribute to this mean. Thus a refined estimate $\mu_h^{(\nu+1)}$ for μ_h is given by:

$$\mu_h^{(\nu+1)} = \frac{\mathrm{e}-1}{\mathrm{e}-2} m^{(\nu)} \approx 2.392\, m^{(\nu)} \tag{32}$$

4. Set $\nu = 1$ and repeat step 3.

Usually, only two iterations are necessary to achieve convergence. A modification would be, to take the median of the values $h(r,c)$ as a robust estimate and compensate for the bias caused 1) by taking the median instead of the mean and 2) by the edge pixels (cf. Brügelmann and Förstner 1992).

This procedure can be applied to every channel in a multi channel image, especially in color images or in gradient images of range images.

Fig. 1. *shows three image sections of 300 × 300 pixels from three different aerial images.*

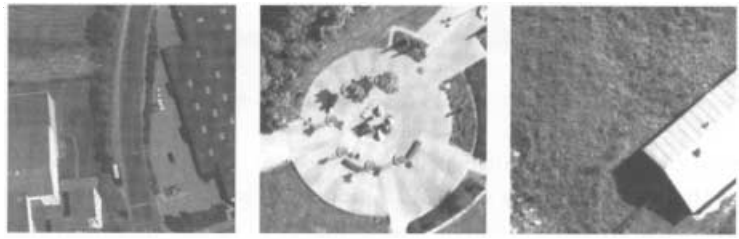


Fig. 2. *shows histograms image sections of 300 × 300 pixels. Observe that the frequencies of the intensities vary heavily. Only for intensities with high frequencies one can expect to obtain reliable estimates for the noise variance (from Waegli 1998).*

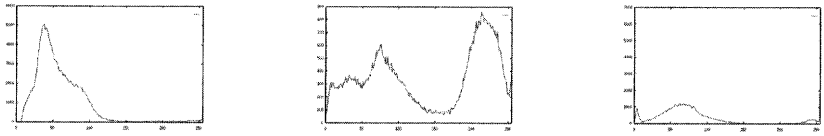


Fig. 3. *shows the noise standard deviation with 32 grey level intervals as a function of the intensity. The estimated variances increase with intensity but not linearly and showing larger variation in areas with low intensity frequency (from Waegli 1998).*

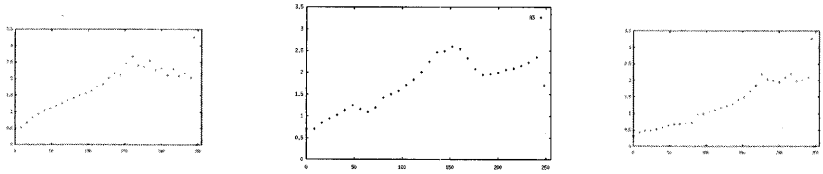
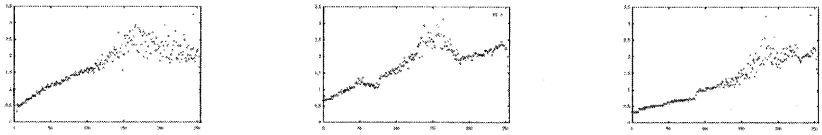


Fig. 4. *shows the noise standard deviation with no partitioning as a function of the intensity. The effect of the intensity frequency onto the reliability of the estimated noise variances is enlarged, but taking groups of 8 intensities is fully acceptable (from Waegli 1998). The small step in the right figure is likely due to the scanner software.*



Case II: General Noise Variance In case the noise variance is not constant over the whole image area and can be assumed only to depend on the intensity, we need to parameterize the noise variance function $\sigma_n^2 = s(g)$ in some way.

The easiest possibility is to assume it to be continuous. Then we can partition the range $[0..G]$ of all intensities g into intervals $I_\gamma, \gamma = 1..\Gamma$ and assume the noise variance to be constant in each interval.

Thus we repeat the procedure of subsection 3.1 for each intensity interval under the condition $g \in I_\gamma$.

The choice of the intervals obviously requires some discussion, as it may significantly influence the solution. Taking a set of constant intervals may lead to intervals where no intensities belong to, even in case one would restrict to the real range $[g_{\min}, g\max]$. Therefore the intervals should be chosen such that

1. they contain enough intensity values..
   The number should be larger than 100 in order to yield precise enough estimates for the noise variances, which in this case has a relative (internal) accuracy better than 10 %. The number of intervals should be chosen in dependency of the expected roughness of $s(g)$. For aerial images we have made good experiences with intervals between 1 and 8 grey values on image patches of 300 × 300 pixels (cf. Waegli 1998 and figs. 3 and 4).
2. they contain an equal number of intensities. This may easily be achieved by using the histogram of the intensities.

Case III: Linear Noise Variance The case of linear noise variance can be handled in a special way. The parameters a and b of the linear noise variance function $\sigma_n^2(g) = a + bg$ can be determined without partitioning the intensity range, but by directly performing a robust weighted linear regression on the gradient magnitudes. The weights, can be derived from the χ_2^2-characteristics of the gradient magnitudes, whereas the robustness can be achieved by excluding all pixels with too large gradient magnitudes, making the threshold dependent on the estimated noise variance function (cf. Brügelmann and Förstner 1992).

3.2 Noise Estimation in Range Images

In range images we assume the curvature to be small. Thus we assume the Hessian of the image function to have zero expectation:

$$\mathsf{E}(\boldsymbol{H}(\underline{f})|H_0) = \mathsf{E}\left(\begin{pmatrix} \underline{f}_{rr} & \underline{f}_{rc} \\ \underline{f}_{cr} & \underline{f}_{cc} \end{pmatrix} \middle| H_0 \right) = \mathbf{0} \tag{33}$$

Then using the kernels

$$c_1 = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \quad c_2 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \tag{34}$$

which both measure the local torsion lead to the homogeneity measure

$$h_\tau = (c_1 * g)^2 + (c_2 * g)^2 \tag{35}$$

which again is χ^2_2-distributed in case the mean curvature is locally zero. The expectation of h_τ is

$$\mathsf{E}(h_\tau) = 8\sigma_n^2 \tag{36}$$

allowing to use the same procedure for estimating σ_n^2 from the histogram of h_τ.

4 Variance Equalization

The estimated noise variance function may be used when thresholding on functions of the intensity image. This was the line of thought in the feature extraction up to now (cf. Förstner 1994, Fuchs 1998 and sect. 6).

The disadvantage is the need to refer the threshold to some center pixel of an operating window. E. g. when thresholding the gradient magnitude determined with the Sobel operator, one needs to use a representative intensity in the 3×3-window for determining the signal dependent noise variance. Actually all 9 intensity values have different noise variance, making a rigorous error propagation tedious, without being sure that the rigorous calculations are much better than the approximate calculations.

In order to avoid this situation one can transform the image such that the intensities in the resulting image have equal noise variance.

4.1 Principle

The idea is to find a pixel wise transformation

$$\dot{g} = T(g) \tag{37}$$

such that $\sigma_{\dot{g}} = \sigma_0 = const.$ is independent on g. From

$$\sigma_{\dot{g}}^2 = \sigma_0^2 = \left(\frac{dT}{dg}\right)^2 \sigma_g^2 = \left(\frac{dT}{dg}\right)^2 s(g) \tag{38}$$

we find

$$dT = \sigma_0 \frac{dg}{\sqrt{s(g)}} \tag{39}$$

and obtain

$$\dot{g} = T(g) = \sigma_0 \int_{l=0}^{g} \frac{dl}{\sqrt{s(l)}} + C \tag{40}$$

where the two parameters σ_0 and C can be chosen freely, e. g. that $T(g_{\min}) = g_{\min}$ and $T(g_{\max}) = g_{\max}$ resulting in an image $\dot{g} = T(g)$ with the same intensity range.

This *variance equalization* simplifies subsequent analysis steps as the noise characteristics is homogeneous throughout the image.

4.2 Linear Variance Function

In case $\sigma_n^2(g) = s(g) = a + b\,g$ we can evaluate the integral and obtain:

$$\dot{g} = t(g) = \frac{2\sigma_0}{b}\sqrt{a + b\,g} + C \tag{41}$$

again with the possibility to freely choose the two parameters σ_0 and C.

4.3 General Variance Function

In the case of general $s(g)$ and piecewise linear approximation the integral can also be evaluated algebraically.

Alternatively, one could sample $s(g)$ and determine

$$\dot{g} = t(g) = \sigma_0 \sum_{l=0}^{g} \frac{1}{\sqrt{s(l)}} + C \tag{42}$$

5 Information Preserving Filtering

After having characterized the image noise and possibly transformed the image we now want to increase the signal to noise ratio by filtering the image. Usually signal and noise cannot completely be separated. Therefore any filtering meant to suppress noise at the same time suppresses signal. Depending on the signal and the noise model different filters may be optimal in increasing the signal to noise ratio.

We want to develop a filter which locally leads to a best restoration, i. e. prediction of the underlying signal. We start with the most simple case, where both signals are Gaussian with known homogeneous and isotropic statistics, requiring a Wiener filter as optimal filter. We then modify this filter to locally adapt to the signal content, which we estimate from the given data. The presentation follows (Förstner 1991).

5.1 The Wiener Filter

The Wiener filter starts from the model eq. (2)

$$\underline{\boldsymbol{g}} = \underline{\boldsymbol{f}} + \underline{\boldsymbol{n}} \tag{43}$$

with

$$\mathsf{E}(\underline{\boldsymbol{f}}) = \mathsf{E}(\underline{\boldsymbol{n}}) = \mathbf{0} \tag{44}$$

$$\mathsf{D}(\underline{\boldsymbol{f}}) = \boldsymbol{\Sigma}_{ff} \qquad \mathsf{D}(\underline{\boldsymbol{n}}) = \boldsymbol{\Sigma}_{nn} \qquad \mathsf{Cov}(\underline{\boldsymbol{f}}, \underline{\boldsymbol{n}}) = \boldsymbol{\Sigma}_{fn} = \mathbf{0} \tag{45}$$

Under these conditions the filter leading to best, i. e. most precise results is the Wiener filter (Wiener 1948, Moritz 1980, p. 80):

$$\underline{\widehat{\boldsymbol{g}}} = \boldsymbol{\Sigma}_{fg}\boldsymbol{\Sigma}_{gg}^{-1}\underline{\boldsymbol{g}} = \boldsymbol{\Sigma}_{ff}(\boldsymbol{\Sigma}_{ff} + \boldsymbol{\Sigma}_{nn})^{-1}\underline{\boldsymbol{g}} = \boldsymbol{W}\underline{\boldsymbol{g}} \tag{46}$$

A slightly different approach has been developed in and (Weidner 1994a and Weidner 1994b). It integrates a variance component estimation for determining the local statistical behavior of the signal and the noise (cf. Förstner 1985). The approximate solution proposed in (Förstner 1991) and given and generalized for range images below, has proven to be similar in performance but computationally much more efficient.

5.2 Approximation of the Autocovariance Function

The statistics of the image function theoretically can be described by its auto covariance function. This, however, would not capture the locally varying structure. This structure also shows severe orientation dependencies at least at edges but also in textured areas, e. g. within agricultural areas.

Therefore we want to describe the local structure of the image function f by an inhomogeneous and anisotropic auto covariance function $k(\boldsymbol{x})$, depending on the difference $\boldsymbol{x} = (r_2 - r_1, c_2 - c_1)^\mathsf{T}$.

The auto covariance function can be assumed to decay smoothly with increasing distance from $\boldsymbol{x} = \boldsymbol{0}$. In a second order approximation it can be characterized by the curvature at $\boldsymbol{x} = \boldsymbol{0}$ (cf. Moritz 1980, p. 175). As an example, assume the auto covariance function has the form:

$$k(\boldsymbol{x}) = \sigma_f^2 \mathrm{e}^{-\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{S}\boldsymbol{x}} \tag{47}$$

with a symmetric positive definite matrix $\boldsymbol{S}$, then its Hessian containing the second derivatives at $\boldsymbol{x} = \boldsymbol{0}$ is given by

$$\boldsymbol{H}_k = \begin{pmatrix} k_{rr} & k_{rc} \\ k_{rc} & k_{cc} \end{pmatrix} = -\sigma_f^2 \boldsymbol{S} \tag{48}$$

Now, due to the moment theorem (cf. Papoulis 1984), the Hessian of the auto covariance function is directly related to the variances and covariances of the gradient of the true signal f by

$$\boldsymbol{H}_k = -\mathsf{D}(\nabla f) = -\begin{pmatrix} \sigma_{f_r}^2 & \sigma_{f_r f_c} \\ \sigma_{f_r f_c} & \sigma_{f_c}^2 \end{pmatrix} = -G_t * \begin{pmatrix} f_r^2 & f_r f_c \\ f_c f_r & f_c^2 \end{pmatrix} = -\overline{\nabla f \, \nabla^\mathsf{T} f} \tag{49}$$

This relation allows to capture the essential part of the auto covariance function from a quite local computation, namely the local dispersion of the gradient of the image function.

As we do have no access to the true image f, but the noisy image g we need to estimate $\mathsf{D}(\nabla f)$. This either can be achieved by iteratively estimating f or by using the relation between $\mathsf{D}(\nabla f)$ and $\mathsf{D}(\nabla g)$:

$$\mathsf{D}(\nabla g) = \mathsf{D}(\nabla f) + \sigma_{n'}^2 \boldsymbol{I} \tag{50}$$

The covariance matrix $\mathsf{D}(\nabla g)$ and $\sigma_{n'}^2$ can be estimated from the given image using (49) and the procedure given in sect. 3.1. With the eigenvalue decomposition $\widehat{\mathsf{D}}(\nabla g) = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^\mathsf{T}$ we obtain

$$\widehat{\mathsf{D}}(\nabla f) = \boldsymbol{C}\widehat{\boldsymbol{\Lambda}}\boldsymbol{C}^\mathsf{T} \tag{51}$$

with

$$\widehat{\lambda}_i = \max(\lambda_i - \sigma_{n'}^2, 0) \tag{52}$$

5.3 An Adaptive Wiener Filter for Intensity Images

In principle the vectors $\underline{g}$ etc. in eqs. (43) – (46) contain the complete image. This is due to the correlation structure of $\underline{f}$, which links all pixels, leading to a full covariance matrix $\boldsymbol{\Sigma}_{ff}$.

The filter we propose is realized as a, possibly iterated, convolution with a small but signal dependent kernel.

In order to motivate this setup, assume the vector $\underline{g}$ contains only the 9 values of a 3×3 window. In a first instance, the covariance between intensity values is assumed to only depend on the distance, thus is homogeneous and isotropic, and can be represented by an auto covariance function

$$\mathsf{Cov}(g(r_1, c_1), g(r_2, c_2)) = k(d_{12}) \tag{53}$$

where $d_{12} = \sqrt{(r_2 - r_1)^2 + (c_2 - c_1)^2}$ is the distance between the pixels. Assuming the correlation to fall off rapidly enough, the Wiener filter with isotropic covariance function can be written as convolution with a small completely symmetric kernel

$$\boldsymbol{w}^{(i)} = \begin{pmatrix} w_1 & w_2 & w_1 \\ w_2 & w_0 & w_2 \\ w_1 & w_2 & w_1 \end{pmatrix} \tag{54}$$

In case the covariance function is anisotropic the kernel will only be point symmetric:

$$\boldsymbol{w}^{(a)} = \begin{pmatrix} w_1 & w_2 & w_3 \\ w_4 & w_0 & w_4 \\ w_3 & w_2 & w_1 \end{pmatrix} \tag{55}$$

As $\mathsf{E}(\boldsymbol{f}) = \boldsymbol{0}$ the weights do not sum to 1 in general. In order to be able to handle signals with $\mathsf{E}(\boldsymbol{f}) \neq \boldsymbol{0}$ the elements need to sum to 1, thus $w_0 + 4w_1 + 4w_2 = 1$, in order that a constant signal is reproduced by the filtering.

The idea now is to locally determine the covariance function and approximately determine the 5 parameters w_0 to w_4 in $\boldsymbol{w}^{(a)}$

We propose a simple expression for the convolution kernel $\boldsymbol{w}$. It should fulfill the following conditions:

1. It should allow strong smoothing in homogeneous areas of the image or, equivalently, in case the signal has the same or a lower variance than the noise. Therefore it needs to depend on the variance of the image noise.
2. It should preserve the sharpness of edges and corners. Therefore it needs to depend on the local auto covariance function.

Thus the idea is to make the coefficients w_i dependent on the curvature $\boldsymbol{S}$ of the auto covariance function and the noise variance σ_n^2.

We propose a simple weight function

$$w(\boldsymbol{x}) = \frac{C}{1 + \frac{1}{2}\frac{\boldsymbol{x}^\mathsf{T}\overline{\nabla f\,\nabla^\mathsf{T} f}\boldsymbol{x}}{2\sigma_{n'}^2}} \tag{56}$$

the constant C being used for normalization.

It has the following properties:

1. For a constant noisy signal $\underline{g} = f + \underline{n}$ the estimated covariance matrix $\widehat{\mathsf{D}}(\nabla f) = \overline{\nabla\widehat{f}\,\nabla^\mathsf{T}\widehat{f}} \approx \mathbf{0}$, thus all weights are equal. The weighting kernel is the box filter R_3

$$\boldsymbol{w} = \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \tag{57}$$

2. in case of a steep straight edge in r-direction we will have

$$\mathsf{D}(\nabla f) = \begin{pmatrix} \sigma_{f_r}^2 & 0 \\ 0 & 0 \end{pmatrix} \tag{58}$$

thus for $\sigma_{f_r}^2 >> \sigma_{n'}^2$ we obtain the weight matrix

$$\boldsymbol{w} = \frac{1}{3}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{59}$$

3. Finally, in case of a sharp symmetric corner we will have

$$\mathsf{D}(\nabla f) = \frac{\sigma_{f'}^2}{2}\boldsymbol{I} \tag{60}$$

thus for $\sigma_{f'}^2 >> \sigma_{n'}^2$ we obtain

$$\boldsymbol{w} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{61}$$

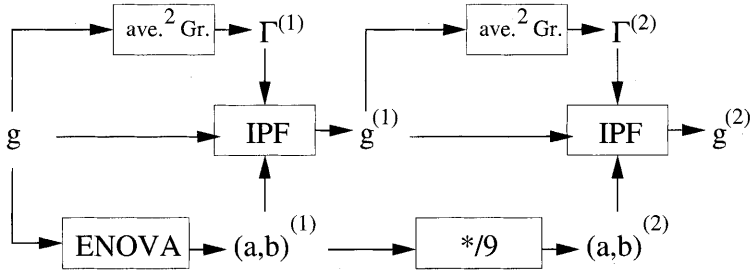
which does not change the signal. The same situation holds for an isolated bright or dark spot.

Thus the smoothing properties are those intended.

Three points are worth mentioning when implementing this filter:

1. As mentioned above, the true signal is not available. A possibility to obtain the characteristics of the true signal f would be to use the estimated covariance matrix of the gradients of f in eq. (52).
2. In order to avoid the eigenvalue decomposition at each pixel, one could alternatively use an estimate $\widehat{f}$ to determine the covariance matrix $\mathsf{D}(\nabla f)$. In a first implementation one could use the signal g as an estimate for f.

Fig. 5. *shows an iterative version of the information preserving filter. The matrix* $\boldsymbol{\Gamma}$ *denotes the dispersion* $\mathsf{D}(\nabla g \nabla^{\mathsf{T}} g)$ *or* $\mathsf{D}(\boldsymbol{H}^2(d))$ *of the gradients or the Hessian resp..* ENOVA *denotes the estimation of the noise variance, here assumed to yield the two parameters a and b of a linear noise variance model.* IPF *is the simple version information preserving filtering with locally adaptive weights. Observe, we reduce the variances by a factor of 9, which is the maximum reduction which can be achieved by filtering with a 3 × 3-kernel.*



A better approximation of the true signal could be obtained by starting with a smoothed version $\overline{f} = G_s * f$ for determining the the covariance matrix $\mathsf{D}(\nabla \overline{f})$. Then adaptive weights are more realistic. This especially holds, in case the noise is not very small. Otherwise, the covariance matrix would not show the full structure of the signal.

3. In case the noise is very large, we would like to smooth more than with a box filter (57). Then we could apply the filter iteratively. This is indicated in Fig. 5.

In our implementation we use the filter kernels (Roberts gradient):

$$\left(\frac{\partial}{\partial r}\right)_R = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \qquad \left(\frac{\partial}{\partial c}\right)_R = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \tag{62}$$

for determining the gradients and the filter kernel

$$R_4 = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{63}$$

instead of G_t for the integration according to eq. (49).

5.4 An Adaptive Wiener Filter for Range Images

In range images the trend $\mathsf{E}(\underline{d})$ cannot be assumed to be locally constant, or, equivalently, $\mathsf{E}(\nabla \underline{d}) \neq \mathbf{0}$. The most simple assumption would be to assume a linear trend, or, equivalently $\mathsf{E}(\boldsymbol{H}(\underline{d})) = \mathbf{0}$ (cf. above).

Starting from the range image d, we then can use the gradient image ∇d as two channel image: $\boldsymbol{f} = \nabla d$. Using the relation $\boldsymbol{H}(d) = \nabla\nabla d$ the mean quadratic

gradient $G_t * (\nabla \boldsymbol{f}\, \nabla^\mathsf{T} \boldsymbol{f})$ then is given by:

$$\overline{\boldsymbol{H}^2(d)} = G_t * (\nabla \boldsymbol{f}\, \nabla^\mathsf{T} \boldsymbol{f}) = G_t * \boldsymbol{H}^2(d) \tag{64}$$

Thus we can use the weight function

$$w(\boldsymbol{x}) = \frac{1}{1 + \frac{1}{2} \frac{\boldsymbol{x}^\mathsf{T} \overline{\boldsymbol{H}^2(d)} \boldsymbol{x}}{2\,\sigma_{n''}^2}} \tag{65}$$

6 Fusing Channels: Extraction of Linear Features

We now want to discuss one of the many algorithms for extracting geometric features from digital images, namely the extraction of edge pixels, which then may be grouped to linear image features (Fuchs 1998).

Edge pixels are meant to be borders of homogeneous regions. Thus they show two properties:

1. The homogeneity is significantly larger than in homogeneous regions.
2. The homogeneity is locally maximum across the edge.

Therefore we need to discuss detection and localization of edge pixels.

The problem we want to solve is how to integrate the information of different channels of a multi-channel image.

6.1 Detecting Edge Pixels

The detection of edge pixels can be performed as a hypothesis test, namely calling all pixels edge pixels in case the homogeneity $h_1 = |\nabla g|^2$ is significant with respect to the noise in the image. Thus edge pixels actually are pixels which are non-homogeneous, thus indicate places where there may be signal, e. g. and edge.

Detecting Edge Pixels in Intensity Images We first assume constant noise variance.

In order to capture the full information in the neighborhood of a pixel we test the averaged homogeneity $\overline{h_1} = G_t * h_1$ for significance. This at the same time allows to detect thin lines, not only edges, as the smoothing caused by the convolution with G_t then covers both sides of a thin line.

We use the test statistic

$$z_1 = \frac{\overline{h_1}}{\sigma_{n'}^2} = \frac{\overline{|\nabla g|^2}}{\sigma_{n'}^2} = \frac{tr\left(G_t * (\nabla \underline{g} \nabla^\mathsf{T} \underline{g})\right)}{\sigma_{n'}^2} \tag{66}$$

which in case the pixels lies in a homogeneous region is χ_2^2-distributed. Observe, we need to use the same differentiation kernel for determining the partial derivatives ∇g of the signal as for determining $\sigma_{n'}^2$ (cf. eq. (26)).

Fig. 6. *shows first a section of a digitized aerial image. The large compound image below shows left the result of the information preserving filter (IPF)and below the difference to the original, and right the result of a smoothing with a Gaussian with* $\sigma = 1.4$ *and below the difference to the original. Observe: the IPF smoothes the vegetation areas and the roof tilings quite strongly without smearing out the edges and the line being a hair of the analog film. The Gaussian filter does not smooth the vegetation areas and the tilings too much but at the same time smoothes the edges. The difference images demonstrate that the edges are better preserved using IPF. The notion* information preserving *obviously needs an application dependent discussion, as a user might be interested in the roof tilings.*

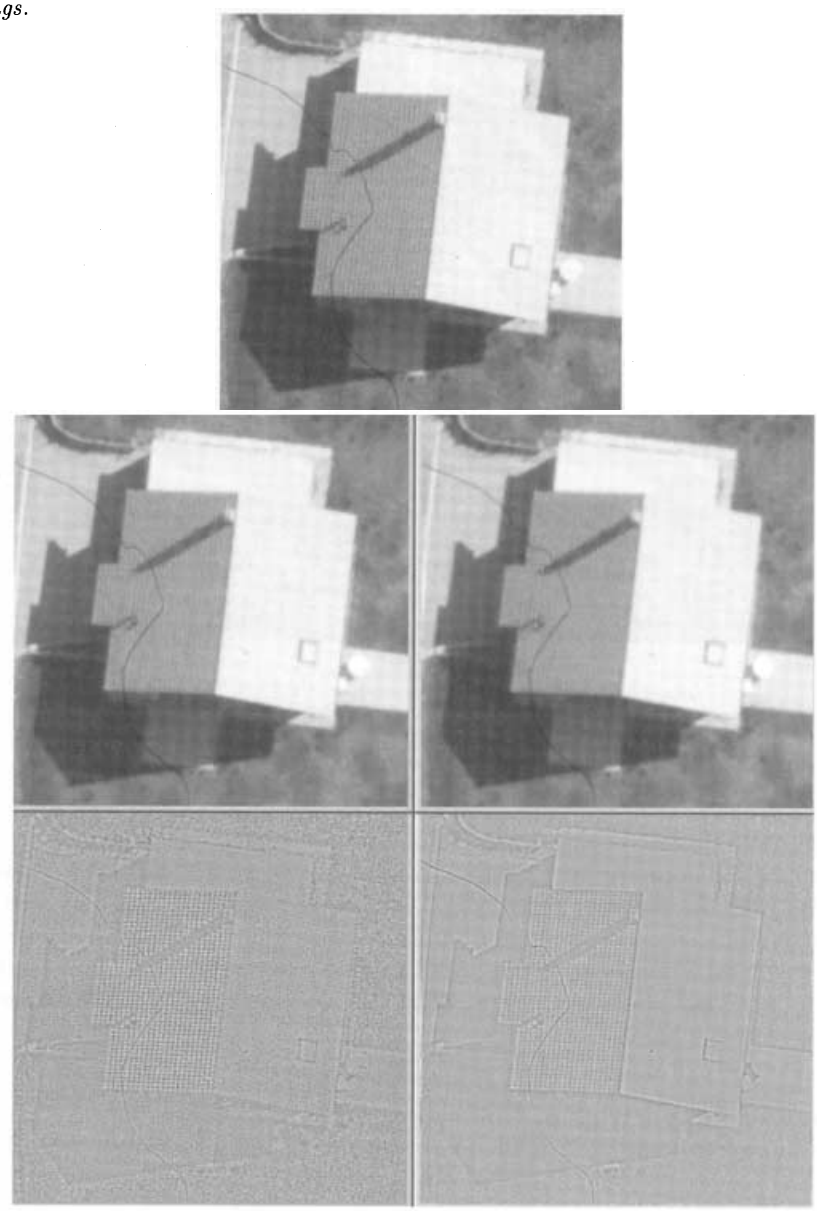
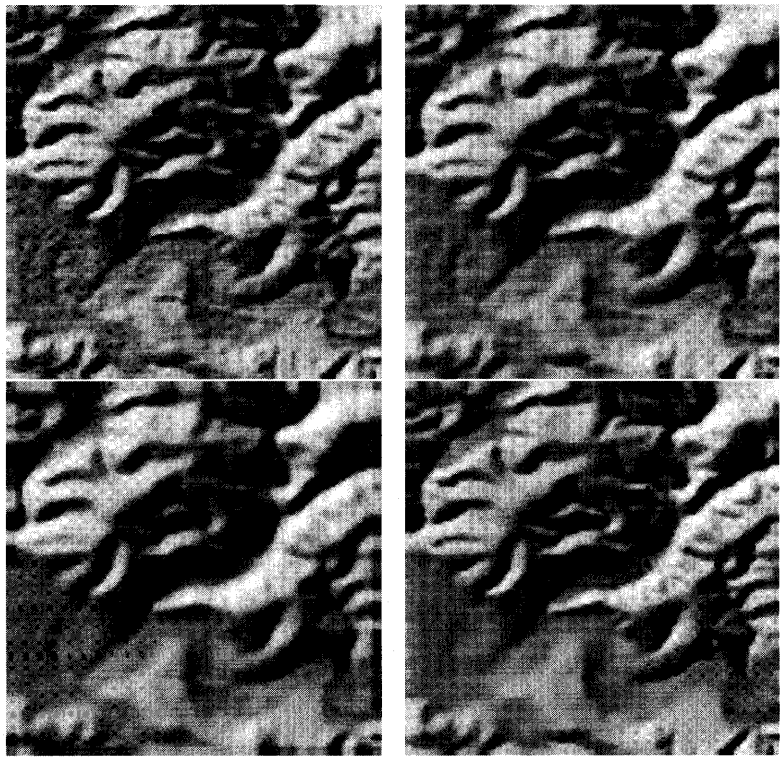


Fig. 7. *shows top left the original of a DHM, shaded. Top right shows the effect of a linear filter with $G_{0.7}$ below left the effect of a linear filter with $G_{1.2}$ and below left shows the effect of the information preserving filter, in the rigorous version from Weidner 1994a). Observe the smoothing effect in the homogeneous areas and the presservation of structure lines.*



Thus pixels with

$$\frac{\overline{|\nabla g|^2}}{\sigma_{n'}^2} > \chi_{2,\alpha}^2 \tag{67}$$

are significantly non-homogeneous, thus likely to be edge pixels.

In case the noise variance is signal dependent the determination of $\sigma_{n'}^2$ would need to take the different variances of the neighboring pixels into account. An approximation would be to assume all pixels involved in the determination of the gradient to have the noise variance of the center pixel. However, first applying a variance equalization, leads to an efficient procedure as $\sigma_{n'}^2$ is constant for the complete image $\dot{g}$.

Detecting Edge Pixels in Color Images In color images we need to integrate the information in the different channels. Applying the tests individually

in general would lead to conflicts, as there may be an edge in one channel where there is no edge in the other image. The significance of the individual results is difficult to evaluate.

We therefore proceed differently and determine a homogeneity measure on the multi valued image.

Again we first assume the different channels to have constant noise variance. Then we can measure the homogeneity by

$$\underline{z} = \sum_{k=1}^{K} \frac{\overline{\underline{h}_{k;1}}}{\sigma^2_{n'_k}} = \sum_{k=1}^{K} \frac{\overline{|\nabla \underline{g}_k|^2}}{\sigma^2_{n'_k}} \tag{68}$$

which now is χ^2_{2K}-distributed in case the multi-channel image is homogeneous. This is due to the fact that $\underline{z}$ is the sum of squares of $2K$ normally distributed variates. Thus pixels with

$$\sum_{k=1}^{K} \frac{\overline{|\nabla g_k|^2}}{\sigma^2_{n'_k}} > \chi^2_{2K,\alpha} \tag{69}$$

are significantly non-homogeneous, thus likely to be edge pixels.

Assuming general noise behavior again reveals the noise equalization to simplify matters. Here we obtain an even more simple expression for the test statistic:

$$\underline{z} = \frac{1}{\sigma^2_{\dot{n}'}} \sum_{k=1}^{K} \overline{\underline{\dot{h}}_k} = \frac{1}{\sigma^2_{\dot{n}'}} \sum_{k=1}^{K} \overline{|\nabla \underline{\dot{g}}_k|^2} \tag{70}$$

which shows that we just add the homogeneities of the normalized channels and refer to the common noise variance $\sigma^2_{\dot{n}'}$ of the gradients.

Detecting Edges in Range Images Edges in range images are pixels which do not lie on flat surfaces, thus are expected to be pixels where the curvature is significant compared to the noise.

In range images we start with the gradient image $\boldsymbol{g} = \nabla d$ as two-channel image. We want to use the same argument for fusing the channels here.

Assuming constant noise variance we instead of the homogeneity h_1 derived from the first derivatives, obtain for the homogeneity h_2 derived from the second derivatives, using $h_1(z) = z_x^2 + z_y^2$

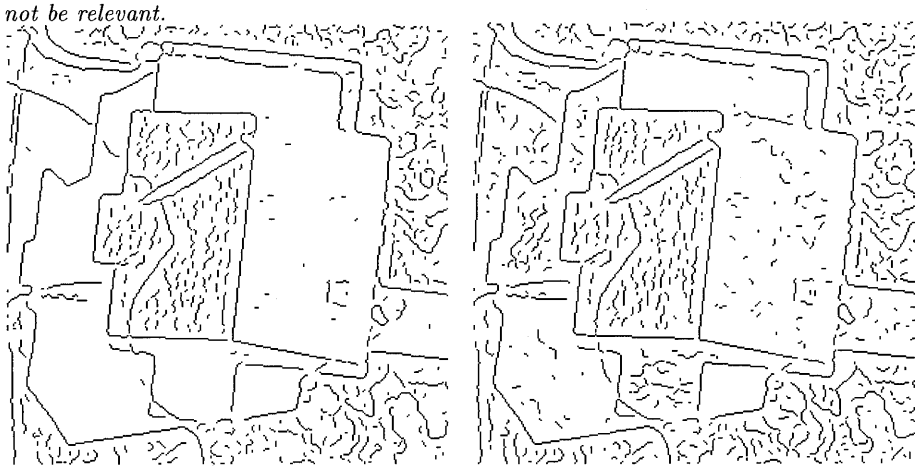
$$h_2 = h_1(d_x) + h_1(d_y) = (d_{xx}^2 + d_{xy}^2) + (d_{yx}^2 + d_{yy}^2) \tag{71}$$

$$= d_{xx}^2 + 2d_{xy}^2 + d_{yy}^2 \tag{72}$$

$$= tr\boldsymbol{H}^2(d) = \lambda_1^2(\boldsymbol{H}) + \lambda_2^2(\boldsymbol{H}) = \kappa_1^2 + \kappa_2^2 \tag{73}$$

in analogy to (64). Observe, h_2 is the quadratic variation. It only is zero in case the pixel's surrounding is flat, as only if the two principle curvature a both zero the homogeneity measure h_2 is zero.

Fig. 8. *shows the edges from the original image (left) and from the information preserving filtered image. Observe the low contrast edges, detected in the image filterd with IPF ate the expense of getting additional, statistically significant edges, which might not be relevant.*



Thus the generalization of the homogeneity measure to multi-channel images together with the use of the gradient as two channel-image for describing the form in a range image leads to a very meaningful result.

The differentiation kernels for determining the second derivatives could be approximations of the corresponding Gaussian's

$$\frac{\partial^2}{\partial x^2} G_s(x,y) = \frac{x^2 - s^2}{s^4} G_s(x,y) \tag{74}$$

$$\frac{\partial^2}{\partial x \partial y} G_s(x,y) = \frac{xy}{s^4} G_s(x,y) \tag{75}$$

$$\frac{\partial^2}{\partial y^2} G_s(x,y) = \frac{y^2 - s^2}{s^4} G_s(x,y) \tag{76}$$

which are orthogonal. Thus normalization with the noise variances leads to the test statistics

$$\underline{z} = \frac{d_{xx}^2}{\sigma_{n_{xx}}^2} + \frac{d_{xy}^2}{\sigma_{n_{xy}}^2} + \frac{d_{yy}^2}{\sigma_{n_{yy}}^2} \tag{77}$$

The noise variances in the denominators can be explicitly derived for the case of the Gaussian kernels eq. (74):

$$\sigma_{n_{xx}}^2 = \sigma_{n_{yy}}^2 = \int\int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial x^2} G_s(x,y) \right)^2 dx\, dy\, \sigma_n^2 = \frac{3}{16\pi s^6} \sigma_n^2 \tag{78}$$

$$\sigma_{n_{xy}}^2 = \int\int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial x \partial y} G_s(x,y) \right)^2 dx\, dy\, \sigma_n^2 = \frac{1}{16\pi s^6} \sigma_n^2 \tag{79}$$

which has shown to be a good approximation for discrete kernels approximating the Gaussian.

The test statistic is χ^2_3-distributed in case the region is homogeneous, which can be used to perform a statistical test for edge detection in range images.

6.2 Localizing Edge Pixels

The statistical test on pixels which are significantly non-homogeneous leads to edge areas. Within these edge areas the edge or the boundary between homogeneous regions is to be expected.

A classical approach to detect edges is to take those pixels where the slope lines show an inflection point. In a one dimensional signal these positions are given by $g'' = 0$, and $g'''g' < 0$, where the second condition is needed to avoid non valid inflection points. The generalization to two dimensional signals motivates the zero crossings of the Laplacian $g_{rr} + g_{cc}$ as edges, with a similar constraint to avoid false edges.

Obviously this technique cannot be generalized to multichannel images. As – in one dimension – edges can be defined by the maxima of g'^2 we easily can generalize this into two dimensions, by looking for local maxima of $|\nabla g|^2 = tr(\nabla g \nabla^\mathsf{T} g)$ in the direction of the gradient. We actually use the locally averaged squared gradient $G_t * |\nabla g|^2 = tr[G_t * (\nabla g \nabla^\mathsf{T} g)]$. This has the advantage of higher stability, and at the same time allows to detect bright or dark *lines*. The orientation of the gradient can be determined by the eigenvectors of $G_t * (\nabla g \nabla^\mathsf{T} g)$.

Generalization now is easy, as we only need to take the possibly weighted sum of these averaged squared gradients of the individual channels:

$$\sum_{k=1}^{K} \frac{G_t * (\nabla g_k \nabla^\mathsf{T} g_k)}{\sigma^2_{\dot{n}'_k}} = \frac{1}{\sigma^2_{\dot{n}'}} \sum_{k=1}^{K} G_t * (\nabla \dot{g}_k \nabla^\mathsf{T} \dot{g}_k) \tag{80}$$

where in the second expression the normalization with $\sigma^2_{\dot{n}'}$ is not necessary for location.

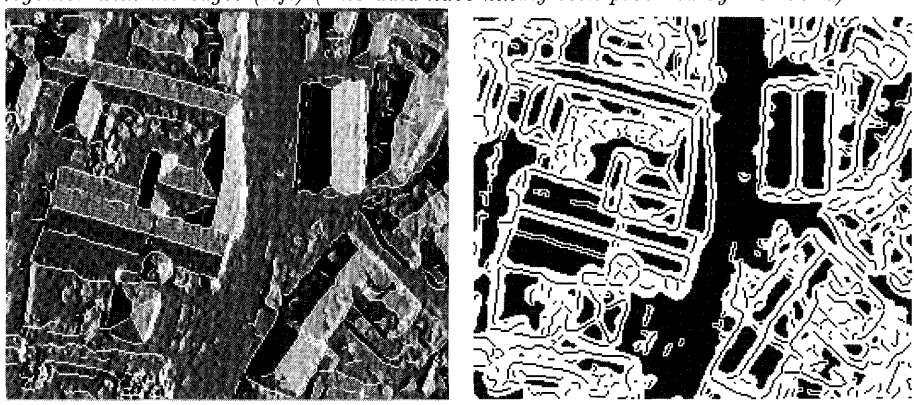
In range images the squared gradient $\nabla g \nabla^\mathsf{T} g$ just needs to be replaced by the squared Hessian $\boldsymbol{H}^2(d)$, which again demonstrates the simplicity of the approach.

7 Outlook

The paper presented tools for preprocessing intensity, color and range images for feature extraction. The idea was to exploit the full knowledge about the image as far as possible. The examples demonstrated the feasibility of the approach.

Obviously the individual steps may be conceptually integrated to a much larger extent. The iterative version of the information preserving filter obviously is an approximation, which needs to be analyses and possibly overcome by a more rigorous solution. The edge detection should be linked with the information

Fig. 9. *shows the first derivative image of a range image having a density of appr. 0.6 m (acknowledging* TOPOSYS*) with the edges overlaid (left) and homogeneous regions together with the edges (left) (The data have kindly been provided by* TOPOSYS*)*



preserving filter in order to exploit the *inhomogeneous* accuracy of the intensity values after the adaptive filter. The detection of edges in range images may take advantage of the range images in homogeneous, i. e. flat regions. Finally methods should be developed which allow to integrate preknowledge about the form of edges in order to increase the accuracy and the resolution of feature extraction.

References

BRÜGELMANN, R.; FÖRSTNER, W. (1992): Noise Estimation for Color Edge Extraction. In: FÖRSTNER, W.; RUWIEDEL, S. (Eds.), *Robust Computer Vision*, pages 90–107. Wichmann, Karlsruhe, 1992.

DAINTY, J. C.; SHAW, R. (1974): *Image Science.* Academic Press, 1974.

FÖRSTNER, W. (1985): Determination of the Additive Noise Variance in Observed Autoregressive Processes Using Variance Component Estimation Techinque. *Statistics & Decisions*, Supplement Issue 2:263–274, 1985.

FÖRSTNER, W. (1991): *Statistische Verfahren für die automatische Bildanalyse und ihre Bewertung bei der Objekterkennung und -vermessung*, Band 370 der Reihe C. Deutsche Geodätische Kommission, München, 1991.

FÖRSTNER, W. (1994): A Framework for Low Level Feature Extraction. In: EKLUNDH, J. O. (Ed.), *Computer Vision - ECCV 94, Vol. II*, Band 802 der Reihe LNCS, pages 383–394. Springer, 1994.

FUCHS, C.; LANG, F.; FÖRSTNER, W. (1994): On the Noise and Scale Behaviour of Relational Descriptions. In: EBNER, HEIPKE, EDER (Ed.), *Int. Arch. f. Photogr. and Remote Sensing, Vol. 30, 3/2*, Band XXX, pages 257–267, 1994.

FUCHS, C. (1998): *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung.* DGK, Bayer. Akademie der Wissenschaften, Reihe C, Heft 502, 1998.

MAYER, H. (1999): Automatic Object Extraction from Aerial Imagery – A Survey Focussing on Buildings. *Computer Vision and Image Understanding*, 74(2):138–149, 1999.

MORITZ, H. (1980): *Advanced Physical Geodesy.* Herbert Wichmann Verlag, Karlsruhe, 1980.

PAPOULIS, A. (1984): *Probability, Random Variables, and Stochastic Processes.* Electrical Engineering. McGraw-Hill, 2. edition, 1984.

WAEGLI, B. (1998): Investigations into the Noise Characteristics of Digitized Aerial Images. In: *Int. Arch. for Photogr. and Remote Sensing, Vol. 32-2*, pages 341–348, 1998.

WEIDNER, U. (1994): Information Preserving Surface Restoration and Feature Extraction for Digital Elevation Models. In: *ISPRS Comm. III Symposium on Spatial Information from Digital Photogrammetry and Computer Vision, Proceedings*, pages 908–915. SPIE, 1994.

WEIDNER, U. (1994): Parameterfree Information-Preserving Surface Restoration. In: EKLUNDH, J.-O. (Ed.), *Computer Vision - ECCV 94, Vol. II, Proceedings*, pages 218–224, 1994.

WIENER, N. (1948): *Cybernetics.* MIT Press, 1948.

Optimization-Based Approaches To Feature Extraction from Aerial Images

P. Fua[1], A. Gruen[2] and H. Li[2]

[1] Computer Graphics Lab (LIG), Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland, fua@lig.di.epfl.ch
[2] Institute of Geodesy and Photogrammetry Swiss Federal Institute of Technology, CH-8093 Zürich, Switzerland, armin@p.igp.ethz.ch

Abstract. Extracting cartographic objects from images is a difficult task because aerial images are inherently noisy, complex, and ambiguous. Using models of the objects of interest to guide the search has proved to be an effective approach that yields good results.
In such an approach, the problem becomes one of fitting the models to the image data, which we phrase as an optimization problem. The appropriate optimization technique to use depends on the exact nature of the model. In this paper, we review and contrast some of the approaches we have developed for extracting cartographic objects and present the key aspects of their implementation.
Using these techniques, rough initial sketches of 2–D and 3–D objects can automatically be refined, resulting in accurate models that can be guaranteed to be consistent with one another. We believe that such capabilities will prove indispensable to automating the generation of complex object databases from imagery, such as the ones required for high-resolution mapping, realistic simulations or intelligence analysis.

1 Introduction

This paper reviews and contrasts a number of optimization techniques that we have developed to automate the extraction of 2–D and 3–D features from images. This is a difficult task because image data, even when it is of very high quality, typically is noisy, complex and ambiguous. As a result, context-free image-processing techniques are seldom robust enough. However, using models to guide the feature extraction algorithms typically leads to a vast improvement in performance. The problem then becomes one of fitting these models to the image data, which is an optimization problem.

The choice of the optimization technique is conditioned by the nature of the model used and the constraints it imposes:

- **Local properties of the contours:** Linear structures, such as boundaries or roads in low-resolution images can be modeled as lines with, or without, thickness that satisfy local regularity constraints and maximize the integral of an image measure along their length. Such models can be fitted effectively

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 190–228, 2000.

and fast using dynamic programming. Section 2 uses the example of roads to present this approach in detail and shows that good results can be obtained even under very unfavorable conditions.

- **Global properties of the contours:** More complex structures, such as buildings, require more global models for which dynamic programming is less well adapted. For those, in Section 3, we introduce the so-called "snake-based" techniques that use the Euler-Lagrange formalism. They handle global constraints, but at a cost: Unlike dynamic programming, they do not find global optima, only local ones. They, therefore, require fairly good starting points. This, however, does not constitute a major handicap because other algorithms, or simple manual intervention, can supply these initializations. We will show that this approach can be used to model many classes of cartographic objects and can be extended to terrain modeling. Furthermore, it can deal with multiple objects simultaneously, and ensure that they are both accurate and consistent with each other. This is critical when generating complex object databases from imagery, such as the ones required for realistic simulations or intelligence analysis.
- **Global photometric properties:** There is more to objects than just their outlines: Their internal structure and corresponding gray-level patterns, that is, their photometric structure, are also very important. In Section 4, we introduce the LSB-Snakes that are related to those of Section 3 but take advantage of the least-squares formalism, instead of the the Euler-Lagrange one, to model internal structure as well as contours. We come back to the example of roads to illustrate their implementation. There the advantage is robustness as more of the image is taken into account. Furthermore, the least squares approach allows for precision and reliability assessment of the estimated 3-D feature via covariance matrix evaluation. The price to pay is added computational complexity but this can be mitigated by using the other techniques to initialize this one.

We view these methods as complementary because they are most effective in different contexts and, also, because the output of the ones can be used as the input to the others. Although the current level of automation on most digital photogrammetric stations is still fairly low, implementations of a number of these techniques are starting to make their way into commercial systems [Gruen, 1996, Miller *et al.*, 1996, Walker and Petrie, 1996] and the trend should accelerate in years to come.

2 Dynamic Programming

Dynamic programming is a technique for solving optimization problems when not all variables in the evaluation function are interrelated simultaneously [Ballard and Brown, 1982]. It is a solution strategy for combined optimization problems which involve a sequential decision-making process. It is an optimization process, expressed as a recursive search [Bellman and Dreyfus, 1962].

2.1 Generic Road Model

A road is a well-defined concept. The definition of the word "road" as given by a dictionary states: "A smooth prepared track or way along which wheeled vehicles can travel". This is more or less a functional description of a road, but cannot be directly applied to identify roads in a digital image. It is necessary for road extraction that a generic road model be formed which describes the appearance of a road in the digital image and whose generation makes the task programmable.

Let $\mathcal{C}$ be a curve on a digital image, corresponding to a road in the object space. We make the following assumptions:

1. The curve $\mathcal{C}$ can be represented by a vector function $\mathbf{f}(s)$, which maps the arc-length s to points (x, y) in the image;
2. The curve $\mathcal{C}$ has continuous derivatives and, for each s, there exists a unit vector $\mathbf{n}(s)$ that is normal to the curve.
3. The preprocessed digital image is represented by a 2-D function $G(x, y)$, which is measurable and has a finite energy and continuous derivatives;

We state the properties of our generic road model, according to our knowledge of the "road" objects, first in words and then using the equivalent mathematical formulation.

1. A road pixel in the image is lighter than its neighbours on both road sides. In other words, a road on a digital image is identified by a continuous and narrow region of high intensity with regions of lower intensity on each side. This suggests that a squared sum of the grey values (or their second derivatives in the direction normal to the road) along the curve attains a maximum. This can be expressed as

$$E_{P1} = \int [G(f(s))]^2 ds \Rightarrow \text{Maximum} \tag{1}$$

2. Grey values along a road usually do not change very much within a short distance. This stems from the knowledge that the road materials do not vary much and their spectral properties are similar within a short distance. This is equivalent to

$$E_{P2} = \sum_i \int_{\Delta s} [G(f(s)) - G_m(\Delta s_i)]^2 ds \Rightarrow \text{Minimum} \tag{2}$$

where Δs_i is a short segment of the curve $\mathcal{C}$, and $G_m(\Delta s_i$ is the average value of G on Δs_i, that is

$$G_m(\Delta s_i) = \int_{\Delta s} G(f(s)) ds / \Delta s_i \tag{3}$$

3. A road is a light linear feature. In fact, this is a generalization of the two previous properties, and as such a more global formulation. This property can be considered through the appropriate evaluation of the formula

$$E_{P^3} = \int w(d(s))[G(f(s)) + d(s)\mathbf{n}(s)]^2 ds \Rightarrow \text{Maximum} \tag{4}$$

where $d(s)$ is the distance between curve $\mathcal{C}$ and the linear feature near it. $w(d(s))$ is a Gaussian weight function that decreases as the distance $d(s)$ increases.

4. In terms of basic geometric property, a road is usually smooth and does not have small wiggles. In fact, most roads consist of straight lines connected with smooth curves, normally circular arcs. This property can be represented mathematically as

$$E_g = \int |f''(s)|^2 \Rightarrow \text{Maximum} \tag{5}$$

This implies that $\mathcal{C}$ can be represented as a cubic spline.

5. The local curvature of a road has an upper bound, which means that the local change of direction is upward bounded. This property follows from smooth traffic flow requirement.

$$C_g = |f''(s)| < T_1 \tag{6}$$

where T_1 is a given threshold.

6. The width of a road does not change significantly. This property is not used in this section because, here, we treat roads as lines without width. We will come back to it in Section 3.1 when we model roads as ribbons.

These six properties—three expressed in photometric terms and three expressed in geometric ones, form the generic road model used in all the road extraction schemes presented in this paper. Although this formulation of a road model is strictly used only in our dynamic programming approach, it also forms the basis for the objective function of the road-snakes of Section 3.1 and observation equations of the LSB-Snakes of Section 4.

2.2 Road Delineation

The general idea of road extraction by dynamic programming is the following.

- A curve $\mathcal{C}$ is described as a polygon with n vertices. The first four properties of the generic road model developed in the last section can be made discrete and combined into a merit function with the property of Equation 5 as a hard constraint;
- Each point or vertex moves around its initial position (x_i^0, y_i^0), for instance in a 5×5 window, forming a number of polygons. The candidate among them for which the maximum of the merit function is achieved under the constraints is considered a road;

- Thus a road extraction can be treated as an optimization problem or multi-stage decision process in which the coordinate candidates of the ith vertex form the choice nodes of the ith decision stage. This can be solved using dynamic programming in the form of a time-delayed algorithm. For more details, we refer the interested reader to our earlier publications [Gruen and Li, 1995, Li, 1997].

Objective function In our implementation a road segment is described as a polygon with n vertices $P = \{p_1, p_2, ..., p_n\}$, $p_i = (x_i, y_i)$. We discretize the generic road model introduced above as follows. The terms of Equations 1, 2 and 4 are rewritten as terms of the form

$$E = \sum_i E(p_i, p_{i+1}) \ . \tag{7}$$

Discretization Equations 5 and 6 yields

$$E_g = \sum_i [2 - 2\cos(\alpha_i - \alpha i + 1)]/|\Delta s_i| \ , \tag{8}$$

and

$$C_g = |\alpha_i - \alpha i + 1| < T_1 \tag{9}$$

where α_i is the direction of vector between points p_{i-1} and p_i, and Δs_i is the distance between them, that is

$$\begin{aligned} \alpha_i &= atan[(y_i - y_{i-1})/(x_i - x_{i-1})] \\ \Delta s_i &= \sqrt{((x_i - x_{i-1})^2 + (y_i - y_{i-1})^2)} \end{aligned} \tag{10}$$

The generic road model can be formulated by the following merit function and an inequality constraint as follows:

$$\begin{aligned} E &= \sum_i [E_{P^1}(p_i, p_{i+1}) - \beta E_{P^2}(p_i, p_{i+1}) + \gamma E_{P^3}(p_i, p_{i+1})] \qquad (11) \\ &\quad \times (1 + \cos(\alpha_i - \alpha i + 1))/|\Delta s_i| \\ &= \sum_i E(p_{i-1}, p_i, p_{i+1}) \qquad (12) \\ C_i &= |\alpha_i - \alpha i + 1| < T_1 \end{aligned}$$

where β and γ are two positive constants and T_1 is a threshold for direction change between two adjacent vectors.

The optimization problem defined by the objective function of Equation 12 can be solved by the "time-delayed" algorithm [Gruen and Li, 1995]. To reduce the computational complexity and make the algorithm more efficient, the number of vertices used to approximate a curve and the number of candidates for each vertex should be reduced to as few as possible.

Coarse to Fine One-Dimensional Vertex Search Recall that we write a polygon with n as $P = \{p_1, p_2, ..., p_n\}$ where p_i denotes the image coordinates of the ith vertex, that is, $p_i = (x_i, y_i)$. The road extracted by the algorithm is represented as another polygon with n vertices $\{\overline{p}_1, \overline{p}_2, ..., \overline{p}_n\}$. The ith vertex $\overline{p}_i$ in the new polygon is selected from a finite set of candidates in the neighbourhood of p_i. One can simply take a window around the vertex p_i, with the candidates inside. To get a large convergence range, a large window has to be used. However, this may greatly increase the computational complexity because it is in the order of $O(nm^3)$. To reduce the number of candidates, two strategies can be employed.

1. Restrict the search in one dimension. Here the candidates are selected only in the direction normal to the initial curve at point p_i. In such a way, the same convergence range can be obtained with a much lower computational complexity. For instance, if the window size is $5x5$, then the number of candidates will be 25, but it is 5 for the one-dimensional search.
2. Select the candidates from a coarse to fine pyramid. Instead of using an image pyramid explicitly, we can simply select the candidates with a certain interval, e.g. every 3 pixels, at the beginning. This results in a large convergence range. To get high accuracy the interval is decreased subsequently. One can even use an interval smaller than 1 pixel to achieve the sub-pixel accuracy, but this was not implemented in this investigation.

Dynamic Vertex Insertion and Deletion We could describe the curve as a polygon with equidistant vertices. This strategy, however, is not optimal because a large number of vertices are needed and their positions are not related to the shape of the curve. In our approach, a few starting vertices are given coarsely by the operator. Connecting these seed points, an initial polygon is formed. After the first iteration of the optimization procedure by dynamic programming on this polygon, two equidistant new vertices are inserted by a linear interpolation between every two adjacent vertices whose distance is larger than a threshold. Then the second iteration is applied to this new polygon. Each new inserted vertex is checked. Points that are collinear with their neighbors or that cause a "blunder" in the form of a wiggle, are removed. The iteration scheme runs until convergence is reached. By this dynamic vertex insertion and deletion control strategy the computational complexity is reduced and at the same time the algorithm is more robust in case of small gaps and other distortions.

As depicted by Figure 1, each internal vertex is checked after each iteration for three conditions

$$|\Delta s| > T_d \; ; \; h > T_c \; ; \; h/|\Delta s| < T_b \tag{13}$$

The first condition requires that the distance between the two neighbour vertices has to be larger than a threshold. The second condition ensures that every vertex is necessary for the polygon. And the third one, together with the constraint of the objective function, ensures that the polygon is an approximation of a smooth curve and makes the algorithm more robust to bridge small gaps and resist the influence of distortions.

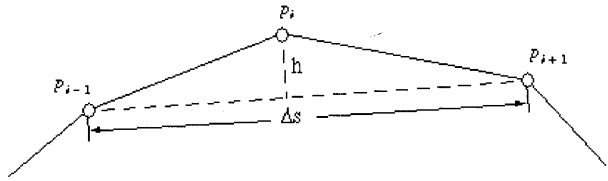


Fig. 1. Vertices and associated length used to perform the test of Equation 13.

Implementation and Experimental Results This dynamic programming approach has been implemented in a monoplotting mode. We have performed a number of experiments for road extraction using SPOT and small scale aerial images [Gruen and Li, 1995, Li, 1997]. Figure 2(a) shows a portion of a SPOT ortho-image of Kefalonia, Greece draped over its underlying DTM. The extracted road segments are overlaid in black. Figure 2(b) shows a road network extracted from a SPOT panchromatic image of Sydney, Australia. The contrast of roads is very low in this image. These results demonstrate that our algorithm works very well even under these unfavourable conditions.

3 Model Based Optimization

Model-Based Optimization (MBO) is a paradigm in which an objective function is used to express both geometric and photometric constraints on features of interest. A parametric model of a feature (such as a road, a building, or coastline) is extracted from one or more images by automatically adjusting the model's state variables until a minimum value of the objective function is obtained. The optimization procedure yields a description that simultaneously satisfies (or nearly satisfies) all constraints, and, as a result, is likely to be a good model of the feature.

The deformable models we use here are extensions of traditional snakes [Terzopoulos *et al.*, 1987, Kass *et al.*, 1988, Fua and Leclerc, 1990]. They are polygonal curves or facetized surfaces to which is associated an objective function combining an "image term" that measures the fit to the image data and a regularization term that enforces geometric constraints.

Because features and surfaces are all uniformly modeled, we can refine several models simultaneously and enforce geometric and semantic constraints between objects, thus increasing not only the accuracy but also the consistency of the reconstruction. The ability to apply such constraints is essential for the accurate modeling of complex sites in which objects obey known geometric and semantic constraints. In particular, when dealing with multiple objects, it is crucial that the models be both accurate and consistent with each other. For example, individual components of a building can be modeled independently, but to ensure realism, one must guarantee that they touch each other in an architecturally feasible way. Similarly, when modeling a cartographic site from aerial imagery,

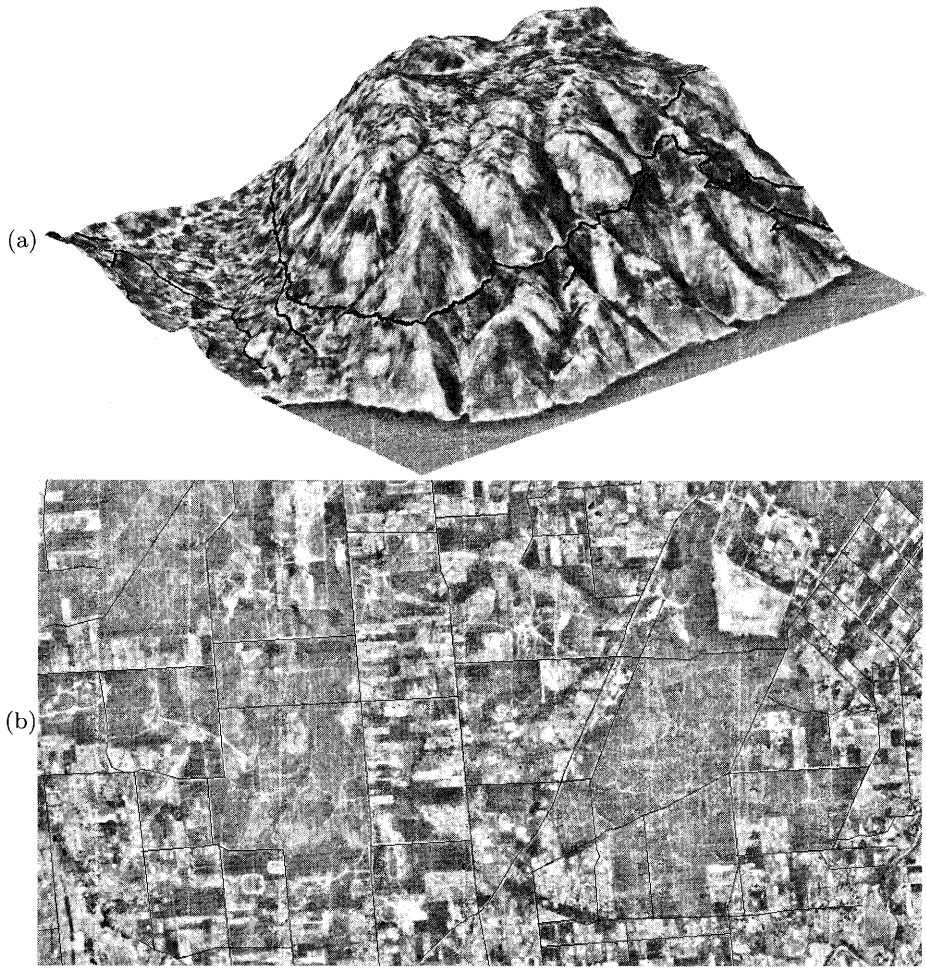


Fig. 2. Dynamic programming approach to road extraction. (a) A portion of a SPOT ortho-image of Kefalonia, Greece draped over its underlying DTM. The extracted road segments are overlaid in black. (b) A road network extracted from a SPOT panchromatic image of Sydney, Australia.

one must ensure that the roads lie on the terrain—and not above or below it—and that rivers flow downhill. To that end, we have developed a constrained-optimization scheme that allows us to impose hard constraints on our snakes at a very low computational cost while preserving their convergence properties.

In the remainder of this section, we first introduce our generalized snakes. We then present our constrained-optimization scheme. Finally, we demonstrate its ability to enforce geometric constraints upon individual snakes and consistency constraints upon multiple snakes to produce complex and consistent site models.

3.1 Generalized Snakes

We model linear features as polygonal curves that may be described either as a sequential list of vertices, or, for more complex objects such as a road network or a 3–D extruded object, described by the network topology. In the latter case, to describe the object completely, one must supply not only the list of vertices but also a list of "edges" that defines the connectivity of those vertices. In addition, with some of these complex objects, one can also define "faces," that is, circular lists of vertices that must be constrained to remain planar.

Similarly, we model the terrain on which these features rest as triangulated surface meshes whose shape is defined by the position of vertices and can be refined by minimizing an objective function.

Our ultimate goal is to accommodate the full taxonomy of those "generalized snakes" described by Table 1. The algorithms described here are implemented within the RADIUS Common Development Environment (RCDE) [Mundy *et al.*, 1992]. The system has been installed at the National Exploitation Laboratory where it is used in a quasi-operational mode by professional image analysts.

Furthermore, we have evaluated the effectiveness of our package by instrumenting the code to record the amount of user intervention. We have found that using our snakes to model high-resolution roads leads to a fivefold reduction in effort as measured by the number of mouse clicks or mouse travel-time [Fua, 1997].

Constraints/Type	Simple curve	Ribbon curve	Network	Meshes
Smooth	Low res. roads, rivers	High res. roads	Road network	Terrain
Polygonal	Man-made structures	City streets	Street networks	
Planar	Planar structures	City streets	Street networks	
Rectilinear	Roof tops, parking lots	City streets	Buildings	

Table 1. Snake taxonomy. The columns represent different types of snakes and the rows different kinds of constraints that can be brought to bear. The table entries are examples of objects that can be modeled using these combinations.

Polygonal Snakes A simple polygonal snake, $\mathcal{C}$, can be modeled as a sequential list of vertices, that is, in two dimensions, a list of 2–D vertices $\mathcal{S}_2$ of the form

$$\mathcal{S}_2 = \{(x_i\ y_i),\ i = 1, \ldots, n\}\ , \tag{14}$$

and, in three dimensions, a list of 3–D vertices $\mathcal{S}_3$ of the form

$$\mathcal{S}_3 = \{(x_i\ y_i\ z_i),\ i = 1, \ldots, n\}\ . \tag{15}$$

In this paper, we refer to S, the vector of all x, y, and z coordinates of the 2–D or 3–D vertices that define the deformable model's shape as the model's *state vector.*

In the 2–D case, the "image energy" of these curves—the term we try to minimize when we perform the optimization is taken to be

$$\mathcal{E}_I(\mathcal{C}) = -\frac{1}{|\mathcal{C}|}\int_0^{|\mathcal{C}|} |\nabla \mathcal{I}(\mathbf{f}(s))|\, ds, \tag{16}$$

where I represents the image gray levels, s is the arc length of $\mathcal{C}$, $\mathbf{f}(s)$ is a vector function mapping the arc length s to points (x, y) in the image, and $|\mathcal{C}|$ is the length of $\mathcal{C}$. In practice, $\mathcal{E}_I(\mathcal{C})$ is computed by integrating the gradient values $|\nabla \mathcal{I}(\mathbf{f}(s))|$ in precomputed gradient images along the line segments that connect the polygonal vertices.

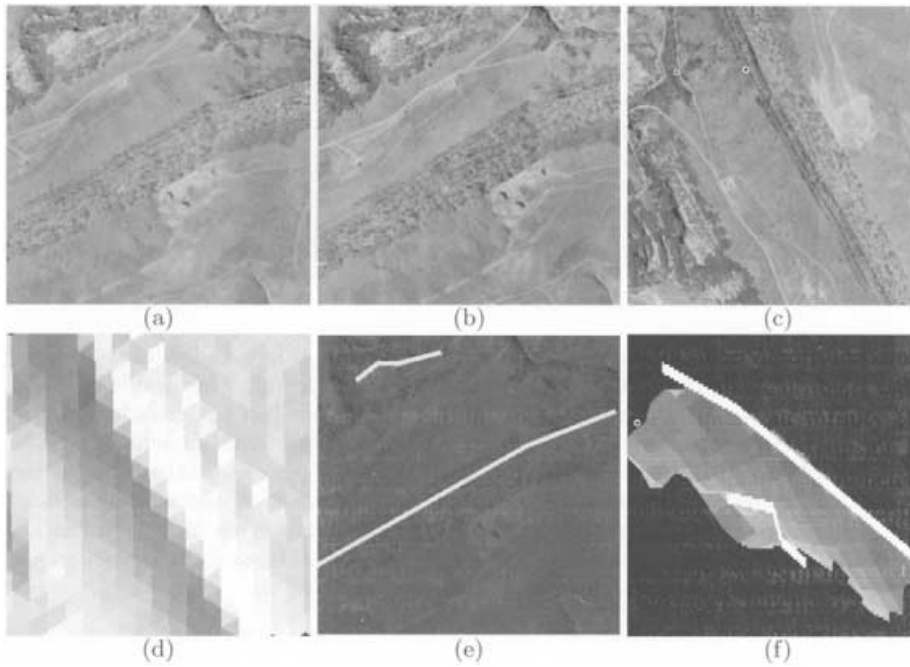


Fig. 3. Rugged terrain with sharp ridge lines. (a,b,c) Three images of a mountainous site. (d) Shaded view of an initial terrain estimate. (e) Rough polygonal approximation of the ridgelines overlaid on image (a). (f) The terrain and ridgeline estimates viewed from the side (the scale in z has been exaggerated).

In the 3–D case, illustrated by Figures 3 and 4(a), $\mathcal{E}_I(\mathcal{C})$ is computed by projecting the curve into a number of images, computing the image energy of each projection, and summing these energies.

Smooth Snakes and Ribbons These snakes are used to model smoothly curving features such as roads or ridgelines.

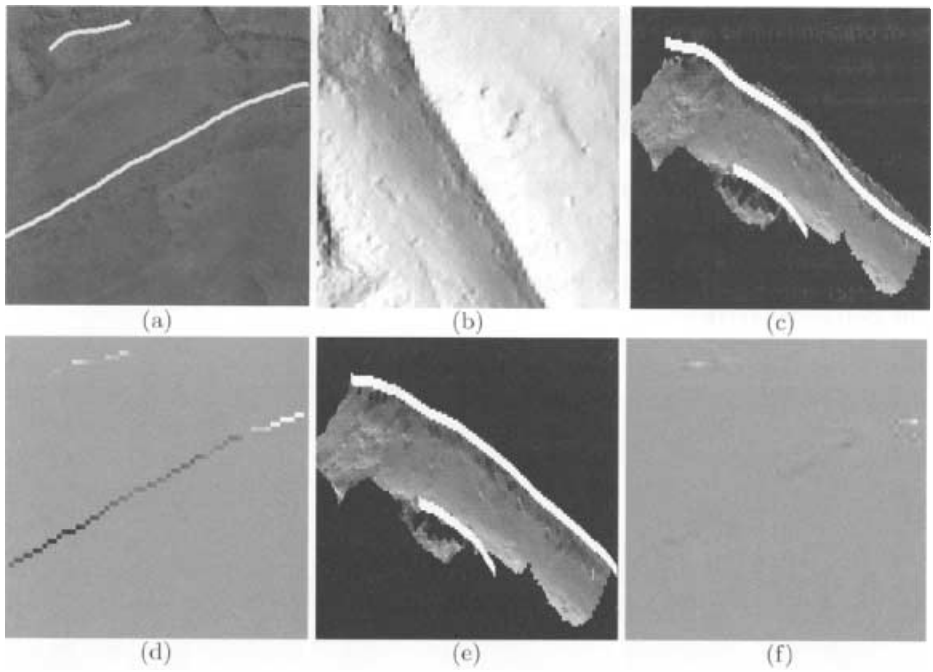


Fig. 4. Recovering the 3–D geometry of both terrain and ridges. (a) Refined ridgeline after 3–D optimization. (b) Shaded view of the terrain after refinement. (c) Side view of the ridgeline and terrain after independent optimization of each one. Note that the shape of the ridgeline does not exactly match that of the terrain. (d) Differences of elevation between the recovered ridgeline and the underlying terrain. The image is stretched so that black and white represent errors of minus and plus 80 feet, respectively. (e) Side view after optimization under consistency constraints. (f) Corresponding difference of elevation image stretched in the same fashion as (d).

2–D curves. Following Kass *et al.* [1988], we choose the vertices of such curves to be roughly equidistant and add to the image energy $\mathcal{E}_I$ a regularization term $\mathcal{E}_D$ of the form

$$\begin{aligned}\mathcal{E}_D(\mathcal{C}) = \mu_1 \textstyle\sum_i (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \\ + \mu_2 \textstyle\sum_i (2x_i - x_{i-1} - x_{i+1})^2 + (2y_i - y_{i-1} - y_{i+1})^2\end{aligned} \tag{17}$$

and define the "total energy" $\mathcal{E}_T$ as

$$\mathcal{E}_T(\mathcal{C}) = \mathcal{E}_D(\mathcal{C}) + \mathcal{E}_I(\mathcal{C}) \ . \tag{18}$$

The first term of $\mathcal{E}_D$ approximates the curve's tension, and the second term approximates the sum of the square of the curvatures, assuming that the vertices are roughly equidistant. In addition, when starting, as we do, with regularly

spaced vertices, this second term tends to maintain that regularity. To perform the optimization we could use the steepest or conjugate gradient, but it would be slow for curves with large numbers of vertices. Instead, it has proven much more effective to embed the curve in a viscous medium and solve the equation of the dynamics

$$\frac{\partial \mathcal{E}}{\partial S} + \alpha \frac{dS}{dt} = 0, \qquad (19)$$
$$\text{with } \frac{\partial \mathcal{E}}{\partial S} = \frac{\partial \mathcal{E}_D}{\partial S} + \frac{\partial \mathcal{E}_I}{\partial S},$$

where $\mathcal{E}$ is the energy of Equation 18, α the viscosity of the medium, and S the state vector that defines the current position of the curve. Since the deformation energy $\mathcal{E}_D$ in Equation 17 is quadratic, its derivative with respect to S is linear, and therefore Equation 19 can be rewritten as

$$K_S S_t + \alpha(S_t - S_{t-1}) = -\left.\frac{\partial \mathcal{E}}{\partial S}\right|_{S_{t-1}}$$
$$\Rightarrow (K_S + \alpha I) S_t = \alpha S_{t-1} - \left.\frac{\partial \mathcal{E}}{\partial S}\right|_{S_{t-1}}, \qquad (20)$$

where

$$\frac{\partial \mathcal{E}_D}{\partial S} = K_S S,$$

and K_S is a sparse matrix. Note that the derivatives of $\mathcal{E}_D$ with respect to x and y are decoupled so that we can rewrite Equation 20 as a set of two differential equations of the form

$$(K + \alpha I) V - t = \alpha V_{t-1} - \left.\frac{\partial \mathcal{E}_I}{\partial V}\right|_{V_{t-1}}, \qquad (21)$$

where V stands for either X or Y, the vectors of the x and y vertex coordinates, and K is a pentadiagonal matrix. Because K is pentadiagonal, the solution to this set of equations can be computed efficiently in $O(n)$ time using LU decomposition and backsubstitution. Note that the LU decomposition need be recomputed only when α changes.

In practice, α is computed in the following manner. We start with an initial step size Δ_p, expressed in pixels, and use the following formula to compute the viscosity:

$$\alpha = \frac{\sqrt{2n}}{\Delta_p} \left|\frac{\partial \mathcal{E}}{\partial S}\right|, \qquad (22)$$

where n is the number of vertices. This ensures that the initial displacement of each vertex is on the average of magnitude Δ_p. Because of the nonlinear term, we must verify that the energy has decreased from one iteration to the next. If, instead, the energy has increased, the curve is reset to its previous position, the step size is decreased, and the viscosity recomputed accordingly. This procedure

is repeated until the step size becomes less than some threshold value. In most cases, because of the presence of the linear term that propagates constraints along the whole curve in one iteration, it takes only a small number of iterations to optimize the initial curve.

3–D curves. To extend the smooth snakes to three dimensions, we add one term in z to the deformation energy of Equation 17. Since the derivatives of $\mathcal{E}_D$ with respect to x, y, and z are still decoupled, we can rewrite Equation 20 as a set of three differential equations of the form of Equation 21, where V now stands for either X, Y, or Z, the x, y, or z vertex coordinates.

The only major difference with the 2–D case is the use of the images' camera models. In practice, $\mathcal{E}_I(\mathcal{C})$ is computed by summing gradient values along the line segments linking the vertices' projections. These projections, and their derivatives, are computed from the state vector S by using the camera models. Similarly, to compute the viscosity, we use the camera models to translate the average initial step Δ_p, a number of pixels, into a step Δ_w expressed in world units and use the latter in Equation 22.

Ribbons 2–D snakes can also be extended to describe ribbon-like objects such as roads in aerial images. A ribbon snake is implemented as a polygonal curve forming the center of the road. Associated with each vertex i of this curve is a width w_i that defines the two curves that are the candidate road boundaries. The list of vertices can be written as

$$\mathcal{S}_2 = \{(x_i\ y_i\ w_i)\},\ i = 1, \ldots, n\} \ . \tag{23}$$

The state vector S becomes the vector of all x, y, and w and the average edge strength the sum of the edge strengths along the two boundary curves. Since the width of roads tends to vary gradually, we add an additional energy term of the form

$$\begin{aligned} \mathcal{E}_W(\mathcal{C}) &= \sum_i (w_i - w_{i-1})^2 \\ \Rightarrow \frac{\partial \mathcal{E}_W}{\partial W} &= LW, \end{aligned} \tag{24}$$

where W is the vector of the vertices' widths and L a tridiagonal matrix. The total energy can then be written as

$$\mathcal{E}_(\mathcal{C}) = \lambda_D \mathcal{E}_D(\mathcal{C}) + \lambda_W \mathcal{E}_W(\mathcal{C}) + \lambda_G \mathcal{E}_I(\mathcal{C}) \ ,$$

where λ_D and λ_W weigh the contributions of the two geometric terms. At each iteration the system must solve the three differential equations in the form of Equation 21, where V now stands for either X, Y, or W, the x, y, or w vertex coordinates.

2–D ribbons can be turned into 3–D ones in exactly the same way 2–D snakes are turned into 3–D ones. The state vector S becomes the vector of all x, y, z, and w and, at each iteration, the system must solve four differential equations, one for each coordinate.

Network Snakes The 2–D and 3–D "network snakes" are direct extensions of the polygonal snakes of Section 3.1.

In the 2–D case, the extension is straightforward. A network snake is now defined by a list of n vertices $\mathcal{S}$ as before and a list of edges $\mathcal{A} = \{(i,j)$ where $1 \leq i \leq n$ and $1 \leq j \leq n\}$. Figure 5 depicts such a network snake. $\mathcal{E}_I(\mathcal{C})$ is computed as

$$\mathcal{E}_I(\mathcal{C}) = \sum_{(i,j)\in\mathcal{A}} \mathcal{E}_I^{i,j} / \sum_{(i,j)\in\mathcal{A}} L^{i,j} \quad , \tag{25}$$

where $\mathcal{E}_I^{i,j}$ is the sum of the edge gradients along the $((x_i, y_i)(x_j, y_j))$ segment and $L^{i,j}$ is its length. The snake is optimized using either steepest gradient descent or conjugate gradient.

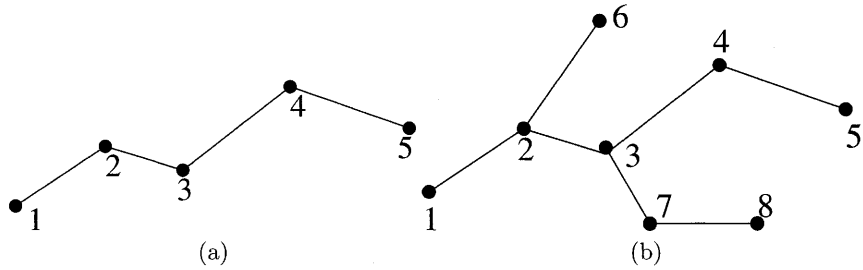
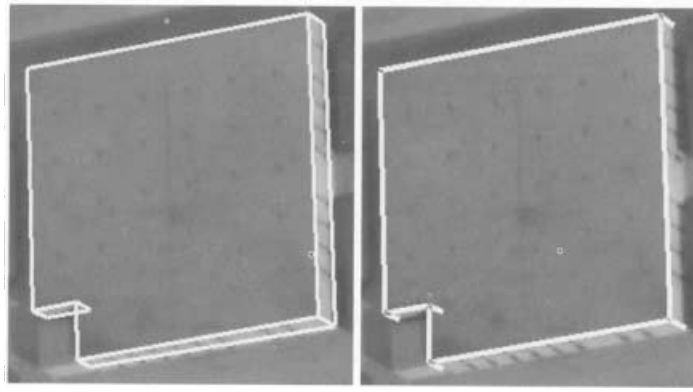


Fig. 5. Snake topology. (a) A simple polygonal curve described by a sequential list of vertices v_i, $1 \leq i \leq 5$. (b) A network described by a list of vertices v_i, $1 \leq i \leq 8$, and a list of edges—((1 2) (2 3) (3 4) (4 5) (2 6) (3 7) (7 8)).

In the 3–D case, one must take into account the fact that not all the network's edges are visible in all views. As a result one must also provide, for each projection of the snake into all the images, a list of visible edges. We compute this list by using the face-visibility methods embedded in RCDE and depicted by Figure 6. Only the visible faces—that is, those whose normal is oriented toward the viewer—are drawn. Note that this heuristic does not account for nonconvexity. As a result, the faces in the lower left corner of Figure 6(a) are improperly drawn. The network snake used to optimize the extruded object is shown in Figure 6(b). It includes roof edges and vertical wall edges. The edges at the back of the building are not drawn—and not used during the computations involving these views—because they belong to hidden faces. The edges at the base of the building are treated as invisible because their appearance is unreliable in typical imagery.

The number of degrees of freedom of generic 3–D networks can be reduced by forcing them to be planar. We do this either by defining a plane of equation

$$z = ax + by + c \tag{26}$$



(a) (b)

Fig. 6. Edge visibility. (a) Visible faces of an RCDE "extruded object." (b) The network snake generated to optimize the object.

and imposing that the vertices lie on such a plane or imposing planar constraints on sets of four vertices using the constrained-optimization approach introduced in Section 3.2. In both cases, we replace the n degrees of freedom necessary to specify the elevation of each vertex by the three degrees of freedom required to define the plane.

These 3–D networks can be further specialized to handle objects that are of particular interest in urban environments: trihedral corners found on building roofs and extruded objects that are used in RCDE to model building outlines. In Figure 7, we show several buildings modeled by roughly entering their outlines within RCDE and optimizing the shapes in three views simultaneously by using our extruded snakes. The use of the snakes has allowed us to perform this task much faster than we would have if we had to precisely delineate all five buildings by hand. To produce this result, we have used the constrained-optimization technique of Section 3.2 to constrain the "wall" edges to remain vertical. We can also constrain the "roof outline" to be planar and the "roof edges" to form 90-degree angles. These constraints greatly reduce the number of degrees of freedom and allow for better convergence properties.

3–D Surface Meshes Given the task of reconstructing a surface from multiple images whose vantage points may be very different, we need a surface representation that can be used to generate images of the surface from arbitrary viewpoints, taking into account self-occlusion, self-shadowing, and other viewpoint-dependent effects. Clearly, a single image-centered representation is inadequate for this purpose. Instead, an object-centered surface representation is required.

Many object-centered surface representations are possible. However, practical issues are important in choosing an appropriate one. First, the representation

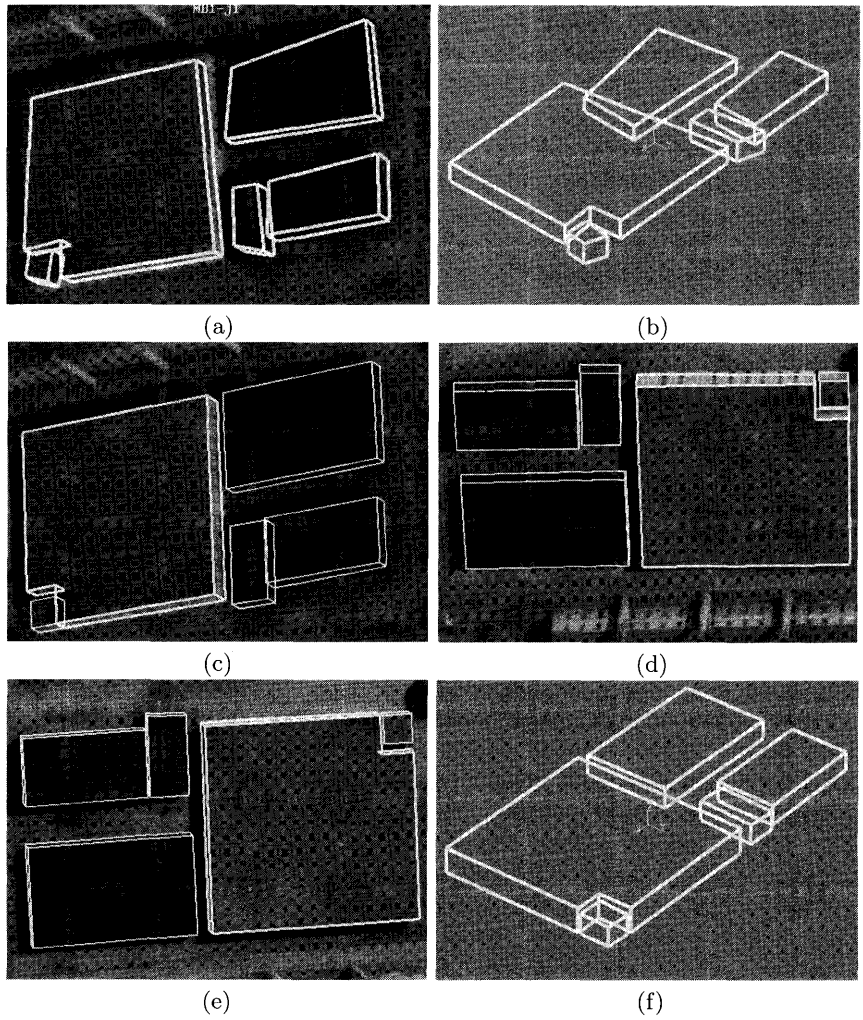


Fig. 7. Buildings modeled by entering rough models within RCDE and optimizing them using the extruded snakes. (a) Rough initial sketches overlaid on one of the images. (b) A view from a different perspective. (c,d,e) Final building outlines overlaid on the three images we used to perform the 3–D optimization. (f) A view of the buildings from the perspective of (b).

should be general-purpose in the sense that it should be possible to represent any continuous surface, closed or open, and of arbitrary genus. Second, it should be relatively straightforward to generate an instance of a surface from standard data sets such as depth maps or clouds of points. Finally, there should be a computationally simple correspondence between the parameters specifying the

surface and the actual 3-D shape of the surface, so that images of the surface can be easily generated, thereby allowing the integration of information from multiple images.

A regular 3–D triangulation is an example of a surface representation that meets the criteria stated above, and is the one we have chosen for our previous work. In our implementation, all vertices except those on the edges have six neighbors and are initially regularly spaced. Such a mesh defines a surface composed of three-sided planar polygons that we call triangular facets, or simply facets. Triangular facets are particularly easy to manipulate for image and shadow generation; consequently, they are the basis for many 3-D graphics systems. These facets tend to form hexagons and can be used to construct virtually arbitrary surfaces. Finally, standard triangulation algorithms can be used to generate such a surface from noisy real data [Fua and Sander, 1992, Szeliski and Tonnesen, 1992].

Sources of information. A number of information sources are available for the reconstruction of a surface and its material properties. Here, we consider two classes of information.

The first class comprises those information sources that do not require more than one image, such as texture gradients, shading, and occlusion edges. When using multiple images and a full 3-D surface representation, however, we can do certain things that cannot be done with a single image. First, the information source can be checked for consistency across all images, taking occlusions into account. Second, when the source is consistent and occlusions are taken into account, the information can be fused over all the images, thereby increasing the accuracy of the reconstruction.

The second class comprises those information sources that require at least two images, such as the triangulation of corresponding points between input images (given camera models and their relative positions). Generally speaking, this source is most useful when corresponding points can be easily identified and their image positions accurately measured. The ease and accuracy of this correspondence can vary significantly from place to place in the image set, and depend critically on the type of feature used. Consequently, whatever the type of feature used, one must be able to identify where in the images that feature provides reliable correspondences, and what accuracy one can expect.

The image feature that we have chosen for correspondence (although it is by no means the only one possible) is simply intensity in radiometrically corrected images—for example, by filtering them. Clearly, intensity can be a reliable feature only when the albedo varies quickly enough on the surface and, consequently, the images are sufficiently textured.

Simple correlation-based stereo methods often use fixed-size windows in images to measure disparities, which will in general yield correct results only when the surface is parallel to the image plane. Instead, we compare the intensities as projected onto the facets of the surface. Consequently, the reconstruction can be significantly more accurate for slanted surfaces. Some correlation-based algorithms achieve similar results by using variable-shaped windows in the images

[Quam, 1984, Nishihara, 1984, Kanade and Okutomi, 1990, Baltsavias, 1991, Devernay and Faugeras, 1994]. However, they typically use only image-centered representations of the surface.

Our approach is much more closely related to the least-squares approaches advocated by Wrobel [1991] and Heipke [1992], who both use a 2.5–D representation of the surface, with recent extensions to full 3–D modeling [Schlueter, 1999].

As for the monocular information source, we have chosen to use shading, where shading is the change in image intensity due to the orientation of the surface relative to a light source. We use this method because shading is most reliable when the albedo varies slowly across the surface; this is the natural complement to intensity correspondence, which requires quickly varying albedo. The complementary nature of these two sources allows us to accurately recover the surface geometry and material properties for a wide variety of images.

In contrast to our approach, traditional uses of shading information assume that the albedo is constant across the entire surface, which is a major limitation when applied to real images. We overcome this limitation by improving upon a method to deal with discontinuities in albedo alluded to in the summary of Leclerc and Bobick [1991]. We compute the albedo at each facet by using the normal to the facet, a light-source direction, and the average of the intensities projected onto the facet from all images. We use the local variation of this computed albedo across the surface as a measure of the correctness of the surface reconstruction. To see why albedo variation is a reasonable measure of correctness, consider the case when the albedo of the real surface is constant. When the geometry of the mesh is correct, the computed albedo should be approximately the same as the real albedo, and hence should be approximately constant across the mesh. Thus, when the geometry is incorrect, this will generally give rise to variations in the computed albedo that we can take advantage of. Furthermore, by using a *local* variation in the computed albedo, we can deal with surfaces whose albedo is not constant, but instead varies slowly over the surface.

Implementation. The triangulated 3–D mesh of vertices that represents a surface, $\mathcal{S}$, is a hexagonally connected set of vertices such as the one shown in Figure 8(a). The position of a vertex v_j is specified by its Cartesian coordinates (x_j, y_j, z_j). The mesh can be deformed by varying these coordinates to minimize an objective function that includes terms derived from stereo and shading information. Its state vector S is the vector of all x,y, and z coordinates.

The stereo component of the objective function is derived by comparing the gray levels of the points in all the images for which the projection of a given point on the surface is visible. It is similar to the term proposed by Wrobel [1991]. As shown in Figure 8(b), this comparison is done for a uniform sampling of the surface. We take the stereo component to be the sum of the variances of the gray level of the projections of these sample points. This method allows us to deal with arbitrarily slanted regions and to discount occluded areas of the surface.

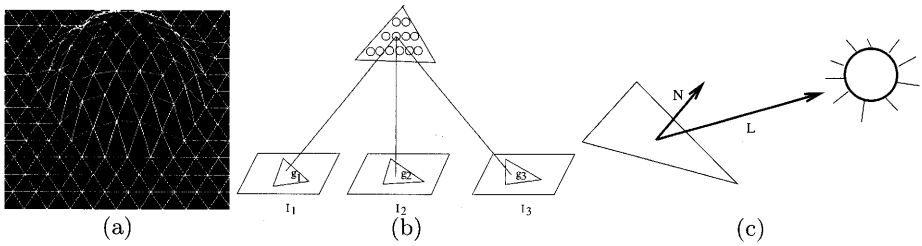


Fig. 8. Mesh representation and objective function: (a) Wireframe representation of the mesh. (b) To compute the stereo component of the objective function, facets are sampled at regular intervals; the circles represent the sample points. (c) Each facet's albedo is estimated using its normal N, the light source direction L, and the average gray level of the projection of the facet into the images.

The shading component of the objective function is computed by using a method that does not invoke the traditional constant albedo assumption. Instead, it attempts to minimize the variation in albedo across the surface, and can therefore deal with surfaces whose albedo varies slowly. This term is depicted by Figure 8(c). We take it to be the sum of the squared differences in estimated albedo across neighboring facets. Each facet's albedo is estimated using its normal N, the light source direction L, and the average gray level of the projection of the facet into the images.

The stereo term is most useful when the surfaces are highly textured. Conversely, the shading term is most reliable where the surfaces have little or no texture. To account for this phenomenon, we take the complete objective function, $\mathcal{E}(\mathcal{S})$, to be a weighted average of these two components where the weighting is a function of texture within the projections of individual facets.

In general, $\mathcal{E}(\mathcal{S})$ is a highly nonconvex function of the vertex positions. To minimize $\mathcal{E}(\mathcal{S})$, we use the "snake-type" [Kass *et al.*, 1988] optimization technique of Section 3.1. We define the total energy of the mesh, $\mathcal{E}_T(\mathcal{S})$, as

$$\mathcal{E}_T(\mathcal{S}) = \mathcal{E}_D(\mathcal{S}) + \mathcal{E}(\mathcal{S}) \tag{27}$$

where $\mathcal{E}_D(\mathcal{S})$ is a regularization term analogous to the one of Equation 18. In practice, we take $\mathcal{E}_D$ to be a measure of the curvature or local deviation from a plane at every vertex. Because the mesh is regular, $\mathcal{E}_D$ can be approximated by using finite differences as a quadratic form [Fua and Leclerc, 1995]

$$\mathcal{E}_D(\mathcal{S}) = 1/2(X^T K X + Y^T K Y + Z^T K Z)\,, \tag{28}$$

where X,Y, and Z are the vectors of the x,y, and z coordinates of the vertices, and K is a sparse and banded matrix. This regularization term serves a dual purpose. First, as before, it "convexifies" the energy landscape when λ_D is large and improves the convergence properties of the optimization procedure. Second,

in the presence of noise, some amount of smoothing is required to prevent the mesh from overfitting the data, and wrinkling the surface excessively.

To speed the computation and prevent the mesh from becoming stuck in undesirable local minima, we typically use several levels of mesh sizes—three in the example of Figure 4(b)—to perform the computation. We start with a relatively coarse mesh that we optimize. We then refine it by splitting every facet into four smaller ones and reoptimizing. Finally, we repeat the split and optimization processes one more time.

3.2 Enforcing Consistency

We now turn to the enforcing of geometric and consistency constraints on the multiple objects that may compose a complex site.

A traditional way to enforce such constraints is to add a penalty term to the model's energy function for each constraint. While this may be effective for simple constraints, this approach rapidly becomes intractable as the number of constraints grows, for two reasons. First, it is well known that minimizing an objective function that includes such penalty terms constitutes an ill-behaved optimization problem with poor convergence properties [Fletcher, 1987, Gill *et al.*, 1981]: the optimizer is likely to minimize the constraint terms while ignoring the remaining terms of the objective function. Second, if one tries to enforce several constraints of different natures, the penalty terms are unlikely to be commensurate and one has to face the difficult problem of adequately weighing the various constraints.

Using standard constrained optimization techniques is one way of solving these two problems. However, while there are many such techniques, most involve solving large linear systems of equations and few are tailored to preserving the convergence properties of the snake-like approaches of Sections 3.1 and 3.1. Exceptions are the approach proposed by Metaxas and Terzopoulos [1991] to enforce holonomic constraints by modeling the second-order dynamics of the system and the technique proposed by Amini *et al.* [1988] using dynamic programming.

Here, we propose a new approach to enforcing hard constraints on our snakes without undue computational burden while retaining their desirable convergence properties.

Constrained Optimization in Orthogonal Subspaces Formally, the constrained optimization problem can be described as follows. Given a function f of n variables $S = \{s_1, s_2, .., s_n\}$, we want to minimize it under a set of m constraints $C(S) = \{c_1, c_2, .., c_m\} = 0$. That is,

$$\text{minimize } f(S) \text{ subject to } C(S) = 0 \ . \tag{29}$$

While there are many powerful methods for nonlinear constrained minimization [Gill *et al.*, 1981], we know of none that are particularly well adapted to snake-like optimization: they do not take advantage of the locality of interactions

that is characteristic of snakes. We have therefore developed a robust two-step approach [Brechbühler *et al.*, 1995, Fua and Brechbühler, 1996] that is closely related to gradient projection methods first proposed by Rosen [1961] and can be extended to snake optimization.

Solving a constrained optimization problem involves satisfying the constraints and minimizing the objective function. For our application, it has proved effective to decouple the two and decompose each iteration into two steps:

1. Enforce the constraints by projecting the current state onto the constraint surface. This involves solving a system of nonlinear equations by linearizing them and taking Newton steps.
2. Minimize the objective function by projecting the gradient of the objective function onto the subspace tangent to the constraint surface and searching in the direction of the projection, so that the resulting state does not stray too far away from the constraint surface.

Figure 9 depicts this procedure. Let C and S be the constraint and state vectors of Equation 29 and A be the $n \times m$ Jacobian matrix of the constraints. The two steps are implemented as follows:

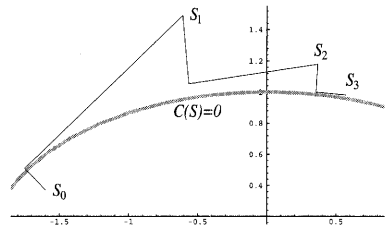


Fig. 9. Constrained optimization. Minimizing $(x - 0.5)^2 + (y - 0.2)^2$ under the constraint that $(x/2)^2 + y^2 = 1$. The set of all states that satisfy the constraint $C(S) = 0$, i.e. the constraint surface, is shown as a thick gray line. Each iteration consists of two steps: orthognal projection onto the constraint surface followed by a line search in a direction tangent to the surface. Because we perform only one Newton step at each iteration, the constraint is fully enforced after only a few iterations.

1. To project S, we compute dS such that $C(S + dS) \approx C(S) + A^t dS = 0$ and increment S by dS. The shortest possible dS is found by writing dS as AdV and solving the equation $A^t AdV = -C(S)$.
2. To compute the optimization direction, we first solve the linear system $A^T(S)A(S)\lambda = A^T(S)\nabla f$ and take the direction to be $\nabla f - A\lambda$. This amounts to estimating Lagrange multipliers, that is, the coefficients that can be used to describe ∇f as closely as possible as a linear combination of constraint normals.

These two steps operate in two locally orthogonal subspaces, in the column space of A and in its orthogonal complement, the null space of A^T. Note that $A^T(S)A(S)$ is an $m \times m$ matrix and is therefore small when there are more variables than constraints, which is always the case in our application.

This technique has been used to enforce the geometric constraints in the example of Figure 7. Furthermore, it can be generalized to handle inequality constraints by introducing an "active set strategy." The inequality constraints that are strictly satisfied are deactivated, while those that are violated are activated and treated as equality constraints. This requires additional bookkeeping but does not appear to noticeably slow down the convergence of our constrained-optimization algorithm.

Constraining Snake Optimization We could trivially extend the technique of Section 3.2 to the refinement of smooth curves and surfaces by taking the objective function f to be the total energy $\mathcal{E}_T$ of Equation 18. However, this would be equivalent to optimizing an unconstrained snake by using gradient descent as opposed to performing the implicit Euler steps that so effectively propagate smoothness constraints.

In practice, propagating the smoothness constraints is key to forcing convergence toward desirable answers. When a portion of the snake deforms to satisfy a hard constraint, enforcing regularity guarantees that the remainder of the snake also deforms to preserve it and that unwanted discontinuities are not generated. This is especially true in our application because many of the constraints we use can be satisfied by moving a small number of vertices, thereby potentially creating "kinks" in the curve or surface that subsequent optimization steps may not be able to remove without getting stuck in local minima.

Therefore, for the purpose of optimizing constrained smooth snakes, we decompose the second step of the optimization procedure of Section 3.2 into two steps. We first solve the unconstrained Dynamics Equation (Equation 20) as we do for unconstrained snakes. We then calculate the component of the snake step vector—the difference between the snake's current state and its previous one—that is perpendicular to the constraint surface and subtract it from the state vector. The first step regularizes, while the second prevents the snake from moving too far away from the constraint surface.

As in the case of unconstrained snakes, α, the viscosity term of Equation 20, is computed automatically at the start of the optimization and progressively increased as needed to ensure a monotonic decrease of the snake's energy and ultimate convergence of the algorithm.

Let S be the snake's state vector as described in Sections 3.1 and 3.1. An iteration of the optimization procedure involves the following three steps:

1. Take a Newton step to project S_{t-1}, the current state vector, onto the constraint surface.

$$S_{t-1} \leftarrow S_{t-1} + AdV where A^T AdV = -C(S_{t-1}) \ .$$

If the snake's total energy has increased, back up and increase viscosity.

2. Take a normal snake step by solving

$$(K_S + \alpha I)S_t = \alpha S_{t-1} - \left.\frac{\partial \mathcal{E}}{\partial S}\right|_{S_{t-1}} .$$

3. Ensure that dS, the snake step from S_{t-1} to S_t, is in the subspace tangent to the constraint surface.

$$S_t \leftarrow S_t - A\lambda \text{ where } A^t A\lambda = A^T(S_t - S_{t-1}) \ ,$$

so that the snake step dS becomes

$$\begin{aligned} dS &= (S_t - A\lambda) - S_{t-1} \\ \Rightarrow A^T dS &= 0. \end{aligned}$$

Multiple Snakes Our technique can be further generalized to the simultaneous optimization of several snakes under a set of constraints that bind them. We concatenate the state vectors of the snakes into a composite state vector S and compute for each snake the viscosity coefficient that would yield steps of the appropriate magnitude if each snake was optimized individually. The optimization steps become

1. Project S onto the constraint surface as before and compute the energy of each individual snake. For all snakes whose energy has increased, revert to the previous position and increase the viscosity.
2. Take a normal snake step for each snake individually.
3. Project the global step into the subspace tangent to the constraint surface.

Because the snake steps are taken individually, we never have to solve the potentially very large linear system involving all the state variables of the composite snake but only the smaller individual linear systems. Furthermore, to control the snake's convergence via the progressive viscosity increase, we do not need to sum the individual energy terms. This is especially important when simultaneously optimizing objects of a different nature, such as a surface and a linear feature, whose energies are unlikely to be commensurate so that the sum of these energies would be essentially meaningless.

In effect, the optimization technique proposed here is a decomposition method and such methods are known to work well [Gill *et al.*, 1981] when their individual components, the individual snake optimizations, are well behaved, which is the case here.

3.3 Consistent Site Modeling

We demonstrate the ability of our technique to impose geometric constraints on 2–D and 3–D deformable models using real imagery. More specifically, we address the issue of optimizing the models of 3–D linear features such as roads, ridgelines, rivers, and the terrain on which they lie under the constraint that

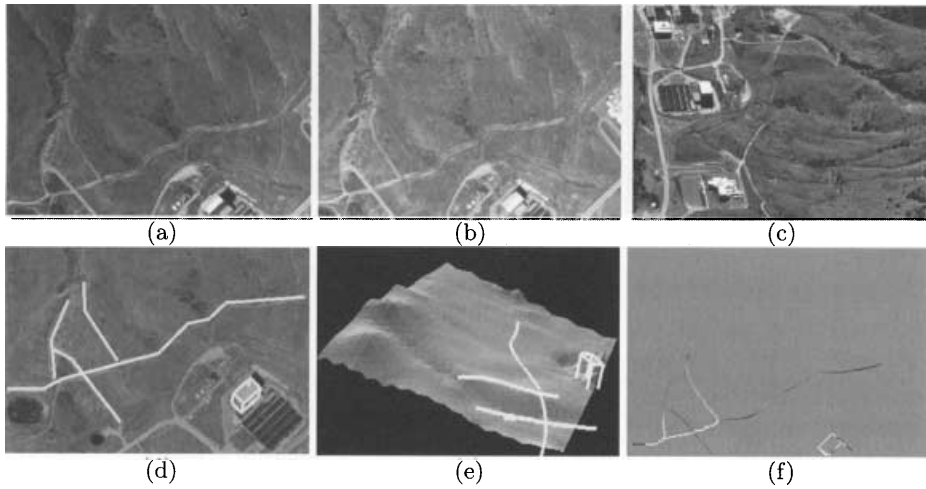


Fig. 10. Building a site model. (a,b,c) Three images of a site with roads and buildings. (d) A rough sketch of the road network and of one of the buildings. (e) Shaded view of the terrain with overlaid roads after independent optimization of each. Note that the two roads in the lower right corner appear to be superposed in this projection because their recovered elevations are inaccurate. (f) Differences of elevation between the optimized roads and the underlying terrain. The image is stretched so that black and white represent errors of minus and plus 5 meters, respectively.

they be consistent with one another. In Figures 3 and 10 we present two such cases where recovering the terrain and the roads independently of one another leads to inconsistencies.

Because we represent the terrain as a triangulated mesh and the features as 3–D polygonal approximations, consistency can be enforced as follows. For each edge $((x_1, y_1, z_1), (x_2, y_2, z_2))$ of the terrain mesh and each $((x_3, y_3, z_3), (x_4, y_4, z_4))$ segment of a linear feature that intersect when projected in the (x, y) plane, the four endpoints must be coplanar so that the segments also intersect in 3–D space. This can expressed as

$$\begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 0 \ , \tag{30}$$

which yields a set of constraints that we refer to as *consistency constraints.*

In Figures 4 and 11, we show that the optimization under the constraints of Equation 30 avoids the discrepancies that result from independent optimization of each feature.

In the example of Figure 4, the "ridge snake" attempts to maximize the average edge gradient along its projections in all three images. In the case of

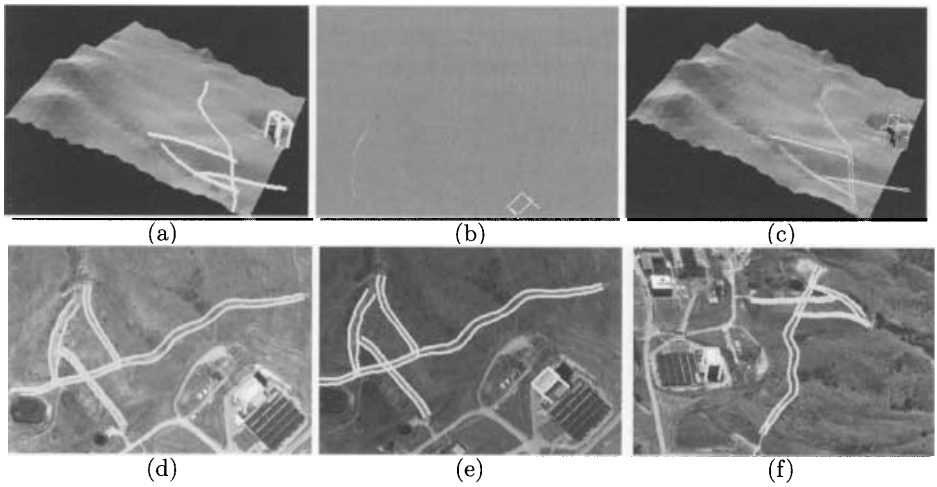


Fig. 11. Recovering the 3–D geometry of both terrain and roads. (a) Shaded view of the terrain with overlaid low-resolution roads after optimization under consistency constraints. (b) Corresponding differences of elevation between features and underlying terrain. The image is stretched like the one of Figure 10(f). Note that only the roof of the building is significantly above the terrain. (c) The roads modeled as ribbons overlaid on the terrain. (d,e,f) The optimized roads overlaid on the original images.

Figures 10 and 11, the roads are lighter than the surrounding terrain. At low resolution, they can effectively be modeled as white lines, and the corresponding snakes attempt to maximize image intensity along their projections. At higher resolution, they are better modeled using the 3–D ribbon snakes of Section 3.1. We also introduce a building and use its base to further constrain the terrain. Figures 11(a,b) depict the result of the simultaneous optimization of the terrain and low-resolution roads. By supplying an average width for the roads, we can turn the lines into ribbons and reoptimize terrain and features under the same consistency constraints as before, yielding the result of Figure 11(c).

The case of rivers is somewhat more complex. Like roads, rivers are represented as linear features that must lie on the terrain. But, in addition, the system must ensure that they flow downhill and at the bottom of valleys. By introducing the active set strategy described at the end of Section 3.2, we have been able to impose such constraints and to generate the more complete site model of of Figure 12.

These examples illustrate the ability of our approach to model different kinds of features in a common reference framework and to produce consistent composite models.

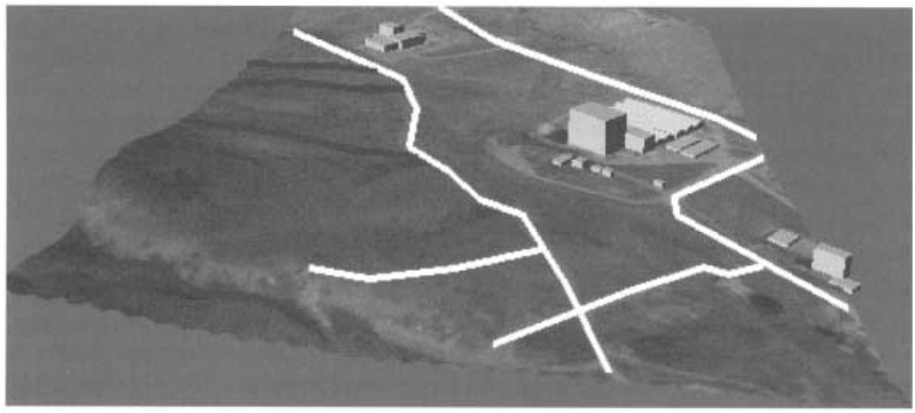


Fig. 12. Texture mapped view of the composite model. The drainage pattern appears as dark lines, the roads as white lines.

4 LSB-Snakes

LSB-Snakes derive their name from the fact that they are a combination of least squares template matching [Gruen, 1985] and B-spline Snakes. B-spline Snakes have been applied to satellite and aerial images [Trinder and Li, 1995, Li, 1997] and are an alternative to the polygonal curves of Section 3.1.

For LSB-Snakes we use three types of observations, which are also based on the generic road model of Section 2. These observations can be divided in two classes, photometric observations, that represent the gray level matching of images with the object model, and geometric observations that express the geometric constraints and the a priori knowledge of the location and shape of the feature to be extracted.

4.1 Photometric Observation Equations

Assume the photometric model of the feature to be extracted is formulated as a discrete two dimensional function $PM(x, y)$. It may be formed by a real or synthetic image. Its values can be the intensities or other quantities derived from them, for instance, the gradient of the intensities. It can also be a vector function whose components express the different aspects of the a priori photometric knowledge (for examples, intensities, gradients, moments and other derivatives) of the feature to be extracted. Suppose an image patch is given as a discrete two dimensional function $I(x, y)$. In terms of least squares adjustment, this patch can be interpreted as an observation vector of the photometric model $PM(x, y)$. A nonlinear observation equation is established as

$$PM(x, y) - e(x, y) = T(I(x, y)) \tag{31}$$

where $e(x, y)$ is a true error function and $T(...)$ is an abstract transform which represents the radiometric relationship between the function values of the photometric model and grey levels of the image patch. This transform basically consists of two parts. The first part represents the functional operations for computation of the values of the photometric model by grey levels. This part can be determined exactly based on the knowledge of the model and applied to the image prior to or during the least squares adjustment. The second part models the corrections of radiometric distortions. The investigations on issues with regard to flexibility of implementation, control of the parameters, stability of the solution, quality analysis and computational speed led to the conclusion that the radiometric corrections should not be included in the estimation model [Baltsavias, 1991]. Instead, they should be applied to the grey levels prior to the adjustment. Assume the transformed image patch is given as $G(x, y)$. Applying Taylor's series to Equation 31, dropping second and higher order terms, the linearized form of the observation equation becomes

$$-e(x,y) = G_x(x^0, y^0)\Delta x + G_y(x^0, y^0)\Delta y + (g(x^0, y^0) - PM(x,y)) \qquad (32)$$

The relationship between the template and the image patch needs to be determined in order to extract the feature: The corrections Δx and Δy of equation 32 have to be estimated. In the conventional least squares template matching applied to feature extraction, an image patch is related to a template through a geometrical transformation, formulated normally by a six parameter affine transformation to model the geometric deformation [Gruen, 1985]. The template is typically square or rectangular and sizes range from 5×5 to 25×25 pixels. Originally the LSM technique is only a local operator used for high precision point measurement. It was extended to an edge tracking technique to automatically extract edge segments [Gruen and Stallmann, 1991], and a further extension was made through the introduction of object-type-dependent neighbourhood constraints between individual patches to enforce the local continuity in a global solution [Gruen and Agouris, 1994]. In another approach least squares template matching was combined with Kalman filtering of the parameters of the road for road tracking [Vosselman and Knecht, 1995].

Instead of a square or rectangular template, we extend the least squares template matching technique into our LSB-Snakes approach by using a deformable contour as the template. Figure 13 shows an initial template and the corresponding final solution. Its center line defines the location of the extracted feature, in the example of the figure the center line of a road segment. Suppose a linear feature, the center line of the template is approximated by a spline curve and represented in parametric form as

$$\begin{aligned} x(s) &= \sum_{i=1}^{n} N_i^m(s) X_i \\ y(s) &= \sum_{i=1}^{n} N_i^m(s) Y_i \end{aligned} \qquad (33)$$

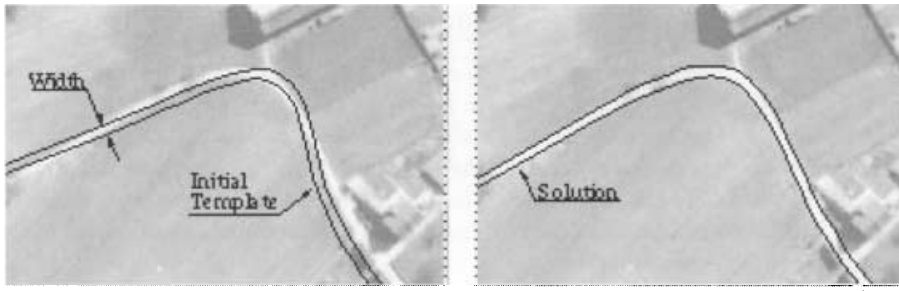


Fig. 13. LSB-Snakes. Initial template and final solution.

where X_i and Y_i are the coefficients of the B-spline curve in the x and y direction respectively. In terms of the B-spline concept, they form the coordinates of the control polygon of the curve. $N_i^m(s)$ is the normalized mth B-spline between knots u_i and u_{i+1} [Bartels and R. H., 1987] . While the knot sequence is given, feature extraction can be treated as a problem of estimation of the coefficients X_i and Y_i of the spline curve. The first order differentials of the B-spline curve can be obtained as

$$\begin{aligned} \Delta x &= \sum_{i=1}^{n} N_i^m(s) \Delta X_i \\ \Delta y &= \sum_{i=1}^{n} N_i^m(s) \Delta Y_i \end{aligned} \tag{34}$$

Substituting the terms in Equation 32, we obtain the linearized photometric observation equations with respect to the coefficients of the B-splines. The linearization of the observation equations for all involved pixels can be expressed in matrix form as

$$-e_m = \mathbf{G}_x \mathbf{N} \Delta \mathbf{X} + \mathbf{G}_y \mathbf{N} \Delta \mathbf{Y} - \mathbf{l}_m \; ; \; \mathbf{P}_m \tag{35}$$

with

$$\mathbf{N} = [N_0^m(s) N_1^m(s) ... N_n^m(s)] \; . \tag{36}$$

Since a linear feature is essentially unidirectional, the template would slide along it during matching. To ease this problem and simplify the implementation, the above equations are changed to

$$\begin{aligned} -e_{mx} &= \mathbf{G}_x \mathbf{N} \Delta \mathbf{X} - \mathbf{l}_{mx} \; ; \; \mathbf{P}_{mx} \\ -e_{my} &= \mathbf{G}_y \mathbf{N} \Delta \mathbf{Y} - \mathbf{l}_{my} \; ; \; \mathbf{P}_{my} \end{aligned} \tag{37}$$

A pair of independent observation equations are thus formed for the x and y directions. The observation vectors $\mathbf{l}_{mx}$ and $\mathbf{l}_{my}$ contain the differences of conjugate pixels. $\mathbf{P}_{mx}$ and $\mathbf{P}_{my}$ are the corresponding weight matrices, which are introduced as diagonal matrices.

4.2 Geometric Observation Equations

In our semi-automatic feature extraction scheme, a set of seed points near the feature of interest is given by a human operator or other preprocessing procedures. In terms of least squares adjustment, these seed points can be interpreted as the control points which determine the location of the feature to be extracted. Because they are only coarsely given, a correction has to be estimated. Therefore they should be considered as observations. Thus the second type of observation equations can be established as

$$\begin{aligned} -e_{cx} &= x - x_0 \; ; \; \mathbf{P}_{cx} \; , \\ -e_{cy} &= y - y_0 \; ; \; \mathbf{P}_{cy} \; , \end{aligned} \tag{38}$$

where x_0 and y_0 are the observation vectors of coordinates of the seed points in the x and y direction respectively, and $\mathbf{P}_{cx}$ are $\mathbf{P}_{cy}$ the corresponding weight matrices, introduced as diagonal matrices. The linearization of the coordinates with respect to the coefficients of the B-splines can be expressed in matrix form as

$$\begin{aligned} -e_{cx} &= \mathbf{N}\Delta\mathbf{X} - \mathbf{t}_{cx} \; ; \; \mathbf{P}_{cx} \; , \\ -e_{cy} &= \mathbf{N}\Delta\mathbf{Y} - \mathbf{t}_{cy} \; ; \; \mathbf{P}_{cy} \; , \end{aligned} \tag{39}$$

where $\mathbf{t}_{cx}$ and $\mathbf{t}_{cy}$ are given by

$$\begin{aligned} \mathbf{t}_{cx} &= \mathbf{x}^0 - \mathbf{x}_0 = \mathbf{N}\mathbf{X}^\mathbf{0} - \mathbf{x}_0 \; , \\ \mathbf{t}_{cy} &= \mathbf{y}^0 - \mathbf{y}_0 = \mathbf{N}\mathbf{Y}^\mathbf{0} - \mathbf{y}_0 \; . \end{aligned} \tag{40}$$

With the seed points an initial curve is formed as a first shape approximation of the feature. In order to stabilize the local deformation of the template we introduce the following smoothness constraints. Assume the initial curve is expressed by $x^0(s)$ and $y^0(s)$. We establish the third type of observation equations based on the first and second derivatives of the curve as

$$\begin{aligned} -e_{sx} &= x_s(s) - x_s^0(s) \; ; \; P_{sx} \; , \\ -e_{sy} &= y_s(s) - y_s^0(s) \; ; \; P_{sy} \; , \end{aligned} \tag{41}$$

$$\begin{aligned} -e_{ssx} &= x_{ss}(s) - x_{ss}^0(s) \; ; \; P_{ssx} \; , \\ -e_{ssy} &= y_{ss}(s) - y_{ss}^0(s) \; ; \; P_{ssy} \; , \end{aligned} \tag{42}$$

Linearizing them with respect to the coefficients of the B-spline they can be expressed in matrix form as

$$\begin{aligned} -e_{sx} &= \mathbf{N_s}\Delta\mathbf{X} - \mathbf{t}_{sx} \; ; \; \mathbf{P}_{sx} \; , \\ -e_{sy} &= \mathbf{N_s}\Delta\mathbf{Y} - \mathbf{t}_{sy} \; ; \; \mathbf{P}_{sy} \; , \end{aligned} \tag{43}$$

$$\begin{aligned} -e_{ssx} &= \mathbf{N_{ss}}\Delta\mathbf{X} - \mathbf{t}_{ssx} \; ; \; \mathbf{P}_{ssx} \; , \\ -e_{ssy} &= \mathbf{N_{ss}}\Delta\mathbf{Y} - \mathbf{t}_{ssy} \; ; \; \mathbf{P}_{ssy} \; , \end{aligned} \tag{44}$$

where $\mathbf{N_s}$ and $\mathbf{N_{ss}}$ are the first and second derivatives of $\mathbf{N}$ defined in Equation 36, and the $\mathbf{t}$ terms are given by

$$\begin{aligned} \mathbf{t}_{sx} &= \mathbf{N_s}\mathbf{X^0} - \mathbf{x}_s^0 \,, \\ \mathbf{t}_{sy} &= \mathbf{N_s}\mathbf{Y^0} - \mathbf{y}_s^0 \,, \end{aligned} \tag{45}$$

$$\begin{aligned} \mathbf{t}_{ssx} &= \mathbf{N_{ss}}\mathbf{X^0} - x_{ss}^0 \,, \\ \mathbf{t}_{ssy} &= \mathbf{N_{ss}}\mathbf{Y^0} - y_{ss}^0 \,. \end{aligned} \tag{46}$$

Any other a priori geometric information of the feature can be formulated in this manner. A joint system is formed by all of these observation Equations 37, 39, 43 and 44.

4.3 Solution of LSB-Snakes

In our least squares approach linear feature extraction is treated as the problem of estimation of the unknown coefficients $\mathbf{X}$ and $\mathbf{Y}$ of the B-spline curve. This is achieved by minimizing a goal function which measures the differences between the template and the image patch and which includes the geometrical constraints. The goal function to be minimized in this approach is the $L_2 - norm$ of the residuals of least squares estimation. It is equivalent to the snake's total energy of and can be written as

$$\begin{aligned} \mathbf{v}^t\mathbf{P}\mathbf{v} &= (\mathbf{v_s}^t\mathbf{P_s}\mathbf{v_s} + \mathbf{v_{ss}}^t\mathbf{P_{ss}}\mathbf{v_{ss}}) + \mathbf{v_m}^t\mathbf{P_m}\mathbf{v_m} + \mathbf{v_c}^t\mathbf{P_c}\mathbf{v_c} \\ &= E_l + E_X + E_C \Rightarrow \text{Minimum} \,. \end{aligned} \tag{47}$$

E_l denotes the internal (geometric) energy of the snakes derived from smoothness constraints, E_X denotes the external (photometric) energy derived from the object model and the image data, and E_C represents the control energy which constrains the distance between the solution and its initial location. At a minimum of this total snake energy, the following necessary conditions must hold:

$$\frac{\partial \mathbf{v}^t\mathbf{P}\mathbf{v}}{\Delta\mathbf{X}} = \frac{\partial \mathbf{v}^t\mathbf{P}\mathbf{v}}{\Delta\mathbf{Y}} = 0 \,. \tag{48}$$

A further development of these formulae will result in a pair of normal equations used for the estimation of $\Delta\mathbf{X}$ and $\Delta\mathbf{Y}$ respectively. Because of the local support property of B-splines, it can be shown that the normal equations are banded (bandwidth $b = m+1$) and the solution can be efficiently computed. The various tools of least squares estimation with their familiar and well established mathematical formulations can be profitably used for the statistical analysis of the results and the realistic evaluation of the algorithm's performance. We can evaluate the estimated parameters' covariance matrix and derive quantities, such as a system noise estimate, from it. As shown in Figure 17, in conjunction with traditional least squares estimation, robust estimation effectively avoids blunders.

4.4 LSB-Snakes with Multiple Images

If a feature is extracted from more than one image, its coordinates in 3–D object space can be derived. There are two main ways to perform multi-image matching:

1. Object-space correlation methods [Wrobel, 1987, Ebner and Heipke, 1998, Helava, 1988]. The methods relate two or more digital images to an object space model. The object is modelled by two surfaces, a geometric terrain model $Z = Z(X, Y)$ and an optical density model $D = D(X, Y)$. The relation between image and object space is through the image formation model which is composed of a geometric camera model and a radiometric model, by which the impact of light waves are strongly expressed. The unknown parameters are estimated by a simultaneous least squares adjustment in which the observations are the pixel intensities. With the maximum configuration of the algorithm, the unknown parameters may include the heights and densities at the nodes of the two grids (ground elements or "groundels" for short), sensor parameters and radiometric model parameters. This method is definitely of theoretical interest. However, it cannot be easily applied to LSB-Snakes without extending the algorithm because we are facing here a truly 3–D problem as opposed to a 2.5–D surface reconstruction problem.
2. Multiphoto geometrically constrained matching (MPGC) [Gruen, 1985, Gruen and Baltsavias, 1985]. This method connects the photometric observation equations of every image by means of external geometrical constraints. One class of the most important constraints is generated by the imaging rays intersection conditions [Gruen, 1985]. Since LSB-Snakes deal with a curve instead of an individual point, a direct analogy to the MPGC technique may introduce many unknowns and therefore increase the complexity of the computation. A modified version of the MPGC algorithm is used and developed into 3–D LSB-Snakes.

Suppose a linear feature in 3–D object space can be approximated by a spline curve and represented in B-spline parametric form as

$$\begin{aligned} X_T(s) &= \mathbf{NX}\,, \\ Y_T(s) &= \mathbf{NY}\,, \\ Z_T(s) &= \mathbf{NZ}\,, \end{aligned} \tag{49}$$

(50)

where $\mathbf{N}$ is defined in Equation 36, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are the coefficient vectors of the B-spline curve in 3–D object space and $X_T(s)$, $Y_T(s)$ and $Z_T(s)$ are the object space coordinates of the feature. In the multi-image case depicted by Figure 14, if the image forming process follows the law of perspective projection, a pair of collinearity conditions in parametric form can be formulated for each image patch as

$$x_i = -c\frac{\alpha_{11}(X_T - X_0) + \alpha_{21}(Y_T - Y_0) + \alpha_{31}(Z_T - Z_0)}{\alpha_{13}(X_T - X_0) + \alpha_{23}(Y_T - Y_0) + \alpha_{33}(Z_T - Z_0)}\,,$$

$$y_i = -c\frac{\alpha_{12}(X_T - X_0) + \alpha_{22}(Y_T - Y_0) + \alpha_{32}(Z_T - Z_0)}{\alpha_{13}(X_T - X_0) + \alpha_{23}(Y_T - Y_0) + \alpha_{33}(Z_T - Z_0)} \,. \tag{51}$$

If the interior and exterior orientation parameters of each image are given or can be derived, the unknowns to be estimated using Equations 49 and 51 are the coefficient vectors of a B-spline curve. The first order differentials can be obtained as

$$\begin{aligned} \Delta x_i &= \frac{\partial x_i}{\partial X_T}\mathbf{N}\Delta\mathbf{X} + \frac{\partial x_i}{\partial Y_T}\mathbf{N}\Delta\mathbf{Y} + \frac{\partial x_i}{\partial Z_T}\mathbf{N}\Delta\mathbf{Z} \,, \\ \Delta y_i &= \frac{\partial y_i}{\partial X_T}\mathbf{N}\Delta\mathbf{X} + \frac{\partial y_i}{\partial Y_T}\mathbf{N}\Delta\mathbf{Y} + \frac{\partial y_i}{\partial Z_T}\mathbf{N}\Delta\mathbf{Z} \,. \end{aligned} \tag{52}$$

The linearization of the observation equations with respect to the coefficient vectors of a 3–D B-spline curve is obtained by substituting Equations 52 in Equations 32.

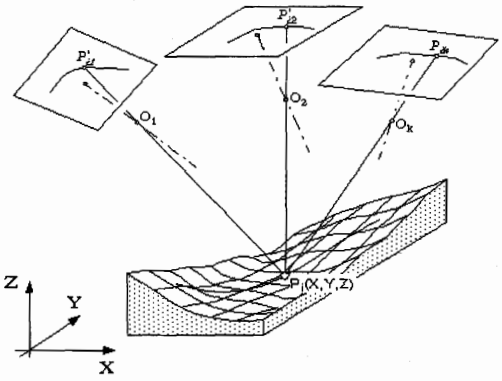


Fig. 14. Multi-image arrangement for 3–D LSB-Snakes. The linear feature is represented by a 3–D B-Spline in object space.

For the same reasons as for 2–D LSB-Snakes the equations are changed such that they can be expressed in matrix form as

$$\begin{aligned} -e_{mx} &= \mathbf{F}_X\mathbf{N}\Delta\mathbf{X} - \mathbf{l}_{mx} \;;\; \mathbf{P}_{mx} \,, \\ -e_{my} &= \mathbf{F}_Y\mathbf{N}\Delta\mathbf{Y} - \mathbf{l}_{my} \;;\; \mathbf{P}_{my} \,, \\ -e_{mz} &= \mathbf{F}_Z\mathbf{N}\Delta\mathbf{Z} - \mathbf{l}_{mz} \;;\; \mathbf{P}_{mz} \,. \end{aligned} \tag{53}$$

$\mathbf{F}_X$, $\mathbf{F}_Y$ and $\mathbf{F}_Z$ are partial derivatives which can be written as

$$F_X = \frac{\partial g(x,y)}{\partial x}\frac{\partial x_i}{\partial X_T} + \frac{\partial g(x,y)}{\partial y}\frac{\partial y_i}{\partial X_T} \,,$$

$$F_Y = \frac{\partial g(x,y)}{\partial x}\frac{\partial x_i}{\partial Y_T} + \frac{\partial g(x,y)}{\partial y}\frac{\partial y_i}{\partial Y_T}\,, \tag{54}$$
$$F_Z = \frac{\partial g(x,y)}{\partial x}\frac{\partial x_i}{\partial Z_T} + \frac{\partial g(x,y)}{\partial y}\frac{\partial y_i}{\partial Z_T}\,.$$

The geometric observation equations 39, 43, 44 can be extended into three dimensions by introducing a new component for the Z-direction. Then the 3–D LSB-Snakes can again be solved by a combined least squares adjustment. That is, a 3–D linear feature is extracted directly from multiple images. The statistical analysis of the obtained results and the realistic evaluation of the algorithmic performance can be done through the use of the covariance matrix of the estimated parameters.

4.5 Road Extraction Experiments

In this section, we present some experimental results of road extraction using our LSB-Snakes approach.

An imaged object is defined and identified by its characteristics, which can be classified into five groups: Photometric, geometric, topological, functional and contextual characteristics. In our semi-automatic feature extraction scheme the high level knowledge, which requires quite some intelligence for the image interpretation process, is used by the human operator to identify and classify the object. The generic object model involved in the model driven feature extraction algorithms consists of some photometric and geometric characteristics. Some of these properties are mathematically formulated and used to generate the template and define the weight functions. For instance, the grey values of the template can be derived from the images through computations of the local contrast according to the first property, while the second property suggests that the weights of the photometric observations should be related to the local changes of the grey levels along the road.

In the current implementation, only one image is displayed through the user interface. The algorithm can run both in a monoplotting and a multi-image mode. In the monoplotting mode, the LSB-Snakes extract linear features in the image space. The 3–D coordinates are obtained in real-time by intersecting the imaging rays through the use of a camera model with an underlying DTM. With multiple images, some very few seed points are given by the operator in one displayed image, the camera model is applied to project them into object space and onto a coarse DTM. Then the 3–D feature is extracted automatically and precisely by the LSB-Snakes algorithm.

Figure 15 shows an example of 2–D LSB-Snakes used for extraction of road segments from a portion of an aerial image of Avenches, Switzerland. The scale of the original photograph is about 1:15,000. The pixel size of the digitized image is 100 microns. Thus, the footprint of the image is about 1.5 meters. Roads in this image have different widths and are disturbed by buildings, trees, cars and other objects. The results show that the algorithm of LSB-Snakes works very well even

Fig. 15. Road extraction in 2-D mode from an aerial image of Avenches. The extracted road center lines are displayed in black, overlaid on the original image. Note that the extracted center lines represent raw data, which, in a second step, has to be refined and converted into a consistent road network structure.

under these unfavourable conditions. Figure 16 shows one example of simultaneous 3–D extraction of a road segment from four images, which are portions of aerial images of Heinzenberg, Switzerland. The four images are from two strips with about 60 % end and side lap. The scale of the original photograph is about 1:15,000. The negative films were scanned at a Leica/Helava scanner with 10 microns and were later subsampled to 40 microns pixel size. Thus the footprint of the images is about 0.6 meters. The seed points (initial polygon), provided manually by the operator, are displayed in black, overlaid on the first image. The extracted road center line is shown as a black curve. Visual tests prove the successful performance of the algorithm. Figure 17 focuses on another portion from the same aerial images. At this time, the problematic areas marked with white rectangles show occlusions in two out of four images caused by trees. In

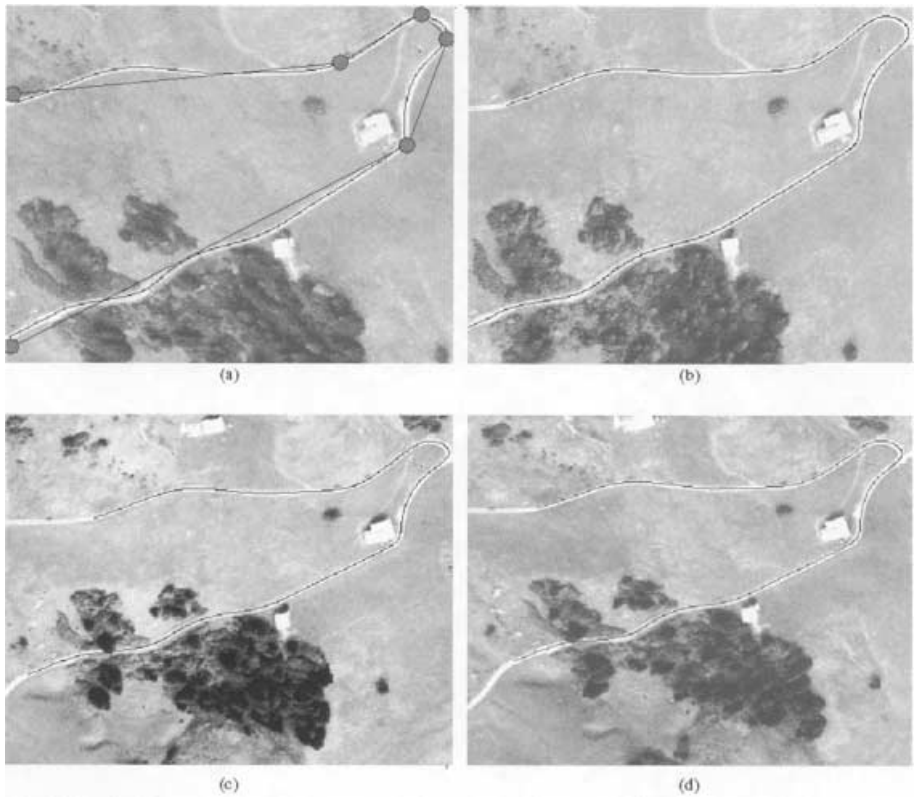


Fig. 16. Simultaneous 3-D extraction of a road segment from four images. (a) Seed points' location (iteration 0) and extracted road center line. (b), (c), (d) Extracted road center line (after ca. 5 iterations).

terms of least squares adjustment, the photometric observations in the problematic areas are blunders. They have to be detected and rejected. This is achieved in our implementation by using adaptive weight functions, in which the weights of observations are related to the ratio of their residuals and the variance factor. To get a large pull-in range the algorithm starts with a flat weight function. To reduce the influence of blunders it becomes steep after three iterations. In such a way, the weights of observations with big residuals will become smaller and smaller. The results shown in Figure 17 prove that the blunders are successfully rejected and the algorithm bridges gaps in a robust manner. For comparison the results without blunder detection are shown in Figure 17(c',d'). Since the road extracted from images (a) and (b) is the same in both cases they are not displayed again. It is also verified by this example that more than two images are required for 3–D linear feature extraction. Using only two images cannot give

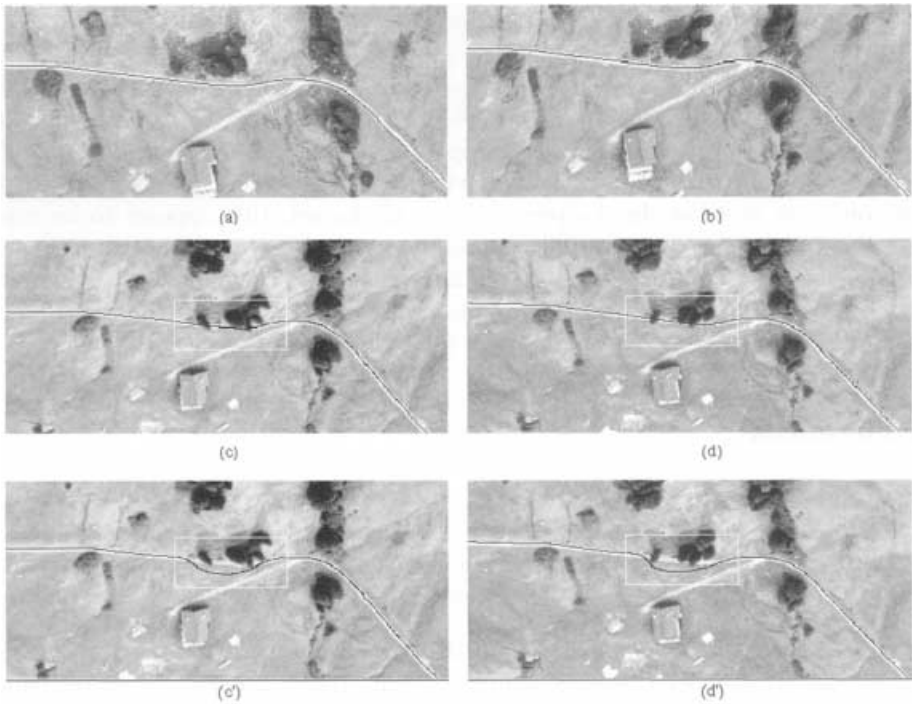


Fig. 17. Simultaneous 3-D extraction of a road segment from four images (a), (b), (c) and (d). The white rectangle denotes the problematic area of occlusion and the extracted road segments are displayed as a black curve. (c') and (d') show the results without blunder detection.

reliable 3–D results. Reliability suffers in places where the linear feature extents in direction of the image base.

5 Conclusion

We have presented optimization-based object modeling techniques for 2–D and 3–D linear features and 3–D surfaces that rely on dynamic-programming, the Euler-Lagrange formalism and least-squares minimization.

Dynamic programming algorithms can be implemented so that they function in real-time on ordinary computers but cannot handle global constraints very effectively. Such constraints are well handled by generalized Euler-Lagrange snakes, but they require reasonable starting points, such as those that can be supplied by dynamic programming algorithms for linear structures or simple stereo-correlation based algorithms for surfaces. Finally, where the contours of the objects to be detected do not suffice, the Euler-Lagrange snakes can be replaced by the LSB-Snakes.

In short, these various algorithms serve different and complementary purposes. We believe that a complete cartographic modeling workstation should include them all and allow a human operator to select the most appropriate one. In parallel to the work reported in this paper, we have developed an expert system based approach to assist the operator in this task [Strat *et al.*, 1997] so that he can perform it without having to be an expert photogrammetrist or computer scientist. In the present state of the art, this appears to be reasonable compromise because fully automatic methods are still far out of reach. Semi-automatic feature extraction methods that allow for human intervention are considered to be a good compromise, combining the mensuration speed and accuracy of a computer algorithm with the interpretation skills of a person.

References

[Amini *et al.*, 1988] A.A. Amini, S. Tehrani, and T.E. Weymouth. Using Dynamic Programming for Minimizing the Energy of Active Contours in the Presence of Hard Constraints. In *International Conference on Computer Vision*, pages 95–99, 1988.

[Ballard and Brown, 1982] D.H. Ballard and C.M. Brown. *Computer Vision.* Prentice-Hall, 1982.

[Baltsavias, 1991] E. P. Baltsavias. *Multiphoto Geometrically Constrained Matching.* PhD thesis, Institute for Geodesy and Photogrammetry, ETH Zurich, December 1991.

[Bartels and R. H., 1987] R. Bartels and Beatty R. H. *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling.* Morgan Kaufmann, 1987.

[Bellman and Dreyfus, 1962] R. Bellman and S. Dreyfus. *Applied Dynamic Programming.* Princeton University Press, 1962.

[Brechbühler *et al.*, 1995] C. Brechbühler, G. Gerig, and O. Kübler. Parametrization of Closed Surfaces for 3-D Shape Description. *Computer Vision and Image Understanding*, 65(2):154–170, March 1995.

[Devernay and Faugeras, 1994] F. Devernay and O. D. Faugeras. Computing Differential Properties of 3–D Shapes from Stereoscopic Images without 3–D Models. In *Conference on Computer Vision and Pattern Recognition*, pages 208–213, Seattle, WA, June 1994.

[Ebner and Heipke, 1998] H. Ebner and C. Heipke. Integration of digital image matching and object surface reconstruction. *International Archives of Photogrammetry and Remote Sensing*, 27(B11):534–545, 1998.

[Fletcher, 1987] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 2nd edition, 1987. A Wiley-Interscience Publication.

[Fua and Brechbühler, 1996] P. Fua and C. Brechbühler. Imposing Hard Constraints on Soft Snakes. In *European Conference on Computer Vision*, pages 495–506, Cambridge, England, April 1996. Available as Tech Note 553, Artificial Intelligence Center, SRI International.

[Fua and Leclerc, 1990] P. Fua and Y. G. Leclerc. Model Driven Edge Detection. *Machine Vision and Applications*, 3:45–56, 1990.

[Fua and Leclerc, 1995] P. Fua and Y. G. Leclerc. Object-Centered Surface Reconstruction: Combining Multi-Image Stereo and Shading. *International Journal of Computer Vision*, 16:35–56, September 1995.

[Fua and Sander, 1992] P. Fua and P. Sander. Segmenting Unstructured 3D Points into Surfaces. In *European Conference on Computer Vision*, pages 676–680, Genoa, Italy, April 1992.

[Fua, 1997] P. Fua. *RADIUS: Image Understanding for Intelligence Imagery*, chapter Model-Based Optimization: An Approach to Fast, Accurate, and Consistent Site Modeling from Imagery. Morgan Kaufmann, 1997. O. Firschein and T.M. Strat, Eds. Available as Tech Note 570, Artificial Intelligence Center, SRI International.

[Gill *et al.*, 1981] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization.* Academic Press, London a.o., 1981.

[Gruen and Agouris, 1994] A. Gruen and P. Agouris. Linear Feature Extraction by Least Squares Template Matching Constrained by Internal Shape Forces. *International Archives of Photogrammetry and Remote Sensing*, 30(3/1):316–323, 1994.

[Gruen and Baltsavias, 1985] A. Gruen and M. Baltsavias. Adaptive Least Squares Correlation with Geometrical Constraints. In *SPIE Proceedings of Computer Vision for Robots*, volume 595, pages 72–82, 1985.

[Gruen and Li, 1995] A. Gruen and H. Li. Road Extraction from Aerial and Satellite Images by Dynamic Programming. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):11–20, 1995.

[Gruen and Stallmann, 1991] A. Gruen and D. Stallmann. High Accuracy Edge Matching with an Extension of the MPGC-Matching Algorithm. In *SPIE Proceedings of Industrial Vision Metrology*, volume 1526, pages 42–45, 1991.

[Gruen, 1985] A. Gruen. Adaptive Least Squares Correlation: a Powerful Image Matching Technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 3(14):175–187, 1985.

[Gruen, 1996] A. Gruen. Digital Photogrammetric Stations Revisited. *International Archives of Photogrammetry and Remote Sensing*, 31(B2):127–134, 1996.

[Heipke, 1992] C. Heipke. Integration of Digital Image Matching and Multi Image Shape From Shading. In *International Society for Photogrammetry and Remote Sensing*, pages 832–841, Washington, D.C., 1992.

[Helava, 1988] U.V. Helava. Object-Space Least-Square Correlation. *Photogrammetric Engineering and Remote Sensing*, 54(6):711–714, 1988.

[Kanade and Okutomi, 1990] T. Kanade and M. Okutomi. A Stereo Matching Algorithm with an Adaptative Window: Theory and Experiment. In *DARPA Image Understanding Workshop.* Morgan Kaufmann, September 1990.

[Kass *et al.*, 1988] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[Leclerc and Bobick, 1991] Y. G. Leclerc and A. F. Bobick. The Direct Computation of Height from Shading. In *Conference on Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 1991.

[Li, 1997] H. Li. *Semi-Automatic Road Extraction from Satellite and Aerial Images.* PhD thesis, Institute of Geodesy and Photogrammetry,ETH-Zuerich, 1997.

[Metaxas and Terzopoulos, 1991] D. Metaxas and D. Terzopoulos. Shape and Nonrigid Motion Estimation through Physics-Based Synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1991.

[Miller *et al.*, 1996] S.B Miller, F. C. Paderes, and A. S. Walker. Automation in Digital Photogrammetric Systems. *International Archives of Photogrammetry and Remote Sensing*, 31(B2):250–255, 1996.

[Mundy *et al.*, 1992] J.L. Mundy, R. Welty, L. Quam, T. Strat, W. Bremmer, M. Horwedel, D. Hackett, and A. Hoogs. The RADIUS Common Development Environment. In *DARPA Image Understanding Workshop*, pages 215–226, San Diego,CA, 1992. Morgan Kaufmann.

[Nishihara, 1984] H.K. Nishihara. Practical Real-Time Imaging Stereo Matcher. *Optical Engineering*, 23(5), 1984.

[Quam, 1984] L.H. Quam. Hierarchical Warp Stereo. In *DARPA Image Understanding Workshop*, pages 149–155, 1984.

[Rosen, 1961] Rosen. Gradient Projection Method for Nonlinear Programming. *SIAM Journal of Applied Mathematics*, 8:181–217, 1961.

[Schlueter, 1999] M. Schlueter. *Von der 2 1/2D- zur 3D-Flaechenmodellierung fuer die photogrammetrische Rekonstruktion im Objektraum.* Deutsche geodaetische kommission, reihe c, nr. 506, isbn-nr.: 3-7696-9545-3, Muenchen, 1999.

[Strat *et al.*, 1997] T.M. Strat, P. Fua, and C. Connolly. *RADIUS: Image Understanding for Intelligence Imagery*, chapter Context-Based Vision. Morgan Kaufmann, 1997. O. Firschein and T.M. Strat, Eds.Available as Tech Note 569, Artificial Intelligence Center, SRI International.

[Szeliski and Tonnesen, 1992] R. Szeliski and D. Tonnesen. Surface Modeling with Oriented Particle Systems. In *Computer Graphics, SIGGRAPH Proceedings*, volume 26, pages 185–194, July 1992.

[Terzopoulos *et al.*, 1987] D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-seeking Models and 3D Object Reconstruction. *International Journal of Computer Vision*, 1:211–221, 1987.

[Trinder and Li, 1995] J. Trinder and H. Li. Semi-Automatic Feature Extraction by Snakes. In Birkhaeuser Verlag, editor, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 95–104, 1995.

[Vosselman and Knecht, 1995] G. Vosselman and J. Knecht. Road tracing by Profile Matching and Kalman filtering. In Birkhaeuser Verlag, editor, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 265–274, 1995.

[Walker and Petrie, 1996] A. S. Walker and G. Petrie. Automation in Digital Photogrammetric Systems. *International Archives of Photogrammetry and Remote Sensing*, 31(B2):384–395, 1996.

[Wrobel, 1987] B.P. Wrobel. Facet Stereo Vision (FAST Vision) - A New Approach to Computer Stereo Vision and to Digital Photogrammetry. In *Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 231–258, Interlaken, Switzerland, 1987.

[Wrobel, 1991] B.P. Wrobel. The evolution of Digital Photogrammetry from Analytical Photogrammetry. *Photogrammetric Record*, 13(77):765–776, April 1991.

Diffraction tomography through phase back-projection

Stefano Valle, Fabio Rocca, Luigi Zanzi
Dipartimento di Elettronica e Informazione, Politecnico di Milano

Abstract

The tomographic imaging can be performed, within the geometrical optics approximation, back-projecting the data along thin (straight) rays. This is the case of traveltime tomography and amplitude tomography. However, the use of thin rays neglects the scattering effect and reduces the resolution capabilities. Diffraction tomography and migration overcome these limits thanks to the "full-wave" approach. They achieve the maximum resolution compatible with a given source. In this paper, a novel diffraction tomography algorithm is presented. Then "full-wave" techniques are compared, focusing the attention on the wavepaths involved in the data back-projection. Finally, numerical and real data tests, designed for the development of a Ground Penetrating Radar tomography technique for Nondestructive Testing, show the successful application of this technique.

1 Introduction

The tomographic techniques based on straight or curved rays (typically traveltime and amplitude tomography) assume that the wave is well approximated by a ray: basically this approximation neglects the band-limited and diffracting character of the electromagnetic (or acoustic) sources. The resolution capabilities can be deduced from two criteria: the Fourier Slice Theorem (indicating the detectable wavenumbers on the basis of the measurement configuration) and the Fresnel zone width (indicating size and spacing of detectable anomalies).

When the scale of the variations is comparable with the wavelength, the ray-approximation is no longer valid since diffraction phenomena become predominant. In this case, the above techniques provide only a partial extraction of the information contained in the data and a full-wave approach is essential to overcome this limit. The diffraction tomography (DT) is one of the most rigorous and well-known full-wave imaging method. DT reconstructs the internal structure of an object from the field diffracted by the refraction index perturbations. The Generalized Fourier Slice Theorem or Fourier Diffraction Theorem contains the DT basic principle. This theorem provides a resolution criterion including both the measurement configuration and the band-limited

A. Dermanis, A. Grün, and F. Sansò (Eds.): LNES 95, pp. 229–256, 2000.

character of the source. Nevertheless, DT requires few approximations to get a linear inversion problem. Born and Rytov approximations are the most common approaches to linearize the relationship between the measurements and the velocity perturbations. They are based on different weak scattering hypotheses.

The most spread DT approach had been proposed by Devaney (1982) and then by Kak and Slaney (1988) and it is based on the scattered field plane-wave decomposition and an inversion in the wavenumber domain. Wu and Toksöz (1987) used this approach to show the advantages of DT with respect to multi-source and multi-frequency holography, that is equivalent to the pre-stack migration (MIG). These methods assume a homogenous background. This assumption must be overcome only when the medium is strongly non-uniform. Miller et al. (1987) used the Generalized Radon Transform to move from the simple imaging migration to a linearized inversion approach that handles variable media. Woodward (1992) proposed a space-domain algorithm to handle irregular geometries and an iterative inversion to take into account inhomogeneous background; besides, this work gives an interesting insight into the wavepaths involved in DT inversion. These results will be used in this lecture to connect DT to MIG. Pratt and Worthington (1990) proposed a non-linear inversion with an optimized forward model to extract the feature of the target in strongly variable media. Gelius (1995) proposed a generalized DT algorithm to handle both irregular geometries and inhomogeneous background; a smooth background is assumed so that iterations are not required. Dickens (1994) and Harris and Wang (1996) showed how to handle layered media without iterations.

In this paper a new DT algorithm is presented. The hypothesis of weak scattering in an homogeneous medium are still assumed. The method estimates the phase of the single wavenumber of the object function from the phase of the scattered field by using the mono-chromatic scattering relationship. The estimate can be done for each source-receiver pair and a constructive sum of each estimate reveals the actual wavenumbers contained in the target. This method can be considered a statistical version of Wu and Toksöz (1987) approach with the advantages that no plane-wave decomposition is required, any geometry can be handled, and the information from each measurement is separately extracted, eventually including the source and receiver radiation patterns.

MIG is one of the most used imaging techniques in seismic exploration. Since MIG and DT are kinematically equivalent, it is worth to compare the methods. The MIG basic principle can be resumed as follows: the field scattered by a point in the model-space can be found in a data-space subset depending on the background velocity; conversely, a data-space point can be related to a model-space subset defined by the background velocity. A simple summation of the contributions collected in the data-space, (or a data smearing in the model-space), if necessary weighted to account for sources and receivers positioning and density, geometrical spreading, attenuation and source-receiver directivity, provides a reconstruction of the object. A section will be devoted to the comparison of the two methods. The attention will be mostly dedicated to the backprojection wavepaths and to the acquisition geometry impact on the final image.

Finally, numerical and experimental tests will provide an interesting applica-

tion of the methods. The examples are carried out with the Ground Penetrating Radar (GPR) technique. GPR is a rather new geophysical technique based on electromagnetic sounding. In this context, we consider its application to Non-Destructive tomographic diagnosis of buildings and ancient monuments.

2 Born approximation and Fourier Diffraction Theorem

Consider the wave equation in a non-homogenous medium in the frequency domain:

$$\left(\nabla^2 + k^2(\mathbf{r})\right) u(\mathbf{r}) = 0 \tag{1}$$

where $u(\mathbf{r})$ is the field complex amplitude and $k(\mathbf{r})$ is the wavenumber. For imaging purposes the wavenumber can be written as product of a constant velocity wavenumber and the perturbations due to the inhomogeneous medium (Kak and Slaney, 1988):

$$k(\mathbf{r}) = k_o n(\mathbf{r}) = k_o (1 + n_\delta) \tag{2}$$

where $k_o = \frac{\omega}{v_o}$, $n(\mathbf{r}) = \frac{v_o}{v(\mathbf{r})}$ and $n_\delta = \frac{v_o - v(\mathbf{r})}{v(\mathbf{r})}$.
The equation (1) now presents a forcing term:

$$\left(\nabla^2 + k_o^2(\mathbf{r})\right) u(\mathbf{r}) = k_o^2 O(\mathbf{r}) u(\mathbf{r}) \tag{3}$$

where the object function $O(\mathbf{r})$ is defined as:

$$O(\mathbf{r}) = [1 - n^2(\mathbf{r})] = \frac{v^2(\mathbf{r}) - v_o^2}{v^2(\mathbf{r})} \tag{4}$$

Consider the field decomposed in incident field $u_o(\mathbf{r})$, solution of the wave equation in the homogeneous medium:

$$\nabla^2 u_o(\mathbf{r}) + k_o^2 u_o(\mathbf{r}) = 0 \tag{5}$$

and scattered field $u_{sca}(\mathbf{r})$ produced by the anomalies. The substitution of $u(\mathbf{r})$ with the sum $u_{sca}(\mathbf{r}) + u_o(\mathbf{r})$, together with the equation (5) leads to the wave equation involving the scattered field:

$$\left(\nabla^2 + k_o^2(\mathbf{r})\right) u_{sca}(\mathbf{r}) = k_o^2 O(\mathbf{r}) u(\mathbf{r}) . \tag{6}$$

The solution of the equation (6) can be expressed in terms of the Green function. The Green function is a solution of the differential equation:

$$\left(\nabla^2 + k_o^2\right) G\left(\mathbf{r}|\mathbf{r}'\right) = -\delta\left(\mathbf{r} - \mathbf{r}'\right) \tag{7}$$

The 3D Green function is

$$G\left(\mathbf{r}|\mathbf{r}'\right) = \frac{\exp\left(jk_o\left|\mathbf{r} - \mathbf{r}'\right|\right)}{4\pi\left|\mathbf{r} - \mathbf{r}'\right|} \tag{8}$$

while the 2D Green function is

$$G\left(\mathbf{r}|\mathbf{r}'\right) = \frac{j}{4} H_0^{(1)}\left(k_o\left|\mathbf{r} - \mathbf{r}'\right|\right) \tag{9}$$

where $H_0^{(1)}$ is the Hankel function of first kind.

On the basis of the equation (6), the Green function can be considered the field resulting from a single point scatterer. By writing the forcing function of equation (6) as an array of impulses (i.e., a summation of point scatterers)

$$O(\mathbf{r})u\left(\mathbf{r}\right) = \int O(\mathbf{r}')u\left(\mathbf{r}'\right)\delta\left(\mathbf{r} - \mathbf{r}'\right)d\mathbf{r}' \tag{10}$$

the total scattered field can be written as a summation of shifted Green's functions weighted by the above forcing function:

$$u_{sca}\left(\mathbf{r}\right) = -k_o^2 \int G\left(\mathbf{r}|\mathbf{r}'\right) O(\mathbf{r}')u\left(\mathbf{r}'\right)d\mathbf{r}'. \tag{11a}$$

The (11a) is not yet useful for the inversion, since $u\left(\mathbf{r}\right)$ in the right side contains $u\left(\mathbf{r}\right)_{sca}$. No direct methods are available to solve the equation (11a). Only approximate solutions, valid in presence of weak scattering, provide a linear relationship between the scattered field and the velocity perturbations. The Born approximation, valid for small-size-high-contrast anomalies (high spatial wavenumber), is better suited to back-project the reflected energy. Rytov approximation is related to the transmitted energy and is valid for slow variations of the velocity field (low spatial wavenumber). DT under this approximation can be considered as a generalization of traveltime ray tomography. In this context, only the Born approximation will be considered, being the simplest.

Considering the total field as sum of incident and scattered field, the equation (11a) can be split into two terms:

$$u_{sca}\left(\mathbf{r}\right) = -k_o^2 \int G\left(\mathbf{r}|\mathbf{r}'\right) O(\mathbf{r}')u_o\left(\mathbf{r}'\right)d\mathbf{r}' - k_o^2 \int G\left(\mathbf{r}|\mathbf{r}'\right) O(\mathbf{r}')u_{sca}\left(\mathbf{r}'\right)d\mathbf{r}'. \tag{12a}$$

If the perturbations in the medium are weak, the second contribution in the equation (12a) can be neglected (Born approximation). The scattered field now depends only on the incident field (i.e., multiple scattering is neglected):

$$u_{sca}(\mathbf{r}) \approx -k_o^2 \int G(\mathbf{r}|\mathbf{r}')\, O(\mathbf{r}') u_o(\mathbf{r}')\, d\mathbf{r}' \tag{13a}$$

The Born equation for a single experiment with a point source in $\mathbf{r}_t$ and receiver in $\mathbf{r}_r$ is expressed by the following equation:

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t) = -k_o^2 \int G(\mathbf{r}_r|\mathbf{r})\, G(\mathbf{r}|\mathbf{r}_t)\, O(\mathbf{r})\, d\mathbf{r} \tag{14a}$$

The Born approximation holds provided that the phase change between the incident field and the field through the object is less than π (Kak and Slaney, 1988). For a circular anomaly of radius a this condition can be expressed in terms of wavelength and refraction index variation:

$$an_\delta < \frac{\lambda}{4}. \tag{15}$$

Under the weak scattering hypothesis, the object function can be also approximated by:

$$O(\mathbf{r}) \approx 2\frac{\Delta v(\mathbf{r})}{v(\mathbf{r})} \tag{16}$$

and the equation (14a) becomes

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t) = -2k_o^2 \int \frac{\Delta v(\mathbf{r})}{v(\mathbf{r})} G(\mathbf{r}_r|\mathbf{r})\, G(\mathbf{r}|\mathbf{r}_t)\, d\mathbf{r}. \tag{17a}$$

The equation (17a) can be Fourier transformed (Wu and Toksöz 1987) assuming that sources and receivers lie on straight lines. Following Wu and Toksöz we suppose perpendicular source and receiver lines at distance d_t and d_r from the origin, respectively (see Figure 1). Fourier transforming along the source and the receiver lines and rearranging the terms we get the following equation:

$$4\gamma_t\gamma_r U_{sca}(\mathbf{K}_t, \mathbf{K}_r) \exp\left[-j\left(\gamma_t d_t + \gamma_r d_r\right)\right] = k_o^2 \widetilde{O}(\mathbf{K}_r - \mathbf{K}_t) \tag{18}$$

where $\widetilde{O}(\mathbf{K}_o)$ is the object function Fourier transform, $\mathbf{K}_t$ and $\mathbf{K}_r$ are the wavenumbers along the source and receiver lines and γ_t and γ_r are the respective perpendicular wavenumbers. The equation (18) is a formulation of the *Fourier*

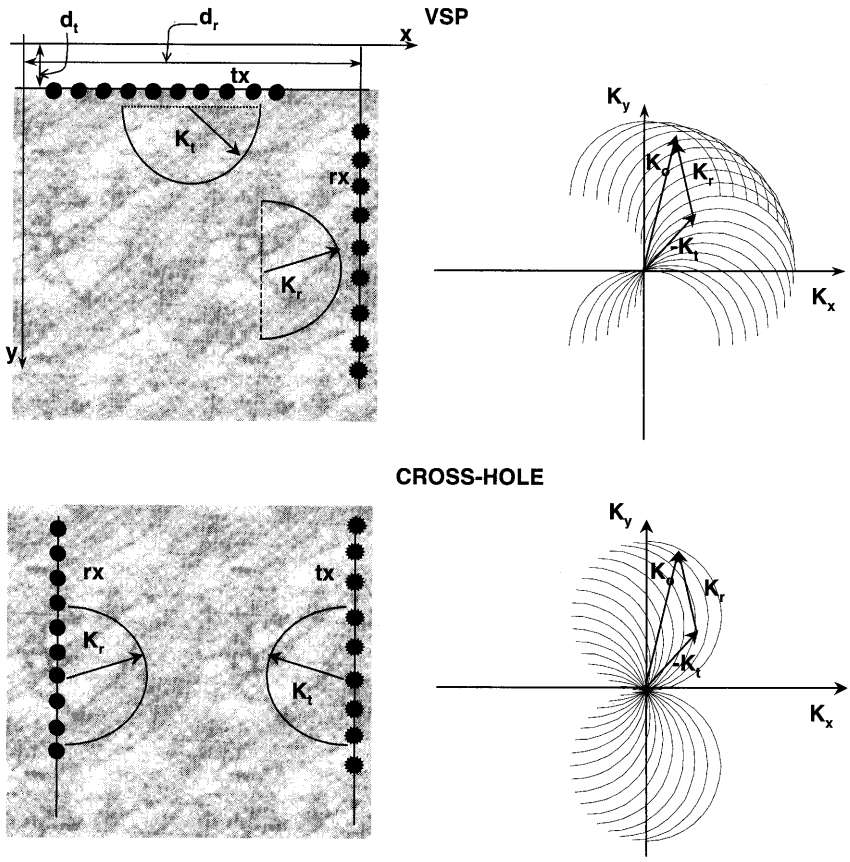


Figure 1: VSP and Cross-hole esperiments and their spectral coverages.

Diffraction Theorem: *the plane-wave scattering response* $U_{sca}(\mathbf{K}_t, \mathbf{K}_r)$ *is proportional to the object function Fourier spectrum evaluated in* $\mathbf{K}_o = \mathbf{K}_r - \mathbf{K}_t$.

The available plane-waves depend on the geometry of the experiment. In any case, for a given source direction, the plane-wave scattering response explores the object function spectrum along a circular arc; its width depends on the directivity of the receivers geometry (i.e., the available receiving plane-waves). The right side of Figure 1 shows the VSP and cross-hole experiment wavenumber coverage. Since real object functions are considered, the Fourier transform property $\widetilde{O}(-\mathbf{K}_o) = \widetilde{O}^*(\mathbf{K}_o)$, where $*$ is the complex conjugate operator, can be used to complete the spectrum.

The use of multi-frequency data improves the wavenumber coverage, since the circular arc radius increases with frequency.

3 Diffraction tomography through phase back-projection

The Fourier Diffraction Theorem presented above has a deterministic approach. The plane-waves are obtained by Fourier transforming along the source-receiver coordinates and each available plane-wave pair yields an object wavenumber. A different approach, inspired from the migration principle, is here considered; it is based on the scattering response of a single source-receiver pair rather than on the plane-wave decomposition involving the entire dataset. Consider the source and the receiver mono-frequency and isotropic: all the plane-waves can be radiated and received; in principle, all the wavenumbers can be seen (actually, the detectable wavenumber are contained in the circular zone of radius $|\mathbf{K}_o| = \frac{2\omega}{v_o}$ and the physical path condition must be satisfied). This paragraph shows how to estimate the phase of each detectable wavenumber from the scattered field phase through a sort of back-projection; the sum of the wavenumber estimates from each source-receiver pair gives the required object spectrum (Woodward and Rocca, 1988).

3.1 Theory

Consider a source-receiver pair. Following equation (14a) the expression of the scattered field involves the Green functions from source to generic point and from it to the receiver. The Weyl integral allows the decomposition of the Green function in plane-waves summation (see Appendix A); this decomposition can be applied both to source and receiver. A single point source plane-wave is represented by the wavenumber $\mathbf{K}_t(k_{xt}, k_{yt})$ (2D hypothesis is used for simplicity; the extension to the 3D case is straightforward). Note that the wavenumber components must satisfy the condition $k_{xt}^2 + k_{yt}^2 = \left(\frac{\omega}{v_o}\right)^2$. The field at the generic position $\mathbf{r}(x, y)$ is:

$$u(\mathbf{r}, \mathbf{r}_t) = e^{j\left[\mathbf{K}_t \bullet (\mathbf{r}-\mathbf{r}_t)+\frac{\pi}{2}\right]} \tag{19}$$

where all the amplitude factors are neglected; temporally we omit the outgoing wave conditions that will be accomplished later(see Appendix A and Appendix B).

Let the object function be a single spatial wavenumber $\mathbf{K}_o(k_{xo}, k_{yo})$:

$$O(\mathbf{r}) = A_o \exp(j\phi_o) \exp(j\mathbf{K}_o \bullet \mathbf{r}) \tag{20}$$

with amplitude A_o and phase ϕ_o. Under the assumption of weak scattering, the scattered field at the receiver is:

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t) = A_o e^{j[\phi_o + \mathbf{K}_t \bullet (\mathbf{r}_r - \mathbf{r}_t) + \mathbf{K}_o \bullet \mathbf{r}_r + \frac{\pi}{2}]} \tag{21}$$

Figure 2 shows the described forward model: the velocity perturbation is a single wavenumber (the real object function sketched in Figure 2 is obtained

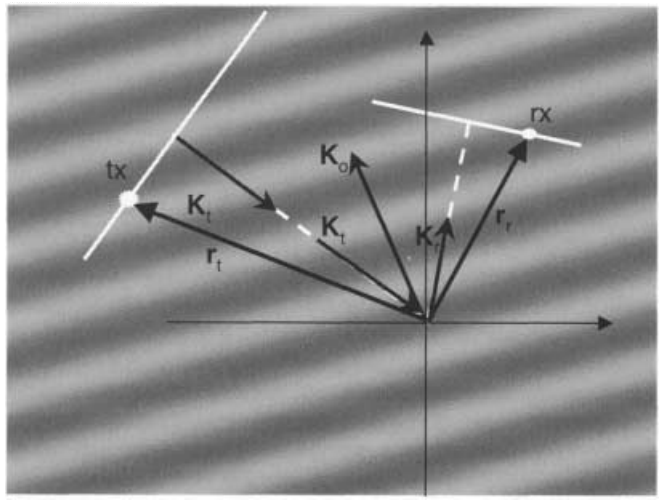


Figure 2: Forward model: a single wavenumber of the object function is considered.

by considering also the complex conjugate wavenumber). By imposing $\mathbf{K}_r = \mathbf{K}_t + \mathbf{K}_o$, the equation(21) becomes:

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t) = A_o e^{j(\phi_o + \mathbf{K}_r \bullet \mathbf{r}_r - \mathbf{K}_t \bullet \mathbf{r}_t + \frac{\pi}{2})} \tag{22}$$

The spectrum phase estimate is obtained by rearranging the above equation:

$$\phi_o = \phi_r - \left[\mathbf{K}_r \bullet \mathbf{r}_r - \mathbf{K}_t \bullet \mathbf{r}_t - \frac{\pi}{2}\right] \tag{23}$$

where ϕ_r is the scattered field phase.

The equation (23) is a Fourier Diffraction Theorem formulation: *the phase of the object function wavenumber* $\mathbf{K}_o = \mathbf{K}_r - \mathbf{K}_t$ *is obtained by subtracting to the received phase the phase corresponding to the paths from source to origin along* $\mathbf{K}_t$ *direction and from origin to receiver along* $\mathbf{K}_r$ *direction.*

An intuitive insight of the statement is given also by the Bragg's law.

An isotropic source-receiver pair will see all the wavenumbers in the circle of radius $|\mathbf{K}_o| = \frac{2\omega}{v_o}$, where v_o is the background velocity. Nevertheless, few wavenumbers within the circle are undetectable; they correspond to non causal paths from the source to the receiver. In case of the source-receiver pair of Figure 3, the undetectable wavenumbers belong to the two empty circles and they have be to excluded in the inversion process (Figure 3b). The empty circles orientation depends on the reciprocal source/receiver position.

The detectable wavenumbers, (the arcs plotted in Figure 3c), are obtained with a regular sampling of the plane-wave directions (α_r, α_t) defined as follows:

$$\begin{aligned} \mathbf{K}_r &= \frac{\omega}{v_o}(\cos\alpha_r, \sin\alpha_r) \\ \mathbf{K}_t &= \frac{\omega}{v_o}(\cos\alpha_t, \sin\alpha_t) \end{aligned} \tag{24}$$

The corresponding object function wavenumber is:

$$\mathbf{K}_o = \frac{\omega}{v_o}(\cos\alpha_r - \cos\alpha_t, \sin\alpha_r - \sin\alpha_t) \tag{25}$$

It is evident that the single experiment explores the wavenumber domain with a non-uniform density implied in the equation $\mathbf{K}_o = \mathbf{K}_r - \mathbf{K}_t$: denser zones are around $|\mathbf{K}_o| = \frac{2\omega}{v_o}$ and $|\mathbf{K}_o| = 0$. The inverse of the Jacobian of the coordinate transformation $(\alpha_r, \alpha_t) \leftrightarrow (k_{xo}, k_{yo})$, i.e. $\left|\frac{\partial(k_{xo}, k_{yo})}{\partial(\alpha_r, \alpha_t)}\right|$, should be required to modulate the amplitude of the spectrum estimate to account for the denser sampling zone. The inverse of the Jacobian is given by:

$$J(\mathbf{K}_o) = \left(\frac{\omega}{v_o}\right)^2 \sin|\alpha_r - \alpha_t| = \frac{\omega}{v_o}|\mathbf{K}_o|\sqrt{1 - \frac{v_o^2|\mathbf{K}_o|^2}{4\omega^2}} \tag{26}$$

Note that the Jacobian is a pure amplitude filter depending on the modulus of the spatial wavenumber $\mathbf{K}_o$. Actually, our implementation does not require the Jacobian, since the $\widetilde{O}(\mathbf{K}_o)$ estimate is already performed on a regular (k_{xo}, k_{yo}) domain; the proper angles (α_r, α_t) are previously computed by inverting relation (25).

Finally, by considering the amplitude of the scattered field, the estimate of the object function spectrum from a single source-receiver pair is:

$$\widetilde{O}(\mathbf{K}_o) = u_{sca}(\mathbf{r}_r, \mathbf{r}_t)\, e^{-j(\mathbf{K}_r \bullet \mathbf{r}_r - \mathbf{K}_t \bullet \mathbf{r}_t - \frac{\pi}{2})} F(\mathbf{K}_o, \mathbf{r}_r, \mathbf{r}_t). \tag{27}$$

where the amplitude filter $F(\mathbf{K}_o, \mathbf{r}_r, \mathbf{r}_t)$ provides the rejection of the undetectable wavenumbers.

All the wavenumber estimates can be summed: the actual wavenumbers in the object function will be revealed by a constructive sums. Thus $\widetilde{O}(\mathbf{K}_o)$ is obtained according to:

$$\widetilde{O}(\mathbf{K}_o) = \sum_{m=1}^{M} u_{sca}(\mathbf{r}_{r_m}, \mathbf{r}_{t_m})\, e^{-j(\mathbf{K}_r \bullet \mathbf{r}_{r_m} - \mathbf{K}_t \bullet \mathbf{r}_{t_m} - \frac{\pi}{2})} \cdot F_m(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m}) \frac{N}{N^2 + N_{\min}^2}. \tag{28}$$

where $m = 1, ...M$ is the number of data. The last term in (28) takes into account the geometry of the acquisition that causes a different estimate density for each wavenumber; the weight is inversely proportional to the number of actual observations

$$N = \sum_{m=1}^{M} F_m \left(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m} \right), \tag{29}$$

provided that this number is sufficiently high ($N > N_{\min}$); when this condition is not satisfied the estimation is considered unreliable and the estimate of this wavenumber is rejected according to the Wiener filtering technique. Wu and Toksöz (1987) remove the information redundancy caused by a wavenumber estimate superposition by filtering in the (α_r, α_t) domain.

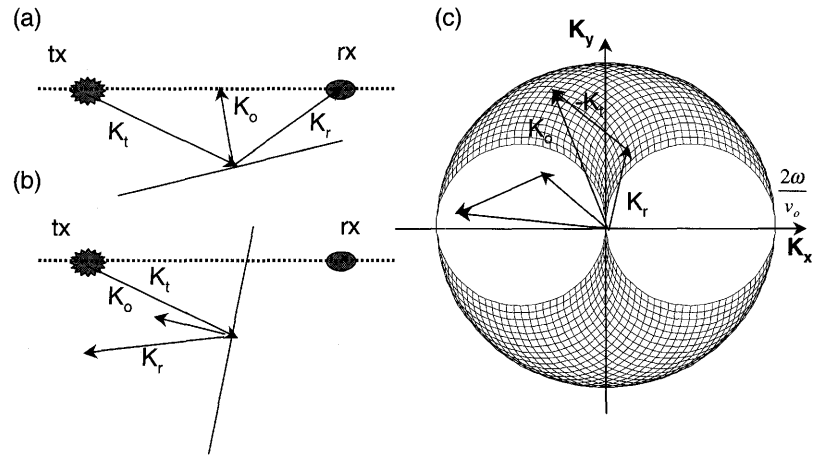


Figure 3: Single source-receiver experiment (the source and the receiver lie on the same ordinate); (a) detectable wavenumber scattering; (b) undetectable wavenumber scattering; (c) wavenumber coverage; the empty zones within the circle correspond to the undetactable wavenumbers.

In case of multi-frequency inversion the respective amplitude factor $0.5 \left(\frac{\omega}{v_o}\right)^{-2}$ has to be included and the estimate on the whole frequency range can be obtained according to:

$$\widetilde{O}\left(\mathbf{K}_o\right) = \frac{1}{2} \frac{N}{N^2 + N_{\min}^2} \sum_{p=1}^{P} \left(\frac{\omega_p}{v_o}\right)^{-2} \sum_{m=1}^{M} F_{mp} u_{sca_{mp}} e^{-j\phi_{mp}}. \tag{30}$$

where $p = 1, ...P$ is the frequency index and

$$N = \sum_{p=1}^{P} \sum_{m=1}^{M} F_{mp}\left(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m}\right) \tag{31}$$

represents the amount of actual observations.

It is important to stress that DT inversion automatically takes into account the distribution of sources and receivers, even for irregular geometries thanks to the Wiener filter. This algorithm is especially suited for irregular geometries since no reference plane-waves are previously computed; as a consequence, the information contained in each signal is optimally extracted even in noisy conditions. This formulation does not require the interpolation in the wavenumber domain since the angles (α_r, α_t) are computed by using a regular grid sampling of the wavenumber domain itself. The appendix B presents few details on the algorithm implementation. In Appendix C we explain how to include the directivity function of sources and receivers.

4 Diffraction tomography and pre-stack migration

Pre-stack migration is strictly related to diffraction tomography. In this chapter, we compare the two methods. Firstly, we examine the kinematical resemblance of the methods, concentrating on the respective wavepaths. The results of Woodward (1992) will help us in the discussion. Then, we show that MIG in the simplest version (multi-frequency holography) suffers from irregular geometry acquisition being an imaging method rather than an inversion process. Only more sophisticated approaches, like the one presented by Miller et al. (1987) and here considered, can make MIG fully comparable to DT. In the final paragraph we try to connect the novel DT algorithm to the MIG algorithm proposed by Miller et al. (1987).

4.1 Diffraction tomography wavepath

We move from the Born equation. The object function $O(\mathbf{r})$ can be obtained by inverting a discrete version of equation (14a). Let be i the measurement index, with $i = 1....N$, where N is the number of source-receiver pairs and k the object pixel index ($k = 1....D$). Then we get the linear system:

$$u_{sca}\left(\mathbf{r}_{r_i}, \mathbf{r}_{t_i}\right) = 2k_o^2 \sum_{k=1}^{D} G\left(\mathbf{r}_{r_i} | \mathbf{r}_k\right) G\left(\mathbf{r}_k | \mathbf{r}_{t_i}\right) \frac{\Delta v\left(\mathbf{r}_k\right)}{v\left(\mathbf{r}_k\right)} \tag{32a}$$

or, equivalently:

$$\mathbf{U}_{sca} = \mathbf{L}\Delta\mathbf{v}_{/\mathbf{v}} \tag{33}$$

The object function results from the LSQR inversion:

$$\Delta \mathbf{v}_{/\mathbf{v}} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{U}_{sca} \tag{34}$$

Usually, this approach does not have a practical interest requiring a pseudo-inverse computation that is very expensive even for small-medium size dataset. However, we use this spatial version of DT because can be easily compared to MIG. The matrix $(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T$ represents the filtered-back-projection of the scattered field along wavepaths L_i given by the product of the Green functions:

$$L_i = 2k_o^2 G\left(\mathbf{r}_{r_i}|\mathbf{r}\right) G\left(\mathbf{r}|\mathbf{r}_{t_i}\right) \tag{35}$$

Note that the novel DT algorithm uses the same wavepath as can be seen by Fourier transforming a single phase projection, i.e. the forward version. By the way, the Jacobian in equation (26) can be obtained as convolution of the Green function spectra involved in (35) (Woodward and Rocca 1988a).

In case of band-limited sources, many frequencies are available and the system (32a) can be extended to a multi-frequency domain, leading to a system of $N * P$ equations where P is the number of frequencies.

A single band-limited source-receiver pair is represented by the following system of P equations:

$$u_{sca}\left(\mathbf{r}_r, \mathbf{r}_t, \omega_p\right) = 2\left(\frac{\omega_p}{v_o}\right)^2 \sum_{k=1}^{D} G\left(\mathbf{r}_r|\mathbf{r}_k, \omega_p\right) G\left(\mathbf{r}_k|\mathbf{r}_t, \omega_p\right) \frac{\Delta v\left(\mathbf{r}_k\right)}{v\left(\mathbf{r}_k\right)} \tag{36a}$$

or

$$\mathbf{U}_{sca}\left(\mathbf{r}_r, \mathbf{r}_t\right) = \mathbf{BO} \tag{37}$$

where the coefficients of matrix $\mathbf{B}$ are:

$$b_{pk} = 2\left(\frac{\omega_p}{v_o}\right)^2 G\left(\mathbf{r}_r|\mathbf{r}_k, \omega_p\right) G\left(\mathbf{r}_k|\mathbf{r}_t, \omega_p\right) \tag{38}$$

The resulting band-limited wavepath can be estimated by Fourier transforming equation (36a):

$$L(t, \mathbf{r}_k) = \int 2\left(\frac{\omega}{v_o}\right)^2 G\left(\mathbf{r}_r|\mathbf{r}_k, \omega\right) G\left(\mathbf{r}_k|\mathbf{r}_t, \omega\right) e^{j\omega t} d\omega \tag{39}$$

The wavepath is an ellipse defined by:

$$\mathbf{r} : t = \frac{1}{v_o}\left(|\mathbf{r}_r - \mathbf{r}| + |\mathbf{r} - \mathbf{r}_t|\right) \tag{40}$$

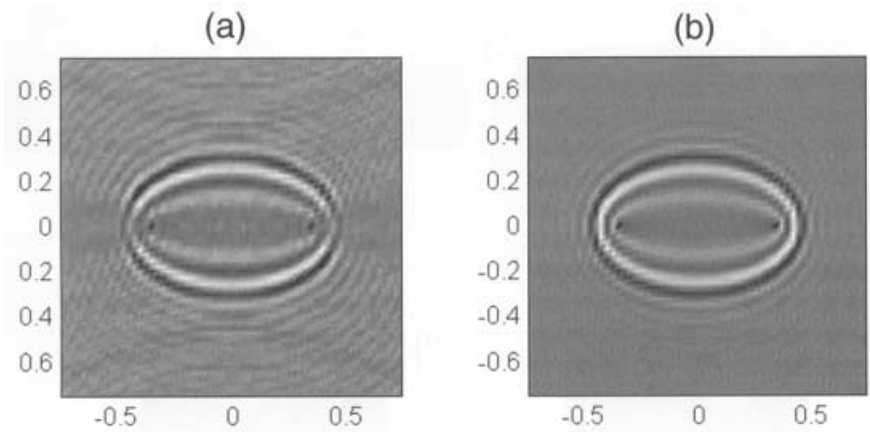


Figure 4: Multifrequency DT wavepath: (a) as computed by phase projection, (b) as computed by Green's function product summation. Source-receiver offset 0.7 m, time delay 6 ns, background velocity 15 cm/ns, source frequency 750MHz with 100% bandwidth.

and with foci in the source and receiver position, respectively.

In Figure 4 the multifrequency DT wavepath is shown as computed with the phase-backprojection method (a) and with the equation (28) (b). The respective spectra are shown in Figure 5.

4.2 Migration wavepath

While DT inverts the difference between the background medium parameters and the actual medium (i.e., $\frac{\Delta v(\mathbf{r})}{v(\mathbf{r})}$), migration images the variations of the actual medium parameters (i.e., the local reflection coefficient $R(\mathbf{r})$). However, the kinematical kernel is practically the same. The migration forward model contains the propagation operators from source to anomaly and from anomaly to receiver. The scattered field is a summation of delayed copies of the source wavelet proportional to the local reflection coefficient, if necessary weighted by an amplitude correction factor (geometrical spreading, angle dependence, source directivity).

The MIG basic principle can be resumed as follows: the field scattered by a point in the model-space can be found in a data-space subset depending on the background velocity called *Reflection Curve*; conversely, a data-space point can be related to a model-space subset defined by the background velocity called *Isochrone Curve*. A simple summation of the contributions collected in the data-space, (or a data smearing in the model-space), if necessary weighted to account for source and receiver positioning and density, geometrical spreading, attenu-

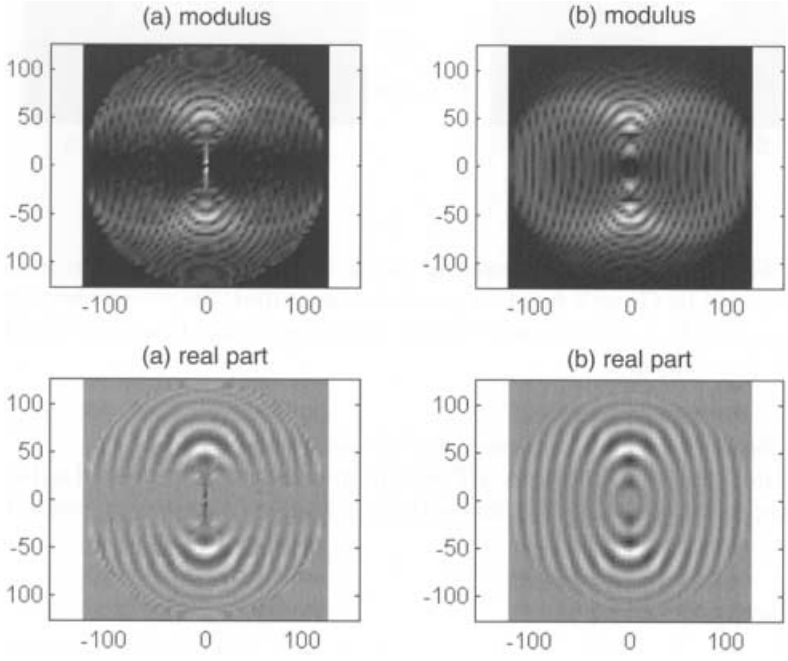


Figure 5: Multifrequency DT wavepath spectra (a) as computed by phase projection, (b) as computed by Green's function product summation. Differences are due to the different domain and process synthesis. Other details in the description of Figure 2.1.

ation and source-receiver directivity, provides a reconstruction of the object. Figure 6 shows the Curves in the data space and in the model space.

The simplest version of the forward migration model in a homogeneous background is:

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t, t) = \sum_{k=1}^{D} \left[s(t) * \delta(t - \frac{1}{v_o}(|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|)) R(\mathbf{r}_k) A(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \right] \quad (41)$$

or in the frequency domain:

$$u_{sca}(\mathbf{r}_r, \mathbf{r}_t, \omega) = \sum_{k=1}^{D} \left\{ S(\omega) e^{\left(-j\frac{\omega}{v_o}(|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|)\right)} R(\mathbf{r}_k) A(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \right\} \quad (42)$$

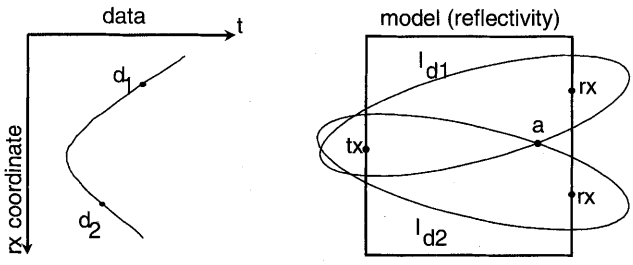


Figure 6: Migration principle scheme.

Here we perform the scattered field back-projection by smearing of the data along the proper ellipse. The imaged object function is a simple summation of the back-projections of each source-receiver pair. The scattered field back-projection for a single source-receiver pair is:

$$R(\mathbf{r}_k) = u_{sca}\left(\mathbf{r}_r, \mathbf{r}_t, \frac{1}{v_o}(|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|)\right) A^{-1}(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t). \quad (43a)$$

The insight in the relationship between this imaging approach and the DT inversion is clearer considering the equation (43a) in the frequency domain:

$$R(\mathbf{r}_k) = \sum_p u_{sca}(\mathbf{r}_r, \mathbf{r}_t, \omega_p) e^{j(\frac{\omega_p}{v_o}[|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|])} A^{-1}(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \quad (44a)$$

or:

$$\mathbf{R} = \mathbf{M}\mathbf{U}_{sca} \tag{45}$$

where the matrix $\mathbf{M}$ elements are:

$$m_{kp} = e^{j\left(\frac{\omega_p}{c_o}[|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|]\right)} A^{-1}(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \tag{46}$$

The LSQR inversion of equation (37) involved in multi-frequency diffraction tomography, for a single source-receiver pair is:

$$\mathbf{O} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{U}_{sca} \tag{47}$$

Note that for 3D case the elements of matrix $\mathbf{B}$ are:

$$b_{pk} = 2\left(\frac{\omega}{v_o}\right)^2 \frac{e^{-j\left(\frac{\omega_p}{c_o}[|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|]\right)}}{16\pi^2 |\mathbf{r}_r - \mathbf{r}_k| \, |\mathbf{r}_k - \mathbf{r}_t|}. \tag{48}$$

It is evident that the kinematical kernels are the same.

Now, we focus on the subtle differences between the "dynamics" implied in the two approaches. To be honest, we compare DT with a more sophisticated migration algorithm proposed by Miller et al. (1987). They moved from the Born approximation equation (17a) and show that the forward model implied in this approximation is very similar to a Radon transform; besides, a complete analogy is determined if the diffraction ellipse are locally approximated with its tangent; this is possible thanks to the stationary phase approximation. An inversion formula is then obtained by using the inverse Radon Transform. For simplicity, the mathematical theory is omitted. We report the 3D migration formula in a homogeneous background by using the present work notation:

$$O(\mathbf{r}_k) = \frac{1}{\pi^2} \int \frac{\left|\cos^3 \alpha \, (\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t)\right|}{v_o^3} A^{-1}(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \\ u_{sca}\left(\mathbf{r}_r, \mathbf{r}_t, \frac{1}{v_o}\left(|\mathbf{r}_r - \mathbf{r}_k| + |\mathbf{r}_k - \mathbf{r}_t|\right)\right) d\boldsymbol{\xi}\,(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t) \tag{49}$$

where α is the incident angle to the ellipse tangent (see Figure 7) and $\boldsymbol{\xi}\,(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t)$ is the unitary vector normal to the ellipse tangent. The integral (49) requires the Jacobian $d\boldsymbol{\xi}$ evaluation, being the data mapped in the coordinates $\mathbf{r}_r, \mathbf{r}_t$. The evaluation can be performed analytically in special cases, while irregular acquisition geometries require a numerical approach or an integral splitting. The kinematical kernel is unchanged while new weights take into account for the wave behavior and for the acquisition geometry.

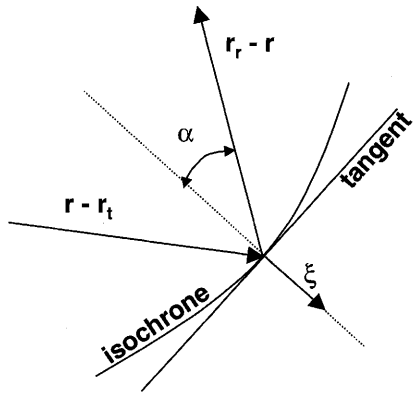


Figure 7: The approximation of the isochrone surface with its plane tangent. ξ is the normal vector and it is in the same direction of the total traveltime gradient $\bigtriangledown \left[\frac{1}{v_0}\left(|\mathbf{r}_r - \mathbf{r}| + |\mathbf{r} - \mathbf{r}_t|\right)\right]$.

4.3 Diffraction tomography and migration: wavepath and inversion process comparison

Now, it is worth to connect the novel algorithm developed in the ω/k (frequency/wavenumber) domain with the migration in the t/x (time/space) domain according to Miller et al. (1987).

The weight

$$\frac{\cos^3 \alpha \left(\mathbf{r}_k, \mathbf{r}_r, \mathbf{r}_t\right)}{v_o^3} \tag{50}$$

takes into account the incidence angle; normal reflections are enhanced with respect to the grazing reflections. The Jacobian (26) plays the same role in DT. We note here that the wavenumbers in the outer zone of the circle $|\mathbf{K}_o| \leq \frac{2\omega}{v_o}$ represent the Born region , i.e., the regions explored by the secondary arrivals or small offset data (more than 10% later than the first arrival); the inner part of the circle is explored by the primary arrivals or large offset data (transmitted energy); hence, we call that zone Rytov region. The two region limits are discussed in details in Woodward and Rocca (1988b).

The term $\left(\frac{\omega}{v_o}\right)^2$, missed in the first migration formula (equation (44a)), is included Miller et al. (1987) approach in the Radon transform analogy proof.

DT handles the acquisition geometry by using the Wiener filter approach; all the information in the data (provided that the Born approximation is satisfied) is

recovered and automatically equalized according to the number of observations. The same role should be played by the Jacobian $d\boldsymbol{\xi}$, that is a local-spatial version of the Wiener filter in DT. According to this Jacobian, the isochrone contributions in $\mathbf{r}$ are proportional to the inverse of their local density. A noise treatment is not included in this approach.

In conclusion, if this weighting factor is omitted, MIG remains an imaging method suffering strong artifacts from irregular geometries. Wu and Toksöz (1987) showed this effect inverting few synthetic dataset in VSP geometry with DT and with the multifrequency holography (or MIG).

Finally, we point out that, once the acquisition geometry effects are well-compensated and no artifact is present, the imaged object function still remains filtered by the specific acquisition space-variant filter. Only a very expensive space-variant deconvolution could help to recover as much details as possible.

5 Numerical and experimental results

In this section we present synthetic and real data applications of DT and MIG algorithms. Physical parameters and geometries of the synthetic data resemble typical GPR conditions. The synthetic data obtained under the Born approximation and with an EM simulator (Carcione 1996) offer reliable tests of the algorithms. Finally, their application to real data acquired with a GPR system on a laboratory model and on an ancient building confirm the algorithm capabilities.

5.1 Data pre-processing

Both methods require the knowledge of the scattered field. Synthetic data obtained with the EM simulator allows the extraction of the scattered field with two simulations with and without the velocity perturbations. In the case of real data the direct wave must be estimated and then subtracted from the data. The data are picked and then aligned on the first arrival. An average of the whole traces, eventually windowed, gives an estimate of the direct wave. The direct wave is then subtracted from the single trace with the proper amplitude and shift that minimize the energy of the difference in the first arrival window. More sophisticated estimates of the direct wave can be obtained following Pratt and Worthington, 1988; they propose an average of the common-offset-common-angle traces. It is worth to apply this step for both the algorithms, but results with MIG show that, at least in case of uniform coverage, the direct wave influence does not compromise the imaging result. Direct wave estimates are necessary to remove the signature phase characteristic from the scattered field. This subtraction is essential when multi-frequency inversion is performed. MIG also requires to select which part of the wavelet must be focused on the scattering interface. Few tests suggest to introduce a focusing time-shift to focus the maximum of the wavelet; the time shift partially substitutes a more expensive signature deconvolution. However, a zero-phase wavelet is always preferred.

Attenuation and geometrical spreading effects can be previously compensated with a simple time-gain function; MIG includes this compensation.

5.2 Numerical examples

The first example shows DT and MIG performances in a very simple limited view condition. We designed a common-offset acquisition along a straight lines and few point scatterers plotted in Figure 8a. We used 181 source-receiver pairs (offset = 10cm) along the dotted horizontal line with 1cm spacing. The entire acquisition line ranges in the interval $-0.9 \div 0.9$ m. The source signature is:

$$s(t) = \cos\left(2\pi f_c \left(t - \frac{3}{2f_c}\right)\right) \exp\left(-2f_c^2 (t - t_o)\right) \tag{51}$$

and the respective spectrum is:

$$S(f) = \sqrt{\frac{\pi}{8f_c^2}} \exp\left(-j\frac{3\pi f}{f_c}\right) \exp\left(-\frac{\pi^2 (f - f_c)^2}{2f_c^2}\right) \tag{52}$$

where the central frequency f_c is 1GHz. The background velocity is 15cm/ns. The data are computed under the Born approximation (equation (14a)). In Figure 8b the data are smeared along the Isochrones without any weighting factor (equation (43a)); in Figure 8c we take into account the weighting factors (equation(49)); the offset is small enough to compute the Jacobian assuming a zero-offset acquisition (see Miller et al. 1987 for the Jacobian computation). Figure 8d shows the multi-frequency DT result (equation(30)); in this case we inverted all the available frequencies to have exactly the same input data. The results confirm the conclusion of the previous Chapter. MIG without weighting factors is affected by strong artifacts; they are strongly reduced by the weights and the imaging effects is practically the same of DT. DT and weighted MIG still present few subtle differences requiring further investigations.

The second synthetic example scheme is presented in Figure 9. Dimensions and EM properties resemble a GPR experiment on a small square structure accessible from all the sides. Transmitter and receiver measure step is 5cm; measurements cover the entire section with all the view angles (2166 measurements). Data are computed with a 2D EM simulator (Carcione 1996) based on a pseudo-spectral method for solving spatial derivatives and a Runge-Kutta method for temporal derivatives. Equation (51) with $f_c = 750MHz$ is used as input wavelet for the magnetic line source. A small amount of attenuation is introduced in the background (velocity is 15cm/ns and its conductivity 0.01S/m). Two opposite sign one wavelength sized anomalies are introduced; the upper anomaly has velocity 16cm/ns, the lower 14cm/ns.

The multi-frequency DT result (520, 640, 760, 880, 1000MHz), the mono-frequency DT result and the MIG result are in Figure 9b,c,d respectively. The

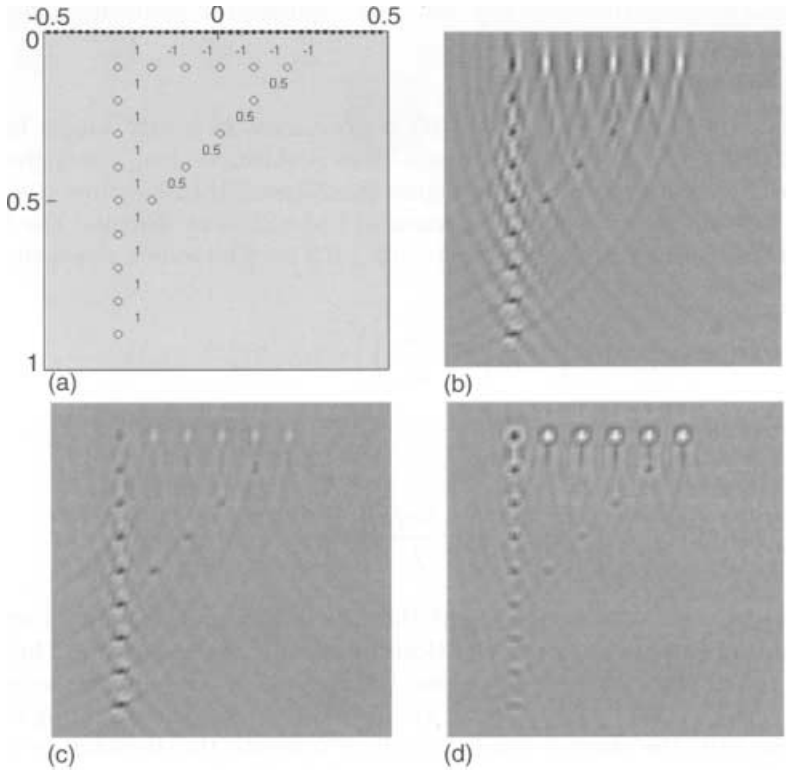


Figure 8: Common-offset reflection profile: (a) horizontal source-receiver line (dotted) and point scatterer design; (b) MIG result without weigths; (c) MIG result with weigths; (d) multi-frequency DT result.

object function spectrum is entirely explored thanks to the full coverage acquisition; for this reason the mono-frequency result is already satisfactory and the multi-frequency result basically improves the SNR. MIG returns a bit less sharp edges and a distorted background. Note that no weighting factor is computed for this numerical experiment; thus the MIG result should be further improved.

5.3 Laboratory model and real case examples

The first test of the algorithms is performed with real data acquired on a laboratory model shown in Figure 10a. 1080 data were acquired with a complete coverage acquisition. The MIG result (Figure 10b) a mono-frequency DT result (Figure 10d) are comparable. The steel bar is the most evident anomaly while the others are partially reconstructed. Traveltime tomography on straight

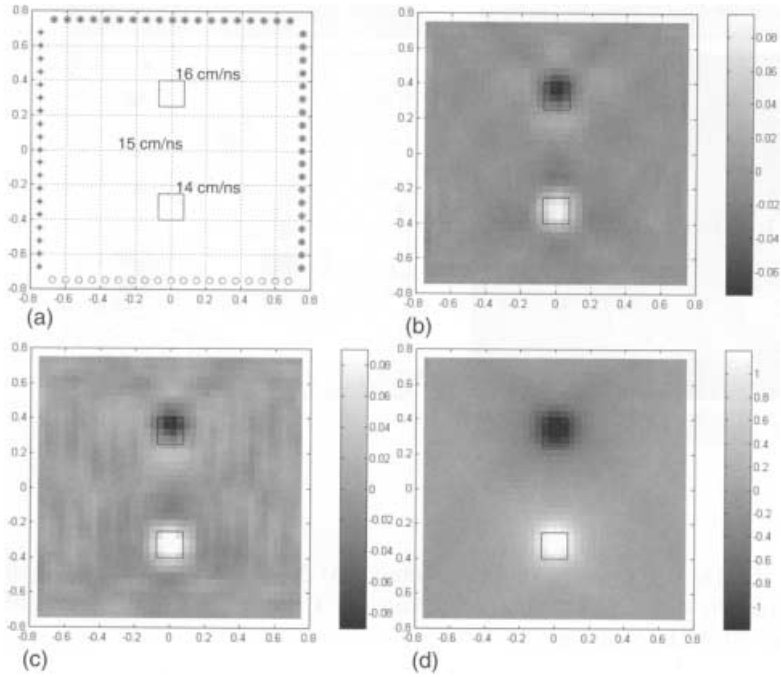


Figure 9: (a) Synthetic example design; source(*)/receiver(o) pair are chosen to have a complete angular coverage; (b) Multi-frequency DT result (520, 640, 760, 880, 1000MHz) (c) Single frequency DT result (760 MHz); (d) MIG result.

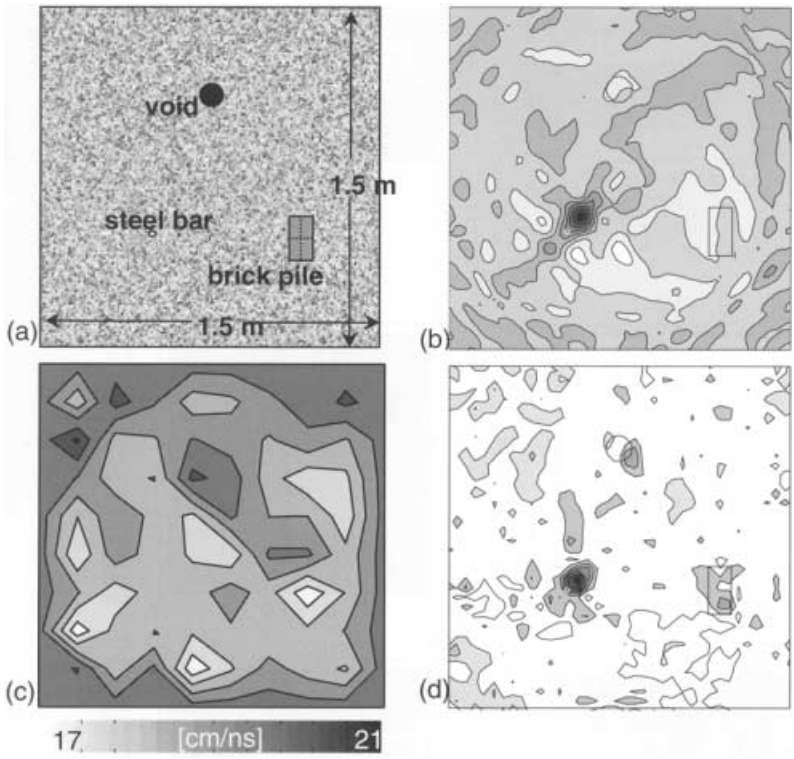


Figure 10: (a) Laboratory model scheme; (b) MIG result obtained with 1080 data acquired with a full-coverage scheme; (c) TT result; (d) DT result obtained with the 700MHz component.

rays provides only an almost useless image (Figure 10c); in fact only the void is partially detected with a very low contrast. The small size of this anomaly and the very low velocity contrast of the brick-pile are close to the traveltime tomography resolution limits (the steel bar can not be detected by traveltime tomography).

Finally, the algorithm was also test on a pier of the Malpaga Castle. The village of Malpaga is situated between Bergamo and Brescia in the north western part of Italy. The Castle (Figure 11) was built in the XIVth century. The interior was very simple since its function was essentially defense. The morphology of the stone-masonry wall sections is still practically unknown, being the stone walls of the castle built in different times and with different building technology.

870 data were acquired with a measure step of 10cm by using 900MHz antenna. The traveltime tomography result (Figure 12a) shows a quite evident fast

Figure 11: View of the Castle of Malpaga.

anomaly located at the center of one side of the pier. The echo profile along the south side of the pier confirms the presence of the anomaly; the events indicated by the arrows should correspond to the reflection from the front and the back sides of the anomaly (Figure 12b).This may be due to a local separation of the wall from the pier. The south side of the pier also shows a high velocity zone that has been interpreted as an artifact produced by the low data coverage in that area. In the MIG result of Figure 12c energy is focused where the anomaly is located, but a strong smearing reduces the quality of the reconstruction. This result suffers from the effects of the irregular acquisition geometry that is not compensated by a proper Jacobian. The DT result is the most satisfactory since the resolution is much higher than traveltime tomography and no artifact affects the image (Figure 12d). Further echo profiles acquired at different heights did not show this anomaly; as a result, it can be considered as practically confined.

Acknowledgments

The authors thank ISMES S.p.A. for providing radar equipment and laboratory models and especially Dr. G. Lenzi and L. Gerosa for conducting the surveys and for sharing their experience. The authors are grateful to Dr. J.M. Carcione for sharing the GPR modeling method that was used to test the algorithms and to Prof. L. Binda and Dr. A. Saisi for the useful discussions on the requirements of NDT survey on masonry structures.

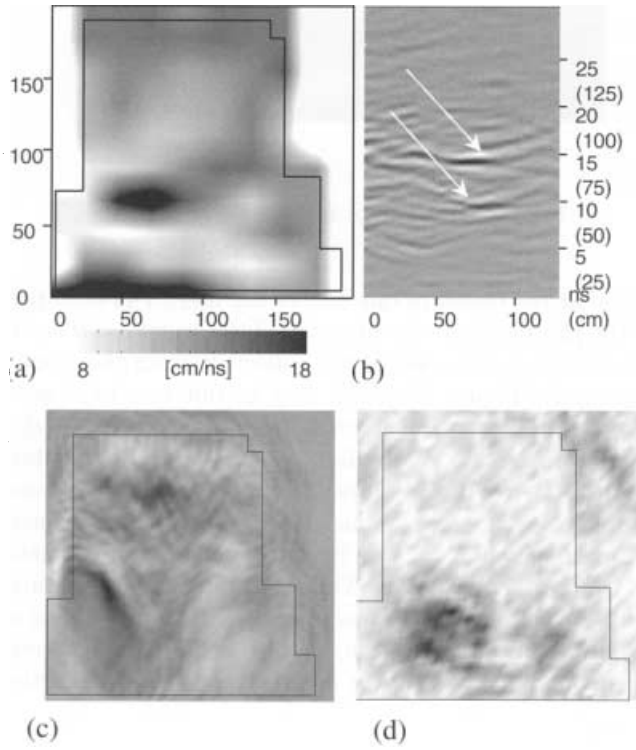


Figure 12: Malpaga Castle pier: (a) Traveltime tomography result; (b) the echo profile along the bottom side (with respect to the Figure) confirms the presence of the anomaly (indicated by the arrows); (c) MIG result; (d) mono-frequency DT result with the 700MHz component.

Appendix A: The Green Functions

The solution of equation (6) can be expressed in terms of the Green function. The Green function is a solution of the differential equation

$$\left(\nabla^2 + k_o^2\right) G\left(\mathbf{r}|\mathbf{r}'\right) = -\delta\left(\mathbf{r} - \mathbf{r}'\right) \tag{53}$$

The 3D Green function is

$$G\left(\mathbf{r}|\mathbf{r}'\right)_{3D} = \frac{\exp\left(jk_o\left|\mathbf{r} - \mathbf{r}'\right|\right)}{4\pi\left|\mathbf{r} - \mathbf{r}'\right|} \tag{54}$$

while the 2D Green function is

$$G\left(\mathbf{r}|\mathbf{r}'\right)_{2D} = \frac{j}{4} H_0^{(1)}\left(k_o\left|\mathbf{r} - \mathbf{r}'\right|\right) \tag{55}$$

where $H_0^{(1)}$ is the Hankel function of first kind; its asymptotic version is:

$$G\left(\mathbf{r}|\mathbf{r}'\right)_{2D} \approx \sqrt{\frac{1}{8\pi k_o\left|\mathbf{r} - \mathbf{r}'\right|}} \exp\left(jk_o\left|\mathbf{r} - \mathbf{r}'\right| + \frac{\pi}{4}\right) \tag{56}$$

The Green function can be decomposed in plane-waves by Fourier transforming equation (53). The Fourier transform of equation (53) is:

$$\left[-\left(k_x^2 + k_y^2 + k_z^2\right) + k_o^2\right] \widetilde{G}\left(\mathbf{K}|\mathbf{r}'\right) = -e^{-j\mathbf{K}\bullet\mathbf{r}'} \tag{57}$$

where $\mathbf{K}(k_x, k_y, k_z)$ is the plane-wave wavenumber. Rearranging the terms we obtain:

$$\widetilde{G}\left(\mathbf{K}|\mathbf{r}'\right) = -\frac{e^{-j\mathbf{K}\bullet\mathbf{r}'}}{k_o^2 - K^2} \tag{58}$$

where K is the wavenumber modulus. Note that the Green function spectrum has significant amplitude on the spherical shell $|K| = k_o$. Anyway the Green function Fourier Transform is not yet a plane-wave decomposition since the wavenumbers have arbitrary velocity. In order to obtain plane-waves in the medium with velocity v_o, we integrate along one of the wavenumbers (e.g., k_y) (Aki and Richards 1980; Chew 1994) with the condition $\left(k_x^2 + k_y^2 + k_z^2\right) = \left(\frac{\omega}{v_o}\right)^2$. In order to avoid the singularities we add a small amount of loss; thus the wavenumber K is complex and the singularities are off the real axis. The plane-wave decomposition resulting from this integration is:

$$G(\mathbf{r}|\mathbf{r}')_{3D} = \frac{j}{2(2\pi)^2} \int\int \frac{e^{j(k_x x + k_y y + k_z |z|)}}{k_z} dk_x dk_y \tag{59}$$

where $\mathbf{r}' = 0$. To ensure radiation conditions it must be $\Re e[k_z] > 0$ and $\Im m[k_z] > 0$. The modulus $|z|$ is needed to have outgoing waves. With similar consideration the line-source plane-wave decomposition is:

$$G(\mathbf{r}|\mathbf{r}')_{2D} = \frac{j}{4\pi} \int \frac{e^{j(k_x x + k_y |y|)}}{k_y} dk_x \tag{60}$$

A more intuitive version of the plane-wave decomposition (60) can be obtained with the variable transformation $k_x \rightarrow \alpha$ where α is the angle between the plane-wave direction and the x axis, $k_x = \frac{\omega}{v_o} \cos\alpha$ and $k_y = \frac{\omega}{v_o} \sin\alpha$:

$$G(\mathbf{r}|\mathbf{r}')_{2D} = -\frac{j}{4\pi} \int e^{j\frac{\omega}{v_o}(x\cos\alpha + y\sin\alpha)} d\alpha \tag{61}$$

Appendix B: implementation details

The phase back-projection is performed directly in the domain (k_{xo}, k_{yo}), avoiding the interpolation process. The angles (α_r, α_t) are computed by inverting the relations (24). The ambiguity due to the fact that two pairs $\mathbf{K}_r - \mathbf{K}_t$ give a single $\mathbf{K}_o$ (see Figure 13) is resolved with the causal path condition $(\mathbf{K}_r + \mathbf{K}_t) \bullet (\mathbf{r}_r - \mathbf{r}_t) \geq 0$. The Fourier transform property of real object function, $\widetilde{O}(-\mathbf{K}_o) = \widetilde{O}^*(\mathbf{K}_o)$, is applied during the phase back-projection; hence, for a given $\mathbf{K}_o$, i.e. a given $\mathbf{K}_r - \mathbf{K}_t$, the estimated phase is back-projected also in $-\mathbf{K}_o$ with the opposite sign.

Appendix C: DT with source/receiver directivity function

All the real sources (receivers) are not isotropic. Their directivity function can be roughly included in the inversion process by excluding the plane-waves (and thus the corresponding object wavenumbers) that are not present in the source (receiver) spectrum. Thus the filter F is reformulated as follows:

$$F'_m(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m}) = F_m(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m}) \{f_{r_m}(\alpha_{r_m}) f_{t_m}(\alpha_{m t}) > 0\} \tag{62}$$

where (α_r, α_t) are the view angles associated with the object wavenumber by means of equation (25), $f(\alpha)$ is the directivity function, and $\{f_r(\alpha_r) f_t(\alpha_t) > 0\}$

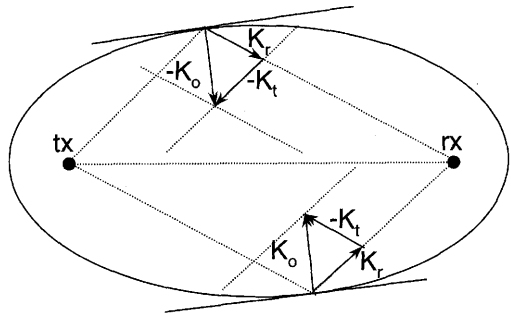


Figure 13: Scheme for the phase back-projection.

is a boolean expression returning a 0/1 value. A more advanced method consists of applying also a weighting factor that really takes into account the source and receiver directivity functions. This can be obtained multiplying each data by $f_{r_m}(\alpha_{r_m}) f_{t_m}(\alpha_{t_m})$ and updating the average factor in equation (28) with $\frac{N_1}{N_1^2+N_{\min}^2}$ where

$$N_1 = \sum_{i=1}^{M} F_m(\mathbf{K}_o, \mathbf{r}_{r_m}, \mathbf{r}_{t_m}) f_{r_m}(\alpha_{r_m}) f_{t_m}(\alpha_{t_m}). \tag{63}$$

Anyway, the optimal estimate is actually obtained with a LSQR approach that will be attempted in the next future.

References

Aki K Richards PG (1980): Quantitative seismology- Theory and methods.W.H. Freeman & Co.

Carcione JM (1996) Ground-penetrating radar: Wave theory and numerical simulation in lossy anisotropic media. Geophysics 61: 1664-1677.

Chew WC (1994): Waves and fields in inhomogeneous media. IEEE Press, New York.

Gelius LJ (1995) Generalized acoustic diffraction tomography. Geophysical Prospecting 43: 3-30.

Devaney AJ (1982) A filtered back propagation algorithm for diffraction tomography. Ultrasonic Imaging 4: 336-350.

Dickens TA (1994) Diffraction tomography for crosswell imaging of nearly layered media. Geophysics 59: 694-706.

Harris JM, Wang GY (1996) Diffraction tomography for inhomogeneities in layered background medium. Geophysics 61: 570-583.

Kak AC, Slaney M (1988): Principles of Computerized Tomographic Imaging. IEEE Press, New York.

Miller D, Oristaglio M, Beylkin G (1987) A new slant on seismic imaging: Migration and integral geometry. Geophysics 52: 943-964.

Pratt GR, Worthington MH (1988) The application of diffraction tomography to cross-hole seismic data Geophysics 53: 1284-1294.

Pratt GR, Worthington MH (1990) Inverse theory applied to multi-source cross-hole tomography. Part 1 and part 2. Geophysical Prospecting 38: 287-329.

Woodward MJ, Rocca F (1988a) Wave equation tomography-II. Stanford Exploration Project 57: 25-47.

Woodward MJ, Rocca F (1988b) Wave-equation tomography. 58th Ann. Internat. Mtg., Soc. Explor. Geophys., Expanded Abstract, pp. 1232-1235.

Woodward MJ (1992) Wave equation tomography. Geophysics 57: 15-26.

Wu RS, Toksöz MN (1987) Diffraction tomography and multisource holography applied to seismic imaging. Geophysics 52: 11-25.