Quality of regularization methods

Johannes Bouman

DEOS Report
no 98.2

Quality of regularization methods

Quality of regularization methods

Johannes Bouman

Delft University Press / 1998

Published and distributed by:

Delft University Press
Mekelweg 4
2628 CD Delft
The Netherlands
Telephone: + 31 15 278 3254
Telefax: + 31 15 278 1661
E-mail: DUP@DUP.TUDelft.NL

ISBN 90-407-1798-2 / CIP

Printed in The Netherlands

Contents

Foreword

This report was written by Johannes Bouman, Ph.D. student at the Faculty of Civil Engineering and Geosciences, Delft, The Netherlands under the supervision of Dr. Radboud Koop and Prof. dr. Roland Klees.

Two Appendices are included for reference. Appendix A deals with introductory functional analysis and in Appendix B spectral decomposition is treated as well as some conventions used in this report.

Summary

The solution of ill-posed problems is non-trivial in the sense that frequently applied methods like least-squares fail. The ill-posedness of the problem is reflected by very small changes in the input data which may result in very large changes in the output data. Hence, some sort of stabilization or regularization is required. Some examples of (geodetic) ill-posed problems are given.

Several regularization methods exist to compute stable solutions, along with several ways of determining the so-called regularization parameter(s). The idea of the regularization methods is discussed as well as the determination of optimal regularization parameters. Moreover, the different methods are compared, emphasizing the quality or accuracy of the methods.

Finally, the differences between methods and parameter choice rules are illuminated by an example from airborne gravimetry.

Acknowledgements

First of all I thank Radboud Koop and Roland Klees for their useful remarks and the fruitful discussions we had. Martin Jutte prepared Figures 3.2 and 4.1, and Axel Smits prepared Figure 3.3. Frank Kleijer, Clare Macfarlane and Peiliang Xu read earlier versions of this report and made suggestions which certainly improved it. Finally, the development of the Matlab Package 'Regularization Tools' by Per Christian Hansen is gratefully acknowledged. The examples of Chapter 5 are computed with this package and the corresponding report/software, Hansen (1997), are public domain and clearly written, which is the best combination one could hope for.

Notation

Roman upper-case

A	1) compact operator mapping an element $\mathbf{f}$ from Hilbert space F to an element $\mathbf{g}$ of Hilbert space G 2) matrix $\in \mathbb{R}^{m \times n}, m \geq n$
A_k	approximation of A in TSVD and TGSVD
A_w	transformed design matrix, $A_w = WA$
C	signal weight matrix
$C[a,b]$	class of continuous functions on $[a,b]$
D	diagonal matrix with elements d_i
E	matrix with ones on the anti diagonal
F	Hilbert space
G	1) Hilbert space 2) $I - A^*A$
H	(semi-)orthogonal matrix in QR factorization
H^p	Sobolev space
I	indentity operator
I_n	indentity matrix of dimension n
$J_\alpha(\mathbf{x})$	function to be minimized with respect to $\mathbf{x}$ for fixed α
K	(semi-)orthogonal matrix in QR factorization
K_k	Krylov subspace
$K(x,y)$	integration kernel
L	1) differential operator, regularization matrix 2) strict lower triangular matrix
L^2	Hilbert space of square integrable functions
M	1) diagonal matrix with elements μ_i in GSVD 2) square invertible matrix 3) average operator
N	1) size of perturbation $\in \mathbb{R}$ 2) normal matrix, $N = A^*A$
$\mathbb{N}$	natural numbers
P	weight matrix of errors $\underline{\varepsilon}$
P_k	matrix with vectors $\mathbf{p}_k$
Q	(semi-)orthogonal matrix in QR factorization
Q_x	error covariance matrix
R	1) upper triangular matrix in QR factorization 2) $R > 1$ in discrepancy principle

$\mathbb{R}$	real numbers
$\mathbb{R}_0^+$	positive real numbers, zero included
$\mathbb{R}^{m\times n}$	space to which matrix A belongs
S	1) symmetric positive definite matrix 2) Choleski decomposition of $C = SS^T$
T	1) compact, symmetric and semi-positive definite operator 2) upper triangular matrix in QR factorization
U	(semi-)orthogonal matrix with singular vectors $\mathbf{u}_j$
V	orthogonal matrix with singular vectors $\mathbf{v}_i$
W	Choleski decomposition of $P = W^TW$
X	nonsingular n-by-n matrix
$Z(\alpha)$	function to become zero for α

Roman lower-case

In general lower-case letters with a roman index (i, j, k, m or n) are real numbers, for example $d_i \in \mathbb{R}$. Printed in bold face, however, lower-case letters with such an index are vectors or functions, for example $\mathbf{u}_n$.

a	minimum of interval, $K(x,y) : [a,b] \rightarrow [a,b]$
a_n	Fourier coefficients
b	maximum of interval, $K(x,y) : [a,b] \rightarrow [a,b]$
$\mathbf{b}$	$\mathbf{b} = A^T\mathbf{y}$, $\mathbf{b} = A^*\mathbf{g}$
c	constant $\in \mathbb{R}_0^+$
$\mathbf{c}_i$	positive function
d_i	i-th diagonal entry of matrix D
$\mathbf{d}_k^\varepsilon$	stability error after k iterations
$\mathbf{e}_k$	approximation error after k iterations
$\mathbf{f}$	∞-vector with exact solution, $\mathbf{f} = \mathbf{f}(x)$, $a \le x \le b$
$\mathbf{f}_s$	any solution of $A\mathbf{f} = \mathbf{g}$
$\mathbf{f}^\varepsilon$	∞-vector with approximate solution from $\mathbf{g}^\varepsilon$, sometimes ε is not written
$\mathbf{f}_\alpha^\varepsilon$	∞-vector with regularized solution from $\mathbf{g}^\varepsilon$, sometimes ε is not written
$\mathbf{g}$	∞-vector with exact observations, $\mathbf{g} = \mathbf{g}(x)$, $a \le x \le b$
$\mathbf{g}^\varepsilon$	∞-vector of observations with error $\underline{\varepsilon}$, sometimes ε is not written
$\mathbf{h}$	1) $\mathbf{h} = \mathbf{h}(x)$, $a \le x \le b$ 2) element of Krylov space
i	index for finite dimension, $\mathbf{x} = (x_1, x_2, \ldots, x_i, \ldots, x_n)^T$
j	index for finite dimension, $\mathbf{y} = (y_1, y_2, \ldots, y_j, \ldots, y_m)^T$
k	1) perturbation frequency $\in \mathbb{N}$ 2) truncation level for truncated singular value decomposition, number of iterations
l	distance from $\mathbf{x}$ to solution $\mathbf{x}_s$
m	number of measurements $< \infty$, $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$
n	1) number of unknowns $< \infty$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ 2) index for infinite dimensions, $\mathbf{f} = (f_1, f_2, \ldots, f_n, \ldots)^T$
o	$n - p$

p	1) number of generalized singular values 2) index of Sobolev space H^p
$\mathbf{p}_k$	general direction
q	$m-(n-p)$
$\mathbf{r}_k$	residual after k iterations
$\mathbf{s}$	vector with side constraints
$\mathbf{u}_n$	singular vector, eigenvector of AA^* (column vector)
$\mathbf{v}_n$	singular vector, eigenvector of A^*A (column vector)
$\mathbf{w}_n$	eigenvector of T (column vector)
x	variable of function, $\mathbf{g}(x), \mathbf{f}(x)$
$\mathbf{x}$	n-vector with exact solution
$\mathbf{x}_g$	generalized biased estimate
$\mathbf{x}_i$	column vector of X
$\mathbf{x}_s$	any approximate solution of $\mathbf{x}$
$\mathbf{x}^\varepsilon$	n-vector with approximate solution from $\mathbf{y}^\varepsilon$, sometimes ε is not written
$\mathbf{x}_\alpha^\varepsilon$	n-vector with regularized solution from $\mathbf{y}^\varepsilon$, sometimes ε is not written
$\hat{\mathbf{x}}$	least-squares estimate
y	variable of function, $\mathbf{f}(y)$, y is the integration variable
$\mathbf{y}$	m-vector with exact observations
$\mathbf{y}_w$	transformed observation, $\mathbf{y}_w = W\mathbf{y}$
$\mathbf{y}^\varepsilon$	m-vector of observations with error $\underline{\varepsilon}$, sometimes ε is not written
$\mathbf{z}$	vector with elements $\in \mathbb{R}$

Greek upper-case

Δ	diagonal matrix with elements α_i
Δx	bias
Σ	diagonal matrix with singular values σ_i
Ω	general Tikhonov penalty term

Greek lower-case

α	regularization parameter, $\alpha \in \mathbb{R}_0^+$ or $\alpha^{-1} = k \in \mathbb{N}$
β	relaxation parameter
γ_i	generalized singular value
δ	$\max\lvert\mathbf{f}-\mathbf{f}^\varepsilon\rvert$, $\max\lvert\mathbf{x}-\mathbf{x}^\varepsilon\rvert$
δ_i	filter factor i
δ_{ij}	Kronecker delta
$\delta\mathbf{x}$	$\delta\mathbf{x} = \mathbf{x}_\alpha^\varepsilon - \mathbf{x}$
$\underline{\varepsilon}$	m-vector or ∞-vector with measurement errors or perturbation
ε	$\lVert\underline{\varepsilon}\rVert$
ζ_i	element of $\mathbf{z}$
$\eta(\alpha)$	$\log\lVert\mathbf{x}_\alpha^\varepsilon\rVert_2$
λ_i	eigenvalue
μ_i	i-th diagonal element of M in GSVD

$\xi(\alpha)$	$\log \|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$
σ^2	variance of unit weight
σ_i	singular value
τ	approximately one, $\tau > 1$
ψ	$\|A\mathbf{x} - \mathbf{y}\|\|\mathbf{x}\|$

Operations

A^T	transpose of $A \in \mathbb{R}^{m\times n}$
A^*	conjugate or adjoint of operator A
A^{-1}	inverse of A
A^+	generalized inverse of A
A_α^+	regularized generalized inverse of A
$D(A)$	domain of A
$E\{\mathbf{y}\}$	expectation of $\mathbf{y}$
$N(A)$	null space of A
$N(A)^\perp$	space orthogonal to the null space of A
$R(A)$	range of A
$\overline{R(A)}$	closure of the range of A
$\|a\|$	absolute value of $a \in \mathbb{R}$
$\langle \mathbf{a}, \mathbf{b}\rangle$	innner product of $\mathbf{a}$ and $\mathbf{b}$, $\langle \mathbf{a}, \mathbf{b}\rangle = \mathbf{b}^T\mathbf{a}$
$d(\mathbf{f}^\varepsilon, \mathbf{f})$	distance from $\mathbf{f}$ to $\mathbf{f}^\varepsilon$, d is the metric defined on F
$\mathbf{f}'$	first derivative of $\mathbf{f}$ with respect to its argument
$\|\mathbf{f}\|_F$	Hilbert space norm of $\mathbf{f}$
$\sum_i \mathbf{u}_i\mathbf{v}_i^T$	matrix
$\{\mathbf{v}_n, \mathbf{u}_n; \sigma_n\}$	singular system
$\sigma(T)$	spectrum of T

The end of examples, definitions and theorems is marked with a ●.

List of abbreviations

BE	Biased Estimation
CG	Conjugate Gradients
DGSVD	Damped Generalized Singular Value Decomposition
DSVD	Damped Singular Value Decomposition
GBE	Generalized Biased Estimation
GCV	Generalized Cross Validation
GPS	Global Positioning System
gsv	generalized singular values
GSVD	Generalized Singular Value Decomposition
l.s.	least-squares
MSE	Mean Square Error
MSEM	Mean Square Error Matrix
PCG	Preconditioned Conjugate Gradients
SNR	Signal-to-Noise Ratio
SST	Satellite-to-Satellite Tracking
SVD	Singular Value Decomposition
TGSVD	Truncated Generalized Singular Value Decomposition
TR	Tikhonov Regularization
TSVD	Truncated Singular Value Decomposition

1 | *INTRODUCTION*

1.1 Background and problem description

An accurate and high resolution knowledge of the earth's gravity field is needed in several earth oriented sciences. In geodesy, for example, the gravity field is needed for levelling with GPS, in oceanography it is important for studying ocean circulation and last but not least in geophysics a better knowledge of the earth's gravity field yields better boundary conditions in the study of the earth's interior.

A model of the earth's gravity field may be determined by means of satellite observations. Examples of satellite measurement techniques for global gravity field determination are satellite tracking from stations at the earth's surface (ranges, range-rates, directions), satellite gradiometry and satellite-to-satellite tracking (SST).

It is well known that only the long wavelengths (about 600 km at the equator, corresponding to spherical harmonic degree 70) of the gravity field are revealed by currently available satellite tracking data, Nerem *et al.* (1994); Schwintzer *et al.* (1997). Gravity field models from satellite tracking data are called satellite-only models. The combination of satellite tracking data with gravimetry and satellite altimetry allows for solving shorter wavelengths down to about 100 km at the equator, corresponding to spherical harmonic degree 360, cf. Rapp *et al.* (1991); Gruber *et al.* (1995).

The computation and the combination process of the satellite-only models is hampered by the lack of a proper quality description of the solutions. On the one hand model errors are responsible for this, e.g. insufficient modelling of drag for satellites and datum connection problems for gravity data, Nerem *et al.* (1994); Heck (1990). On the other hand there is concern that the quality of the satellite-only solutions is not described properly: although it is generally recognized that the solutions are biased, this bias is not accounted for, Marsh *et al.* (1989); Xu (1992b).

In the near future several dedicated gravity field missions might be launched, such as Grace using low-low SST, Tapley (1996), and Goce using a combination of high-low SST and gradiometry, ESA (1996). The purpose of these missions is to determine very ac-

curately high resolution gravity field models (Goce), and time-varying gravity signatures (Grace). However, as noted above, it is unclear how the accuracy has to be described. When the bias is taken into account the accuracy description might be different from the conventional (least-squares collocation) accuracy description. Furthermore, collocation probably is no longer the optimal estimation procedure as it is based on unbiased assumptions. Hence, it is of interest to look at other estimation methods that take the bias into account.

1.2 Purpose and limitations

The *purpose* of this study is to describe different regularization methods and their consequences for the quality of the solution. By quality we mean the deviation of the estimated from the true function.

In geodesy one type of regularization, Tikhonov-regularization, has been interpreted as a kind of collocation, Rummel *et al.* (1979); Moritz (1980); Marsh *et al.* (1989). Xu (1992a) introduced biased estimation in geodesy as an alternative. It is shown here that basically Tikhonov-regularization and biased estimation are equivalent with the same description of quality, whereas collocation has a different quality description. Other regularization methods are studied, most of them have been used in geodesy.

Although a reasonable amount of geodetic literature on regularization exists, an overview of the different methods together with their implications for the quality description has to our knowledge not been given yet. This overview is given here and the mean square error is considered to be a suitable measure of the quality. For a comparison of methods see, e.g. Schwarz (1979); Rummel *et al.* (1979); Sansò (1989); Rauhut (1992), concerning the quality description cf. Moritz (1980); Xu (1992b); Schwarz (1973); Gerstl and Rummel (1981); Neyman (1985); Xu (1992a); Xu and Rummel (1994a). For a comparison of methods in non-geodetic literature see Louis (1989); Groetsch (1993); Engl *et al.* (1996); Hansen (1997); Phillips (1962); Tikhonov (1963b, 1963a); Tikhonov and Arsenin (1977); Nashed (1976); Groetsch (1984); Wahba (1990).

The errors considered in this report are restricted to data errors, that is, model errors are not part of the discussion. One reason is that it would unnecessarily complicate matters. Moreover, physical models, relating the measurements to the unknowns, are usually well known, Wing (1991). Especially when the model error is small compared to the data error it causes no real additional problems, Morozov (1984). A geodetic example is the computation of the geoid height in a certain point from global gravity data using Stokes' formula. This relation is valid in spherical, constant radius approximation, which produces a model error of less than 1% with respect to a reference ellipsoid. It is assumed that model errors can be overcome by iteration if it converges, and these errors are assumed to have equal influence on the quality when comparing different regularization methods.

One important inverse problem that we have in mind is the determination of the global gravity field from satellite gradiometric measurements. Typically, these measurements are not sampled on a global basis, since the satellite moves in a non-polar orbit, Blaser *et al.* (1996); ESA (1996). At a first glance one might think that the determination of the global gravity field from these 'local' measurements has an inherent model error. However, the

computation of the second derivatives of the potential at height h above the earth's surface in a region $0 \leq \lambda \leq 2\pi$, $pg \leq \theta \leq \pi - pg$ is perfectly legitimate, there is no model error (λ stands for longitude, θ stands for co-latitude and pg is half the size of the polar gap, the region without observations). The inverse computation suffers from, among others, a lack of uniqueness.

Note that aliasing caused by the finite sampling interval has to be dealt with when performing practical computations. Aliasing is neglected here, however.

The observation model used is linear. The regularization of non-linear models can not be treated with such generality as that of linear models, Engl *et al.* (1996). Moreover, non-linear models are usually linearized, iteration should account for the approximation, see for example Van Gelderen (1992). Snieder (1998) discusses problems related to non-linear inverse problems.

1.3 Outline

In *Chapter 2* ill-posed problems are introduced via integral equations and the spectral decomposition of the operator equation should further clarify the ill-posedness. Moreover, some examples of ill-posed problems are given. Several regularization methods as well as their quality are discussed in *Chapter 3*. With a few exceptions, all these methods have had applications in geodesy and they will be compared with each other. In *Chapter 4* the determination of the regularization parameter(s) and other computational aspects are discussed. A better idea of similarities of and differences between methods is obtained by considering airborne gravimetry, *Chapter 5*, as an example. Finally, the conclusions and recommendations can be found in *Chapter 6*.

2 |

DEFINITION AND EXAMPLES OF ILL-POSED PROBLEMS

2.1 Introduction

In this Chapter it is shown why the linear integral equation

$$\int_a^b K(x,y)\mathbf{f}(y)dy = \mathbf{g}(x)\ ,\quad a \le x \le b \tag{2.1}$$

or symbolical

$$A\mathbf{f} = \mathbf{g} \tag{2.2}$$

is ill-posed. This becomes especially clear using the spectral decomposition of the compact operator A. Ill-posedness is illustrated with some (geodetic) examples.

Mainly Kress (1989); Groetsch (1993) and Tikhonov and Arsenin (1977) are used, but see also Hansen (1997) and Rummel *et al.* (1979). Two introductory books on inverse problems are Wing (1991) and Groetsch (1993).

Before we proceed some important concepts are treated, cf. Kress (1989), see also Appendix A. Equation (2.1) is a *Fredholm integral equation of the first kind*, where the function $\mathbf{f}$ is unknown and the Kernel K and the right hand side $\mathbf{g}$ are given functions. The operator $A : F \to G$ in equation (2.2) is a single valued mapping with domain F and whose range is contained in G, that is, for every $\mathbf{f} \in F$ the mapping A assigns a unique element $A\mathbf{f} \in G$. The *range* $R(A)$ is the set $R(A) \equiv \{A\mathbf{f} : \mathbf{f} \in F\}$ of all image elements.

Injective, surjective and bijective. If for each $\mathbf{g} \in R(A)$ there is only one element $\mathbf{f} \in F$ with $A\mathbf{f} = \mathbf{g}$ then A is said to be *injective* and its inverse mapping $A^+ : R(A) \to F$ is defined by $A^+\mathbf{g} \equiv \mathbf{f}$. The inverse mapping has domain $R(A)$ and range F. It satisfies $A^+A = I$ on F and $AA^+ = I$ on $R(A)$ where I is the identity operator. If $R(A) = G$ then the mapping is said to be *surjective*. If it is injective and surjective the mapping is called *bijective*, that is, the inverse mapping $A^+ : G \to F$ exists.

Bounded operators. An operator $A : F \to G$ mapping a linear space F into a linear space G is called *linear* if

$$A(c_1\mathbf{f}_1 + c_2\mathbf{f}_2) = c_1 A\mathbf{f}_1 + c_2 A\mathbf{f}_2$$

for all $\mathbf{f}_1, \mathbf{f}_2 \in F$ and all $c_1, c_2 \in \mathbb{R}$.

Definition (bounded). A linear operator $A : F \to G$ from a normed space F into a normed space G is called *bounded* if there exists a positive number c such that

$$\|A\mathbf{f}\|_G \leq c\|\mathbf{f}\|_F$$

for all $\mathbf{f} \in F$. Each number c for which this inequality holds is called a *bound* for the operator A. ●

Definition (continuous). Consider the mapping $A : F \to G$ where F and G are metric spaces with d and $\tilde{d}$ the metrics defined. The problem of determining the solution $\mathbf{f} \in F$ from $\mathbf{g} \in G$ is said to be *stable* or *continuous* on the spaces (F, G) if for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\tilde{d}(A\mathbf{f}_1, A\mathbf{f}_2) = \tilde{d}(\mathbf{g}_1, \mathbf{g}_2) < \varepsilon \quad \forall\, \mathbf{f} \text{ satisfying } \; d(\mathbf{f}_1, \mathbf{f}_2) < \delta \tag{2.3}$$

where $\mathbf{f}_1, \mathbf{f}_2 \in F$ and $\mathbf{g}_1, \mathbf{g}_2 \in G$. ●

A linear operator is continuous if and only if it is bounded. Hence, for a linear operator boundedness and continuity mean the same.

Compact operators. A linear operator $A : F \to G$ is compact if and only if for each bounded sequence $\{\mathbf{f}_n\}$ in F the sequence $\{A\mathbf{f}_n\}$ contains a convergent subsequence in G. Compact linear operators are bounded.

Theorem 2.1. Let F, G, H be normed spaces and let $A : F \to G$ and $B : G \to H$ be bounded linear operators. Then the product $BA : F \to H$ is compact if one of the two operators is compact. ●

Theorem 2.2. The indentity operator $I : F \to F$ is compact if and only if F has finite dimension. ●

Therefore, the compact operator A cannot have a bounded inverse unless its range is finite. ($A^+A = I$ is not compact in infinite dimensions due to Theorem 2.2 and because A is compact A^+ has to be unbounded in view of Theorem 2.1.) For a proof of these theorems see (Kress, 1989, Ch. 2).

2.2 Ill-posed problems

There are inverse or indirect problems which imply that there are also direct problems: given a cause $\mathbf{f}$ and a model A find the effect $\mathbf{g}$, $\mathbf{g} = A\mathbf{f}$. A is assumed to be linear and compact, therefore there is a unique effect $\mathbf{g}$ for each cause $\mathbf{f}$ (the mapping A is injective) and small changes in $\mathbf{f}$ result in small changes in $\mathbf{g}$. In addition to the direct problem there are two types of inverse problems:

causation given A and $\mathbf{g}$, determine $\mathbf{f}$,

model identification given $\mathbf{f}$ and $\mathbf{g}$, determine A.

One might hope that the first inverse problem, causation, is the most important one, since it is more complicated to determine an unknown model. However, often only an approximate model is known and with the causation some model parameters should be identified as well. For example, in the case of determination of the gravity field from satellite measurements models for drag and solar-pressure acting on a satellite are not always accurate enough. Although this is an interesting subject it is not treated here, as stated before, to avoid unnecessary complications. We will concentrate on causation and will not look at model identification.

2.2.1 Existence, uniqueness and stability

In the direct problem the solution is assumed to exist and to be unique and stable. For inverse problems some or all of these properties may not hold. When a solution for the problem of determining $\mathbf{f}$ exists and this solution is unique and stable the problem is said to be *well-posed* or *properly posed*:

Definition (well-posed, ill-posed). Let $A : F \to G$ be an operator from a normed space F into a normed space G. The equation

$$A\mathbf{f} = \mathbf{g} \tag{2.4}$$

with $\mathbf{f} \in F, \mathbf{g} \in G$ is called *well-posed* if A is bijective and the inverse operator $A^+ : G \to F$ is continuous. Otherwise the problem is said to be *ill-posed* or *improperly posed.* ●

According to this definition three types of ill-posedness can be distinguished. If A is not surjective then (2.4) is not solvable for all $g \in G$ (*nonexistence*). If A is not injective then (2.4) may have more than one solution (*nonuniqueness*). Finally, if A^+ exists but is not continuous then the solution $\mathbf{f}$ of Eq. (2.4) does not depend continuously on the data $\mathbf{g}$ (*instability*).

The well-posedness of a problem is a property of the operator A together with the solution space F and the data space G including the norms on F and G. Therefore maybe, instability could be overcome by changing the spaces F and G and their norms. However, this approach is inadequate since the spaces F and G including their norms are determined by practical needs, Kress (1989).

As mentioned above, usually at least one of the conditions is not satisfied in inverse problems, therefore inverse problems are usually ill-posed. The problem can for instance be unstable, that is small changes in the data $\mathbf{g}$ result in large changes in the solution $\mathbf{f}$ and that is of course undesirable. For example, if one wants to determine a function $\mathbf{f}$ from measurements $\mathbf{g}^\varepsilon$, the solutions $\mathbf{f}^\varepsilon$ should preferably be close to the 'true' function $\mathbf{f}$. When the small difference between $\mathbf{g}$ and $\mathbf{g}^\varepsilon$ causes a large difference between $\mathbf{f}$ and $\mathbf{f}^\varepsilon$ one obviously has a problem. To overcome this difficulty some type of stabilization or *regularization* is applied, compare the next Chapter.

The existence of the solution will not be a matter of great concern in this report. Naturally, it is an important requirement that a solution exists for exact data, but for

perturbed data the problem has to be changed (regularized) and the notion of a solution can be relaxed. In practice the existence of an *approximate* solution is required. As side remark we note that even in the presence of exact data no solution may exist since every model contains simplifications and approximations.

The amount of data that is available for the determination of the solution $\mathbf{f}$ is usually finite. Because this is a continuous function there are infinitely many degrees of freedom, and therefore the approximate solution is never unique in practice. Also when continuous, exact data is available the null space of A may not be zero. However, we will assume injectivity unless stated otherwise, and we will not discuss nonuniqueness due to the finite amount of data, see for example Backus and Gilbert (1967, 1968); Parker (1994); Trampert and Snieder (1996).

2.2.2 Inverse problems and integral equations

Equation (2.1) is a special form of the more general equation with $\mathbf{h}(x)\mathbf{f}(x)$ added to the right-hand side:

$$\mathbf{g}(x) = \int_a^b K(x,y)\mathbf{f}(y)dy + \mathbf{h}(x)\mathbf{f}(x).$$

If $\mathbf{h}(x) \equiv 0$ then we have an integral equation of the *first kind*, if $\mathbf{h}(x) \neq 0$ for $a \leq x \leq b$, the equation is of the *second kind*, and if $\mathbf{h}(x)$ vanishes somewhere but not identically, the equation is of the *third kind*, Phillips (1962). Here we consider

$$\begin{aligned} \mathbf{f} - A^*A\mathbf{f} &= A^*\mathbf{g} \\ \mathbf{f} - T\mathbf{f} &= \mathbf{b} \end{aligned}$$

where $T : F \to F$ is a compact linear operator. If $N(I-T) = \{0\}$, then $I-T$ is injective and the solution $f \in F$ is unique and depends continuously on $\mathbf{b}$, Groetsch (1984); Kress (1989). In our case, therefore, the integral equation of the second kind is known to be well-posed, a unique and stable solution exists, Groetsch (1993). This fact will be useful later.

From Theorems 2.1 and 2.2 one knows that a compact operator cannot have a bounded inverse. The following example, taken from Tikhonov and Arsenin (1977), shows the instability of (2.1).

Example. In equation (2.1), $\mathbf{f}(y)$ is the unknown function in $F, \mathbf{g}(x)$ is a known (or measured) function in G. The solution $\mathbf{f}(y) \in C[a,b]$, i.e. $\mathbf{f}$ is an element of the class C of functions that are continuous on the closed interval $[a,b]$. Note that the linear space $C[a,b]$ of continuous functions defined on an interval $[a,b] \subset I\!R$ is complete with respect to the maximum norm

$$\|\mathbf{f}\|_\infty = \max_{a \leq y \leq b} |\mathbf{f}(y)|$$

but not with respect to the mean square norm L^2. The Hilbert space $L^2[a,b]$ is the completion of $C[a,b]$ with respect to the inner product

$$\langle \mathbf{g}_1, \mathbf{g}_2 \rangle = \int_a^b \mathbf{g}_1(x)\mathbf{g}_2(x)dx,$$

see Kreyszig (1989). Therefore, changes in the left-hand member of the equation are measured with the L^2-metric

$$\tilde{d}(\mathbf{g}_1, \mathbf{g}_2) = \sqrt{\int_a^b [\mathbf{g}_1(x) - \mathbf{g}_2(x)]^2 \, dx}$$

while changes in $\mathbf{f}(y)$ are measured with the metric

$$d(\mathbf{f}_1, \mathbf{f}_2) = \max_{y \in [a,b]} |\mathbf{f}_1(y) - \mathbf{f}_2(y)| .$$

Consider the function $\mathbf{f}^\varepsilon(y) = \mathbf{f}(y) + N \sin ky$ which is a solution of equation (2.1) with left-hand member

$$\mathbf{g}^\varepsilon(x) = \mathbf{g}(x) + N \int_a^b K(x, y) \sin ky dy$$

with $k \in \mathbb{N}, N \in \mathbb{R}$. The *Riemann-Lebesgue lemma*, Groetsch (1993), states that if the kernel is square-integrable,[1] then

$$\mathbf{g}_k \equiv \int_a^b K(x, y) \sin ky dy \to 0 \text{ as } k \to \infty$$

where the convergence is in the sense of the mean square norm.

Thus for any N

$$\lim_{k \to \infty} \tilde{d}(\mathbf{g}^\varepsilon, \mathbf{g}) = \lim_{k \to \infty} |N| \|\mathbf{g}_k\|_G = 0$$

but

$$d(\mathbf{f}^\varepsilon, \mathbf{f}) = \max_{y \in [a,b]} |\mathbf{f}(y) - \mathbf{f}^\varepsilon(y)| = \max_{y \in [a,b]} |N \sin ky| = |N|$$

independent of k. The distance between the solutions $\mathbf{f}$ and $\mathbf{f}^\varepsilon$ is therefore arbitrarily large. This makes the instability fundamental, and not just a consequence of some special form of the kernel. Very small changes in $\mathbf{g}(x)$ can be accounted for by large changes in $\mathbf{f}(y)$, Groetsch (1993). ●

2.2.3 Spectral decomposition

The spectral form of $A\mathbf{f}$ is

$$A\mathbf{f} = \sum_{n=1}^{\infty} \sigma_n \langle \mathbf{f}, \mathbf{v}_n \rangle \mathbf{u}_n$$

and is called the *singular value decomposition* (SVD) of A. The orthonormal eigenvectors $\mathbf{u}_n$ and $\mathbf{v}_n$ form a complete orthonormal set for $\overline{R(AA^*)}$ and $\overline{R(A^*A)}$ respectively. The numbers σ_n are called *singular values*, which decrease towards zero for increasing n if A is a compact operator, cf. Appendix B.

[1] Which is the case for the larger majority of integral equations of the first kind in real applications. Then the operator on L^2 is compact, Groetsch (1993).

Generalized inverse, instability. The operator A^+ in $A^+\mathbf{g} = \mathbf{f}$ is also called the *generalized inverse* of A, and $A^+\mathbf{g}$ is the l.s. solution of $\min \|A\mathbf{f} - \mathbf{g}\|_G^2$, cf. Appendix B. The generalized inverse A^+ can be written in spectral form as well. For $\mathbf{g} \in D(A^+)$:

$$A^+\mathbf{g} = \sum_{n=1}^{\infty} \frac{\langle \mathbf{g}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \tag{2.5}$$

Equation (2.5) shows that the inverse becomes unstable when errors are present. Errors in $\mathbf{g}$ corresponding to high frequencies, i.e. large n, are amplified by large factors $1/\sigma_n$. If $\dim R(A) < \infty$ the amplification stays bounded, but might be unacceptably large. However, if $\dim R(A) = \infty$, then $\lim_{n\to\infty} \sigma_n = 0$ holds, so that data errors of a fixed size (e.g. white noise) can be amplified without bounds. For example, if $\mathbf{g}^\varepsilon = \mathbf{g} + \varepsilon \mathbf{u}_n$, then $\|\mathbf{g}^\varepsilon - \mathbf{g}\|_G = \varepsilon$, but, due to (2.5)

$$\|A^+\mathbf{g} - A^+\mathbf{g}^\varepsilon\|_F = \left\| \frac{\langle \varepsilon \mathbf{u}_n, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n \right\|_F = \frac{\varepsilon}{\sigma_n} \to \infty \text{ as } n \to \infty.$$

This also suffices to show that in case of finite dimensional problems we may formally not speak about ill-posed problems since the error always stays bounded. However, the error may be unacceptably large and the finite dimensional equation is usually derived from the original infinite dimensional problem $A\mathbf{f} = \mathbf{g}$. With increasing order of approximation (with increasing n) the problem of solving $A\mathbf{x} = \mathbf{y}$ becomes therefore more and more ill-posed:

$$\lim_{n,m\to\infty} (A\mathbf{x} - \mathbf{y}) = (A\mathbf{f} - \mathbf{g}).$$

Statistical approach. The considerations so far are primarily based on the deterministic approach, the statistical properties of the measurements and the solution were not given much attention. A different point of view on the same problem is the statistical approach, which classifies the problem of solving $\min \|A\mathbf{x} - \mathbf{y}^\varepsilon\|_2^2$ as *nonorthogonal*, that is A^TA is not nearly an indentity matrix, Hoerl and Kennard (1970).

Consider the linear regression model $\mathbf{y} = A\mathbf{x} + \underline{\varepsilon}$, where it is assumed that A is $m \times n$, $m \geq n$ and of rank n, $\mathbf{x}$ is $n \times 1$ and unknown, $E\{\underline{\varepsilon}\} = 0$, and $E\{\underline{\varepsilon}\underline{\varepsilon}^T\} = \sigma^2 I_m$. The best linear unbiased estimate of $\mathbf{x}$ is

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

and minimizes the quadratic form

$$\phi(\hat{\mathbf{x}}) = (\mathbf{y} - A\hat{\mathbf{x}})^T(\mathbf{y} - A\hat{\mathbf{x}}). \tag{2.6}$$

The matrix A^TA is said to be in correlation form, Hoerl and Kennard (1970), and we are concerned with cases for which it is not nearly a unit matrix. The effects of this condition on the estimation of $\mathbf{x}$ can be demonstrated by considering the error variance-covariance matrix of $\hat{\mathbf{x}}$ and its distance from the expected value. The first is given by

$$\text{var}(\hat{\mathbf{x}}) = Q_{\hat{\mathbf{x}}} = \sigma^2(A^TA)^{-1}.$$

Let the distance from $\mathbf{x}$ to $\hat{\mathbf{x}}$ be l, then:

$$l^2 = (\hat{\mathbf{x}} - \mathbf{x})^T(\hat{\mathbf{x}} - \mathbf{x}) \tag{2.7}$$

and

$$E\{l^2\} = \sigma^2 \text{trace}(A^T A)^{-1} \tag{2.8}$$

or equivalently from (2.7) and (2.8)

$$E\{\hat{\mathbf{x}}^T \hat{\mathbf{x}}\} = \mathbf{x}^T \mathbf{x} + \sigma^2 \text{trace}(A^T A)^{-1}. \tag{2.9}$$

When the error $\underline{\varepsilon}$ is normally distributed, then

$$\text{var}(l^2) = 2\sigma^4 \text{trace}(A^T A)^{-2}.$$

These properties show the uncertainty in $\hat{\mathbf{x}}$ when $A^T A$ moves from a unit matrix to an ill-conditioned one, if we look at the spectral decomposition of $A^T A$. Let the ordering of the eigenvalues be as usual (that is, decreasing, cf. Appendix B), then the average value of the squared distance from $\hat{\mathbf{x}}$ to $\mathbf{x}$ is given by

$$E\{l^2\} = \sigma^2 \sum_{i=1}^{n} \frac{1}{\lambda_i}$$

and the variance when the error has a normal distribution is given by

$$\text{var}(l^2) = 2\sigma^4 \sum_{i=1}^{n} \frac{1}{\lambda_i^2}$$

compare Hoerl and Kennard (1970). Lower bounds for the average and the variance are σ^2/λ_n and $2\sigma^4 \lambda_n^2$ respectively. Hence, if $A^T A$ has small eigenvalues, the distance from $\hat{\mathbf{x}}$ to $\mathbf{x}$ tends to be large. The probability that $\hat{\mathbf{x}}$ is close to $\mathbf{x}$ is therefore small.

2.3 Examples of ill-posed problems

To illustrate the above we shall look at some examples. Many more examples can be found in the textbooks of Tikhonov and Arsenin (1977); Louis (1989); Wing (1991); Anger *et al.* (1993); Groetsch (1993); Engl *et al.* (1996).

Density or mass anomaly. To geodesists perhaps the most famous inverse problem is that of determining the Earth's mass distribution from the exterior gravitational potential. As *Stokes theorem* states this is impossible (Rummel, 1992, p. 2.22): "a function V harmonic outside Σ is uniquely determined by its values on the boundary. On the other hand, however, there are infinitely many mass distributions, which have the given V as exterior potential." Hence, the lack of uniqueness makes this an ill-posed problem. See also Parker (1994). ●

Backwards heat equation. (From Groetsch (1993).)

Consider a bar of length π with heat flow in the x-direction. The temperature $\mathbf{u}(x,t)$ satisfies the partial differential equation

$$\frac{\partial \mathbf{u}}{\partial t} = \frac{\partial^2 \mathbf{u}}{\partial x^2}, \quad 0 < x < \pi.$$

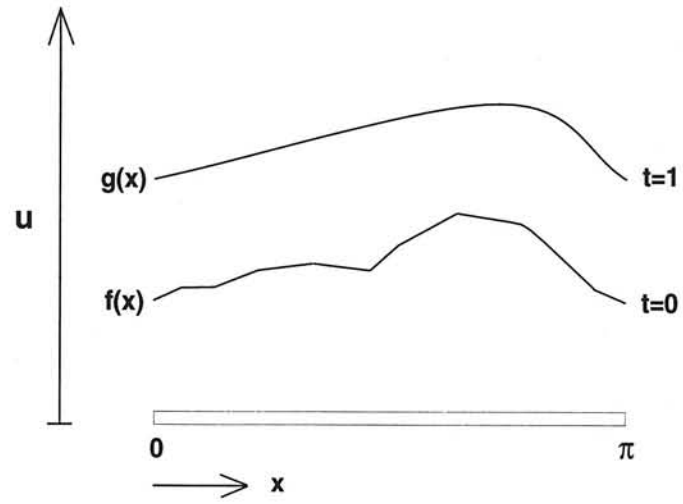

Figure 2.1: *Temperature distribution between 0 and π at $t = 0$ and $t = 1$. The temperature at both ends is kept zero.*

Assume the boundary and initial conditions

$$\mathbf{u}(0,t) = \mathbf{u}(\pi,t) = 0,\ \mathbf{u}(x,0) = \mathbf{f}(x)$$

hold, i.e. the temperature at the ends is kept zero and the initial temperature distribution is some function $\mathbf{f}(x)$, $0 \leq x \leq \pi$. The time interval t of the function $\mathbf{u}$ is bounded by zero and one: $0 \leq t \leq 1$. More generally the upper bound is T, see (Kress, 1989, p. 222-223).

The method of separation of variables leads to

$$\mathbf{u}(x,t) = \sum_{n=1}^{\infty} a_n e^{-n^2 t} \sin nx \tag{2.10}$$

with coefficients

$$a_n = \frac{2}{\pi} \int_0^{\pi} \mathbf{f}(y) \sin nydy. \tag{2.11}$$

Let $\mathbf{g}(x) = \mathbf{u}(x,1)$, see Figure 2.1. Substituting (2.11) into (2.10) and interchanging the summation and integration one arrives at

$$\mathbf{g}(x) = \int_0^{\pi} K(x,y)\mathbf{f}(y)dy \tag{2.12}$$

with

$$K(x,y) = \frac{2}{\pi} \sum_{n=1}^{\infty} e^{-n^2} \sin nx \sin ny$$

compare Kress (1989); Groetsch (1993).

The inverse problem is to determine the initial temperature distribution $\mathbf{f}(x)$ that gives rise to $\mathbf{g}(x)$. Since the initial temperature distribution $\mathbf{f}$ is highly diffused at the

later time $t = 1$, recovering this detailed information from measurements of $\mathbf{g}$ will be extremely difficult. Specifically, high frequencies (large n) are severely damped by the very small factor e^{-n^2}, Groetsch (1993).

Suppose $\mathbf{f}$ and $\mathbf{g}$ satisfy (2.12). Let $\varepsilon > 0$ (small) and $N > 0$ (large) and let $f_N(y) = N \sin ky$. The perturbation $\mathbf{f}_N$ is arbitrarily large (C-metric)

$$\max_{y\in[0,\pi]} |N \sin ky| = |N| .$$

The perturbation in $\mathbf{g}$ becomes

$$\begin{aligned} \mathbf{g}_N(x) &= \int_0^\pi K(x,y)\mathbf{f}_N(y)dy \\ &= \frac{2}{\pi}N\int_0^\pi \left[\sum_{n=1}^{\infty}\left[e^{-n^2}\sin nx \sin ny\right]\sin ky\right] dy \\ &= \frac{2}{\pi}N\sum_{n=1}^{\infty}\left[e^{-n^2}\sin nx\int_0^\pi [\sin ny \sin ky]\, dy\right]. \end{aligned} \tag{2.13}$$

The integral in (2.13) is zero for $n \neq k$. If $n = k$ then $\mathbf{g}_N$ becomes

$$\mathbf{g}_N(x) = Ne^{-k^2}\sin kx$$

which amplitude is

$$|\mathbf{g}_N(x)| = Ne^{-k^2}\sqrt{\left[\int_0^\pi \sin kx dx\right]^2}.$$

For large k the amplitude is arbitrarily small:

$$\begin{aligned} |\mathbf{g}_N(x)| &= \lim_{k\to\infty} Ne^{-k^2}\sqrt{\left[\int_0^\pi \sin kx dx\right]^2} \\ &= N\lim_{k\to\infty} e^{-k^2}\left|\int_0^\pi \sin kx dx\right| \\ &< N\lim_{k\to\infty}\left|\int_0^\pi \sin kx dx\right| < \varepsilon. \end{aligned}$$

since the limit approaches zero (Riemann-Lebesgue lemma). Hence, a large disturbance in the solution $\mathbf{f}$ can be accounted for by a small disturbance in the measurements $\mathbf{g}$. This is therefore an ill-posed problem. ●

This may not be a geodetic example, but it is a nice one and it resembles downward continuation which is treated next.

Downward continuation. The determination of the gravity potential at the earth's surface from the potential at satellite altitude is an ill-posed problem. Assume that one has a sphere at height h above the earth's surface where the potential is observed continuously.

The gravity potential expressed in spherical harmonics at $r = R + h$ is[2]

$$V(\theta_P, \lambda_P, r_P) = \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \sum_{m=0}^{n} \left(\bar{C}_{nm} \cos m\lambda_P + \bar{S}_{nm} \sin m\lambda_P\right) \bar{P}_{nm} (\cos \theta_P) . \tag{2.14}$$

The potential at the earth's surface equals (2.14) with $r_P = R$. The coefficients $\bar{C}_{nm}$ and $\bar{S}_{nm}$ are, Rummel (1992)

$$\begin{aligned} \bar{C}_{nm} &= \frac{1}{4\pi} \int_{\sigma_Q} V(\theta_Q, \lambda_Q, R) \cos m\lambda_Q \bar{P}_{nm} (\cos \theta_Q) \, d\sigma_Q, \\ \bar{S}_{nm} &= \frac{1}{4\pi} \int_{\sigma_Q} V(\theta_Q, \lambda_Q, R) \sin m\lambda_Q \bar{P}_{nm} (\cos \theta_Q) \, d\sigma_Q. \end{aligned} \tag{2.15}$$

Inserting (2.15) into (2.14), interchanging summation and integration and using the *addition theorem*, Rummel (1992)

$$(2n+1) P_n (\cos \Psi_{PQ}) = \sum_{m=0}^{n} \bar{P}_{nm} (\cos \theta_P) \, \bar{P}_{nm} (\cos \theta_Q) \cos m (\lambda_P - \lambda_Q)$$

yields

$$V(P) = \frac{1}{4\pi} \int_{\sigma_Q} K (P - Q) \, V(Q) d\sigma_Q \tag{2.16}$$

where

$$K (P - Q) = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} (2n+1) P_n (\cos(P - Q)) . \tag{2.17}$$

Analogous to the previous example high frequencies are damped by the factor $(R/r_P)^{n+1}$.

Consider a disturbance $V_N(Q) = N \bar{P}_{kk}(\cos \theta_Q) \sin k\lambda_Q$. Inserting this in (2.16) and rewriting (2.17) in its extended form to separate P and Q parts, interchanging summation and integration and finally taking all λ_Q terms together, one obtains:

$$\begin{aligned} V_N(P) \;=\; & \frac{N}{4\pi} \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \sum_{m=0}^{n} \bar{P}_{nm}(\mathrm{c}\theta_P) \int_0^{\pi} \bar{P}_{nm}(\mathrm{c}\theta_Q) \bar{P}_{kk}(\mathrm{c}\theta_Q) \times \\ & \times \left(\mathrm{c}m\lambda_P \int_0^{2\pi} \mathrm{c}m\lambda_Q \mathrm{s}k\lambda_Q d\lambda_Q + \mathrm{s}m\lambda_P \int_0^{2\pi} \mathrm{s}m\lambda_Q \mathrm{s}k\lambda_Q d\lambda_Q\right) \mathrm{s}\theta_Q d\theta_Q \end{aligned}$$

with the abbreviations c = cos and s = sin. The first integral over λ always equals zero, the second equals π for $m = k$ and 0 for $m \neq k$. As long as $n < k$ there are no contributions since m runs from 0 to n. Hence

$$V_N(P) = \frac{N}{4} \sum_{n=k}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \bar{P}_{nk}(\cos \theta_P) \int_0^{\pi} \bar{P}_{nk}(\cos \theta_Q) \bar{P}_{kk}(\cos \theta_Q) \sin \theta_Q d\theta_Q.$$

Because of the orthogonality of the Legendre functions the integral equals 4 if $n = k$ and zero if $n \neq k$. Therefore $V_N(P)$ can be written as:

$$V_N(P) = N \left(\frac{R}{r_P}\right)^{k+1} \bar{P}_{kk}(\cos \theta_P) \tag{2.18}$$

[2]The functions $\bar{P}_{nm}$ are *fully normalized associated Legendre functions of the first kind* and dimensionless. They make the orthogonality relationship fairly simple, see e.g. Heiskanen and Moritz (1967); Rummel (1992).

which amplitude is

$$|V_N(P)| = N\left(\frac{R}{r_P}\right)^{k+1}\left|\bar{P}_{kk}(\cos\theta_P)\right|.$$

Since

$$\|\bar{P}_{kk}(x)\|_{L^2}^2 = \int_{-1}^{1}\bar{P}_{kk}(x)^2dx = 4$$

the amplitude of $\bar{P}_{kk}$ is bounded (which we already used to arrive at (2.18)). Thus

$$\lim_{k\to\infty}|V_N(P)| < \varepsilon,$$

whereas

$$\max_{\theta,\lambda}\left|N\bar{P}_{kk}(\cos\theta)\sin k\lambda\right|$$

can be made arbitrarily large by choosing N large. ●

Laplace equation. (From Tikhonov and Arsenin (1977) and Kress (1989).)

The initial value or Cauchy problem for the two-dimensional Laplace equation consists of finding a solution of the equation $\Delta\mathbf{u}(x,y) = 0$ from the initial data:

$$\mathbf{u}(.,0) = 0, \quad \frac{\partial}{\partial y}\mathbf{u}(.,0) = \mathbf{f}(x), \quad -\infty < x < \infty$$

where $\mathbf{f}(x)$ is a given continuous function and $\mathbf{u}$ a harmonic function. Let the data be

$$\mathbf{f}_k(x) = k^{-1}\sin kx, \quad x \in I\!R$$

for $k \in I\!N$, then one obtains the solution

$$\mathbf{u}_k(x,y) = \frac{1}{k^2}\sin kx\sinh ky, \quad k > 0.$$

Let $\mathbf{f}_0(x) = 0$ with solution $\mathbf{u}_0(x,y) = 0$. The difference in the initial data is

$$d(\mathbf{f}_k,\mathbf{f}_0) = \max_x|\mathbf{f}_k(x) - \mathbf{f}_0(x)| = \frac{1}{k}$$

which can be made arbitrarily small by taking k sufficiently large. However, for any fixed $y > 0$, the difference between the solutions

$$\begin{aligned} d(\mathbf{u}_k,\mathbf{u}_0) &= \max_x|\mathbf{u}_k(x,y) - \mathbf{u}_0(x,y)| \\ &= \max_x\left|\frac{1}{k^2}\sin kx\sinh ky\right| = \frac{1}{k^2}\sinh ky \end{aligned}$$

can be made arbitrarily large for sufficiently large k: $\sinh ky = (e^{ky} - e^{-ky})/2$ behaves like e^{ky} for large k, which goes faster to infinity than k^{-2} goes to zero.

This is somewhat familiar in physical geodesy: the solution of $\Delta\mathbf{u} = 0$ expressed in spherical harmonics has either r^n or $r^{-(n+1)}$ as upward continuation term. The additional constraint of vanishing potential at infinity turns down the first possibility, which leaves us with a properly posed problem. ●

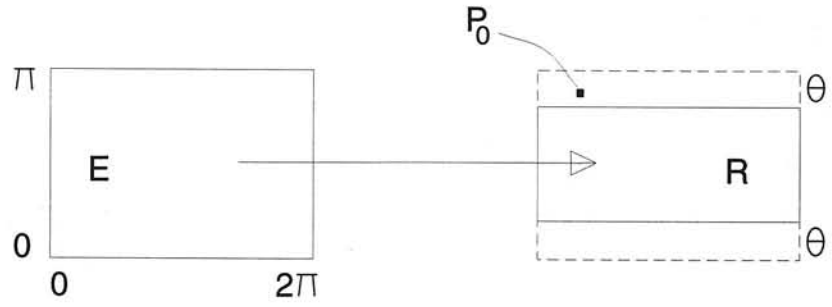

Figure 2.2: *Mapping of a function onto a smaller region.*

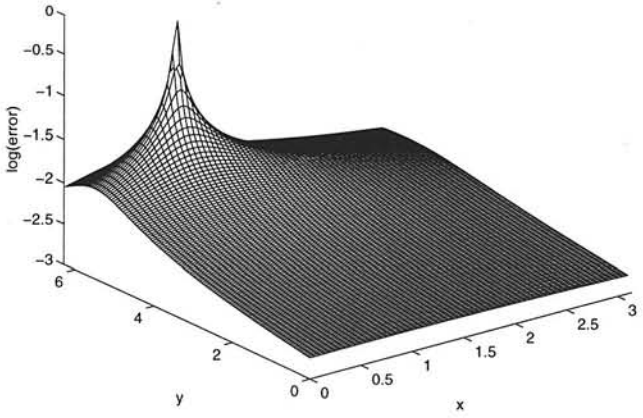

Figure 2.3: *Logarithmic plot of* $\varepsilon/(P - P_0)$. *The coordinates of* P_0 *are* $(1/4\pi, 7/4\pi)$, $\varepsilon = 0.01$ *and the coordinates of* P *vary from* $0 - 2\pi$ *and* $0 - \pi$ *with step size 0.05.*

Polar gap. (See also Tikhonov and Arsenin (1977))

Consider the mapping $A : E \to R$ or $A : [0, \pi] \times [0, 2\pi] \to [\theta, \pi - \theta] \times [0, 2\pi]$ with $0 < \theta < \pi$, compare Figure 2.2. The mapping is from E(arth) to a smaller region R. This means that the function is measured not on the entire region but in a limited part. Think for example of a satellite in a non-polar orbit, an inclination $i \neq 90$ degrees gives two polar gaps of size $\theta = |90 - i|$.

Take a point P_0 on E but not on R with distance Ψ to R. On the total region E the function $\mathbf{f}_1(P)$ is defined, where the function depends on the location P. Another function is $\mathbf{f}_2(P) = \mathbf{f}_1(P) + \varepsilon/(P - P_0)$. These two functions differ by $\varepsilon/(P - P_0)$, which on R does not exceed ε/Ψ. The ratio ε/Ψ, hence the difference between the two functions, on R can be made arbitrarily small by choosing a sufficiently small ε. However, the difference $\mathbf{f}_2(P) - \mathbf{f}_1(P) = \varepsilon/(P - P_0)$ is unbounded on the region E as a whole, see Figure 2.3. It is therefore an ill-posed problem. ●

2.4 Summary

It is inherent to integral equations of the first kind that the determination of the unknown function $\mathbf{f}$ from the data $\mathbf{g}$ is ill-posed. The direct problem, the determination of $\mathbf{g}$ from $\mathbf{f}$, is continuous if the operator A is compact and bounded. However, the immediate consequence is that the inverse operator is not compact and bounded: the solution does not continuously depend on the data, the inverse problem is unstable.

This becomes especially clear when examining the spectral decomposition of the inverse operator A^+. For high frequencies the singular values become smaller and smaller, which means that their inverses become larger and larger. Therefore, any error at high frequencies is greatly amplified and tends to infinity if n, the frequency, goes to infinity.

The examples of ill-posed problems show that we are indeed dealing with physical meaningful problems. Furthermore, it becomes clear that a large variety of ill-posed problems exists. Although examples of ill-posed problems have been given whose cause is lack of uniqueness or stability, the emphasis in the remainder of this report is only on stability.

3 | METHODS OF REGULARIZATION

3.1 Introduction

Several methods exist to compute stable solutions of inverse problems. In this Chapter the methods more or less familiar in geodesy are discussed:

- Tikhonov-Phillips regularization
- biased estimation
- collocation
- truncated singular value decomposition (TSVD)
- iteration

An example of the application of Tikhonov-Phillips regularization can be found in Rummel *et al.* (1979), biased estimation e.g. in Xu (1992a). The principal reference for collocation is Moritz (1980). Lerch *et al.* (1993) apply TSVD and Wenzel (1985) and Schuh (1996) apply iteration to compute a (high degree) gravity field solution.

The principle of each method is explained and the distance between the true and approximate solution, which is measured in the mean square norm, is derived. We distinguish two groups of regularization methods, the direct and the iteration methods. This classification is not as strict as it may seem since the computation of a direct solution involves iteration as well. The computation of the so-called regularization parameter causes the iteration, compare Chapter 4. However, an approximate solution computed with an iteration method can be expressed as the previous solution with some additional terms, Section 3.7.

The differences and similarities of the methods are discussed. Especially it is shown that biased estimation and Tikhonov-Phillips regularization are equivalent for the most simple case. Both methods are very similar to collocation looking at the formulas, however the underlying line of thought is different.

Requirements a regularization method should fulfil

Suppose perfect, continuous measurements $\mathbf{g}$ are available. Then one would like that the solution $\mathbf{f}$ is such that $A\mathbf{f} = \mathbf{g}$ holds. The generalized inverse A^+ provides such a solution. Unfortunately, this solution is unstable as was shown in the preceding Chapter.

Therefore, in presence of measurement errors, the generalized inverse can not be used as such, regularization is necessary. This regularization in general looks like

$$A_\alpha^+ \mathbf{g}^\varepsilon = \mathbf{f}_\alpha^\varepsilon$$

with

$$\lim_{\alpha \to 0} A_\alpha^+ = A^+$$

and if $\varepsilon \to 0$ then $\alpha \to 0$, hence $\alpha = \alpha(\varepsilon)$. With these two requirements it follows that exact data give the exact solution. This can be formalized by the following definition given by Kress (1989).

Definition. The choice of the regularization parameter $\alpha = \alpha(\varepsilon)$ depending on the error level ε for a regularization scheme $A_\alpha^+, \alpha > 0$ is called *regular* if for all $\mathbf{g} \in R(A)$ and all $\mathbf{g}^\varepsilon \in G$ with $\|\mathbf{g}^\varepsilon - \mathbf{g}\| \leq \varepsilon$ there holds

$$A_{\alpha(\varepsilon)}^+ \mathbf{g}^\varepsilon \to A^+ \mathbf{g}, \quad \varepsilon \to 0.$$

●

In the sequel it is assumed that the linear operator A is injective. This is not a principle loss of generality since uniqueness for a linear equation can always be achieved by a suitable modification of the solution space F, Kress (1989).

Regularization scheme in spectral form. We have seen in the preceding Chapter, Eq. (2.5), that the ill-posedness of an equation of the first kind with compact operator stems from the behaviour of the singular values $\sigma_n \to 0, n \to \infty$. An obvious idea is to filter out the influence of the factor $1/\sigma_n$. To this end, consider the filter $\delta : (0, \infty) \times (0, \|A\|] \to I\!R$ which is defined as a bounded function satisfying the conditions:

1. For each $\alpha > 0$ there exists a positive constant $c(\alpha)$ such that

$$|\delta(\alpha, \sigma)| \leq c(\alpha)\sigma \tag{3.1}$$

for all $0 < \sigma \leq \|A\|$.

2. There holds

$$\lim_{\alpha \to 0} \delta(\alpha, \sigma) = 1 \tag{3.2}$$

for all $0 < \sigma \leq \|A\|$.

Then the operator $A_\alpha^+ : G \to F, \alpha > 0$, defined by

$$A_\alpha^+ \mathbf{g} \equiv \sum_{n=1}^{\infty} \frac{\delta(\alpha, \sigma_n)}{\sigma_n} \langle \mathbf{g}, \mathbf{u}_n \rangle \mathbf{v}_n$$

for all $\mathbf{g} \in G$, describes a regularization scheme with

$$\|A_\alpha^+\| \leq c(\alpha).$$

Thus, A_α^+ is a bounded linear operator with bound c, Kress (1989). It is not allowed to use any arbitrary filter since conditions (3.1) and (3.2) have to be satisfied. For all regularization methods in this Chapter the filter is derived.

3.2 Tikhonov-Phillips regularization

Tikhonov-Phillips regularization was developed independently in the early sixties by (of course) Tikhonov (1963b, 1963a) and Phillips (1962) and is also called Tikhonov regularization (TR) for short.

In geodetic literature Tikhonov regularization has been studied with emphasis on the connection with collocation, e.g. Rummel *et al.* (1979). However, the regularization error is mostly neglected. Tikhonov regularization does not necessarily give an unbiased answer, whereas collocation does.

3.2.1 Principle of the method

Regularization with signal constraint

Consider the integral equation of the first kind

$$(A\mathbf{f})(x) = \int_a^b K(x,y)\mathbf{f}(y)dy = \mathbf{g}(x)x, \quad a \leq x \leq b.$$

The kernel $K(x,y)$ is continuous and integrable. As shown earlier this is an ill-posed integral equation, in the sense that a small change or error in $\mathbf{g}$ may cause a large change in the solution $\mathbf{f}$. Imposing an additional condition on $\mathbf{f}$ provides a stable solution:

$$\|\mathbf{f}\|_F^2 \leq c < \infty$$

where c is a constant. This additional condition is also called the constraint or the penalty term. Instead of minimizing

$$J(\mathbf{f}) = \|A\mathbf{f} - \mathbf{g}\|_G^2$$

the functional

$$J_\alpha(\mathbf{f}) = \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f}\|_F^2 \tag{3.3}$$

has to be minimized, where α is the positive Lagrange multiplier. Whenever a minimization problem has the above appearance (3.3) it is said to be in *standard form.* Equation (3.3) is a short-hand notation of

$$\begin{aligned} J_\alpha(\mathbf{f}) &= \int_a^b \left((A\mathbf{f})(x) - \mathbf{g}(x)\right)^2 dx + \alpha \int_a^b \mathbf{f}(x)^2 dx \\ &= \int_a^b \left(\int_a^b K(x,y)\mathbf{f}(y)dy - \mathbf{g}(x) \right)^2 dx + \alpha \int_a^b \mathbf{f}(x)^2 dx. \end{aligned}$$

The minimizer $\mathbf{f}_\alpha$ of (3.3) is given by the unique solution of the equation

$$(A^*A + \alpha I)\mathbf{f}_\alpha = A^*\mathbf{g}$$

or

$$\mathbf{f}_\alpha = (A^*A + \alpha I)^{-1}A^*\mathbf{g}. \tag{3.4}$$

and depends continuously on $\mathbf{g}$, where A^* is the adjoint of A. The operator $A^*A + \alpha I$ (for $\alpha > 0$) is bijective, and the inverse is bounded. As $\alpha \to 0, \mathbf{f}_\alpha \to A^+\mathbf{g}$, Nashed (1976), where A^+ is the generalized inverse of A. Note that the above equation is an integral equation of the second kind and therefore well-posed, Groetsch (1984); Kress (1989), which is equivalent to the continuous dependence of $\mathbf{f}_\alpha$ on $\mathbf{g}$. Hence, an interpretation is that the original ill-posed integral equation of the first kind is replaced by a nearby well-posed integral equation, Tikhonov and Arsenin (1977); Nashed (1976); Groetsch (1993).

The above can be summarized by the following theorem which is proven in, for example, Kress (1989).

Theorem (Continuity and uniqueness of the regularized solution). Let $A : F \to G$ be a bounded linear operator with F and G Hilbert spaces, and let $\alpha > 0$. Then for each $\mathbf{g} \in G$ there exists a unique $\mathbf{f}_\alpha \in F$ such that

$$\|A\mathbf{f}_\alpha - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f}_\alpha\|_F^2 = \inf_{\mathbf{f}\in F}\{\|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f}\|_F^2\}.$$

The minimizer $\mathbf{f}_\alpha$ is given by the unique solution of the equation

$$\alpha\mathbf{f}_\alpha + A^*A\mathbf{f}_\alpha = A^*\mathbf{g}$$

and depends continuously on $\mathbf{g}$. ●

Spectral decomposition. Since the operator A is compact it has a singular system $\{\mathbf{v}_n, \mathbf{u}_n; \sigma_n\}$. The regularized solution reads

$$\mathbf{f}_\alpha^\varepsilon = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha}\langle \mathbf{g}^\varepsilon, \mathbf{u}_n\rangle \mathbf{v}_n = \sum_{n=1}^{\infty} \delta_n \frac{\langle \mathbf{g}^\varepsilon, \mathbf{u}_n\rangle}{\sigma_n}\mathbf{v}_n. \tag{3.5}$$

A comparison with (2.5) shows the stabilization: the errors are not propagated with σ_n^{-1} but with bounded factors $\sigma_n/(\sigma_n^2 + \alpha)$, which is a filter δ_n:

$$\delta_n = \frac{\sigma_n^2}{\sigma_n^2 + \alpha}. \tag{3.6}$$

For a specific n the regularized solution is written as

$$f_{\alpha,n}^\varepsilon = \delta_n f_n + \delta_n \frac{\varepsilon_n}{\sigma_n} \tag{3.7}$$

where f_n is the exact solution from exact data, see (2.5), and ε_n represents the data errors. This, for one thing, shows that an optimal α should be as small as possible to obtain a solution close to $A^+\mathbf{g}$ ($\delta_n \to 1$), the first term in (3.7). However, α should be as large as possible to reduce the influence of the data error on the solution ($\delta_n \to 0$), represented by the second term. More is said in the next Chapter about choosing a regularization parameter.

General Tikhonov regularization

First order regularization. In his original paper on integral equations of the first kind Tikhonov (1963b) proposed to damp out highly oscillating parts (which are manifestations of the instability) in the approximate solution by including the first derivative into the penalty term:

$$\int_a^b ((A\mathbf{f})(x) - \mathbf{g}(x))^2\, dx + \alpha \int_a^b \left(\mathbf{c}_0(x)\mathbf{f}(x)^2 + \mathbf{c}_1(x)\mathbf{f}'(x)^2\right) dx \tag{3.8}$$

where $\mathbf{c}_i, i = 0, 1$ are strictly positive functions, that is $\mathbf{c}_0(x) > 0$ and $\mathbf{c}_1(x) > 0$. Therefore, the functional

$$J_\alpha(\mathbf{f}) = \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f}\|_{F^1}^2 \tag{3.9}$$

has to be minimized, where A is a compact operator from a real Hilbert space F into a real Hilbert space G, and is also called *first order* Tikhonov regularization. The minimizer $\mathbf{f}_\alpha$ of (3.9) can be shown to be unique, compare Tikhonov (1963b); Groetsch (1984); Kress (1989).

The norm $\|.\|_{F^1}$ is associated with *Sobolev spaces*, which are defined as follows, Kress (1989); Martensen and Ritter (1997).

Definition (Sobolev space). Let $0 \leq p < \infty$. The space $H^p[0, 2\pi]$ of all functions $\mathbf{f} \in L^2[0, 2\pi]$ with the property

$$\sum_{n=-\infty}^{\infty} (1 + n^2)^p |a_n|^2 < \infty \tag{3.10}$$

for the Fourier coefficients a_n of $\mathbf{f}$ is called a *Sobolev space.* ●

The Fourier coefficients of $\mathbf{f}$ are

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{f}(x) e^{-inx} dx$$

and the Fourier series of $\mathbf{f}$ is

$$\sum_{n=-\infty}^{\infty} a_n e^{inx}.$$

From (3.10) one sees that the Sobolev spaces $H^p[0, 2\pi]$ are subspaces of $L^2[0, 2\pi]$. A function $\mathbf{f}$ can only be an element of $H^p[0, 2\pi]$ when the Fourier coefficients a_n decay quickly enough as $|n| \to \infty$. Note that $H^0[0, 2\pi]$ coincides with $L^2[0, 2\pi]$. Furthermore, note that the interval $[0, 2\pi]$ is taken for convenience, the generalization to the interval $[a, b]$ is straight forward.

Higher order regularization. An immediate generalization of first order TR is to consider the constraint

$$\alpha\Omega^{(p)} = \alpha \int_a^b \sum_{i=0}^{p} \mathbf{c}_i(x) \left(\frac{d\mathbf{f}^i(x)}{dx^i}\right)^2 dx, \quad p \in I\!N$$

where $\mathbf{c}_i(x) \in C^i[a, b]$ are given positive functions, Tikhonov (1963a); Groetsch (1984). The condition

$$\alpha\Omega^{(p)} \leq c$$

can be interpreted as a smoothing condition because the norm of the function and of a selected number of its derivatives must be bounded, Schwarz (1979).

Regularization with seminorm

The above regularization techniques all constrain the derivatives in the H^p-norm. Phillips (1962) suggested to penalize only by the L^2-norm of the derivative, that is, to minimize

$$J_\alpha(\mathbf{f}) = \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f}''\|_F^2 \tag{3.11}$$

which differs from general Tikhonov regularization in that the regularization term is a seminorm rather than a norm. Also, general TR, in contrast to (3.11), always contains a term which tends to minimize the mean of the approximate solution, which may be undesirable, Groetsch (1984).

Equation (3.11) may not have a unique solution if $N(A)$ contains a nonzero linear function $\mathbf{f}_s$ as $A\mathbf{f} = A\mathbf{f}_s$ and $\mathbf{f}'' = \mathbf{f}_s''$. Phillips (1962) minimized a discrete version of (3.11) and removed the nonuniqueness by imposing zero boundary conditions on the approximate solution.

Generally, one obtains a regularized solution by minimizing the functional

$$J_\alpha(\mathbf{f}) = \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|L\mathbf{f}\|_F^2, \quad \mathbf{f} \in D(L), \tag{3.12}$$

with L a differential operator. If $N(A) \cap N(L) = \{0\}$ then the minimizer $\mathbf{f}_\alpha$ of (3.12) is unique and satisfies

$$A^*A\mathbf{f}_\alpha + \alpha L^*L\mathbf{f}_\alpha = A^*\mathbf{g} \tag{3.13}$$

see (Groetsch, 1984, Ch. 3) and (Engl *et al.*, 1996, Ch. 8). Note that in the finite dimensional space often A is assumed to have full rank, $\text{rank}(A) = n$. Hence, $N(A) = \{0\}$ and therefore $N(A) \cap N(L) = \{0\}$.

We may rewrite (3.13) as

$$(A^*A - \alpha I)\mathbf{f} + \alpha(L^*L + I)\mathbf{f} = A^*\mathbf{g}.$$

As $L^*L + I$ has a symmetric compact inverse B, Groetsch (1984), this is equivalent to

$$B(A^*A - \alpha I)\mathbf{f} + \alpha\mathbf{f} = BA^*\mathbf{g}$$

which is a Fredholm equation of the second kind with a stable solution.

It is rather straightforward to show that

$$\begin{aligned}\mathbf{x}_\alpha &= (A^TA + \alpha L^TL)^{-1}A^T\mathbf{y} \\ &= X\begin{pmatrix} D_\alpha\Sigma_P^+ & 0 \\ 0 & I_o \end{pmatrix}U^T\mathbf{y} = \sum_{i=1}^{p}\delta_i\frac{\mathbf{u}_i^T\mathbf{y}}{\sigma_i}\mathbf{x}_i + \sum_{i=p+1}^{n}\mathbf{u}_i^T\mathbf{y}\mathbf{x}_i\end{aligned}$$

using the GSVD of (A, L), Hansen (1990), cf. Appendix B. Here D_α is a diagonal matrix with elements

$$\delta_i = \frac{\gamma_i^2}{\gamma_i^2 + \alpha}, \quad i = 1, \ldots, p. \tag{3.14}$$

Alternatively, this filter can be written as

$$\delta_i = \begin{cases} \gamma_i^2/(\gamma_i^2 + \alpha), & i = 1, \ldots, p \\ \sigma_i, & i = p+1, \ldots, n \end{cases} \tag{3.15}$$

with solution

$$\mathbf{x}_\alpha = \sum_{i=1}^{n}\delta_i\frac{\mathbf{u}_i^T\mathbf{y}}{\sigma_i}\mathbf{x}_i.$$

3.2.2 Mean square error

Regularization error

Because the generalized inverse of (2.2)

$$\mathbf{f} = A^{+}\mathbf{g}$$

is not a continuous operator such a solution is unstable. Replacing this inverse by a continuous approximate solution

$$\mathbf{f}_\alpha = A_\alpha^{+}\mathbf{g}$$

is called regularization (under the condition $\lim_{\alpha\to 0} A_\alpha^{+} = A^{+}$). This regularization, however, introduces a *regularization error.* Obviously the data are not free of errors and we have a *data error* as well. The difference between the solution $\mathbf{f}$ from the error free data $\mathbf{g}$ and the regularized solution $\mathbf{f}_\alpha^\varepsilon$ from erroneous data $\mathbf{g}^\varepsilon$ is:

$$\mathbf{f}_\alpha^\varepsilon - \mathbf{f} = A_\alpha^{+}(\mathbf{g}^\varepsilon - \mathbf{g}) + (A_\alpha^{+} - A^{+})\mathbf{g}$$

where $A_\alpha^{+} = (A^{*}A + \alpha L^{*}L)^{-1}A^{*}$. The first term on the right hand side is called the data error, the second the regularization error, Louis (1989). Later on it is shown that the latter term equals the *bias* as studied in Xu (1992a, 1992b).

Finite dimensional case. Suppose one has a number of measurements $\mathbf{y}$ with weight matrix P and the parameters to be determined $\mathbf{x}$, have a signal weight matrix C. This is nothing but changing (or describing more properly) the metric of the spaces one works with. Minimizing

$$J_\alpha(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_P^2 + \alpha\|\mathbf{x}\|_C^2$$

gives

$$\mathbf{x}_\alpha = (A^{T}PA + \alpha C)^{-1}A^{T}P\mathbf{y}. \tag{3.16}$$

The total error for the finite dimensional case is

$$\mathbf{x}_\alpha^\varepsilon - \mathbf{x} = A_\alpha^{+}(\mathbf{y}^\varepsilon - \mathbf{y}) + (A_\alpha^{+} - A^{+})\mathbf{y} \tag{3.17}$$

or

$$\delta\mathbf{x} = (A^{T}PA + \alpha C)^{-1}A^{T}P\underline{\varepsilon} + \left((A^{T}PA + \alpha C)^{-1} - (A^{T}PA)^{-1}\right)A^{T}P\mathbf{y}.$$

The expectation yields

$$\begin{aligned} E\{\delta\mathbf{x}\} = \Delta\mathbf{x} &= 0 + \left((A^{T}PA + \alpha C)^{-1} - (A^{T}PA)^{-1}\right)A^{T}PA\mathbf{x} \\ &= (A^{T}PA + \alpha C)^{-1}(A^{T}PA + \alpha C - \alpha C)\mathbf{x} - I\mathbf{x} \\ &= -(A^{T}PA + \alpha C)^{-1}\alpha C\mathbf{x} \neq 0. \end{aligned} \tag{3.18}$$

Hence, the expectation of the total error (and the regularization error) is not zero. Equation (3.18) is needed in the discussion on biased estimation.

Error propagation

By means of the total error, which is the sum of the data error and the regularization error, we wish to assess the quality of the regularization methods. It is always possible through proper transformations to consider the standard form, compare Chapter 4 and Section B.1.3. Hence, the total error is derived for this case only. Unless stated otherwise, it is assumed from here on that the measurement errors are normally distributed with equal variance, $\underline{\varepsilon} \sim N(0, \sigma^2 I)$ or $P = \sigma^{-2} I$ and also $C = I$, yielding

$$\mathbf{x}_\alpha = (A^T A + \alpha' I)^{-1} A^T \mathbf{y}$$

with $\alpha' = \alpha\sigma^2$. We will write α instead of α' in the sequel.

Recall the difference between the approximate and true solution (3.17), where the first term on the right-hand side is the data error and the second the regularization error. In spectral form the difference is

$$\begin{aligned} \delta\mathbf{x} &= \sum_{i=1}^{n} \delta_i \frac{\langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i + \sum_{i=1}^{n} (\delta_i - 1) \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i \\ &= \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + \alpha} \langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle \mathbf{v}_i + \sum_{i=1}^{n} \frac{-\alpha}{\sigma_i^2 + \alpha} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i \end{aligned} \tag{3.19}$$

with filter δ_i defined in (3.6).

Equation (3.19) is of course not very useful in practical computations since the difference between $\mathbf{y}^\varepsilon$ and the exact observation $\mathbf{y}$ is involved. Instead we will look at the mean square error (MSE) which is defined as the sum of the trace of the propagated error and the squared bias:

$$MSE = \text{trace}(Q_x) + \Delta\mathbf{x}^T \Delta\mathbf{x}.$$

The MSE is the expectation of the squared distance between the true solution $\mathbf{x}$ and its estimate $\mathbf{x}_\alpha$, compare next Section.

Error propagation applied to (3.4) in finite dimensions, assuming that $Q_y = \sigma^2 I$, leads to

$$\begin{aligned} Q_x &= \sigma^2 (A^T A + \alpha I)^{-1} A^T A (A^T A + \alpha I)^{-1} \\ &= \sigma^2 (V\Sigma^2 V^T + \alpha V V^T)^{-1} V\Sigma^2 V^T (V\Sigma^2 V^T + \alpha V V^T)^{-1} \\ &= \sigma^2 V (\Sigma^2 + \alpha I)^{-1} \Sigma^2 (\Sigma^2 + \alpha I)^{-1} V^T \end{aligned}$$

or $Q_x = \sigma^2 V(\Lambda + \alpha I)^{-1} \Lambda (\Lambda + \alpha I)^{-1} V^T$ where Λ contains the eigenvalues of $A^T A$. The trace of this matrix is

$$\text{trace}(Q_x) = \sigma^2 \sum_{i=1}^{n} \frac{\lambda_i}{(\lambda_i + \alpha)^2}$$

see Xu and Rummel (1994a); Bouman (1993). The bias term can be shown to be

$$\Delta\mathbf{x}^T \Delta\mathbf{x} = \sum_{i=1}^{n} \frac{\alpha^2 \langle \mathbf{x}, \mathbf{v}_i \rangle^2}{(\lambda_i + \alpha)^2}.$$

Thus, the MSE is

$$\begin{aligned} MSE &= \sum_{i=1}^{n} \frac{\sigma^2 \lambda_i}{(\lambda_i + \alpha)^2} + \sum_{i=1}^{n} \frac{\alpha^2 \langle \mathbf{x}, \mathbf{v}_i \rangle^2}{(\lambda_i + \alpha)^2} \\ &= \sum_{i=1}^{n} \frac{\sigma^2 \lambda_i + \alpha^2 \langle \mathbf{x}, \mathbf{v}_i \rangle^2}{(\lambda_i + \alpha)^2} \end{aligned} \tag{3.20}$$

compare with (3.19). This is still not a very practical equation since $\mathbf{x}$ is needed. Using $\mathbf{x}_\alpha$ instead of $\mathbf{x}$, one can estimate the MSE, although the bias is underestimated, Xu (1992b).

3.3 Biased estimation

The idea is to add an arbitrary positive-definite matrix to the system of normal equations. This matrix is chosen such that the total error, bias and noise, is minimal. *Biased estimation* as studied by Xu(1992a,1992b) is usually called *ridge regression*, for example Vinod and Ullah (1981). Both terms are used here.

It is shown here that TR and ordinary biased estimation (BE) are equivalent, although they originate from different research fields. Consequently, the underlying ideas are different. The point of departure for TR is the integral equation of the first kind and its approximate solution by a nearby well-posed integral equation. As we will see in Chapter 4 the regularization parameter is chosen such that an a priori noise or signal bound is not violated, or chosen such that there is a certain compromise between data error and regularization error. On the other hand BE directly starts with the normal matrix A^TA. If this matrix has small eigenvalues then the least-squares solution may fail and one tries to get closer to the 'true' solution $\mathbf{x}$ at the expense of some bias. It is tried to minimize the mean square error.

Trying to minimize this error then leads to generalized biased estimation. Multiple regularization parameters are introduced instead of only one. These parameters are arbitrary as long as they produce the smallest error, compare Chapter 4.

3.3.1 Ordinary ridge regression

Principle of the method

Let the unstable least-squares solution of $E\{\mathbf{y}\} = A\mathbf{x}$ be

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{y}.$$

One obtains a stable solution with

$$\mathbf{x}_\alpha = (A^TA + \alpha I)^{-1}A^T\mathbf{y} \tag{3.21}$$

where $\alpha \geq 0$. This method (3.21) is called *biased estimation* or *ridge regression*, see also Golub and van Loan (1996). Comparing equation (3.21) with (3.4) ones sees that the first is the finite dimensional version of the latter.

Introducing the weight matrix for the observations P, equation (3.21) becomes

$$\mathbf{x}_\alpha = (A^TPA + \alpha I)^{-1}A^TP\mathbf{y}. \tag{3.22}$$

Taking the difference between the expectation of the biased estimator, $E\{\mathbf{x}_\alpha\}$, and the exact solution, $\mathbf{x}$, gives the magnitude of the bias, Vinod and Ullah (1981); Xu (1992b)

$$\Delta\mathbf{x} = -(A^TPA + \alpha I)^{-1}\alpha I\mathbf{x} \tag{3.23}$$

which resembles the regularization error (3.18).

Again, let $P = \sigma^{-2}I$. Then the spectral decompositions of

$$\begin{aligned} A &= U\Sigma V^T \\ A^TA &= V\Sigma^2V^T = V\Lambda V^T \end{aligned}$$

yield for (3.22)

$$\mathbf{x}_\alpha = V(\Lambda + \alpha I)^{-1}\Lambda^{1/2}U^T\mathbf{y}$$

since $V^T = V^{-1}$, or

$$\mathbf{x}_\alpha = \sum_{i=1}^{n} \delta_i \frac{\mathbf{u}_i^T\mathbf{y}}{\lambda_i^{1/2}}\mathbf{v}_i, \quad \delta_i = \frac{\lambda_i}{\lambda_i + \alpha} \text{ (cf. 3.6).}$$

Some further properties of the ridge estimator. Hoerl and Kennard (1970) derive some interesting properties of the biased estimator (3.21). Left multiplication of $\mathbf{x}_\alpha$ with $(A^TA)^{-1}(A^TA + \alpha I)$ gives the least-squares estimate $\hat{\mathbf{x}}$ and therefore

$$\begin{aligned} \mathbf{x}_\alpha &= \left(I + \alpha(A^TA)^{-1}\right)^{-1}\hat{\mathbf{x}} \\ &= D\hat{\mathbf{x}} \end{aligned} \tag{3.24}$$

assuming that $E\{\varepsilon\varepsilon^T\} = \sigma^2 I$. The eigenvalues of D are $\delta_i = \lambda_i/(\lambda_i + \alpha)$ where λ_i are the eigenvalues of A^TA. Of course δ_i is again equal to (3.6).

The length of the biased estimator $\mathbf{x}_\alpha$ for $\alpha > 0$ is shorter than the length of $\hat{\mathbf{x}}$:

$$\mathbf{x}_\alpha^T\mathbf{x}_\alpha < \hat{\mathbf{x}}^T\hat{\mathbf{x}}$$

which follows readily from (3.24), or

$$\mathbf{x}_\alpha^T\mathbf{x}_\alpha \leq \frac{\lambda_1}{\lambda_1 + \alpha}\hat{\mathbf{x}}^T\hat{\mathbf{x}}, \quad \alpha > 0. \tag{3.25}$$

From (3.25) and (3.24) it is seen that $D = I$ for $\alpha = 0$ and that D approaches 0 as $\alpha \to \infty$.

Let $\mathbf{x}_s$ be any estimate of the vector $\mathbf{x}$. Then the residual sum of squares can be written as

$$\begin{aligned} \phi &= (\mathbf{y} - A\mathbf{x}_s)^T(\mathbf{y} - A\mathbf{x}_s) \\ &= (\mathbf{y} - A\hat{\mathbf{x}})^T(\mathbf{y} - A\hat{\mathbf{x}}) + (\mathbf{x}_s - \hat{\mathbf{x}})^TA^TA(\mathbf{x}_s - \hat{\mathbf{x}}) \\ &= \phi_{min} + \phi(\mathbf{x}_s) = \phi(\hat{\mathbf{x}}) + \phi_0 \end{aligned}$$

see also (2.6). Contours of constant ϕ are the surfaces of the hyperellipsoids centered at $\hat{\mathbf{x}}$. The value of ϕ is the minimum value plus the value of the quadratic form $(\mathbf{x}_s - \hat{\mathbf{x}})$, Hoerl and Kennard (1970). The biased estimate will therefore give a larger residual sum of squares than the least-squares estimate. However, the worse the conditioning of A^TA, the more $\hat{\mathbf{x}}$ can be expected to be too long, Section 2.2.3, and the further one can move from $\hat{\mathbf{x}}$ without a large increase in the residual sum of squares. In view of (2.9) it seems reasonable that if one moves away from the point where the sum of squares is minimal, the movement should be in the direction which will shorten the length of the regression vector, and this is exactly what $\mathbf{x}_\alpha$ does as shown by Hoerl and Kennard (1970).

The problem of minimizing $\mathbf{x}_s^T\mathbf{x}_s$ subject to

$$(\mathbf{x}_s - \hat{\mathbf{x}})^T A^T A(\mathbf{x}_s - \hat{\mathbf{x}}) = \phi_0 \tag{3.26}$$

is equivalent to

$$J = \mathbf{x}_s^T\mathbf{x}_s + \alpha^{-1}((\mathbf{x}_s - \hat{\mathbf{x}})^T A^T A(\mathbf{x}_s - \hat{\mathbf{x}})^T - \phi_0) = \text{ min.}$$

Then

$$\frac{\partial J}{\partial \mathbf{x}_s} = 2\mathbf{x}_s + \alpha^{-1}(2(A^TA)\mathbf{x}_s - 2(A^TA)\hat{\mathbf{x}}) = 0$$

or

$$\mathbf{x}_s = \mathbf{x}_\alpha = (A^TA + \alpha I)^{-1}A^T\mathbf{y}$$

where α is chosen such that (3.26) is satisfied. In practice it is of course easier to choose an $\alpha \geq 0$ and then compute ϕ_0. The above derivation shows that for a fixed ϕ a single value of $\mathbf{x}_s$ is chosen with minimum length.

Mean square error

Although the biased estimate gives a sum of squares that is larger than that of the least-squares estimate, its distance from $\mathbf{x}$ is smaller than the distance of $\hat{\mathbf{x}}$ from $\mathbf{x}$. The distance from $\mathbf{x}_\alpha$ to $\mathbf{x}$ as expressed by $E\{l^2(k)\}$ is defined as the mean square error. Straightforward application of the expectation operator and (3.24) gives

$$\begin{aligned}
E\{l^2(k)\} &= E\{(\mathbf{x}_\alpha - \mathbf{x})^T(\mathbf{x}_\alpha - \mathbf{x})\} \\
&= E\{(\hat{\mathbf{x}} - \mathbf{x})^T D^T D(\hat{\mathbf{x}} - \mathbf{x})\} + (D\mathbf{x} - \mathbf{x})^T(D\mathbf{x} - \mathbf{x}) \\
&= \sigma^2 \text{trace}(A^TA)^{-1}D^TD + \mathbf{x}^T(D - I)^T(D - I)\mathbf{x} \\
&= \sigma^2\left(\text{trace}(A^TA + \alpha I)^{-1} - \text{trace}(A^TA + \alpha I)^{-2}\alpha\right) + \alpha^2\mathbf{x}^T(A^TA + \alpha I)^{-2}\mathbf{x} \\
&= \sigma^2\sum_{i=1}^{n}\frac{\lambda_i}{(\lambda_i + \alpha)^2} + \alpha^2\mathbf{x}^T(A^TA + \alpha I)^{-2}\mathbf{x}
\end{aligned} \tag{3.27}$$

which equals (3.20). The first term on the right-hand side of (3.27) is the sum of the variances of the parameter estimates, see Section 3.2.2. The second term is the squared distance from $D\mathbf{x}$ to $\mathbf{x}$. Since it equals zero when $\alpha = 0$ it can be considered the square of the bias. As α increases the sum of variances decreases and the bias increases: the first is a monotonic decreasing function of α, while the second is monotonic increasing, compare Figure 3.1. It can be shown that it is always possible to choose an $\alpha > 0$ such that the mean square error is smaller than that of the least-squares estimate, Hoerl and Kennard (1970).

3.3.2 Generalized ridge regression

Principle of the method

Instead of using one ridge parameter α, we could introduce more parameters in order to reduce the total error (see Section 4.3.2):

$$\mathbf{x}_g = V(\Lambda + \Delta)^{-1}\Lambda^{1/2}U^T\mathbf{y} \tag{3.28}$$

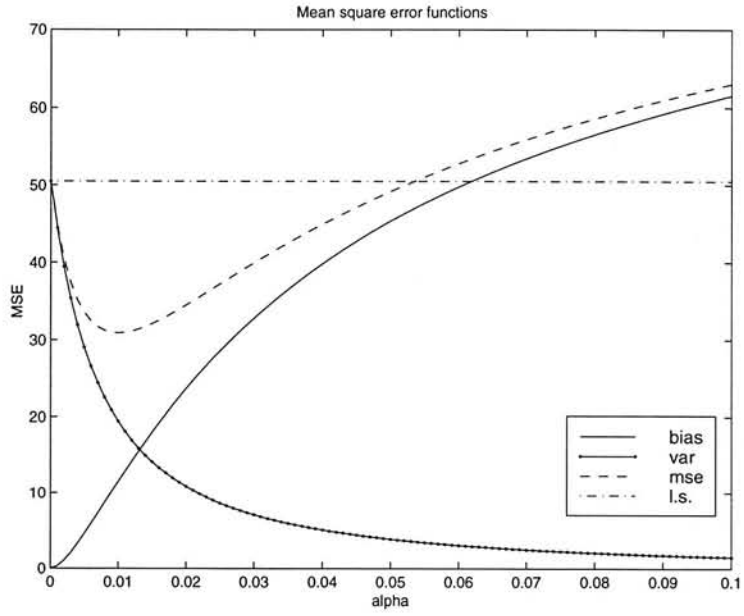

Figure 3.1: *Example of the mean square error, the eigenvalues* λ_i *are* $1/i$, $n = 100$, $\sigma^2 = 10^{-2}$ *and* $\langle \mathbf{x}, \mathbf{v}_i \rangle = 1$.

where Δ is a diagonal matrix with positive elements $\alpha_1, \cdots, \alpha_n$ to be determined. The corresponding filter is

$$\delta_i = \lambda_i/(\lambda_i + \alpha_i). \tag{3.29}$$

Generalized biased estimation (GBE) can also be written in terms of minimizing the sum of noise, and some additional constraint on the signal. Rewrite (3.28) as

$$\begin{aligned} \mathbf{x}_g &= V(\Sigma^2 + \Delta)^{-1}\Sigma U^T \mathbf{y} \\ &= (V\Sigma^2 V^T + V\Delta V^T)^{-1} V\Sigma U^T \mathbf{y} \\ &= (A^T A + M)^{-1} A^T \mathbf{y}. \end{aligned}$$

In general M is a full and positive definite matrix. Therefore, the generalized ridge regression solution minimizes

$$\|A\mathbf{x} - \mathbf{y}\|_P^2 + \|L\mathbf{x}\|_2^2, \quad M = L^T L.$$

However, L can no longer be identified as a differential operator. The elements of Δ are arbitrary (as long as they produce the least error) and L could be a combination of operators. Hence, we can not directly compare generalized ridge regression and Tikhonov regularization. Further discussion of similarities and differences can be found in Section 3.8.

Mean square error

Again the total error, that is the difference between the estimated and true function, should be described by summing the bias and the noise. The propagated error is, Xu and

Rummel (1994a):

$$Q_x = \sigma^2 V(\Lambda + \Delta)^{-1}\Lambda(\Lambda + \Delta)^{-1}V^T.$$

The bias term is

$$\Delta\mathbf{x}\Delta\mathbf{x}^T = V(\Lambda + \Delta)^{-1}\Delta V^T\mathbf{x}\mathbf{x}^T V\Delta(\Lambda + \Delta)^{-1}V^T.$$

Now the mean square error matrix (MSEM) is $Q_x + \Delta\mathbf{x}\Delta\mathbf{x}^T$. In this case the MSE, which is the trace of the MSEM, is:

$$\begin{aligned} MSE &= \text{trace}(Q_x) + \text{trace}(\Delta\mathbf{x}\Delta\mathbf{x}^T) \\ &= \text{trace}(Q_x) + \Delta\mathbf{x}^T\Delta\mathbf{x} \\ &= \sum_{i=1}^{n} \frac{\sigma^2\lambda_i}{(\lambda_i + \alpha_i)^2} + \sum_{i=1}^{n} \frac{\alpha_i^2\langle\mathbf{x}, \mathbf{v}_i\rangle^2}{(\lambda_i + \alpha_i)^2}. \end{aligned} \tag{3.30}$$

3.4 Least-squares collocation

3.4.1 Principle of the method

In geodesy least-squares collocation is the technique that gives the best approximation of any linear functional of the disturbing potential at any place on or above the earth's surface, given a set of observations being linear functionals of the disturbance potential. Best means that any other method would on the average give a larger least-squares difference with the (unknown) true functional. There seems to be therefore no need to consider other (regularization) methods. However, l.s. collocation only deserves the label 'best' in case of unbiasedness assumptions. The ideas of collocation will now be elaborated in more detail, compare for example Moritz (1980) for a complete treatment.

Error free data

Let the disturbing potential T be

$$T = V - U$$

with V the earth's gravitational potential and U the reference potential, for example that of a reference ellipsoid. Suppose observations $\mathbf{y}$ are given which are linear functionals of the disturbance potential T from which one wants to estimate unknowns $\mathbf{x}$, linear functionals of T as well. The linear relation between $\mathbf{y}$ and $\mathbf{x}$ is $\mathbf{y} = A\mathbf{x}$. The condition of minimum error leads to the solution

$$\mathbf{x} = C_{xy}C_{yy}^{-1}\mathbf{y} \tag{3.31}$$

where C_{ij} are signal covariance matrices between i and j. These covariances give the relation between values at different locations. The signal covariances are related as

$$\begin{aligned} C_{xy} &= C_{xx}A^T \\ C_{yy} &= AC_{xx}A^T \end{aligned}$$

see for example Moritz (1980).

A condition the signals $\mathbf{x}$ and $\mathbf{y}$ must fulfil is that their average M over the whole sphere is zero:

$$M\{\mathbf{x}\} = M\{\mathbf{y}\} = 0.$$

Data with observation errors

Instead of error free observations, consider data $\mathbf{y}^\varepsilon = \mathbf{y} + \underline{\varepsilon}$. The signal and the noise are assumed to be uncorrelated and

$$C_{\varepsilon\varepsilon} = \sigma^2 I = P^{-1}$$

which gives

$$C_{y^\varepsilon y^\varepsilon} = C_{yy} + \sigma^2 I.$$

The noise-equivalent of (3.31) then becomes

$$\mathbf{x}^\varepsilon = C_{xy}(C_{yy} + \sigma^2 I)^{-1}\mathbf{y}^\varepsilon \tag{3.32}$$

which is the fundamental formula for least-squares collocation with noise.

The collocation equation (3.32) is related to TR as follows, Rummel *et al.* (1979):

$$\begin{aligned} \mathbf{x}^\varepsilon &= C_{xx}A^T(AC_{xx}A^T + \sigma^2 I)^{-1}\mathbf{y}^\varepsilon \\ &= (\sigma^{-2}A^TA + C_{xx}^{-1})^{-1}\sigma^{-2}A^T\mathbf{y}^\varepsilon \\ &= (A^TA + \sigma^2 C_{xx}^{-1})^{-1}A^T\mathbf{y}^\varepsilon. \end{aligned}$$

The latter minimizes

$$J_{\alpha=1}(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_P^2 + \|\mathbf{x}\|_{C_{xx}^{-1}}^2.$$

When $\alpha = 1$, Tikhonov regularization and least-squares collocation are therefore equal, compare equation (3.16), $C = C_{xx}^{-1}$. The spectral decomposition of the filter therefore is

$$\delta_i = \frac{\lambda_i}{\lambda_i + 1} \tag{3.33}$$

when $C_{xx} = I$.

3.4.2 Committed error

The least-squares collocation solution is equivalent to the least-squares solution of

$$E\{\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_0 \end{pmatrix}\} = \begin{pmatrix} A \\ I \end{pmatrix}\mathbf{x}, \quad D\{\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_0 \end{pmatrix}\} = \begin{pmatrix} P^{-1} & 0 \\ 0 & C_{xx} \end{pmatrix} \tag{3.34}$$

with $\mathbf{y}_0$ zero observations for all unknowns $\mathbf{x}$ and C_{xx} their variance matrix. The least-squares solution of (3.34) yields the l.s. collocation solution and is unbiased under the assumption that $\mathbf{x}$ is expected to be zero. The error variance matrix of $\mathbf{x}$ simply is

$$Q_x = (\sigma^{-2}A^TA + C_{xx}^{-1})^{-1}$$

which follows from error propagation. The trace of Q_x is

$$\text{trace}(Q_x) = \sum_{i=1}^{n} \frac{1}{\lambda_i/\sigma^2 + 1} \tag{3.35}$$

again when $C_{xx} = I$.

3.5 Truncated singular value decomposition

3.5.1 Principle of the method

Let $A \in \mathbb{R}^{m \times n}$ be a rectangular matrix with $m \geq n$. The SVD of A is

$$A = U\Sigma V^T = \sum_{i=1}^{n} \mathbf{u}_i \sigma_i \mathbf{v}_i^T$$

where $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ are matrices with orthonormal columns, $U^T U = V^T V = I_n$, and where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$ with σ_i the singular values, see (B.8). The condition number of A is equal to the ratio σ_1/σ_n.

The truncated singular value decomposition (TSVD) is obtained by approximating A by A_k

$$A_k = \sum_{i=1}^{k} \mathbf{u}_i \sigma_i \mathbf{v}_i^T, \quad k < n.$$

The smallest singular values are left out improving the condition number, i.e. making it smaller.

The TSVD solution is given by

$$\mathbf{x}_k = \sum_{i=1}^{k} \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i$$

or

$$\mathbf{x}_k = \sum_{i=1}^{n} \delta_i \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i$$

with filter δ_i defined as

$$\delta_i = \begin{cases} 1 & \text{for } i = 1, \ldots, k \\ 0 & \text{for } i = k+1, \ldots, n \end{cases} . \tag{3.36}$$

This is therefore an *ideal lowpass filter*, Oppenheim *et al.* (1983). Despite its name, the ideal lowpass filter is not optimal, since it does not produce the smallest error, Rummel (1997).

Using the SVD one obtains the solution $\mathbf{x}$ of $\min \|A\mathbf{x} - \mathbf{y}\|_2$ with smallest norm. Hence, the TSVD solves the problem

$$\min \|A_k \mathbf{x} - \mathbf{y}\|_2 \text{ subject to } \min \|\mathbf{x}\|_2.$$

For an application in geodesy cf. Lerch *et al.* (1993).

3.5.2 Mean square error

The difference between the regularized and true parameters is

$$\begin{aligned} \mathbf{x}_k^\varepsilon - \mathbf{x} &= A_k^+ (\mathbf{y}^\varepsilon - \mathbf{y}) + (A_k^+ - A^+)\mathbf{y} \\ &= \sum_{i=1}^{k} \frac{\langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i + \sum_{i=k+1}^{n} \frac{-\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i. \end{aligned}$$

The data error stays bounded because the large values σ_i^{-1}, $i = k+1, \ldots, n$ are left out. One may have the impression that the regularization error is unbounded, however, the last term on the right-hand side equals $\sum_{i=k+1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i$.

Comparing (3.19) with (3.20) the MSE for TSVD must be

$$\begin{aligned} MSE &= \sum_{i=1}^{k} \frac{\sigma^2}{\sigma_i^2} + \sum_{i=k+1}^{n} \langle \mathbf{x}, \mathbf{v}_i \rangle^2 \\ &= \sum_{i=1}^{n} \left[\frac{\sigma^2 \delta_i}{\sigma_i^2} + (1 - \delta_i) \langle \mathbf{x}, \mathbf{v}_i \rangle^2 \right] \end{aligned} \tag{3.37}$$

with filter δ_i defined in (3.36), see also Xu (1997).

3.6 Generalizations of TSVD

The SVD minimizes $\|\mathbf{x}\|_2$ subject to $\min \|A\mathbf{x} - \mathbf{y}\|_2$, whereas truncated SVD is subject to $\min \|A_k \mathbf{x} - \mathbf{y}\|_2$. A generalization of the latter method is obviously

$$\min \|L\mathbf{x}\|_2 \text{ subject to } \min \|A_k \mathbf{x} - \mathbf{y}\|_2$$

and this leads to truncated GSVD.

Both TSVD and TGSVD use filter factors which become either zero or one. A further generalization is to introduce more smooth filter factors and this leads to damped SVD and damped GSVD.

3.6.1 Truncated GSVD

Recall that the GSVD of the matrix pair (A, L) is

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & I_o \end{pmatrix} X^{-1}, \quad L = V \begin{pmatrix} M & 0 \end{pmatrix} X^{-1} \tag{3.38}$$

with $o = n - p$. It is now easy to show that the regularized generalized inverse A_α^+ associated with the minimization problem $\min \|A\mathbf{x} - \mathbf{y}\|_2 + \alpha \|L\mathbf{x}\|_2$ becomes

$$A_\alpha^+ = X \begin{pmatrix} D & 0 \\ 0 & I_o \end{pmatrix} \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & I_o \end{pmatrix} U^T = X_p D \Sigma^{-1} U_p^T + X_o U_o^T$$

where the filter matrix $D = \mathrm{diag}(\delta_i) \in \mathbb{R}^{p \times p}$ has elements

$$\delta_i = \frac{\gamma_i^2}{\gamma_i^2 + \alpha} \tag{3.39}$$

and $U = (U_p, U_o)$, $X = (X_p, X_o)$ where U_p and X_p have p columns. The truncated GSVD solution $\mathbf{x}_k$ is then obtained by setting k diagonal elements of D, corresponding to the k largest singular values, equal to one and the others equal to zero. Thus, $\mathbf{x}_k = A_k^+ \mathbf{y}$ with

$$A_k^+ \equiv X \begin{pmatrix} \Sigma_k^+ & 0 \\ 0 & I_o \end{pmatrix} U^T = X_p \Sigma_k^+ U_p^T + X_o U_o^T$$

with matrix Σ_k^+ defined by

$$\Sigma_k^+ \equiv \text{diag}(0, \ldots, 0, \sigma_{p-k+1}^{-1}, \ldots, \sigma_p^{-1})$$

or $\Sigma_k^+ = D_k \Sigma^{-1}$ with $D_k = \text{diag}(0, \ldots, 0, 1_{p-k+1}, \ldots, 1_p)$. The above definition was first given in Hansen (1989). In terms of filter factors this is of course

$$\delta_i = \begin{cases} 0 & \text{for } i = 1, \ldots, p-k \\ 1 & \text{for } i = p-k+1, \ldots, p \end{cases} . \tag{3.40}$$

The regularized solution $\mathbf{x}_k$ is

$$\begin{aligned} \mathbf{x}_k &= \sum_{i=1}^{p} \delta_i \frac{\mathbf{u}_i^T \mathbf{y}}{\sigma_i} \mathbf{x}_i + \sum_{i=p+1}^{n} (\mathbf{u}_i^T \mathbf{y}) \mathbf{x}_i \\ &= \sum_{i=p-k+1}^{p} \frac{\mathbf{u}_i^T \mathbf{y}}{\sigma_i} \mathbf{x}_i + \sum_{i=p+1}^{n} (\mathbf{u}_i^T \mathbf{y}) \mathbf{x}_i \end{aligned}$$

with filter factors δ_i as defined in (3.40). A remark on notation: be aware that $\mathbf{x}_i$ is the i-th column of X, whereas $\mathbf{x}_k$ is the TGSVD solution with truncation index k. The second term on the right-hand side is the component that lies in the null space of L, Hansen (1997). A slightly different notation of the regularized solution is

$$\mathbf{x}_k = \sum_{i=1}^{n} \delta_i \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{x}_i$$

with

$$\delta_i = \begin{cases} 0 & \text{for } i = 1, \ldots, p-k \\ 1 & \text{for } i = p-k+1, \ldots, p \\ \sigma_i & \text{for } i = p+1, \ldots, n \end{cases} . \tag{3.41}$$

3.6.2 Damped SVD and GSVD

Principle of the method

Instead of the filter factors zero and one as in TSVD and TGSVD, one can also introduce more smooth filter factors δ_i defined as

$$\delta_i = \frac{\sigma_i}{\sigma_i + \sqrt{\alpha}} \text{ (for } L = I_n) \quad \text{and} \quad \delta_i = \frac{\sigma_i}{\sigma_i + \sqrt{\alpha}\mu_i} \text{ (for } L \neq I_n). \tag{3.42}$$

These two methods are known as damped SVD and damped GSVD respectively, Hansen (1997). The latter equation in (3.42) can also be written as

$$\delta_i = \frac{\gamma_i}{\gamma_i + \sqrt{\alpha}} \tag{3.43}$$

since $\gamma_i = \sigma_i / \mu_i$. Note that these filter factors decay slower than the Tikhonov filter factors and therefore introduce less filtering.

Mean square error

The difference between the approximate DSVD solution and true solution is

$$\mathbf{x}_\alpha^\varepsilon - \mathbf{x} = \sum_{i=1}^{n} \frac{1}{\sigma_i + \sqrt{\alpha}} \langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle \mathbf{v}_i + \sum_{i=1}^{n} \frac{-\sqrt{\alpha}}{\sigma_i + \sqrt{\alpha}} \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i.$$

The total error therefore is

$$MSE = \sum_{i=1}^{n} \frac{\sigma^2 + \alpha \langle \mathbf{x}, \mathbf{v}_i \rangle^2}{(\sigma_i + \sqrt{\alpha})^2}. \tag{3.44}$$

3.7 Iteration methods

Several iteration methods exist to solve the normal equation

$$A^* A\mathbf{f} = A^* \mathbf{g}.$$

The idea is to make as many iteration steps as necessary to extract the low order components and to stop before the solution becomes oscillatory due to magnification of data errors. The number of iterations can thus be considered as the regularization parameter.

The following two iteration methods are considered:

- Landweber iteration
- conjugate gradients

The first method is relatively simple and reveals the basic ideas of iteration. The conjugate gradient method is a non-linear method and has been applied by Schuh (1996) to compute a low degree gravity field model.

The combination of for example TR and conjugate gradients is also feasible, that is, to solve $(A^*A + \alpha I)\mathbf{f} = A^*\mathbf{g}$ iteratively, but this is not discussed here, compare Engl *et al.* (1996); Engl (1997).

3.7.1 Landweber iteration

Error free data

The most straightforward iteration method is that of *Landweber*. The idea is to rewrite the equation $A^*A\mathbf{f} = A^*\mathbf{g}$ as

$$\mathbf{f} = \mathbf{f} + (A^*\mathbf{g} - A^*A\mathbf{f})$$

which suggests the iterative method

$$\mathbf{f}_{k+1} = \mathbf{f}_k + (A^*\mathbf{g} - A^*A\mathbf{f}_k). \tag{3.45}$$

As starting value we may take $\mathbf{f}_0 = 0$, see e.g. Engl *et al.* (1996). To guarantee convergence (3.45) is rewritten as

$$\mathbf{f}_{k+1} = \mathbf{f}_k + \beta(A^*\mathbf{g} - A^*A\mathbf{f}_k) \tag{3.46}$$

where β is the relaxation parameter. A condition for β can be found by considering the error in the approximation $\mathbf{e}_k$ (we use Groetsch (1993)):

$$\mathbf{e}_k = \mathbf{f}_k - \mathbf{f}.$$

From $A^*A\mathbf{f} = A^*\mathbf{g}$ and (3.46), it follows that

$$\mathbf{e}_{k+1} = (I - \beta A^*A)\mathbf{e}_k$$

and therefore

$$\mathbf{e}_k = (I - \beta A^*A)^k \mathbf{e}_0. \tag{3.47}$$

The nonzero eigenvalues of A^*A are $\|A\|^2 = \lambda_1 \geq \lambda_2 \geq \cdots$ with corresponding eigenvectors $\{\mathbf{v}_j\}$. The system $\{\mathbf{v}_j\}$ is a complete orthonormal system for $N(A^*A)^\perp = N(A)^\perp$, and, since $\mathbf{e}_0 = -\mathbf{f} \in N(A)^\perp$, $\mathbf{e}_0$ can be expanded in terms of the eigenvectors $\{\mathbf{v}_j\}$. Then from (3.47)

$$\|\mathbf{e}_k\|_G^2 = \sum_{j=1}^{\infty} (1 - \beta\lambda_j)^{2k} |\langle \mathbf{e}_0, \mathbf{v}_j \rangle|^2.$$

If $0 < \beta < 2/\lambda_1$ then

$$|1 - \beta\lambda_j| < 1$$

for all j and, by Bessel's inequality,

$$\sum_{j=1}^{\infty} |\langle \mathbf{e}_0, \mathbf{v}_j \rangle|^2 \leq \|\mathbf{e}_0\|_G^2.$$

Since $|1 - \beta\lambda_j|^{2k} \to 0$ as $k \to \infty$, for each j, we see that $\|\mathbf{e}_k\|_G^2 \to 0$, that is, $\mathbf{f}_k \to A^+\mathbf{g}$.

Data with observation errors

Consider errors in the observations $\mathbf{g}^\varepsilon$ satisfying

$$\|\mathbf{g} - \mathbf{g}^\varepsilon\|_G \leq \varepsilon.$$

Approximations now are

$$\mathbf{f}_{k+1}^\varepsilon = \mathbf{f}_k^\varepsilon + \beta(A^*\mathbf{g}^\varepsilon - A^*A\mathbf{f}_k^\varepsilon). \tag{3.48}$$

The parameter k plays the role of a regularization parameter. This means that there is some final value $k = k(\varepsilon)$ with the property that, if the iteration is terminated at step $k(\varepsilon)$, then

$$\mathbf{f}_{k(\varepsilon)}^\varepsilon \to A^+\mathbf{g} \text{ as } \varepsilon \to 0.$$

Or in other words for smaller and smaller ε, k becomes larger and larger, without giving an unstable solution. To see this, define the "stability error"

$$\mathbf{d}_k^\varepsilon = \mathbf{f}_k^\varepsilon - \mathbf{f}_k$$

hence from (3.46) and (3.48) we deduce

$$\mathbf{d}_{k+1}^\varepsilon = (I - \beta A^*A)\mathbf{d}_k^\varepsilon + \beta A^*(\mathbf{g}^\varepsilon - \mathbf{g}), \quad \mathbf{d}_0^\varepsilon = 0.$$

Since β is chosen such that $\|I - \beta A^*A\| \leq 1$, we have

$$\|\mathbf{d}_{k+1}^{\varepsilon}\|_F \leq \|\mathbf{d}_k^{\varepsilon}\|_F + \beta\|A\|\varepsilon,$$

hence

$$\|\mathbf{d}_k^{\varepsilon}\|_F \leq k\beta\|A\|\varepsilon.$$

Therefore

$$\begin{aligned} \|\mathbf{f}_k^{\varepsilon} - A^+\mathbf{g}\|_F &\leq \|\mathbf{f}_k - A^+\mathbf{g}\|_F + \|\mathbf{f}_k^{\varepsilon} - \mathbf{f}_k\|_F \\ &\leq \|\mathbf{f}_k - A^+\mathbf{g}\|_F + O(k\varepsilon). \end{aligned}$$

Because it was shown above that $\mathbf{f}_k \to A^+\mathbf{g}$, a sufficient condition for regularity of $\mathbf{f}_{k(\varepsilon)}^{\varepsilon}$ is that the iteration number $k = k(\varepsilon)$ satisfies $k\varepsilon \to 0$ as $\varepsilon \to 0$. For example if $k = \varepsilon^{-1/2}$, then $\|\mathbf{f}_k^{\varepsilon} - A^+\mathbf{g}\|_F = O(\sqrt{\varepsilon})$.

The filtering effect of the iteration. Landweber iteration can also be written in terms of filter factors. Rewrite recursion (3.48) as

$$\begin{aligned} \mathbf{f}_{k+1} &= \beta A^*\mathbf{g} + (I - \beta A^*A)\mathbf{f}_k \\ &= \mathbf{b} + G\mathbf{f}_k \end{aligned}$$

with $\mathbf{b} = \beta A^*\mathbf{g}$ and $G = I - \beta A^*A$. If $\mathbf{f}_0 = 0$ the first few iterations are

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{b} \\ \mathbf{f}_2 &= G\mathbf{f}_1 + \mathbf{b} = G\mathbf{b} + \mathbf{b} \\ \mathbf{f}_3 &= G\mathbf{f}_2 + \mathbf{b} = G^2\mathbf{b} + G\mathbf{b} + \mathbf{b}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{f}_k &= \sum_{j=0}^{k-1} G^j\mathbf{b} \\ &= (I - G^k)(I - G)^{-1}\mathbf{b} \end{aligned}$$

where the second equality is obtained by multiplying $\sum_{j=0}^{k-1} G^j$ with $I - G$ which gives $I - G^k$. Rewriting G as

$$G = (I - \beta A^*A) = V(I - \beta\Sigma^2)V^*$$

and using the fact that $V^{-1} = V^*$ one has

$$\begin{aligned} \mathbf{f}_k &= V\left(I - (I - \beta\Sigma^2)^k\right)\Sigma^{-1}U^*\mathbf{g} \\ &= \sum_{n=1}^{\infty}\left(1 - (1 - \beta\sigma_n^2)^k\right)\frac{\langle\mathbf{g}, \mathbf{u}_n\rangle}{\sigma_n}\mathbf{v}_n \\ &= \sum_{n=1}^{\infty}\delta_{k,n}\frac{\langle\mathbf{g}, \mathbf{u}_n\rangle}{\sigma_n}\mathbf{v}_n \end{aligned}$$

with $\delta_{k,n} = 1 - (1 - \beta\sigma_n^2)^k$. For comparison with conjugate gradients the filter δ_k is rewritten as

$$\delta_{k,n} = 1 - \prod_{j=1}^{k}(1 - \beta\sigma_n^2). \tag{3.49}$$

The effect of the filter becomes clear when considering an example. Let $\beta = \sigma_1^{-2}$, that is the inverse of the largest eigenvalue. Then for small k the filter value is approximately one for the large singular values since $\beta\sigma_n^2 \approx 1$. For increasing n the filter factor becomes zero since $\beta\sigma_n^2 \approx 0$. For large k, in contrast, the filter factors are approximately one for all n since the filter then is: $\lim_{k\to\infty} 1 - \gamma^k = 1$, with $0 < \gamma < 1$. Hence, the number of iterations plays the role of the regularization parameter $\alpha = k^{-1}$.

Mean square error

The difference between the true solution $\mathbf{x}$ and the approximate solution $\mathbf{x}_k^\varepsilon$ is

$$\begin{aligned}
\mathbf{x} - \mathbf{x}_k^\varepsilon &= A_k^+(\mathbf{y}^\varepsilon - \mathbf{y}) + (A_k^+ - A)\mathbf{y} \\
&= \sum_{i=1}^{n} \delta_{k,i} \frac{\langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i + \sum_{i=1}^{n} (\delta_{k,i} - 1) \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i \\
&= \sum_{i=1}^{n} \left(1 - (1 - \beta\sigma_i^2)^k\right) \frac{\langle \mathbf{y}^\varepsilon - \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i + \sum_{i=1}^{n} (1 - \beta\sigma_i^2)^k \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i.
\end{aligned}$$

The propagated error and the bias therefore are

$$\operatorname{trace}(Q_x) = \sum_{i=1}^{n} \frac{\sigma^2 \delta_{k,i}^2}{\sigma_i^2}$$

and

$$\Delta\mathbf{x}^T \Delta\mathbf{x} = \sum_{i=1}^{n} (1 - \delta_{k,i})^2 \langle \mathbf{x}, \mathbf{v}_i \rangle^2$$

respectively with σ^2 the variance of unit weight. Hence, the MSE is

$$MSE = \sum_{i=1}^{n} \left[\frac{\sigma^2 \delta_k^2}{\sigma_i^2} + (1 - \delta_{k,i})^2 \langle \mathbf{x}, \mathbf{v}_i \rangle^2 \right]. \tag{3.50}$$

3.7.2 Conjugate gradients (CG)

More generally, iteration methods, such as Landweber iteration, can be written as

$$\mathbf{f}_{k+1} = \mathbf{f}_k + M^{-1}(\mathbf{b} - N\mathbf{f}_k) \tag{3.51}$$

where $\mathbf{b} = A^*\mathbf{g}$, $N = A^*A$. In case of Landweber iteration $M = \beta^{-1}I$. Other choices for M are $M = D$ (Gauss-Jacobi), $M = D - L$ (Gauss-Seidel) and $M = \beta^{-1}D - L$ (Successive over-relaxation), compare Strang (1986); Golub and van Loan (1996); Van Kan and Segal (1993). Here $A^*A = D - L - L^T$, with D diagonal and L strict lower triangular. The term $\mathbf{r}_k = \mathbf{b} - N\mathbf{f}_k$ in (3.51) is called the k-th residual and one has $\mathbf{r}_k = -N\mathbf{e}_k$.

The first few iterations (3.51) are

$$\begin{aligned} \mathbf{f}_0 &= \mathbf{f}_0 \\ \mathbf{f}_1 &= \mathbf{f}_0 + M^{-1}\mathbf{r}_0 \\ \mathbf{f}_2 &= \mathbf{f}_1 + M^{-1}(\mathbf{b} - N\mathbf{f}_0 - NM^{-1}\mathbf{r}_0) \\ &= \mathbf{f}_0 + 2M^{-1}\mathbf{r}_0 - M^{-1}NM^{-1}\mathbf{r}_0. \end{aligned}$$

Hence

$$\mathbf{f}_k \in \mathbf{f}_0 + \operatorname{span}\{M^{-1}\mathbf{r}_0, (M^{-1}N)M^{-1}\mathbf{r}_0, \ldots, (M^{-1}N)^{k-1}M^{-1}\mathbf{r}_0\}.$$

The space

$$K_k(N, \mathbf{r}_0) = \operatorname{span}\{\mathbf{r}_0, N\mathbf{r}_0, \ldots, N^{k-1}\mathbf{r}_0\}$$

is called the *Krylov subspace*. For the above iteration methods, therefore, $\mathbf{f}_k \in \mathbf{f}_0 + K_k(M^{-1}N, M^{-1}\mathbf{r}_0)$.

The idea of CG

Let $M = I$ and $\mathbf{f}_0 = 0$, so $\mathbf{r}_0 = \mathbf{b}$. Then the iterates are elements of the Krylov space

$$\mathbf{f}_k \in K_k(N, \mathbf{r}_0).$$

The CG method also has its iterates in this Krylov space and tries to minimize the distance between the k-th iterate and the true solution, one way or the other. The best one could achieve is to solve the minimization problem

$$\min_{\mathbf{h} \in K_k(N, \mathbf{r}_0)} \|\mathbf{h} - \mathbf{f}\|_F$$

where $\mathbf{f}$ is the solution of $N\mathbf{f} = \mathbf{b}$ and $\mathbf{h}$ is an element of the iteration space. Since this is not possible, we are iterating towards $\mathbf{f}$, another strategy is required. It turns out that by defining the norm

$$\|\mathbf{v}\|_A = \sqrt{\langle \mathbf{v}, A^*A\mathbf{v} \rangle},$$

the problem

$$\min_{\mathbf{h} \in K_k(N, \mathbf{r}_0)} \|\mathbf{h} - \mathbf{f}\|_A$$

is solvable and leads to the CG-method, as is shown below.

Derivation

We start with the method of steepest descent and arrive at the conjugate gradient method via A-conjugate search directions. The material in this section is based on Golub and van Loan (1996); Strang (1986); Press *et al.* (1992); Schuh (1996); Hansen (1997).

Steepest descent. Minimizing the function

$$J(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T N\mathbf{x} - \mathbf{x}^T\mathbf{b} \tag{3.52}$$

where $\mathbf{b} \in \mathbb{R}^n$ and $N \in \mathbb{R}^{n \times n}$ is symmetric positive definite, is achieved by setting $\mathbf{x} = N^{-1}\mathbf{b}$, Golub and van Loan (1996). Thus, minimizing J and solving $N\mathbf{x} = \mathbf{b}$ (or

$A^TA\mathbf{x} = A^T\mathbf{y}$) are equivalent problems (note that we have switched to finite dimensions, this is not essential).

At a current point $\mathbf{x}_k$ the function J decreases most rapidly in the direction of the negative gradient: $-\nabla J(\mathbf{x}_k) = \mathbf{b} - N\mathbf{x}_k$. This is the *steepest descent* since $\nabla J(x)$ gives the direction of fastest increase of J. If the residual

$$\mathbf{r}_k = \mathbf{b} - N\mathbf{x}_k$$

is nonzero, then a positive β exists such that $J(\mathbf{x}_k + \beta\mathbf{r}_k) < J(\mathbf{x}_k)$. Minimizing

$$J_\beta(\mathbf{x}_k + \beta\mathbf{r}_k) = J(\mathbf{x}_k) - \beta\mathbf{r}_k^T\mathbf{r}_k + \frac{1}{2}\beta^2\mathbf{r}_k^T N\mathbf{r}_k$$

gives $\beta_k = \mathbf{r}_k^T\mathbf{r}_k / \mathbf{r}_k^T N\mathbf{r}_k$.

It can be shown that the method of steepest descent always converges, Golub and van Loan (1996). Unfortunately, the rate of convergence may be very slow since it is governed by the ratio $(\lambda_1 - \lambda_n)/(\lambda_1 + \lambda_n)$ which is very close to 1 for ill-posed problems. Geometrically this means that the level curves of J are very elongated hyperellipsoids and minimizing J leads to travelling back and forth across the valley rather than down the valley to the lowest point, Strang (1986); Golub and van Loan (1996). This is visualized in Figure 3.2, where the curved lines are contours, J is constant. Perpendicular to the contour lines is the direction of steepest descent.

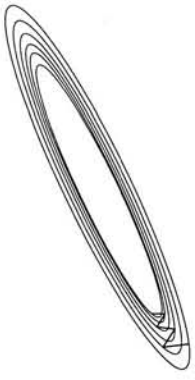

Figure 3.2: *Entering a narrow valley.*

A-conjugate search directions. The disadvantage of the method of steepest descent can be avoided by successive minimization of J along a set of directions $\{\mathbf{p}_1, \mathbf{p}_2, \ldots\}$ not necessarily corresponding to the residuals $\{\mathbf{r}_0, \mathbf{r}_1, \ldots\}$. A new solution $\mathbf{x}_k$ is found by taking information of the shape of the space into account, that is a new search direction $\mathbf{p}_k$ is chosen such that it is N-conjugate or perpendicular to the previous search directions:

$$\langle \mathbf{p}_k, N\mathbf{p}_j \rangle = 0 \ \ \forall\, j \neq k$$

or $\mathbf{p}_k \in \text{span}\{N\mathbf{p}_1, \ldots, N\mathbf{p}_{k-1}\}^\perp$. In most textbooks this is called A-conjugate because A is the symmetric positive definite matrix. A-conjugate is applicable here as well:

$$\langle \mathbf{p}_k, N\mathbf{p}_j \rangle \ = \ \langle \mathbf{p}_k, A^TA\mathbf{p}_j \rangle$$

$$= \langle A\mathbf{p}_k, A\mathbf{p}_j \rangle$$
$$= \langle \mathbf{p}_k, \mathbf{p}_j \rangle_A .$$

Combining steepest descent and A-conjugacy. The new approximate solution vector

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \beta_k \mathbf{p}_k$$

one obtains by choosing the vector $\mathbf{p}_k$ that is A-conjugate to $\mathbf{p}_1, \ldots, \mathbf{p}_{k-1}$ and closest to $\mathbf{r}_{k-1}$.

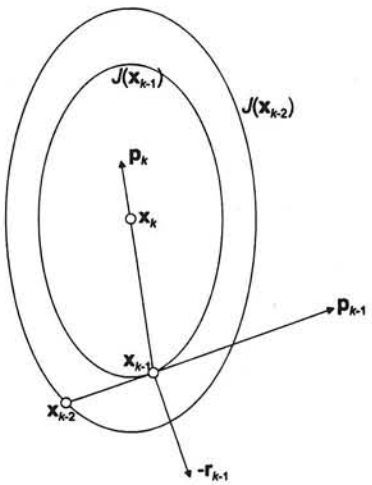

Figure 3.3: *Conjugate gradients (from Schuh (1996)).*

It is the conjugate diameter in the plane spanned by $\mathbf{r}_{k-1}$ and $\mathbf{p}_{k-1}$. The latter is tangent to the ellipse $J(\mathbf{x}_{k-1}) = const.$, and the residual vector $\mathbf{r}_{k-1}$, Figure 3.3:

$$\mathbf{p}_k = \mathbf{r}_{k-1} + \gamma_k \mathbf{p}_{k-1}$$

Schuh (1996). The factor γ_k of the linear combination is determined by the conjugate condition and yields

$$\gamma_k = \frac{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}{\mathbf{r}_{k-2}^T \mathbf{r}_{k-2}}.$$

In Golub and van Loan (1996) it is shown that, using the properties of and the relations between $\mathbf{p}_k$ and $\mathbf{r}_{k-1}$, the conjugate gradient algorithm is as follows:

Algorithm (Conjugate gradients).

$\mathbf{x}_0 = 0 \Rightarrow \mathbf{r}_0 = \mathbf{b}$
$k = 0$
while $\mathbf{r}_k \neq 0$
    $k = k + 1$
    if $k = 1$

$\mathbf{p}_1 = \mathbf{r}_0$
else
$\gamma_k = \mathbf{r}_{k-1}^T \mathbf{r}_{k-1} / \mathbf{r}_{k-2}^T \mathbf{r}_{k-2}$ $(= \|\mathbf{r}_{k-1}\|_2^2 / \|\mathbf{r}_{k-2}\|_2^2)$
$\mathbf{p}_k = \mathbf{r}_{k-1} + \gamma_k \mathbf{p}_{k-1}$
end
$\beta_k = \mathbf{r}_{k-1}^T \mathbf{r}_{k-1} / \mathbf{p}_k^T N \mathbf{p}_k$ $(= \|\mathbf{r}_{k-1}\|_2^2 / \|\mathbf{p}_k\|_A^2)$
$\mathbf{x}_k = \mathbf{x}_{k-1} + \beta_k \mathbf{p}_k$
$\mathbf{r}_k = \mathbf{r}_{k-1} + \beta_k N \mathbf{p}_k$
end

The algorithm as such is not directly applicable since we need to find a proper stopping value k such that the approximate solution is not overwhelmed by oscillations due to the instability.

Filter factors

It can be shown that the iterates $\mathbf{x}_k^\varepsilon$ of the conjugate gradient iteration minimize the residual in the corresponding Krylov subspace:

$$\|\mathbf{y}^\varepsilon - A\mathbf{x}_k^\varepsilon\| = \min\{\|\mathbf{y}^\varepsilon - A\mathbf{x}\| \mid \mathbf{x} \in K_k(A^T(\mathbf{y}^\varepsilon - A\mathbf{x}), A^T A)\}.$$

Consequently, CG requires less iterations than for example the Landweber method, Engl *et al.* (1996). The k-th approximation can be related to the initial values as

$$\mathbf{x}_k = \mathbf{x}_0 - P_{k-1}(A^T A)\mathbf{r}_0$$

where $P_{k-1} \in \prod_{k-1}$ is a polynomial of degree $k-1$ which depends on y, Louis (1989). The CG method is therefore nonlinear.

Hansen (1997) gives the filter factors for the CG method as

$$\delta_{k,i} = 1 - P_k(\sigma_i), \quad i = 1, \ldots, n$$

where P_k is the *Ritz polynomial*:

$$P_k(\sigma) = \prod_{j=1}^{k} \frac{\theta_{k,j}^2 - \sigma^2}{\theta_{k,j}^2}$$

with $\theta_{k,j}^2$ the Ritz values which are the k eigenvalues of $A^T A = N$ restricted to the Krylov subspace $K_k(N, \mathbf{r}_0)$. For small k the Ritz values are approximations to some of the largest eigenvalues and the filter therefore equals one. The filter becomes zero for the smaller eigenvalues because the eigenvalues are negligible with respect to the Ritz values. For increasing k more and more smaller eigenvalues are being approximated resulting in a solution that includes also higher frequencies. A regularized solution one obtains by taking k not too large.

The similarities and differences of the CG and Landweber filter factors are more pronounced when the CG filter is written as

$$\delta_{k,i}(\sigma_i) = 1 - \prod_{j=1}^{k} (1 - \theta_{k,j}^{-2} \sigma_i^2), \quad i = 1, \ldots, n. \tag{3.53}$$

Hansen (1997) states that (3.53) should not be used as such since it is extremely sensitive to rounding errors.

Mean square error

Since the CG method is nonlinear, no simple error propagation exists.

Preconditioned conjugate gradients (PCG)

Although CG has better convergence properties than Landweber iteration, the method only works well on matrices that are either well conditioned or have just a few distinct eigenvalues, Golub and van Loan (1996). The idea behind PCG is to apply CG to the transformed system

$$\bar{N}\bar{\mathbf{x}} = \bar{\mathbf{b}}$$

where $\bar{N} = C^{-1}NC^{-1}$ is well-conditioned, $\bar{\mathbf{x}} = C\mathbf{x}$, $\bar{\mathbf{b}} = C^{-1}\mathbf{b} = C^{-1}A^T\mathbf{y}$, and C is a symmetric positive definite. Golub and van Loan (1996) discuss several preconditioners, and Schuh (1996) applies CG for gravity field determination.

3.8 Comparison of regularization methods

The regularization methods considered in this report are Tikhonov regularization, (generalized) biased estimation, collocation, truncated (generalized) singular value decomposition, damped (generalized) singular value decomposition, Landweber iteration and conjugate gradients. All these methods minimize

$$\|A\mathbf{x} - \mathbf{y}\|_2$$

with constraint

$$\alpha\|L\mathbf{x}\|_2 \leq c$$

for the direct methods and

$$\mathbf{x} \in K_k$$

for the iteration methods. All regularized solutions can be written as a filtered generalized inverse $\mathbf{x} = A_\alpha^+\mathbf{y}$:

$$\mathbf{x} = \sum_{i=1}^{n} \delta_i \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i}\mathbf{v}_i \text{ (for } L = I_n) \quad \text{or} \quad \mathbf{x} = \sum_{i=1}^{n} \delta_i \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\sigma_i}\mathbf{x}_i \text{ (for } L \neq I_n)$$

with filter factors δ_i summarized in Tables 3.1 and 3.2. Note that the filter equation with $\mathbf{v}_i$ should be taken for GBE as well, although $L \neq I_n$. Furthermore, note that the filter for Landweber iteration could also be written as

$$\delta_{k,i} = \beta\lambda_i \sum_{j=0}^{k-1}(1 - \beta\lambda_i)^j = 1 - (1 - \beta\lambda_i)^k.$$

Tikhonov regularization, biased estimation and collocation appear to be equal to a large extent. The differences are: 1) the least-squares collocation solution is unbiased, assuming that the unknowns to be solved for have zero expectation; 2) the 'a priori' information is always a unit matrix in case of biased estimation, no special structure of the solution space, like Kaula's rule in satellite geodesy, is assumed; 3) no regularization parameter needs to be determined using least-squares collocation although sometimes

Table 3.1: *Filter factors for regularization methods with parameter α.*

Method	**Filter** $\delta_{\alpha,i}$	**Eq.**	**Remarks**
TR	$\frac{\lambda_i}{\lambda_i+\alpha}$	(3.6)	$L = I_n$
	$\frac{\gamma_i^2}{\gamma_i^2+\alpha}$	(3.39)	$L \neq I_n$, gsv[a]
BE	$\frac{\lambda_i}{\lambda_i+\alpha}$	(3.6)	$L = I_n$
GBE	$\frac{\lambda_i}{\lambda_i+\alpha_i}$	(3.29)	$L \neq I_n$, L is ?
collocation	$\frac{\lambda_i}{\lambda_i+1}$	(3.33)	$L = I_n$
DSVD	$\frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i}+\sqrt{\alpha}}$	(3.42)	$L = I_n$
DGSVD	$\frac{\gamma_i}{\gamma_i+\sqrt{\alpha}}$	(3.43)	$L \neq I_n$, gsv

[a]gsv = generalized singular values

Table 3.2: *Filter factors for regularization methods with parameter k.*

Method	**Filter** $\delta_{k,i}$	**Eq.**	**Remarks**
TSVD	$\begin{cases} 1 & \text{for } i = 1,\ldots,k \\ 0 & \text{for } i = k+1,\ldots,n \end{cases}$	(3.36)	$L = I_n$
TGSVD	$\begin{cases} 0 & \text{for } i = 1,\ldots,p-k \\ 1 & \text{for } i = p-k+1,\ldots,p \\ \sigma_i & \text{for } i = p+1,\ldots,n \end{cases}$	(3.41)	$L \neq I_n$, gsv
Landweber	$1 - \prod_{j=1}^{k}(1-\beta\lambda_i)$	(3.49)	none
CG	$1 - \prod_{j=1}^{k}(1-\theta_{k,j}^{-2}\lambda_i)$	(3.53)	none

implicit such a parameter is involved. Schwintzer (1990) for example gives an algorithm to determine the a posteriori variance of unit weight $\hat{\sigma}^2$; 4) Tikhonov regularization explicit allows to include constraints on derivatives of the signal.

The main difference between all regularization methods is the filter, which results in different mean square errors, Table 3.3. Since GBE is designed to give the minimum mean square error this might be a very good choice for many problems. However, it is not possible to constrain derivatives of the signal which may be desirable. Furthermore, the method does not give good solutions when parts of the long wavelengths of the signal correspond to small singular values, Bouman and Koop (1997).

For all regularization methods it holds that the regularized solutions have good stability when α is large or k is small. The solutions fit the data well when α is small or k is large. In the next Chapter several methods are discussed that try to find a compromise between data fit and stability.

Table 3.3: *The errors of several regularization methods.*

Method	trace(Q_x)	$\Delta\mathbf{x}^T\Delta\mathbf{x}$	**MSE**	**Eq.**
TR/BE	$\sum_{i=1}^{n} \frac{\sigma^2\lambda_i}{(\lambda_i+\alpha)^2}$	$\sum_{i=1}^{n} \frac{\alpha^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\lambda_i+\alpha)^2}$	$\sum_{i=1}^{n} \frac{\sigma^2\lambda_i + \alpha^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\lambda_i+\alpha)^2}$	(3.20)
GBE	$\sum_{i=1}^{n} \frac{\sigma^2\lambda_i}{(\lambda_i+\alpha_i)^2}$	$\sum_{i=1}^{n} \frac{\alpha_i^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\lambda_i+\alpha_i)^2}$	$\sum_{i=1}^{n} \frac{\sigma^2\lambda_i + \alpha_i^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\lambda_i+\alpha_i)^2}$	(3.30)
collocation	$\sum_{i=1}^{n} \frac{\sigma^2}{\lambda_i+\sigma^2}$	-	$\sum_{i=1}^{n} \frac{\sigma^2}{\lambda_i+\sigma^2}$	(3.34)
TSVD[a]	$\sum_{i=1}^{k} \frac{\sigma^2}{\lambda_i}$	$\sum_{i=k+1}^{n} \langle \mathbf{x}, \mathbf{v}_i\rangle^2$	$\sum_{i=1}^{n} \left[\frac{\sigma^2\delta_i}{\lambda_i} + (1-\delta_i)\langle \mathbf{x}, \mathbf{v}_i\rangle^2\right]$	(3.37)
DSVD	$\sum_{i=1}^{n} \frac{\sigma^2}{(\sqrt{\lambda_i}+\sqrt{\alpha})^2}$	$\sum_{i=1}^{n} \frac{\alpha\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\sqrt{\lambda_i}+\sqrt{\alpha})^2}$	$\sum_{i=1}^{n} \frac{\sigma^2 + \alpha\langle \mathbf{x}, \mathbf{v}_i\rangle^2}{(\sqrt{\lambda_i}+\sqrt{\alpha})^2}$	(3.44)
Landweber	$\sum_{i=1}^{n} \frac{\sigma^2\delta_k^2}{\lambda_i}$	$\sum_{i=1}^{n} (1-\delta_k)^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2$	$\sum_{i=1}^{n} \left[\frac{\sigma^2\delta_k^2}{\lambda_i} + (1-\delta_k)^2\langle \mathbf{x}, \mathbf{v}_i\rangle^2\right]$	(3.50)

[a]Filter δ_i is defined in (3.36).

4 |

SOME COMPUTATIONAL ASPECTS OF REGULARIZATION METHODS

4.1 Introduction

In the preceding Chapter several regularization methods were discussed as well as their mean square error for fixed α or α_i. It was shown that the regularization methods all try to minimize the residual norm $\|A\mathbf{y} - \mathbf{x}\|$ together with the norm of (derivatives of) the signal $\|L\mathbf{x}\|$. In this Chapter we deal with the problem of computing (an) optimal regularization parameter(s). Furthermore it is much easier to treat regularization in standard form, minimize:

$$J_\alpha(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{x}\|_2^2$$

as is shown by Eldén (1977). Here we discuss the transformation to standard form for direct and iterative methods. Finally it is sometimes possible to define additional side constraints on $\mathbf{x}$, for example when parameters or their sum have to be positive.

4.2 Transformation to standard form

It turns out to be a good idea to distinguish between direct and iteration methods when considering the transformation to standard form, Hansen (1997). First, the transformation for the direct methods is discussed in some detail and then the transformation for iteration methods is summarized.

4.2.1 Direct methods

Consider the problem

$$\min_{\mathbf{x}\in B} \|\mathbf{x}\|_L, \quad B = \{\mathbf{x} \,|\, \|A\mathbf{x} - \mathbf{y}\|_W \text{ is minimum}\}$$

where $\|.\|_L$ and $\|.\|_W$ are the seminorms

$$\|\mathbf{x}\|_L^2 = \mathbf{x}^T L^T L \mathbf{x}, \quad \|\mathbf{y}\|_W^2 = \mathbf{y}^T W^T W \mathbf{y}$$

for some matrices L and W. The solution

$$\mathbf{x} = (WA)^+ W\mathbf{y} \tag{4.1}$$

is unique if $N(WA) \cap N(L) = \{0\}$ or equivalently $(WA)^T WA + L^T L$ is positive definite. Since $W^T W = P$ is the weight matrix of the observations which is positive definite, WA has full column rank n. Hence, $N(WA) = \{0\}$ and the solution is unique. Equation (4.1) is nothing but the weighted least-squares solution, compare Appendix B. Eldén (1982) treats the more general case, WA does not have full column rank.

We may therefore conclude that the weight matrix for the observations poses no additional problems and we can further concentrate on the seminorm L.

As Hansen (1997) argues it is simpler to treat problems in standard form because only one matrix, A, is involved instead of two (A, L). Hence, one would like to have a numerically stable transformation method to rewrite

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{y}\|_2^2 + \alpha \|L\mathbf{x}\|_2^2$$

as

$$\min_{\bar{\mathbf{x}}} \|\bar{A}\bar{\mathbf{x}} - \bar{\mathbf{y}}\|_2^2 + \alpha \|\bar{\mathbf{x}}\|_2^2. \tag{4.2}$$

When L is square and invertible, the transformation is simply $\bar{A} = AL^{-1}, \bar{\mathbf{y}} = \mathbf{y}$, the back-transformation becomes $\mathbf{x}_\alpha = L^{-1}\bar{\mathbf{x}}_\alpha$.

However, when L is the discrete approximation, in the space domain, of some derivative operator it is not square and invertible. Typical examples are $L_1 \in \mathbb{R}^{(n-1)\times n}$ and $L_2 \in \mathbb{R}^{(n-2)\times n}$:

$$L_1 = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & 1 & -1 \\ & & & & 1 & -1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \\ & & & & -1 & 2 & -1 \end{pmatrix}.$$

These matrices are approximations of the first and second derivative operators on a uniform net, Hansen (1989).

Now the transformation involves two QR factorizations. Let $o = n - p$ and $q = m - (n - p)$. First compute the QR factorization of L^T

$$L^T = KR = \begin{pmatrix} K_p & K_o \end{pmatrix} \begin{pmatrix} R_p \\ 0 \end{pmatrix} \tag{4.3}$$

where K is an orthogonal matrix and R is upper triangular, compare Appendix B. Since L has full rank p, its generalized inverse is $L^+ = K_p R_p^{-T}$. Moreover, the columns of K_o form an orthonormal basis for the null space of L, Hansen (1997). For L_1 and L_2 one has

$$N(L_1) = \text{span}\{(1,1,\ldots,1)^T\}, \quad N(L_2) = \text{span}\{(1,1,\ldots,1)^T, (1,2,\ldots,n)^T\}.$$

Secondly, compute the QR factorization of $AK_o \in \mathbb{R}^{m\times o}$

$$AK_o = HT = \begin{pmatrix} H_o & H_q \end{pmatrix} \begin{pmatrix} T_o \\ 0 \end{pmatrix}. \tag{4.4}$$

Then the transformed quantities are given by

$$\begin{aligned} \bar{A} &= H_q^T A L^+ = H_q^T A K_p R_p^{-T} \\ \bar{\mathbf{y}} &= H_q^T \mathbf{y}. \end{aligned}$$

Solving (4.2) gives $\bar{\mathbf{x}}_\alpha$, which is related to $\mathbf{x}_\alpha$ as

$$\mathbf{x}_\alpha = L^+ \bar{\mathbf{x}}_\alpha + K_o T_o^{-1} H_o^T (\mathbf{y} - A L^+ \bar{\mathbf{x}}_\alpha) \tag{4.5}$$

see Hansen (1989, 1997).

Relation of TGSVD with TSVD and the transformation to standard form

One would expect that the SVD of the transformed problem in standard form and the GSVD of the original problem in non-standard form are connected one way or the other. This turns out to be true. The proof of the relations in this Section are given in Hansen (1989).

Let the SVD of the transformed matrix be $\bar{A} = \bar{U}\bar{\Sigma}\bar{V}^T$ and the GSVD as in (3.38). Then

$$\bar{U} = H_q^T U_p E, \quad \bar{\Sigma} = E\Sigma M^{-1} E, \quad \bar{V} = VE$$

where $E = \text{antidiag}(1,\ldots,1)$ is the $p \times p$ exchange matrix and H_q is as in (4.4). Further, let $\bar{\mathbf{x}}_k$ denote the TSVD of (4.2)

$$\bar{\mathbf{x}}_k = \bar{A}_k^+ \mathbf{y}, \quad \bar{A}_k^+ \equiv \bar{V}\text{diag}(\bar{\sigma}_1^{-1}, \ldots, \bar{\sigma}_k^{-1}, 0, \ldots, 0)\bar{U}^T.$$

Inserting this solution in (4.5) gives the desired one.

In the case of Tikhonov regularization the transformed solution $\bar{\mathbf{x}}_\alpha$ is

$$\bar{\mathbf{x}}_\alpha = \bar{A}_\alpha^+, \quad \bar{A}_\alpha^+ \equiv \bar{V}\text{diag}\left(\frac{\bar{\sigma}_i^2}{\bar{\sigma}_i^2 + \alpha}\right)\bar{\Sigma}^{-1}\bar{U}^T.$$

Again, inserting $\bar{\mathbf{x}}_\alpha$ in (4.5) gives the proper solution.

When L is well-conditioned, one can compute the GSVD of (A, L) stably from the SVD of $\bar{A}$ without performing the complicated GSVD computation:

$$U = \begin{pmatrix} U_p & U_o \end{pmatrix} = \begin{pmatrix} H_p \bar{U} E & U_o \end{pmatrix}, \quad V = \bar{V}E, \quad X = \begin{pmatrix} M^{-1}V^T L \\ H_o^T A \end{pmatrix}^{-1}$$

where the singular values of $\bar{A}$ and the generalized singular values of (A, L) are related as

$$\bar{\sigma}_i = \gamma_{p-i+1} = \frac{\sigma_{p-i+1}}{\mu_{p-i+1}}.$$

With

$$\sigma_i = \frac{\gamma_i}{\sqrt{\gamma_i^2 + 1}}, \quad \mu_i = \frac{1}{\sqrt{\gamma_i^2 + 1}}, \quad i = 1, \ldots, p \tag{4.6}$$

the matrices Σ and M of the GSVD can be computed. The equalities (4.6) follow from $\sigma_i^2 + \mu_i^2 = 1$, $\forall\, i$, Hansen (1989).

4.2.2 Iteration methods

Define the A-weighted generalized inverse of L as follows

$$L_A^+ = X \begin{pmatrix} M^{-1} \\ \\ 0 \end{pmatrix} V^T.$$

Also, define the vector

$$\mathbf{x}_0 = \sum_{i=p+1}^{n} \mathbf{u}_i^T \mathbf{y} \mathbf{x}_i$$

which is the part of $\mathbf{x}_k$ that lies in the null space of L, Hansen (1997). Then the standard form quantities $\bar{A}$ and $\bar{\mathbf{y}}$ are defined as

$$\bar{A} = AL_A^+, \quad \bar{\mathbf{y}} = \mathbf{y} - A\mathbf{x}_0$$

and the transformation back to the general-form setting is

$$\mathbf{x} = L_A^+ \bar{\mathbf{x}} + \mathbf{x}_0.$$

Using the above relations and the fact that the k iterate belongs to the Krylov space K_k, Hansen (1997) shows that

$$\mathbf{x}_k = \sum_{i=1}^{k-1} c_i \left(L_A^+ (L_A^+)^T A^T A \right)^i L_A^+ (L_A^+)^T A^T \mathbf{y} + \mathbf{x}_0$$

with c_i constants.

4.3 Determination of the regularization parameter(s)

All regularization methods involve one or more regularization parameter(s) to be determined. Several methods to choose a single parameter are discussed as well as one method to determine multiple parameters in case of generalized biased estimation. The relation of the different parameter choice rules with the (minimum) mean square error is treated also.

4.3.1 One regularization parameter

The methods to determine a single regularization parameter discussed here are

- quasi-solutions, Ivanov (1962),
- discrepancy principle, Morozov (1984),
- L-curve, Hansen (1992),
- generalized cross validation (GCV), Wahba (1990),
- quasi-optimality, Morozov (1984).

These methods can be divided into two groups, the a posteriori methods and the heuristic methods. The first two parameter choice rules belong to the first group and the last three choice rules to the second. It can be shown for the a posteriori methods that α goes to zero as ε goes to zero, whereas this formally is not the case for the heuristic methods, Engl *et al.* (1996); Engl (1997).

The parameter choice rules are given here with emphasis on Tikhonov regularization. The application of these rules to other regularization methods is given in Section 4.3.3.

A posteriori methods

Quasi-solutions. The method of quasi-solutions is an a posteriori method for the choice of the regularization parameter α: given a perturbed $\mathbf{g}^\varepsilon$ of $\mathbf{g} \in G$, choose α such that

$$\alpha \mathbf{f}_\alpha^\varepsilon + A^* A \mathbf{f}_\alpha^\varepsilon = A^* \mathbf{g}^\varepsilon \tag{4.7}$$

satisfies $\|\mathbf{f}_\alpha^\varepsilon\|_F = c$, where c is an a priori bound on the norm of the exact solution, Kress (1989). The method of quasi-solutions is derived here for Tikhonov regularization, the application to other regularizations is given in Section 4.3.3.

Numerically the regularization parameter can be obtained by *Newton's method* for solving

$$Z(\alpha) = \|\mathbf{f}_\alpha^\varepsilon\|_F^2 - c^2 = 0.$$

Subsequent α's are related as

$$\alpha_{n+1} = \alpha_n - \frac{Z(\alpha_n)}{Z'(\alpha_n)}$$

see Press *et al.* (1992). The derivative of Z is given by

$$Z'(\alpha) = 2\langle \frac{d\mathbf{f}_\alpha^\varepsilon}{d\alpha}, \mathbf{f}_\alpha^\varepsilon \rangle$$

since $\|\mathbf{f}_\alpha^\varepsilon\|^2 = \langle \mathbf{f}_\alpha^\varepsilon, \mathbf{f}_\alpha^\varepsilon \rangle$, and

$$\frac{d\mathbf{f}_\alpha^\varepsilon}{d\alpha} = -(A^*A + \alpha I)^{-1} \mathbf{f}_\alpha^\varepsilon \tag{4.8}$$

as can be derived from (4.7).

Provided that $\|A^+ \mathbf{g}\|_F \leq c$, one has the estimate

$$\alpha c \leq \|A\| \varepsilon$$

which may serve as a starting value for the iteration to find the desired α for which $\|\mathbf{f}_\alpha^\varepsilon\|_F = c$ holds, Kress (1989).

Discrepancy principle. The discrepancy principle is also an a posteriori method to find the regularization parameter: given a perturbed $\mathbf{g}^\varepsilon$ of $\mathbf{g} \in G$ with a known error level $\|\mathbf{g}^\varepsilon - \mathbf{g}\|_G \leq \varepsilon < \|\mathbf{g}^\varepsilon\|_G$, choose α such that $\|A\mathbf{f}_\alpha^\varepsilon - \mathbf{g}^\varepsilon\|_G = \varepsilon$.

The regularization parameter can be obtained by solving

$$Z(\alpha) = \|A\mathbf{f}_\alpha^\varepsilon - \mathbf{g}^\varepsilon\|_G^2 - \varepsilon^2 = 0$$

with Newton's method. Rewriting the above norm as, using (4.7),

$$\begin{aligned} \|\mathbf{g}^\varepsilon - A\mathbf{f}_\alpha^\varepsilon\|_G^2 &= \langle \mathbf{g}^\varepsilon - A\mathbf{f}_\alpha^\varepsilon, \mathbf{g}^\varepsilon - A\mathbf{f}_\alpha^\varepsilon \rangle \\ &= \langle \mathbf{g}^\varepsilon - A\mathbf{f}_\alpha^\varepsilon, \mathbf{g}^\varepsilon \rangle - \langle A^*(\mathbf{g}^\varepsilon - A\mathbf{f}_\alpha^\varepsilon), \mathbf{f}_\alpha^\varepsilon \rangle \\ &= \|\mathbf{g}^\varepsilon\|_G^2 - \langle \mathbf{f}_\alpha^\varepsilon, A^*\mathbf{g}^\varepsilon \rangle - \alpha \|\mathbf{f}_\alpha^\varepsilon\|_F^2 \end{aligned}$$

one obtains

$$Z(\alpha) = \|\mathbf{g}^\varepsilon\|_G^2 - \langle \mathbf{f}_\alpha^\varepsilon, A^*\mathbf{g}^\varepsilon \rangle - \alpha \|\mathbf{f}_\alpha^\varepsilon\|_F^2 - \varepsilon^2$$

and

$$Z'(\alpha) = -\langle \frac{d\mathbf{f}_\alpha^\varepsilon}{d\alpha}, A^*\mathbf{g}^\varepsilon \rangle - \|\mathbf{f}_\alpha^\varepsilon\|_F^2 - 2\alpha \langle \frac{d\mathbf{f}_\alpha^\varepsilon}{d\alpha}, \mathbf{f}_\alpha^\varepsilon \rangle$$

where the derivative $d\mathbf{f}_\alpha/d\alpha$ is given by (4.8).

Provided that $\|\mathbf{g}^\varepsilon\|_G > \varepsilon$ (SNR > 1), one has the estimate

$$\alpha(\|\mathbf{g}^\varepsilon\|_G - \varepsilon) \leq \|A\|^2 \varepsilon$$

which may serve as starting value for the iteration (until $\|A\mathbf{f}_\alpha^\varepsilon - \mathbf{g}^\varepsilon\|_G = \varepsilon$), see Kress (1989); Groetsch (1984).

The discrepancy principle is widely used, Louis (1989), for example, exclusively applies the discrepancy principle as parameter choice rule. It has to be mentioned that often the criterion $\|A\mathbf{f}_\alpha^\varepsilon - \mathbf{g}^\varepsilon\|_G < R\varepsilon$, with $R > 1$ is used.

The method of quasi-solutions and the discrepancy principle are related as follows, Kress (1989):

- For given $c > 0$ minimize the defect $\|A\mathbf{f} - \mathbf{g}\|_G$ subject to the constraint that the norm is bounded by $\|\mathbf{f}\|_F \leq c$.
- For given $\varepsilon > 0$ minimize the norm $\|\mathbf{f}\|_F$ subject to the constraint that the defect is bounded by $\|A\mathbf{f} - \mathbf{g}\|_G \leq \varepsilon$.

Heuristic methods

A disadvantage of the above two methods is the necessity of a priori bounds on either the signal or the measurement error. Dealing with gravity field determination of the earth, some signal models exist such as Kaula's rule, Kaula (1966) or Tscherning-Rapp, Tscherning and Rapp (1974). However, models are always approximate and the power of the models differs from one model to another, e.g. Rapp (1972); Jekeli (1978); Rapp (1979). Consequently, for quasi-solutions the regularization parameter may be too large or too small, resulting in a too smooth or too rough solution.

The information on the noise level may also be unreliable. Typically, the worst-case bound will be a severe overestimation, while the standard deviation might underestimate the true error, Engl *et al.* (1996).

Therefore, it is necessary to consider alternative a posteriori parameter choice rules that avoid knowledge of the noise level or the signal energy, and to determine a regularization parameter on the basis of the actual performance of the regularization method. Examples are the L-curve, GCV and the quasi-optimality criterion. Strictly speaking these heuristic parameter choice rules cannot provide a convergent regularization method, i.e. $\alpha = \alpha(\varepsilon)$ and $\lim_{\varepsilon \to 0} A_\alpha^+ = A^+$, Engl *et al.* (1996). In practice, however, these methods may work well.

L-curve. The L-curve is a plot, for all valid regularization parameters, of the (semi)norm $\|L\mathbf{x}_\alpha^\varepsilon\|_2$ of the regularized solution versus the corresponding residual norm $\|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$. For discrete ill-posed problems it turns out that the L-curve, when plotted in *log-log* scale, has an L-shaped appearance with a distinct corner separating the vertical and horizontal parts of the curve, Hansen (1997), see Figure 4.1. Originally the use of the L-curve was suggested by Lawson and Hanson (1974).

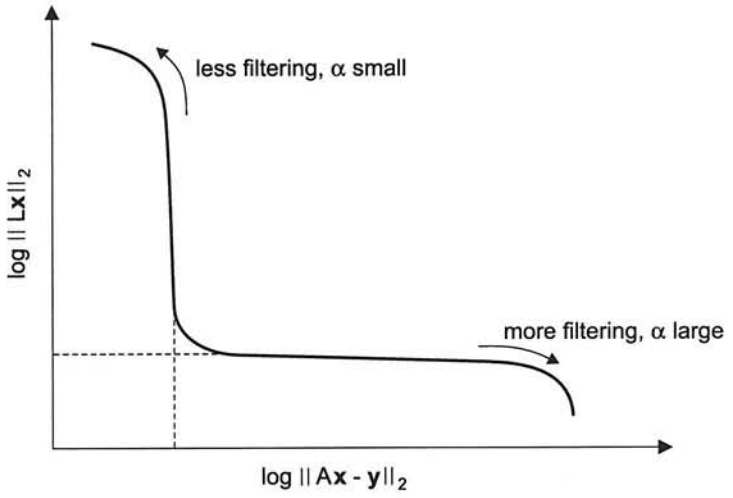

Figure 4.1: *The L-curve in log-log scale (from Hansen (1997)).*

This behaviour can be explained by considering the two error components, that is the perturbation error $\underline{\varepsilon}$ and the regularization error $\Delta\mathbf{x}$. The vertical part of the L-curve corresponds to solutions where $\|L\mathbf{x}_\alpha^\varepsilon\|_2$ is very sensitive to changes in the regularization parameter because the perturbation error $\underline{\varepsilon}$ dominates $\mathbf{x}_\alpha^\varepsilon$ and because $\underline{\varepsilon}$ does not satisfy the discrete Picard condition, Hansen (1997). Stated otherwise, the vertical part corresponds to smaller α. The emphasis of minimizing $J(\alpha)$ is on $\|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$, allowing $\|L\mathbf{x}_\alpha^\varepsilon\|_2$ to become large. The horizontal part of the L-curve corresponds to solutions where the residual norm $\|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$ is most sensitive to the regularization parameter because $\mathbf{x}_\alpha^\varepsilon$ is dominated by the regularization error, as long as $\mathbf{y}$ satisfies the discrete Picard condition (ibid).

The exact location of the corner can be found by maximum curvature. For a continuous regularization parameter α one computes the curvature of the curve

$$(\xi(\alpha), \eta(\alpha))$$

where $\xi(\alpha) = \log \|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$ and $\eta(\alpha) = \log \|\mathbf{x}_\alpha^\varepsilon\|_2$, and finds the point of maximum curvature. When the regularization parameter is discrete, e.g. TSVD, one can approximate the discrete L-curve in *log-log* scale by a 2D spline and compute the point with maximum curvature on the spline. The corner of the L-curve is defined as the point closest to the corner of the spline curve, Hansen (1997).

An alternative for locating the corner of the L-curve is to consider the point $C = (\xi(\alpha_c), \eta(\alpha_c))$ where the L-curve is concave and the tangent at C has slope -1. The concave condition is necessary, because the slope may also be -1 near the endpoints of the curve, compare Figure 4.1. It turns out that point C is a corner of the L-curve if and only if the function

$$\psi(\alpha) = \|\mathbf{x}_\alpha^\varepsilon\|_2 \|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2$$

has a local minimum at $\alpha = \alpha_c$, Regińska (1996); Engl *et al.* (1996).

Although the L-curve method seems to work well in a number of applications, it still lacks a sound mathematical foundation, see (Engl *et al.*, 1996, Section 4.5) and Vogel (1996).

Generalized cross validation. The idea of GCV is that if an arbitrary element y_i of $\mathbf{y}$ is left out, then the corresponding regularized solution should predict this observation well. Moreover, the choice of the regularization parameter should be independent of an orthogonal transformation of $\mathbf{y}$, Wahba (1990); Hansen (1997). This leads to the minimization of:

$$J(\alpha) = \frac{\|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2^2}{(\mathrm{trace}(I_m - AA_\alpha^+))^2}. \tag{4.9}$$

The denominator can be expressed in terms of filter factors:

$$\mathrm{trace}(I_m - AA_\alpha^+) = m - (n-p) - \sum_{i=1}^{p} \delta_i$$

with the filter defined in (3.14), see Hansen (1997).

The range of the operator, $R(A)$, has finite dimension, since the foundation of generalized cross-validation originates from statistical considerations and depends on the assumption that the data perturbation is discrete white noise, Engl *et al.* (1996):

$$E\{\mathbf{y} - \mathbf{y}^\varepsilon\} = 0 \text{ and } E\{(\mathbf{y} - \mathbf{y}^\varepsilon)(\mathbf{y} - \mathbf{y}^\varepsilon)^T\} = \sigma^2 I.$$

This implies that $E\{\|\mathbf{y} - \mathbf{y}^\varepsilon\|_2^2\} = m\sigma^2$, hence $\varepsilon = \sqrt{m}\sigma$.

The assumption of *white* noise is indeed essential as Hansen and O'Leary (1993) show. In case of coloured noise no minimum is found with the GCV method whereas the L-curve works well.

Quasi-optimality. The third heuristic parameter choice rule we discuss is the quasi-optimality method, Morozov (1984); Engl *et al.* (1996); Hansen (1997). This rule also tries to compromise between the data error and the regularization error by minimizing the change in the regularized solution with respect to α. The idea is that if α is too small, $\mathbf{f}_\alpha$ is dominated by the data error which now is sensitive to small changes in α. On the other hand, if α is too large, $\mathbf{f}_\alpha$ is dominated by the regularization error which now is sensitive to small changes in α. The optimal α is obtained when both errors are about equal. Hence, the L-curve and quasi-optimality are alike.

Let us follow the line of Morozov (1984) to derive the quasi-optimality equations. Let $\mathbf{f}_{\alpha_1}$ be the solution of the problem of minimizing

$$J_\alpha(\mathbf{f}, \mathbf{f}_0) \equiv \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f} - \mathbf{f}_0\|_F^2 \tag{4.10}$$

where usually we have $\mathbf{f}_0 = 0$. Since $0 < \alpha < \infty$ the minimizer $\mathbf{f}_{\alpha_1}$ is also called a family of solutions, specifically the *primary family*. As a second step consider the minimization of the same functional (4.10) for $\mathbf{f}_0 = \mathbf{f}_{\alpha_1}$:

$$J_\alpha(\mathbf{f}, \mathbf{f}_{\alpha_1}) = \|A\mathbf{f} - \mathbf{g}\|_G^2 + \alpha\|\mathbf{f} - \mathbf{f}_{\alpha_1}\|_F^2. \tag{4.11}$$

The solutions of (4.11) are denoted $\mathbf{f}_{\alpha_2}$ and are called *two-fold regularized families*, Morozov (1984).

This two-fold regularized family can be expressed in terms of the primary family:

$$\begin{aligned}
\mathbf{f}_{\alpha_1} &= (A^*A + \alpha I)^{-1}(A^*\mathbf{g} + \alpha\mathbf{f}_0) = \mathbf{f}_\alpha \\
\mathbf{f}_{\alpha_2} &= (A^*A + \alpha I)^{-1}(A^*\mathbf{g} + \alpha\mathbf{f}_{\alpha_1}) = (A^*A + \alpha I)^{-1}(A^*\mathbf{g} + \alpha\mathbf{f}_0 + \alpha(\mathbf{f}_{\alpha_1} - \mathbf{f}_0)) \\
&= \mathbf{f}_{\alpha_1} - \alpha(A^*A + \alpha I)^{-1}(\mathbf{f}_0 - \mathbf{f}_{\alpha_1}) \\
&= \mathbf{f}_\alpha - \alpha\frac{d\mathbf{f}_\alpha}{d\alpha}.
\end{aligned} \tag{4.12}$$

The equality

$$\frac{d\mathbf{f}_\alpha}{d\alpha} = (A^*A + \alpha I)^{-1}(\mathbf{f}_0 - \mathbf{f}_\alpha)$$

can be checked by straightforward calculation.

The next step is to choose a mesh in the parameter α, that is, $\alpha_j, j = 0, 1, \ldots, N$ in a neighbourhood of the optimal α. Two consecutive α_j's are almost equal and therefore $\alpha_{j+1} = \tau\alpha_j, \tau \approx 1, \forall\, j$ (and $\tau > 1$). The derivative of $\mathbf{f}_\alpha$ with respect to α can now be approximated by

$$\alpha\frac{d\mathbf{f}_\alpha}{d\alpha} = \alpha\frac{\mathbf{f}_\alpha - \mathbf{f}_{\tau\alpha}}{\alpha - \tau\alpha} = \frac{\mathbf{f}_{\alpha_1} - \mathbf{f}_{\tau\alpha_1}}{1 - \tau}$$

which gives for (4.12)

$$\begin{aligned}
\mathbf{f}_{\alpha_2,j} &= \mathbf{f}_{\alpha,j} - \frac{\mathbf{f}_{\alpha,j} - \mathbf{f}_{\alpha,j+1}}{1 - \tau} \\
&= \frac{\mathbf{f}_{\alpha_1,j}(1 - \tau) - \mathbf{f}_{\alpha_1,j} + \mathbf{f}_{\alpha_1,j+1}}{1 - \tau} \\
&= \frac{\mathbf{f}_{\alpha_1,j+1} - \tau\mathbf{f}_{\alpha_1,j}}{1 - \tau}.
\end{aligned}$$

The above formula shows that the elements of the two-fold regularized family can be approximately computed using the elements of the initial family.

Intuitively, it seems reasonable to choose a value $j = j_0$ for which

$$\|\mathbf{f}_{\alpha_{j+1}} - \mathbf{f}_{\alpha_j}\|_F \tag{4.13}$$

is minimized, which should correspond to balancing the data error and the regularization error. The value α_{j_0} is called the *quasi-optimal value* of the regularization parameter.

Minimizing the distance (4.13) means minimizing $\alpha\|d\mathbf{f}_\alpha/d\alpha\|$ or in finite dimensions

$$\alpha \left\| \frac{d\mathbf{x}_\alpha}{d\alpha} \right\|_2 = \left(\sum_{i=1}^{p} \left(\delta_i(1-\delta_i) \frac{\mathbf{u}_i^T \mathbf{y}}{\gamma_i} \right)^2 \right)^{1/2}$$

evaluated at the (generalized) singular values, Hansen (1997). Note that $d\mathbf{x}_\alpha/d\alpha$ in spectral form simply is $d\delta_i/d\alpha$, compare equations (3.5) and (3.6):

$$\begin{aligned} \alpha \frac{d\delta_i}{d\alpha} &= -\frac{\lambda_i \alpha}{(\lambda_i + \alpha)^2} = -\frac{\lambda_i}{\lambda_i + \alpha} \frac{\lambda_i + \alpha - \lambda_i}{\lambda_i + \alpha} \\ &= -\delta_i(1-\delta_i). \end{aligned}$$

Initial value of α

For the a posteriori parameter choice rules initial values were already given. Press *et al.* (1992) suggest to firstly use

$$\alpha = \frac{\text{trace}(A^T A)}{\text{trace}(L^T L)}$$

which tends to make the two parts of the minimization have comparable weights.

Relation between the parameter choice rules and the mean square error

It is demonstrated in for example Golub *et al.* (1979); Wahba (1990) that the GCV criterion is expected to give a regularization parameter that results in a MSE close to the minimum MSE. Wahba (1990) remarks that the discrepancy principle does not give a minimum MSE but is likely to give too smooth solutions. Kitagawa (1987) showed that the α which minimizes $\alpha\|d\mathbf{x}_\alpha/d\alpha\|$ seeks to minimize the mean square error, compare also Hansen (1992). The relation of the L-curve with the (minimum) MSE is not well solved, although often it gives too smooth solutions, Xu (1997). We expect that the corner of the L-curve is related to the MSE as follows. The horizontal and vertical part correspond to a large change in data error and regularization error respectively. The corner of the L-curve is defined as the point where the change in both errors is about equal. Translated to the quantities trace(Q_x) and $\Delta\mathbf{x}^T\Delta\mathbf{x}$ this means that one seeks α such that the derivative of these two components with respect to α is equal but with opposite sign, and this α does not necessarily lead to a minimum MSE. The L-curve therefore is not expected to give the minimum MSE beforehand, but might be close to it.

4.3.2 Multiple regularization parameters

The GBE solution involves the determination of multiple regularization parameters. Hoerl and Kennard (1970) show that, starting from the l.s. solution, one can iterate towards a set of α_i's with minimum MSE. Later, Hemmerle (1975) found an explicit expression for the set of optimal regularization parameters with respect to the least-squares solution.

Minimum MSE. The set of α_i with minimum MSE is obtained by differentiating the MSE with respect to α_i (see for example Xu and Rummel (1994a)):

$$\frac{\partial MSE}{\partial \alpha_i} = \frac{2\lambda_i(\alpha_i \langle \mathbf{x}, \mathbf{v}_i \rangle^2 - \sigma^2)}{(\lambda_i + \alpha_i)^3}.$$

The minimum is obtained for $\alpha_i = \sigma^2 / \langle \mathbf{x}, \mathbf{v}_i \rangle^2$. With this the MSE now becomes

$$\min(MSE) = \sum_{i=1}^{n} \frac{\sigma^2}{\lambda_i + \sigma^2 / \langle \mathbf{x}, \mathbf{v}_i \rangle^2}.$$

The above equation is not very useful for practical purposes since the x_i that appear are unknown. Having gravity field determination in mind one could for example use approximate coefficients from an existing gravity model such as OSU91A, Rapp *et al.* (1991), instead of the true coefficients $\mathbf{x}$. Iteration gives updated α_i's until the change in the α's is considered to be small enough.

Hoerl and Kennard (1970) suggest to use the least-squares solution as initial value for the iteration:

$$\alpha_{i,0} = \frac{\hat{\sigma}^2}{\langle \hat{\mathbf{x}}, \mathbf{v}_i \rangle^2}$$

where $\hat{\sigma}^2$ and $\hat{\mathbf{x}}$ are least-squares values. However, it may occur in practice that because of numerical instability it is impossible to compute a least-squares solution. Otherwise, one can start the iteration from any (stable) BE solution. The first part on the right-hand side of (3.30) is a continuous, monotonically decreasing function of $\alpha_1, \alpha_2, \ldots, \alpha_n$, whereas the second part on the right-hand side is continuous, monotonically increasing with respect to α_i, Xu and Rummel (1994a). The choice of the initial MSE solution is therefore unimportant.

4.3.3 Explicit application to the regularization methods

The application of the parameter choice rules to the regularization methods is rather straightforward but some additional remarks are necessary. The five parameter choice rules were given above for Tikhonov regularization and can be directly applied to ordinary biased estimation. Applying collocation, no parameter choice has to be made since $\alpha = 1$. However, Schwintzer (1990) does give an algorithm to determine α in the framework of collocation. Schwintzer's idea is as follows. In a least-squares context we have

$$E\{\hat{\mathbf{e}}^T P \hat{\mathbf{e}}\} = m - n \tag{4.14}$$

with $\hat{\mathbf{e}} = \mathbf{y} - A\hat{\mathbf{x}}$ the vector of estimated residuals, $\hat{\mathbf{x}}$ the least-squares estimate, P the weight matrix of the observations $\mathbf{y}$, m the number of observations and n the number

of unknowns, $m > n$. Generally, equation (4.14) does not hold when instead of $\hat{\mathbf{e}}$ the residuals $\mathbf{e}_\alpha = \mathbf{y} - A\mathbf{x}_\alpha$ are used. The correct α now is assumed to be that α for which

$$\mathbf{e}_\alpha^T P \mathbf{e}_\alpha = m - n$$

is true. See also Bouman (1993) for further discussion.

One difficulty associated with the application of the parameter choice rules to TSVD and the iteration methods is that some of the choice rules are defined for a continuous parameter only. As we will see hereafter this problem can be solved.

TSVD

First an a priori method is considered. Suppose that $A^+\mathbf{y} \in R(A^TA)$ and that $\|\mathbf{y} - \mathbf{y}^\varepsilon\|_2 \leq \varepsilon$. The truncation level k is chosen such that

$$\sigma_{k+1}^2 \leq \varepsilon < \sigma_k^2$$

that is, the smallest singular values are above the level of the error variance. For larger k the singular values are below the measurement noise and reveal no signal information. Hence k plays the role of the regularization parameter, $k = k(\varepsilon)$. The difference between the regularized solution and the exact solution is $\|\mathbf{x}_k^\varepsilon - A^+\mathbf{y}\|_2 = O(\sqrt{\varepsilon})$, where $\mathbf{y}$ are the error free observations, Groetsch (1993).

Obviously this is an a priori method when the error level and the singular values can be calculated in advance. If ε is estimated from the actual measurements it becomes an a posteriori method, the measurements then play no further role in determining the regularization parameter, however.

The method of quasi-solutions can be formulated as: minimize $\|A_k\mathbf{x}_k - \mathbf{y}\|$ subject to the constraint $\|\mathbf{x}_k\| \leq c$. Louis (1989) gives the application of the discrepancy principle to TSVD. Suppose that the measurement signal is above the noise. Subtract from the total power in the observed signal the most dominant contributions as given by the largest singular values and singular vectors $\mathbf{u}_i$. At a certain moment the remaining power becomes less than the noise, the SVD should be truncated. In formulas this is: suppose

$$\|\mathbf{y}^\varepsilon\|_2 > R\varepsilon$$

then subtract from $\|\mathbf{y}^\varepsilon\|_2^2$ the term

$$\langle \mathbf{y}^\varepsilon, \mathbf{u}_i \rangle^2$$

until a number $\leq (R\varepsilon)^2$ is found. The corresponding index $i = k$ gives the a posteriori parameter choice

$$\alpha = \alpha(\varepsilon, \mathbf{y}^\varepsilon) = \sigma_k.$$

The computation of the corner of the L-curve by maximum curvature is somewhat problematic since the curve is not continuous for k. Hansen and O'Leary (1993) propose to fit a cubic spline through the discrete point set and use this continuous spline. It is probably easier to use $\psi(\alpha)$ because no derivatives are involved.

The generalized cross validation method causes no difficulties and equation (4.9) may be directly minimized for k.

The quasi-optimality method is not defined for a discrete regularization parameter. However, with some approximations one can use this method for T(G)SVD as well. Let $\alpha = \sigma_k$ and use the approximation

$$\left\| \frac{d\mathbf{x}_\alpha}{d\alpha} \right\|_2 \approx \frac{\|\Delta \mathbf{x}_k\|_2}{|\Delta \sigma_k|}$$

to obtain

$$\alpha \left\| \frac{d\mathbf{x}_\alpha}{d\alpha} \right\|_2 \approx \frac{\langle \mathbf{y}, \mathbf{u}_k \rangle}{|\sigma_k - \sigma_{k-1}|}.$$

Hansen (1997) further introduces the approximation $|\Delta \sigma_k| \approx \sigma_{k-1}$ which is valid when $\sigma_{k-1} \gg \sigma_k$. If the constraint is a differential operator, $L \neq I$, then in the above formula the singular values have to be replaced by generalized singular values.

DSVD

The damped singular value decomposition is a continuous function of α and the parameter choice rules can be applied almost in the same manner as with TR.

Iteration methods

If the discrepancy principle is used to determine the regularization parameter, then the iteration should be terminated when the defect

$$\|A\mathbf{f}_k^\varepsilon - \mathbf{g}^\varepsilon\|_G < R\varepsilon$$

for the first time. Since the defect is monotonically decreasing this is a proper a posteriori parameter choice Louis (1989).

As far as the other choice rules are concerned, the remarks made in the section about TSVD are valid here too, except for quasi-solutions which cannot be applied here.

4.3.4 Approximation of some parameter choice rules

So far the discussion of parameter choice rules fully relied on the SVD. In some applications however, it might be virtually impossible to compute the SVD of A since the dimensions of A are large. In gravity field determination of the Earth, the Moon etc. by satellite tracking the number of unknowns is typically of the order 10^3, while the number of observations is 10^6. Furthermore, usually the design matrix nor the observations are accessible but the inverted matrix $(A^TA + \alpha K)^{-1}$ is, as well as the solution $\mathbf{x}_\alpha^\varepsilon$. If one wishes to assess the quality of these gravity models or wants to compute a regularization parameter not on the basis of trial-and-error, which is common practice, then one can approximate the observations and some of the parameter choice rules as described hereafter.

Suppose the solution $\mathbf{x}_\alpha^\varepsilon$ and the error covariance matrix $(A^TA + \alpha K)^{-1}$ are available. Although this error description is based on assumptions of unbiasedness, the variances of $\mathbf{x}_\alpha^\varepsilon$ as described by $(A^TA + \alpha K)^{-1}$ may serve as a first approximation. Take $\mathbf{x}_\alpha^\varepsilon$ as 'ground truth', that is $\mathbf{x} = \mathbf{x}_\alpha^\varepsilon$ and consider the observations

$$\mathbf{y}^e = A\mathbf{x}^e$$

with $\mathbf{x}^e = \mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is $O(\sigma_i^{-1})$ and σ_i are the singular values. Then the regularized solution reads

$$\begin{aligned} \mathbf{x}_\alpha^e &= (A^TA + \alpha K)^{-1}A^T\mathbf{y} \\ &= (A^TA + \alpha K)^{-1}A^TA\mathbf{x}^e \end{aligned}$$

showing that $\mathbf{y}$ and A are not needed explicitly. The singular values could be obtained by computing the eigenvalues of A^TA which requires little computation time compared to a SVD but the results can be unreliable. Note that $\mathbf{y}^e \in R(A)$ which is not true in general for observations $\mathbf{y}^\varepsilon$, and the least-squares solution is

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{y}^e \\ &= (A^TA)^{-1}A^TA\mathbf{x}^e \\ &= \mathbf{x}^e. \end{aligned}$$

In theory the l.s. solution should therefore exactly be $\hat{\mathbf{x}} = \mathbf{x}^e$. In practice, however, this is not true since the inverse of A^TA is numerically unstable and cannot be computed. Moreover, $\mathbf{x}^e$ has errors behaving like the inverse of the singular values amplify noise.

Since we assume we do not have $\mathbf{y}^\varepsilon$ and A, the norm

$$\|A\mathbf{x}_\alpha^\varepsilon - \mathbf{y}^\varepsilon\|_2 \tag{4.15}$$

needed for several parameter choice rules, cannot be computed. Because of our choice of $\mathbf{y}^e$, we can approximate (4.15) by

$$\begin{aligned} \|A\mathbf{x}_\alpha^e - \mathbf{y}^e\|_2 &= \|A(\mathbf{x}_\alpha^e - \mathbf{x}^e)\|_2 \\ &\leq \|A\|\|\mathbf{x}_\alpha^e - \mathbf{x}^e\|_2 \\ &= \sqrt{\lambda_1}\|\mathbf{x}_\alpha^e - \mathbf{x}^e\|_2 \end{aligned}$$

and λ_1 is the largest eigenvalue of A^TA which can easily be obtained with the *power method*, Kreyszig (1988).

4.4 Regularization with additional side constraint

In addition to the usual problem of minimizing $J_\alpha(\mathbf{x})$ it is possible to define a (linear) side constraint of the form

$$C\mathbf{x} \geq \mathbf{s}$$

where C is an $l \times n$ matrix and $\mathbf{s}$ a known l vector. When for instance x_i is a physical parameter that should be positive then for C a diagonal matrix, $C_{ii} = 1$ and $s_i = 0$, that is, $x_i \geq 0$. Think for example of the light intensity of a pixel in a picture or of the distance between two points. The problem of minimizing $J(\mathbf{x})$, the least-squares problem, with linear inequality constraint and a computation algorithm is treated by Lawson and Hanson (1974). Hemmerle and Brantle (1978) discuss GBE with linear inequality constraint, while (Engl *et al.*, 1996, Section 5.4) characterize the solution as an element of a convex set.

An application in geodesy might be gradiometric analysis. It is well known that the observation of solely the elements of the gravity gradient tensor results in badly determined low order gravitational potential coefficients (coefficients of a spherical harmonic

expansion of the potential), Van Gelderen and Koop (1997). This problem might partially be overcome by fixing the sign of the coefficients. One could for example adopt the sign of an existing gravitational potential model when the SNR of a specific coefficient of this model is larger than some threshold.

4.5 Summary

The implementation of the regularization methods for actual computations becomes easier when it is possible to consider one standard form. This is possible indeed and the transformation to standard form is especially simple for norms weighted with a positive definite matrix, like the error variance matrix of the observations. With some additional effort semi-positive definite matrices, like the matrix corresponding to a seminorm, can be handled as well.

Several methods exist to determine (an) optimal regularization parameter(s). Some of those are directly linked to the minimum mean square error whereas others are not expected to give a minimum mean square error. The explicit application of these parameter choice rules to the regularization methods is rather straightforward although some of the parameter choice rules are defined for continuous methods only.

The major part of this report is devoted to the mathematics of the inverse problems but sometimes it is possible to include additional side constraints on the solution based on the physics of the corresponding problem. Whenever such a situation occurs it is probably wise to use these side constraints.

5 | *EXAMPLE: AIRBORNE GRAVIMETRY*

5.1 Introduction

In the two preceding Chapters several regularization methods were discussed, as well as different methods to choose the regularization parameter(s). It was shown that all regularization methods are in fact low pass filters, and that the filters make the difference between methods. Consequently, the regularized solution and the corresponding mean square error differ from one method to another.

These theoretical comparisons in this Chapter are exampled with airborne gravimetry, that is, scalar gravity is measured at some height h above the earth's surface, for instance in an airplane. A gradiometric example can be found in Bouman and Koop (1998). Also Xu and Rummel (1994b) compare several biased estimators using gradiometric observables.

The outline of the current Chapter is as follows. First, the measurements themselves are discussed as well as their relation with gravity at the earth's surface. Secondly, the Tikhonov regularization method is examined in detail for several parameter choice rules. Then the results for the SVD methods are summarized, since they resemble each other to a great extent.

5.2 Measurement setup and spectral relation

5.2.1 Planar approximation and Fourier series

Consider gravity measurements (magnitude only) at height h above the earth's surface, for example measurements collected in a flying airplane. Then the relation between gravity anomalies at the earth's surface and at height h is given by the convolution equation

$$\Delta \mathbf{g}_h(x, y, h) = P(x, y, h) * \Delta \mathbf{g}_0(x, y)$$

where

$$P(x, y, h) = \frac{1}{2\pi} \frac{h}{(x^2 + y^2 + h^2)^{3/2}}$$

is the Poisson kernel in planar approximation, Hirsch (1996). In the example discussed here the observations lie on a single line, hence $y = 0$, which is equivalent to the assumption that $\Delta\mathbf{g}$ is constant perpendicular to the measurement line (cross-track), Haagmans (1988).

The Fourier series coefficients of $\Delta\mathbf{g}$ are

$$\begin{aligned} \mathcal{F}\{\Delta\mathbf{g}\} = a_k &= \frac{1}{N} \sum_{x=<N>} \Delta\mathbf{g}(x) e^{-jk\frac{2\pi}{N}x}, \quad k = 0, \ldots, N-1 \\ &= \frac{1}{N} \sum_{x=<N>} \Delta\mathbf{g}(x) e^{-jk\omega x}, \quad k = 0, \ldots, N-1 \end{aligned}$$

where ω is the fundamental frequency, hence $\Delta\mathbf{g}$ is assumed to be periodic with period N. The relation between the spectra at $h = 0$ and $h = h$ is, Haagmans (1988)

$$a_k(\Delta\mathbf{g}_h) = e^{-k|\omega|h} a_k(\Delta\mathbf{g}_0) \tag{5.1}$$

or

$$a_k(\Delta\mathbf{g}_0) = e^{k|\omega|h} a_k(\Delta\mathbf{g}_h). \tag{5.2}$$

5.2.2 Measurement synthesis

The ground truth we use is derived from 2^8 or 256 ship measurements in the Indonesian waters. In total three different profiles are used and the end of the profiles are matched with each other so no 'jumps' occur. To achieve periodicity, the values at the beginning and the end of the profile are forced towards zero. The mean is -1.136 mGal, other numbers of interest are listed in Table 5.1. The average spacing between subsequent observations is 1 km, therefore the total length of the profile is approximately 255 km. Figure 5.1 displays the profile.

Table 5.1: *Minimum, maximum and rms of gravity anomalies in mGal.*

	mean	min	max	rms
$\Delta\mathbf{g}_0$	-1.136	-76.5	139.6	57.8
$\Delta\mathbf{g}_h$	-1.136	-71.5	105.9	52.3

The true gravity anomalies at height h are computed as follows. First, the Fourier series of $\Delta\mathbf{g}_0$ is computed. With (5.1) the Fourier coefficients of $\Delta\mathbf{g}_h$ are obtained and the inverse Fourier transform of these coefficients gives $\Delta\mathbf{g}_h$. The mean of $\Delta\mathbf{g}_h$ is -1.136 mGal, the other numbers are listed in Table 5.1 also. To these true anomalies random generated noise is added. The noise has zero mean and a standard deviation of 2.5 mGal. Both the true and noisy observations at height h are displayed in Figure 5.2. The height $h = 2000$ m.

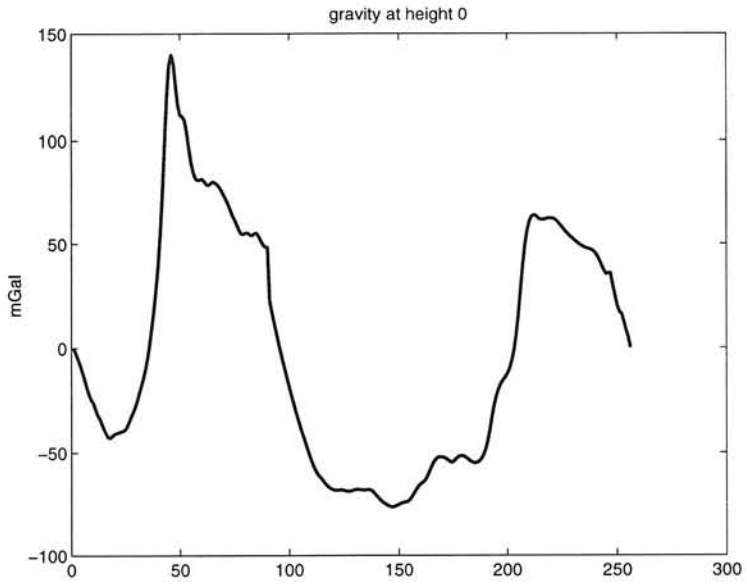

Figure 5.1: *Gravity anomalies along profile at ground level.*

Because the rms difference between true and noisy anomalies is 2.3 mGal (computed) and the rms of the total signal at 2000 m is 52.3 mGal, the best approximation one can expect at ground level is (inverse SNR)

$$\frac{2.3}{52.3} \times 100\% = 4.4\%$$

accurate with respect to rms($\Delta \mathbf{g}_0$). As is shown in Figure 5.3 the least-squares solution certainly does not provide a reasonable answer. The rms difference between the l.s. solution and the true data with respect to the rms of the true solution is 617%.

5.3 Solution with Tikhonov regularization

As we have seen in the previous Section, regularization is obviously necessary because least-squares fails: we are dealing with an ill-posed problem. But how is the least-squares solution computed? Actually, it is the exact inverse of the algorithm sketched above. Thus, compute the Fourier series of $\Delta \mathbf{g}_h^\epsilon$, compute the Fourier coefficients at $h = 0$ through application of the downward continuation factor $e^{k|\omega|h}$ and finally the solution is obtained by inverse Fourier. The observation equation is

$$E\{\mathbf{y}\} = A\mathbf{x}$$

with least-squares solution

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{y}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the Fourier coefficients at $h = 0$ and $h = h$ respectively and A is a diagonal matrix with elements $e^{-k|\omega|h}$.

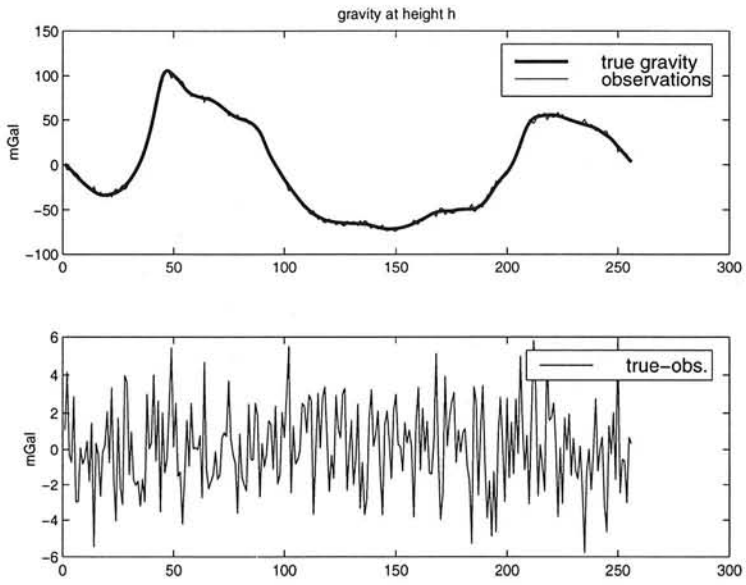

Figure 5.2: *Gravity anomalies along profile at 2000 m.*

Tikhonov regularization is therefore applied in the frequency domain. Three examples will be discussed, that is, regularization with zeroth, first and second derivative constraint.

Regularization with signal constraint

The regularization matrix is an identity matrix, Figures 5.4 - 5.8 show the different regularized solutions. The only difference between the Figures is the determination of the regularization parameter. Shown are the true solution together with the regularized solution as well as their differences. Table 5.2 summarizes these differences in terms of relative norm. The closer a certain percentage is to 4.4%, the better the parameter choice rule performs. Also shown are the various regularization parameters α.

From the Figures and the Table one can conclude that four of the parameter choice rules give approximately the same result (quasi-solution, discrepancy principle, generalized cross validation and quasi-optimality). The regularized solution is acceptable and no further improvement is likely to occur. The L-curve method underestimates α. The regularized solution is a factor of two worse with regard to the other solutions. However, the solution is an order of magnitude better than the least-squares solution (Table 5.2).

A difficulty associated with the discrepancy principle and the quasi-solution method is how to choose $R\|\varepsilon\|$ and $R\|\mathbf{x}\|$ properly. For the current example it holds

$$\|\varepsilon\| = \left(\sum_{i=1}^{256} 2.3^2\right)^{1/2} \approx 37 \text{ mGal}$$

and

$$\|\mathbf{x}\| = \|\Delta\mathbf{g}\| \approx 924 \text{ mGal}.$$

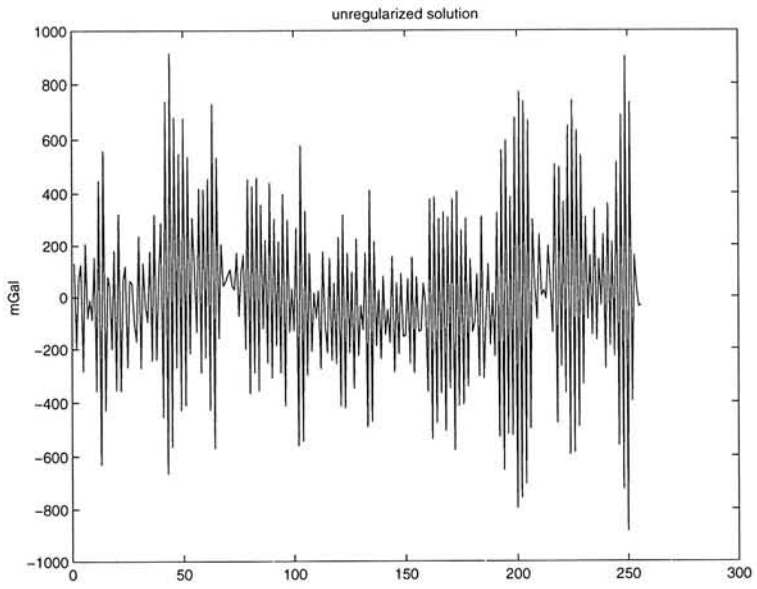

Figure 5.3: *Least-squares solution derived from noisy observations.*

A priori there is not much information about the norm of the signal, whereas the norm of the error is known approximately. Assume that the signal norm is known with some accuracy, even then it is not clear beforehand how to choose R. In the examples above we found that the values of R as given in Table 5.2 yield the best approximation. However, this is only possible when the true solution is known, which usually is not the case. Hence, one has to decide on basis of previous experience, comparison with other solutions or for example visual inspection of solution plots, whether a certain R is acceptable or not. Since these are all subjective methods, we shall not use them in the remainder of this report.

Regularization with first derivative constraint

The regularization with a derivative constraint requires of course the computation of the derivative of the signal. So far only the derivative in the space domain was discussed, but now we need the derivative in the frequency domain since the Fourier coefficients are the estimated signal. Here we use the derivative with respect to the height variable, which is

$$\frac{\partial a_k}{\partial h} = -k|\omega|a_k.$$

Hence, the regularization matrix L^TL has diagonal elements $k^2|\omega|^2$. One slight inconvenience now is that this matrix does not have the proper dimension because k starts at 0, the GSVD cannot be computed with the standard procedure. A way out of trouble is to give the corresponding zero element of L^TL a small positive value (small with respect to the same element of A^TA). We tested values of $10^{-2}, 10^{-4}$ and 10^{-6} and compared the values for the regularization parameter determined with GCV, the L-curve and quasi-optimality. For all three methods the variation in α stays below 7%, which we considered to be satisfactory enough.

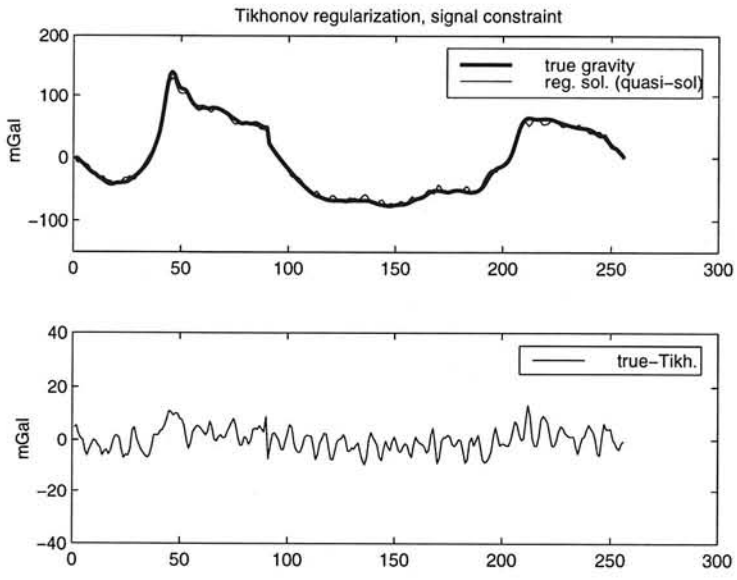

Figure 5.4: *Solution determined with Tikhonov regularization and quasi-solution as parameter choice rule (signal constraint).*

The results of Tikhonov regularization with first derivative constraint are summarized in Table 5.2. All three heuristic methods perform well and apparently the result is not very sensitive with respect to α.

Regularization with second derivative constraint

The second derivative in the frequency domain corresponds to diagonal elements $k^4|\omega|^4$ of L^TL. Again, the three small values replaced the zero diagonal element and now the solutions were the same. The results of Table 5.2 show that the L-curve and GCV perform equally well. The quasi-optimality method did not found a solution, the curve $(\alpha, \|\alpha d\mathbf{x}_\alpha/d\alpha\|)$ is monotonically decreasing. So, the stopping value is arbitrary but larger values would give solutions more and more equal to a straight line. It therefore makes no sense to seek for larger α's.

5.4 Summary of the SVD solutions

TSVD

The truncated (generalized) singular value decomposition works well in all cases except for the signal constraint/L-curve method, and the second derivative/quasi-optimality method, which give too rough and too smooth solutions respectively. Also here tests with $L(1,1) = 10^{-1}, 10^{-2}$ and 10^{-3} gave similar k's.

Table 5.2: *Results for the regularization methods.*

Method	Choice rule	Signal			First derivative			Second derivative			Remark
		α or k	$\frac{\|\Delta\mathbf{g}_0^\varepsilon - \Delta\mathbf{g}_0\|}{\|\Delta\mathbf{g}_0\|}$		α or k	$\frac{\|\Delta\mathbf{g}_0^\varepsilon - \Delta\mathbf{g}_0\|}{\|\Delta\mathbf{g}_0\|}$		α or k	$\frac{\|\Delta\mathbf{g}_0^\varepsilon - \Delta\mathbf{g}_0\|}{\|\Delta\mathbf{g}_0\|}$		
Tikhonov	quasi-sol.	0.197	7.6%	+[a]	-	-		-	-		$R = 10.8$
	discrepancy	0.181	7.6%	+	-	-		-	-		$R = 13.5$
	L-curve	0.047	23.4%	$\pm$	0.485	4.3%	+	1.032	4.1%	+	-
	GCV	0.093	11.6%	$\pm$	0.281	5.3%	+	0.351	5.5%	+	-
	quasi-opt.	0.147	9.0%	+	0.998	5.7%	+	$\pm 10^3$	80.9%	$\uparrow$	-
T(G)SVD	L-curve	64	22.0%	$\pm$	22	4.9%	+	16	6.7%	+	-
	GCV	36	9.0%	+	36	9.0%	+	36	9.0%	+	-
	quasi-opt.	19	5.7%	+	19	5.7%	+	2	77.0%	$\uparrow$	-
D(G)SVD	L-curve	0.006	211%	$\downarrow$	0.260	8.3%	+	1.302	5.0%	+	-
	GCV	0.006	211%	$\downarrow$	0.107	14.7%	$\pm$	0.305	6.3%	+	-
	quasi-opt.	0.203	23.8%	$\pm$	10^3	97.6%	$\uparrow$	$\pm 10^3$	73.5%	$\uparrow$	-

[a]The symbol + denotes a correct solution, $\pm$ is an 'about right' solution (percentage between 10 and 25), $\uparrow$ is a too smooth solution (α too large) and $\downarrow$ indicates that the solution is too rough (α too small).

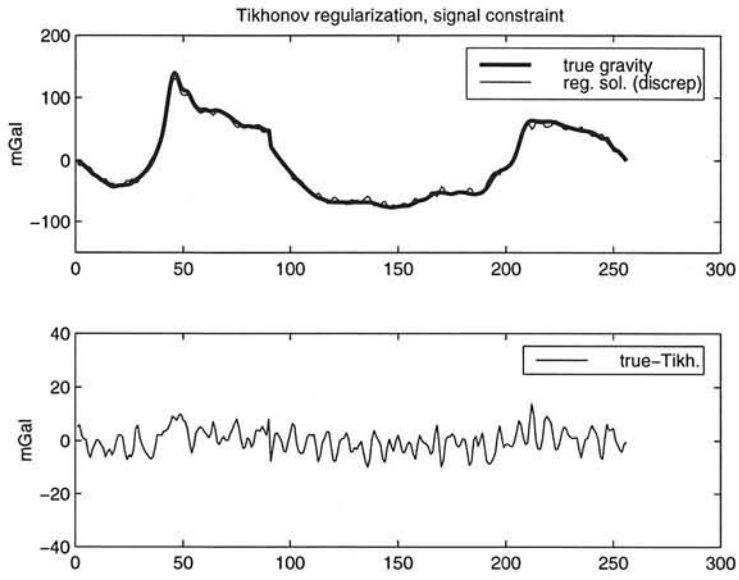

Figure 5.5: *Solution determined with Tikhonov regularization and the discrepancy principle as parameter choice rule (signal constraint).*

DSVD

The damped (generalized) singular value solutions are not satisfactory in general. They become better, however, for higher derivative constraints. One explanation is that the DSVD introduces less filtering with respect to TR and that in this case the signal constraint is not enough. Constraints on the first and second derivative yield smoother solutions compensating for the weaker filtering.

Conclusions

The regularization of airborne gravimetric data with first derivative constraint gives the best solutions, followed closely by the second derivative solutions. The signal constraint does not give many good solutions (using the heuristic choice rules) nor does the DSVD method. Tikhonov regularization and TSVD perform equally well, while TR is maybe slightly better.

The quasi-optimality method gives smoother solutions than the L-curve and GCV method in general. Often the regularization parameter is too large (or k is too small). In many cases the L-curve gives somewhat better results than GCV but the L-curve underestimates α more frequently. We do not recommend to use the a posteriori parameter choice rules because some arbitrary scale factor R has to be chosen which may be difficult in practice.

The most important conclusion probably is that one should not rely on one single regularization method or parameter choice rule but to use several parameter choice rules, different constraints and regularization methods instead. A comparison of the different

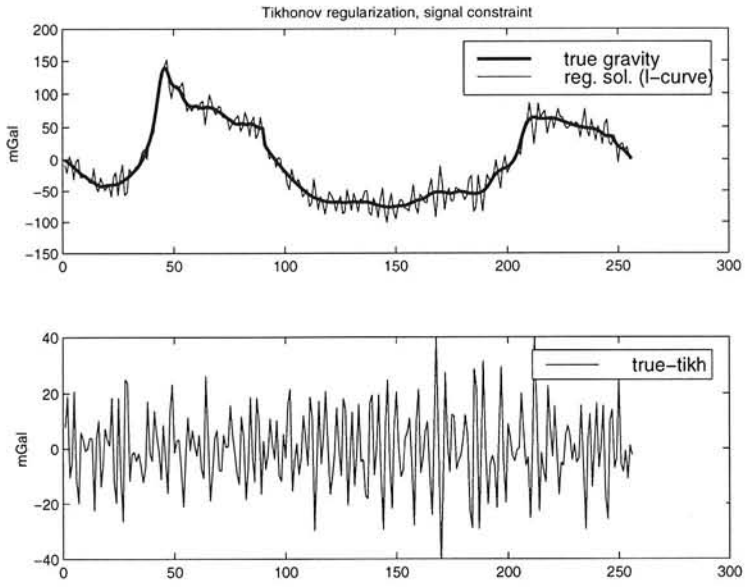

Figure 5.6: *Solution determined with Tikhonov regularization and the L-curve as parameter choice rule (signal constraint).*

solutions may give an idea of those solutions that are definitely too smooth or too rough. Eliminating these leaves us with 'acceptable' solutions.

It has to be stressed that the results in this Chapter are just one example. It illustrates that differences do exist between regularization methods and parameter choice rules. However, in other circumstances (other inverse problems) the above conclusions may not be valid. Also, we did not show results for all regularization methods. It is therefore not legitimate to draw any far-reaching conclusions from the above example.

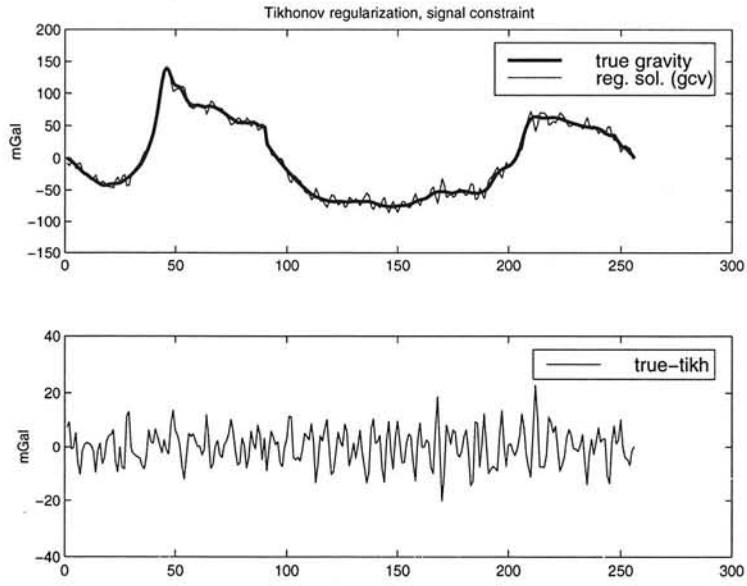

Figure 5.7: *Solution determined with Tikhonov regularization and generalized cross validation as parameter choice rule (signal constraint).*

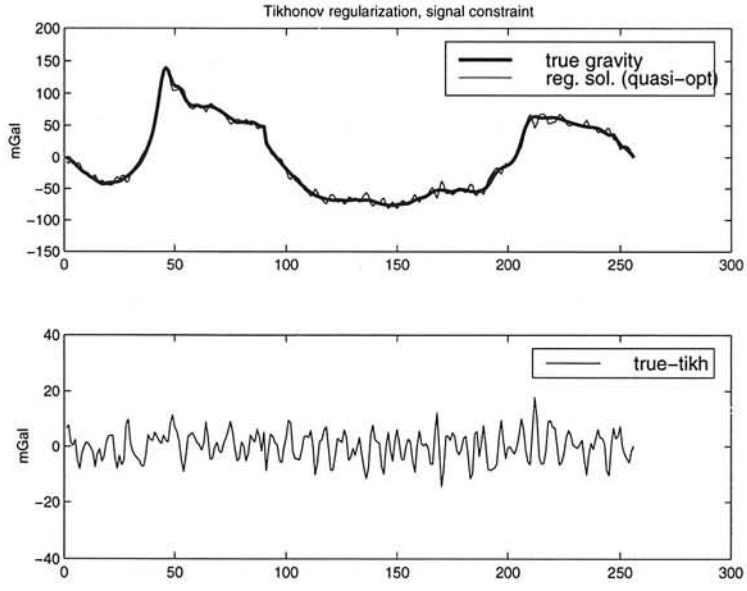

Figure 5.8: *Solution determined with Tikhonov regularization and quasi-optimality as parameter choice rule (signal constraint).*

6 | *CONCLUSIONS AND RECOMMENDATIONS*

Many relations in geodesy and other (earth) sciences can be formulated as Fredholm integral equations of the first kind with a compact operator. The compactness of the operator has the advantage that the integral may be approximated by a finite sum without giving a large discretization error. This is reflected by the fact that the singular values (the spectral representation of the operator) tend to zero for better and better approximations. Consequently, the inverse of this operator, which is associated with inverse problems, becomes unstable. The inverse of the singular values tends to infinity, amplifying noise arbitrarily much.

The solution of inverse problems, therefore, requires some sort of stabilization or regularization. The regularization methods in this report all can be written as a least-squares solution with a filter damping out the high frequencies. The specific form of the filter is distinct from one method to another. Translated to the space domain the regularization corresponds to constraints on the size of (derivatives of) the signal.

The quality of the solution obtained by regularization should not only take into account the data error but the regularization error or bias as well. The expectation of the regularized solution is no longer the 'true' solution because of the filtering. A sufficient measure for the quality description seems to be the mean square error which is the sum of the data error and the bias. This mean square error can be derived for all regularization methods except for conjugate gradients which is a nonlinear method. We have only given the mean square error for regularization with signal constraint but constraints on derivatives of the signal are of interest as well. The mean square error of these solutions can be obtained by straightforward error propagation or by first transforming the original problem to the standard form (with signal constraint).

A disadvantage of these equations is that the computation of the mean square error involves the true unknowns, which is not feasible of course. The mean square error could be approximated by using the regularized unknowns instead of the true unknowns, although

this may lead to too optimistic estimates of the mean square error, see for example Xu (1992a). We did not compare the exact differences of Chapter 5 with those implied by the formulas in Chapter 3. Further research concerning this comparison is of interest.

All regularization methods require the determination of a regularization parameter which is responsible for the balance between bias and data error. A number of parameter choice rules exists and they can be divided in the a posteriori and the heuristic rules. Although the first group has the theoretical advantage that the regularized solution converges to the 'true' solution for decreasing measurement error, they require some (arbitrary) scaling factor on the norm of the solution or the measurement error which are assumed to be known. We therefore prefer the heuristic parameter choice since they do not have this disadvantage. One should always be aware, however, that the solution one obtains is subjective, even when a heuristic parameter choice rule is used (since regularized solutions which are considered 'too smooth' or 'too rough' will not be accepted).

Although the regularization methods are not expected to give the same mean square error and although the heuristic parameter choice rules are not expected to give the same regularization parameter, it is shown that for an airborne gravimetric example several regularization methods and choice rules yield valuable solutions. It is therefore recommended not to rely on one regularization method and parameter choice rule but to use several of them in order to compare solutions.

A | *INTRODUCTION TO FUNCTIONAL ANALYSIS*

Some background on functional analysis may be necessary for reading this report. Inverse problems involve finding unknown functions, inverse mapping etc. This Appendix should make the report more self contained and easier to read for those not too familiar with functional analysis. Geodetic references concerning functional analysis are Meissl (1975, 1976); Tscherning (1978, 1986). Here Kreyszig (1989) is the prime source, but we want to mention Akhiezer and Glazman (1981) and Groetsch (1980) as well.

In the first part definitions of spaces and properties are given with special attention to operators and finite dimension. In the second part the spectral theory of operators is discussed, especially with respect to compact operators. The latter are more simple to deal with. Fortunately the operators usually are compact in geodetic applications, Rummel *et al.* (1979).

Since Kreyszig (1989) is the main reference, his notation is used in Appendix A. It differs from adopted notation elsewhere in this report, but this should cause no difficulties.

A.1 Spaces, definitions and properties

Chapters one to three of Kreyszig (1989) are summarized here. The line followed runs from abstract to more concrete. Proofs are omitted, cf. Kreyszig (1989); Groetsch (1980); Akhiezer and Glazman (1981).

A.1.1 Metric space

Consider the abstract set X, the nature of the elements is left unspecified. They could for example be real numbers or functionals. A distance function on X is defined as follows:

Definition (Metric space, metric). A *metric space* is a pair (X, d), where X is a set and d is a *metric* on X, that is, a function defined on $X \times X$ (the set of all *ordered* pairs of elements of X, i.e. an order can be assigned) such that $\forall\, x, y, z \in X$

$$\begin{aligned} &d \in \mathbb{R},\ 0 \leq d < \infty \\ &d(x, y) = 0 \ \Leftrightarrow\ x = y \\ &d(x, y) = d(y, x) \\ &d(x, y) \leq d(x, z) + d(z, y) \end{aligned}$$

The last property is the triangle inequality, $d(x, y)$ is called the distance from x to y. ●

Examples of metric spaces

Euclidean space $\mathbb{R}^n$. The n-dimensional Euclidean space $\mathbb{R}^n$ is obtained by taking the set of all ordered n-tuples of real numbers written

$$x = (\xi_1, \ldots, \xi_n),\quad y = (\eta_1, \ldots, \eta_n)$$

and the Euclidean metric defined by

$$d(x, y) = \sqrt{(\xi_1 - \eta_1)^2 + \ldots + (\xi_n - \eta_n)^2}.$$

Function space $C[a, b]$. The set X is the set of all real-valued functions $x, y, \ldots$ which are functions of an independent real variable t and are defined and continuous on a given closed interval $[a, b]$. The metric defined by

$$d(x, y) = \max_{t \in [a,b]} |x(t) - y(t)|$$

where max denotes the maximum and $|.|$ the absolute value, leads to the metric space $C[a, b]$. Note that a function becomes a point in a large space.

Continuous mapping and closure

Definition (Continuous mapping). Let $X = (X, d)$ and $Y = (Y, \tilde{d})$ be metric spaces. A mapping $T : X \to Y$ is said to be *continuous at a point* $x_0 \in X$ if for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\tilde{d}(Tx, Tx_0) < \varepsilon \quad \text{for all } x \text{ satisfying} \quad d(x_0, x) < \delta.$$

T is said to be *continuous* if it is continuous at every point of X. ●

Closure. Let M be a subset of a metric space X. Then a point x_0 of X (which may or may not be a point of M) is called an *accumulation point* of M if every ε-neighborhood of x_0 contains at least one point $y \in M$ distinct from x_0. The set consisting of the points of M and the accumulation points of M is called the *closure* of M and is denoted by $\overline{M}$, compare also Figure A.1.

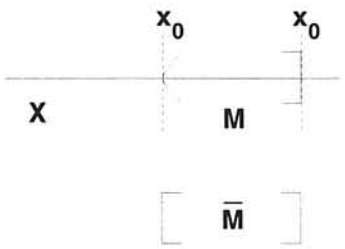

Figure A.1: *Half open set M and its closure $\overline{M}$.*

Definition (Dense set, separable space). A subset M of a metric space X is said to be *dense* in X if

$$\overline{M} = X$$

X is said to be *separable* if it has a countable subset which is dense in X. ●

For example the real line $\mathbb{R}$ is separable since the set Q of all rational numbers is countable and is dense in $\mathbb{R}$.

Convergence, Cauchy sequence and completeness

An important property a metric space may have is that of *completeness*. This means that every *Cauchy sequence* in a space has a limit which is an element of that space, i.e. every Cauchy sequence converges.

Definition (Convergence of a sequence, limit). A sequence (x_n) in a metric space $X = (X, d)$ is said to *converge* if there is an $x \in X$ such that

$$\lim_{n\to\infty} d(x_n, x) = 0$$

x is called the *limit* of x_n and we write

$$\lim_{n\to\infty} x_n = x$$

or $x_n \to x$, x_n *converges* to x. ●

Definition (Cauchy sequence, completeness). A sequence (x_n) in a metric space $X = (X, d)$ is said to be *Cauchy* if for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that

$$d(x_m, x_n) < \varepsilon \quad \text{for every} \quad m, n > N.$$

The space X is said to be *complete* if every Cauchy sequence in X converges, that is, has a limit which is an element of X. ●

For example the real line $\mathbb{R}$ is complete. An example of an incomplete metric space is $X = (0, 1]$ with metric $d(x, y) = |x - y|$, and the sequence x_n, where $x_n = 1/n$ and $n = 1, 2, \ldots$. This is a Cauchy sequence, but it does not converge since the point 0 to which it wants to converge is not a point of X. Another example is the space Q of all rational numbers also with metric $d(x, y) = |x - y|$. The sequences $(2.7, 2.71, 2.718, 2.7182, \ldots)$ and $(3.1, 3.14, 3.141, 3.1415, \ldots)$ want to converge to e and π respectively but this is impossible since these two numbers are not elements of Q.

Theorem (Complete subspace). A subspace M of a complete metric space X is in itself complete if and only if the set M is closed in X.

A.1.2 Normed space

A normed space is a vector space with a metric defined by a norm. Therefore, first a vector space has to be defined.

Definition (vector space). A *vector space over a field* K is a nonempty set X of elements $x, y, \ldots$ (called *vectors*) together with *vector addition* and *multiplication of vectors by scalars*, that is, by elements of K. ●

Here we take $K = \mathbb{R}$, the real numbers.

Vector addition is a mapping $X \times X \to X$, whereas multiplication by scalars is a mapping $K \times X \to X$. For example

$$\begin{aligned} z &= x + y \\ z &= \alpha x \end{aligned}$$

where $x, y, z \in X$ and $\alpha \in \mathbb{R}$ a scalar.

A *subspace* of a vector space X is a nonempty subset Y of X such that for all $y_1, y_2 \in Y$ and all scalars α, β we have $\alpha y_1 + \beta y_2 \in Y$. Hence Y itself is a vector space. A special subspace of X is the *improper subspace* $Y = X$. Every other subspace of $X (\neq \{0\})$ is called *proper*.

A *linear combination* of vectors $x_1, \ldots, x_m$ of a vector space X is an expression of the form

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m$$

where the coefficients $\alpha_1, \ldots, \alpha_m$ are any scalars.

Definition (Linear independence, linear dependence). Linear independence and dependence of a given set M of vectors $x_1, \ldots, x_r$ $(r \geq 0)$ in a vector space X are defined by means of

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_r x_r = 0 \tag{A.1}$$

where $\alpha_1, \ldots, \alpha_r$ are scalars. If the only r-tuple of scalars for which (A.1) holds is $\alpha_1 = \ldots = \alpha_r = 0$, the set M is said to be *linearly independent*, else M is *linearly dependent.* ●

Definition (Finite and infinite dimensional vector spaces). A vector space X is said to be *finite dimensional* if there is a positive integer n such that X contains a linearly independent set of n vectors whereas any set of $n + 1$ or more vectors of X is linearly dependent. n is called the *dimension* of X, written $n = \dim X$. If X is not finite dimensional, it is said to be *infinite dimensional.* ●

If $\dim X = n$, a linearly independent n-tuple of vectors of X is called a *basis for* X. If $\{e_1, \ldots, e_n\}$ is a basis for X, every $x \in X$ has a unique representation as a linear combination of the basis vectors:

$$x = \alpha_1 e_1 + \ldots + \alpha_n e_n.$$

Definition (Normed space, Banach space). A *normed space* X is a vector space with a norm defined on it. A *Banach space* is a complete normed space (complete in the metric defined by the norm). Here a *norm* on a vector space X is a real-valued function on X whose value at $x \in X$ is denoted by

$$\|x\|$$

and which has the properties

$$\begin{aligned} \|x\| &\geq 0 && \text{(A.2)} \\ \|x\| = 0 &\Leftrightarrow x = 0 && \text{(A.3)} \\ \|\alpha x\| &= |\alpha|\|x\| && \text{(A.4)} \\ \|x + y\| &\leq \|x\| + \|y\|. && \text{(A.5)} \end{aligned}$$

Here x and y are arbitrary vectors in X and α is any scalar.

A norm on X defines a metric d on X which is given by

$$d(x, y) = \|x - y\|$$

and is called the metric defined by the norm. ●

A *seminorm* on a vector space X is a mapping $p : X \to \mathbb{R}$ satisfying (A.2), (A.4), (A 5). For (A.3) only the relation from right to left is valid. When the converse is also true, then p is a norm.

An example of a complete normed space (Banach space) is the Euclidean space $\mathbb{R}^n$ with norm defined by

$$\|x\| = \left(\sum_{j=1}^{n} \xi_j^2 \right)^{1/2} = \sqrt{\xi_1^2 + \ldots + \xi_n^2}$$

Thus the norm can be associated with the length of a vector here.

A sequence (x_n) in a normed space X is *convergent* if X contains an x such that

$$\lim_{n \to \infty} \|x_n - x\| = 0.$$

Then we write $x_n \to x$ and call x the *limit* of (x_n). A sequence (x_n) in a normed space X is *Cauchy* if for every $\varepsilon > 0$ there is an N such that

$$\|x_m - x_n\| < \varepsilon \quad \text{for all } m, n > N.$$

Let (x_k) be a sequence in a normed space X. The sequence (s_n) of *partial sums* is

$$s_n = x_1 + x_2 + \ldots + x_n$$

where $n = 1, 2, \ldots$. If (s_n) is convergent, $s_n \to s$, then the *infinite series*

$$s = \sum_{k=1}^{\infty} x_k$$

is said to be *convergent*, s is called the sum. If $\|x_1\| + \|x_2\| + \ldots$ converges, this series is said to be *absolutely convergent.*

If a normed space X contains a sequence (e_n) with the property that for every $x \in X$ there is a unique sequence of scalars (α_n) such that

$$\|x - (\alpha_1 e_1 + \ldots + \alpha_n e_n)\| \to 0 \quad \text{as } n \to \infty$$

then (e_n) is called a *(Schauder) basis* for X. The expansion of x with respect to (e_n) is

$$x = \sum_{k=1}^{\infty} \alpha_k e_k.$$

The above implies that only a separable space X can have a basis and a normed space X which possesses a spanning countable sequence is called separable, Meissl (1976). Also a basis is complete in X (since it is the basis of a Banach space), Akhiezer and Glazman (1981).

Linear operators

In functional analysis metric spaces are considered, and mappings of these spaces. In the case of vector spaces and normed spaces a mapping is called an *operator*.

Definition (Linear operator). A *linear operator* T is an operator such that

(i) the domain $D(T)$ of T is a vector space and the range $R(T)$ lies in a vector space over the same field,

(ii) for all $x, y \in D(T)$ and any scalar α,

$$\begin{aligned} T(x+y) &= Tx + Ty \\ T(\alpha x) &= \alpha Tx \end{aligned}$$

●

The *null space* of T is the set of all $x \in D(T)$ such that $Tx = 0$, denoted as $N(T)$.

Theorem (Range and null space). Let T be a linear operator. Then:

(i) The range $R(T)$ is a vector space.

(ii) If $\dim D(T) = n < \infty$, then $\dim R(T) \leq n$.

(iii) The null space $N(T)$ is a vector space.

●

Theorem (Inverse operator). Let X, Y be vector spaces and $T : D(T) \to Y$ be a linear operator with domain $D(T) \subset X$ and range $R(T) \subset Y$. Then:

(i) The inverse $T^{-1} : R(T) \to D(T)$ exists if and only if

$$Tx = 0 \Rightarrow x = 0.$$

(ii) If T^{-1} exists, it is a linear operator.

(iii) If $\dim D(T) = n < \infty$ and T^{-1} exists, then $\dim R(T) = \dim D(T)$.

●

This means that the inverse of a linear operator exists if and only if the null space of the operator consists of the zero vector only.

Definition (Bounded linear operator). Let X and Y be normed spaces and $T : D(T) \to Y$ a linear operator, where $D(T) \subset X$. The operator T is said to be *bounded* if there is a real number c such that for all $x \in D(T)$,

$$\|Tx\| \leq c\|x\|. \tag{A.6}$$

●

In (A.6) the norms are on Y and X respectively. Formula (A.6) shows that a bounded linear operator maps bounded sets in $D(T)$ onto bounded sets in Y. The smallest number c for which $\|Tx\| \leq c\|x\|$ is true for all x out of the domain of T, is called the norm of T and denoted by $\|T\|$:

$$\|T\| = \sup_{x \in D(T)} \frac{\|Tx\|}{\|x\|}, \quad x \neq 0.^1$$

If a normed space X is finite dimensional, then every linear operator on X is bounded.

Theorem (Continuity and boundedness). Let $T : D(T) \to Y$ be a linear operator, where $D(T) \subset X$ and X, Y are normed spaces. Then T is continuous if and only if T is bounded. ●

Thus, for a linear operator continuity and boundedness become equivalent concepts.

A linear *functional* is an operator whose range lies on the real line $\mathbb{R}$ (or in the complex plane C).

A.1.3 Inner product space

Definition (inner product space, Hilbert space). An *inner product space* is a vector space X with an inner product defined on X. A *Hilbert space* is a complete inner product space. Here, an inner product on X is a mapping of $X \times X$ into the scalar field K of X; that is, with every pair of vectors x and y there is an associated scalar which is written

$$\langle x, y \rangle$$

and is called the *inner product* of x and y, such that for all vectors x, y, z and scalars α

$$\begin{aligned}
\langle x + y, z \rangle &= \langle x, z \rangle + \langle y, z \rangle && \text{(A.7)} \\
\langle \alpha x, y \rangle &= \alpha \langle x, y \rangle && \text{(A.8)} \\
\langle x, y \rangle &= \langle y, x \rangle && \text{(A.9)} \\
\langle x, x \rangle &\geq 0 && \text{(A.10)} \\
\langle x, x \rangle &= 0 \Leftrightarrow x = 0
\end{aligned}$$

[1] Let E be a nonempty subset of $\mathbb{R}$. A number $a \in \mathbb{R}$ is called the supremum of E, written sup E, if (i) a is an upper bound of E; (ii) if $b < a$ then b is not an upper bound of E.

An inner product on X defines a norm on X given by

$$\|x\| = \sqrt{\langle x, x \rangle}$$

and a metric on X given by

$$d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$$

●

Hence, inner product spaces are normed spaces, and Hilbert spaces are Banach spaces. A concept that can be defined in inner product spaces is that of *orthogonality* of functions.

Definition (Orthogonality). An element x of an inner product space X is said to be *orthogonal* to an element $y \in X$ if

$$\langle x, y \rangle = 0$$

x and y are orthogonal, $x \perp y$. The zero vector is orthogonal to all $x \in X$. ●

An inner product and the corresponding norm satisfy the Schwarz inequality

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

where the equality sign holds if and only if $\{x, y\}$ is a linearly dependent set, and the triangle inequality

$$\|x + y\| \leq \|x\| + \|y\|$$

where the equality sign holds if and only if $y = 0$ or $x = cy$ $(c \geq 0)$.

The *orthogonal complement* of a Hilbert space H is

$$Y^{\perp} = \{z \in H | z \perp Y\},$$

which is the set of all vectors orthogonal to Y. For every $x \in H$ there is a $y \in Y$ such that $x = y + z$, $z \in Z = Y^{\perp}$. y is called the *orthogonal projection* of x on Y.

Definition (compact). A metric space X is said to be *compact* if every sequence in X has a convergent subsequence. (Remember that inner product and normed spaces are metric spaces, therefore, this definition is valid for these spaces as well.) ●

Akhiezer and Glazman (1981) proof the following theorem which is a criterion for (strong) compactness.

Theorem. Let X be a separable space, and let $(e_n), n = 1, \ldots, \infty$ be an orthonormal basis in X. Let M be a bounded set of elements x from X, and suppose that, for any $\varepsilon > 0$, there is a natural number $n = n(\varepsilon)$ such that for any $x \in M$,

$$\left\| x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i \right\| < \varepsilon.$$

Then the set M is compact. ●

The key in the proof is that M is bounded, and therefore one can pick out from an arbitrary sequence $(x_n) \subset M$ a weakly convergent subsequence, Akhiezer and Glazman (1981). We think that the practical relevance is that a bounded element, gravitational potential for example, can be approximated by a finite series, a truncated spherical harmonics series for example.

Corollary (Maximum and minimum). A continuous mapping T of a compact subset M of a metric space X into $\mathbb{R}$, $T : M \to \mathbb{R}$, assumes a maximum and a minimum at some points of M. ●

Definition (Hilbert-adjoint operator). Let $T : H_1 \to H_2$ be a bounded linear operator, where H_1 and H_2 are Hilbert spaces. Then the *Hilbert-adjoint operator* T^* of T is the operator

$$T^* : H_2 \to H_1$$

such that for all $x \in H_1$ and $y \in H_2$,

$$\langle Tx, y\rangle = \langle x, T^*y\rangle.$$

●

If T is self-adjoint, that is $T = T^*$, and also $H_1 = H_2$, then $\langle Tx, y\rangle = \langle x, Ty\rangle$.

Let T be a continuous linear operator from a Hilbert space H_1 into a Hilbert space H_2. Recall that the range, $R(T)$, and null space, $N(T)$, of a linear operator with domain $D(T)$ are defined by $R(T) = \{Tx | x \in D(T)\}$ and $N(T) = \{x \in D(T) | Tx = 0\}$ respectively. Then the following Theorem holds.

Theorem. If $T : H_1 \to H_2$ is a continuous linear operator, then $R(T)^\perp = N(T^*)$ and $N(T)^\perp = \overline{R(T^*)}$. Since $T = T^{**}$ we also have $R(T^*)^\perp = N(T)$ and $N(T^*)^\perp = \overline{R(T)}$. ●

A.2 Spectral theory of linear operators in normed spaces

The spectral representation of the operator T gives a great deal of clarity and insight. We begin with finite dimensional vector spaces, which is much simpler than the spectral theory of operators in infinite dimensional spaces. These operators are not considered in general, only *compact* linear operators are discussed. Their properties closely resemble those of operators on finite dimensional spaces. Finally, we look at bounded self-adjoint linear operators. These operators can be associated with the normal matrix T^*T.

The subsequent Sections summarize Chapters 7, 8 and 9 of Kreyszig (1989) respectively. Compare also (Akhiezer and Glazman, 1981, Ch. 5), (Groetsch, 1980, Ch. 4) for example. Again, proofs are omitted in general. Since the spectral decomposition is an important tool it is elaborated in detail in Appendix B.

A.2.1 Finite dimensional normed spaces

For a given $n \times n$ matrix A eigenvalues and eigenvectors are defined in terms of the equation

$$Ax = \lambda x \tag{A.11}$$

as follows.

Definition (Eigenvalues, eigenvectors, eigenspaces, spectrum). An *eigenvalue* of a square matrix A is a number λ such that (A.11) has a solution $x \neq 0$. This x is called an *eigenvector* of A corresponding to that eigenvalue λ. The eigenvectors corresponding to that eigenvalue λ and the zero vector form a vector subspace of X which is called *eigenspace* of A corresponding to that eigenvalue λ. The $\sigma(A)$ of all eigenvalues of A is called the *spectrum* of A. ●

Theorem (Eigenvalues of a matrix). The eigenvalues of an n-rowed square matrix A are given by the solutions of the characteristic equation

$$\det(A - \lambda I) = 0$$

of A. A has at least one eigenvalue and at most n different eigenvalues. ●

Theorem (Eigenvalues of an operator). All matrices representing a given linear operator $T : X \to X$ on a finite dimensional normed space X relative to various bases for X have the same eigenvalues. ●

Therefore, one can speak of the (unique) spectrum etc. of the linear operator T.

Remark: If X is infinite dimensional, then T may have spectral values which are not eigenvalues, cf. (Kreyszig, 1989, Ch. 7).

A.2.2 Compact linear operators on normed spaces

Compact linear operators play a central role in the theory of integral equations. Their properties closely resemble those of operators on finite dimensional spaces.

Definition (compact linear operator) Let X and Y be normed spaces. An operator $T : X \to Y$ is called a *compact linear operator* if T is linear and if for every bounded subset M of X the image $T(M)$ is relatively compact, that is, the closure $\overline{T(M)}$ is compact. ●

It is not required that T is continuous. However, compact linear operators are always continuous and are therefore also called *completely continuous linear operators*, Groetsch (1980).

Lemma (Continuity). Let X and Y be normed spaces. Then every compact linear operator $T : X \to Y$ is bounded, hence continuous. *Proof.* If $T : X \to Y$ is compact and B is the closed unit ball in X, then $\overline{T(B)}$ is compact and therefore bounded. Therefore there is an $M > 0$ such that $\|y\| \leq M$ for all $y \in \overline{T(B)}$. It follows that $\|Tx\| \leq M$ for $\|x\| \leq 1$, that is, $\|T\| \leq M$. ●

Theorem (Adjoint operator). Let $T : X \to Y$ be a linear operator. If T is compact, so is its adjoint operator $T^* : Y \to X$; here X and Y are Hilbert spaces. ●

Theorem (Eigenvalues, null space and range). Let $T : X \to X$ be a compact linear operator on a normed space X. Then:

1. the set of eigenvalues is countable, and the only possible point of accumulation is $\lambda = 0$,
2. for every $\lambda \neq 0$ the null space $N(T_\lambda)$ of $T_\lambda - \lambda I$ is finite dimensional,
3. for every $\lambda \neq 0$ the range of $T_\lambda - \lambda I$ is closed.

●

A.2.3 Bounded self-adjoint linear operators

Let $T : H \to H$ be a bounded linear operator on a complex Hilbert space H. T is said to be *self-adjoint* if $T = T^*$ or

$$\langle Tx, y\rangle = \langle x, Ty\rangle.$$

All the eigenvalues of T (if they exist) are real. Eigenvectors corresponding to (numerically) different eigenvalues of T are orthogonal.

A *compact* self-adjoint operator $T \neq 0$ has at least one eigenvector x corresponding to a non-zero eigenvalue λ, Akhiezer and Glazman (1981). Further properties of bounded self-adjoint linear operators, not directly relevant to this work, can be found in (Akhiezer and Glazman, 1981, Ch. 6) and (Kreyszig, 1989, Ch. 9).

B |

CONVENTIONS AND SPECTRAL DECOMPOSITION

In this Appendix the adopted conventions are explained. Furthermore, the spectral decomposition of compact operators between Hilbert spaces is given, which is an important tool.

B.1 Adopted conventions

B.1.1 Finite and infinite dimension

A clear distinction should be made between finite dimensional spaces and infinite dimensional spaces. The first are related to real world observations and solved parameters, while the latter have to do with 'experiments of thought' and theoretical foundation of methods.

The basic relation to be studied is

$$\mathbf{g} = A\mathbf{f} \tag{B.1}$$

where $A : F \rightarrow G$, F and G are Hilbert spaces, $\mathbf{f} \in F, \mathbf{g} \in G$ (in Appendix A some results from functional analysis are given). The operator or mapping A is assumed to be linear and compact (and therefore bounded), and relates the measurement $\mathbf{g}$ to the unknown $\mathbf{f}$. Equation (B.1) is a short hand notation of the integral equation of the first kind

$$\mathbf{g}(x) = \int_a^b K(x,y)\mathbf{f}(y)dy, \quad a \leq x \leq b$$

where $K(x,y)$ is the kernel of the integral operator A. The intervals, where the functions $\mathbf{g}$ and $\mathbf{f}$ are defined, are equal, which can always be realized by appropriate scaling, Wing (1991).

When the number of parameters to be determined and the data are finite, one is restricted to finite dimensional spaces. Instead of (B.1)

$$\mathbf{y} = A\mathbf{x} \tag{B.2}$$

is written. Now A is a matrix of dimension $m \times n$, where the number of observations m is always larger than or equal to the number of unknowns n: $m \geq n$. This is not strictly necessary but avoids underdetermination as long as A has full column rank, $\text{rank}(A) = n$, cf. Lanczos (1961).

The functions $\mathbf{f}$ and $\mathbf{g}$ as well as the vectors $\mathbf{x}$ and $\mathbf{y}$ are assumed to be real.

B.1.2 Measurement errors, norm and generalized inverse

Relations (B.1) and (B.2) only hold when the data are exact, that is there are no measurement errors (note that the models are assumed to be free of errors). Of course in reality these errors can not be avoided. Measurement errors are denoted with ε:

$$\begin{aligned} \mathbf{g}^\varepsilon &= A\mathbf{f} + \underline{\varepsilon} = \mathbf{g} + \underline{\varepsilon} \\ \mathbf{y}^\varepsilon &= A\mathbf{x} + \underline{\varepsilon} = \mathbf{y} + \underline{\varepsilon} \end{aligned}$$

where $\|\mathbf{g} - \mathbf{g}^\varepsilon\|_G = \|\underline{\varepsilon}\|_G \leq \varepsilon$ and $\|\mathbf{y} - \mathbf{y}^\varepsilon\|_2 = \|\underline{\varepsilon}\|_2 \leq \varepsilon$ and the norms

$$\begin{aligned} \|\mathbf{g}\|_G &= \left(\int_a^b \mathbf{g}(x)^2 dx \right)^{1/2}, \\ \|\mathbf{y}\|_2 &= \left(\sum_{j=1}^m y_j^2 \right)^{1/2}. \end{aligned}$$

These norms are called $L^2[a,b]$ and l^2-norm respectively or 2-norm for short.

The measurements have the property

$$\begin{aligned} E\{\mathbf{g}^\varepsilon\} &= \mathbf{g} \\ E\{\mathbf{y}^\varepsilon\} &= \mathbf{y}. \end{aligned}$$

The superscript ε is frequently dropped since it will be clear from the context whether error-free data or not are considered.

Because of the errors, $\mathbf{g}$ and $\mathbf{y}$ may not be in the range of A and no solution would exist (the redundant system is not compatible). Then, it seems natural to minimize the distance

$$\|A\mathbf{f} - \mathbf{g}\|_G$$

or

$$\|A\mathbf{x} - \mathbf{y}\|_2$$

instead. Minimization with the 2-norm(s) gives the least-squares solution. Minimizing

$$J(\mathbf{x}) = \|A\mathbf{x} - \mathbf{y}\|_2^2 \tag{B.3}$$

results in

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{y}.$$

It can be shown that the same solution is obtained by the generalized inverse of A, A^+

$$\mathbf{x} = A^+\mathbf{y}. \tag{B.4}$$

As long as A is a regular matrix, i.e. $\text{rank}(A) = n$, the solution is unique, Lanczos (1961). When A is not regular, i.e. $\text{rank}(A) < n$, then (B.4) gives the minimum solution with the smallest norm itself, that is $\|\mathbf{x}\|_2 < \|\mathbf{x}_s\|_2$ where $\mathbf{x}_s$ is any other solution of (B.3).

The generalized inverse exists for the continuous case as well.

B.1.3 Weighted norm

When the errors in the measurements are described by a variance-covariance matrix, it is better to compute a weighted least-squares solution as follows:

$$\mathbf{x} = (A^TPA)^{-1}A^TP\mathbf{y}$$

where P is the weight matrix of the observations. The corresponding weighted norm can be written as

$$\|A\mathbf{x} - \mathbf{y}\|_P^2.$$

Since P is positive definite its Choleski decomposition is

$$P = W^TW$$

with W an upper triangular matrix. The transformations

$$A_w = WA$$

and

$$\mathbf{y}_w = W\mathbf{y}$$

lead to the minimization problem

$$J(\mathbf{x}) = \|A_w\mathbf{x} - \mathbf{y}_w\|_2^2$$

with solution

$$\mathbf{x} = (A_w^TA_w)^{-1}A_w^T\mathbf{y}_w \tag{B.5}$$

or

$$\mathbf{x} = A_w^+\mathbf{y}_w$$

and (B.5) equals the weighted least-squares solution. One can therefore conclude that P causes no additional problems, see also Section 4.2.

B.2 Introduction to spectral decomposition

An important tool when dealing with inverse problems is the spectral decomposition of the operator. A spectrum gives clear insight in the behaviour of the operator for different frequencies and further illuminates the ill-posedness of the problem at hand. The main references here are Lanczos (1961); Groetsch (1980); Kreyszig (1989), compare also Nashed (1976); Golub and van Loan (1996); Louis (1989); Groetsch (1993); Engl *et al.* (1996).

Definition (eigenvalues, spectrum). Let $T : F \to G$ be a compact, symmetric (or self-adjoint, $T = T^*$) and semi-positive definite ($\langle T\mathbf{f}, \mathbf{f}\rangle \geq 0 \ \forall \ \mathbf{f} \in F$) linear operator. Then T has a finite or countably infinite number of *eigenvalues* λ_n; in the latter case $\lambda_n \to 0$ as $n \to \infty$ (the only possible point of accumulation is zero which follows from the compactness of the operator, e.g. (Kreyszig, 1989, Sec. 8.3)). The eigenvalues can be arranged in a sequence converging to zero

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq \cdots \geq 0 \tag{B.6}$$

with corresponding (nonzero) orthonormal eigenvectors $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n, \cdots$:

$$T\mathbf{w}_n = \lambda_n \mathbf{w}_n.$$

The set of eigenvectors $\{\mathbf{w}_n : \lambda_n \neq 0\}$ is a Schauder basis for $\overline{R(T)}$.[1]

The set $\sigma(T)$ of numbers λ for which the operator $T - \lambda I$ has no bounded inverse is called the *spectrum* of T. In the case of a compact, symmetric, semi-positive definite operator the spectrum is real, nonempty and every nonzero member of $\sigma(T)$ is an eigenvalue of T. The corresponding eigenspace $N(T - \lambda I)$ is finite dimensional. ●

The norm of T is equal to the *spectral radius*:

$$\|T\| = \max\{\lambda : \lambda \in \sigma(T)\} = \lambda_1.$$

For every $\mathbf{f} \in F$ we may write

$$T\mathbf{f} = \sum_{n=1}^{\infty} \lambda_n \langle \mathbf{f}, \mathbf{w}_n \rangle \mathbf{w}_n.$$

Definition (singular values, singular system). Now consider the compact operators $A : F \to G$, $A^*A : F \to F$ and $AA^* : G \to G$. The latter two are self-adjoint and have the same nonnegative eigenvalues, similar to the operator T from above. The spectra of both operators are the same $\sigma(A^*A) = \sigma(AA^*)$.

Denote the eigenvectors of A^*A as $\mathbf{v}_n$, and the eigenvectors of AA^* as $\mathbf{u}_n$. The eigenvalues, equal for both operators, are λ_n, the ordering is as in (B.6). Let $\sigma_n = \sqrt{\lambda_n}$ and $\mathbf{u}_n = \sigma_n^{-1} A\mathbf{v}_n$. Then

$$A\mathbf{v}_n = \sigma_n \mathbf{u}_n$$

and

$$A^*\mathbf{u}_n = \sigma_n \mathbf{v}_n.$$

The numbers σ_n are called the *singular values* for the operator A, the system $\{\mathbf{v}_n, \mathbf{u}_n; \sigma_n\}$ is called a *singular system* for A. ●

Note that sometimes the singular values are defined as $1/\sqrt{\lambda_n}$.

From the last two equations above it follows that $\mathbf{v}_n$ and $\mathbf{u}_n$ are indeed eigenvectors of A^*A and AA^* respectively. The eigenvectors $\mathbf{v}_n$ are a complete orthonormal system or basis for

$$\overline{R(A^*)} = \overline{R(A^*A)} = N(A)^{\perp}$$

[1] A basis is complete and $R(T)$ might not be complete. Thus the completion of $R(T)$ is needed.

and $\mathbf{u}_n$ are a complete orthonormal system for

$$\overline{R(A)} = \overline{R(AA^*)} = N(A^*)^\perp \tag{B.7}$$

with $N(A)^\perp$ the space perpendicular to the null-space of A.

If, and only if, A has a finite-dimensional range, A has only finitely many singular values. If A is an integral operator with infinitely many singular values, they accumulate (only) at 0

$$\lim_{n\to\infty} \sigma_n = 0$$

as was the case for the eigenvalues. If there are finitely many singular values the kernel of the integral operator is *degenerate.*

Again equivalent to the norm of T we can write

$$\|A\| = \sigma_1$$

where, obviously, σ_1 is the largest singular value.

Theorem (Picard condition). The representation

$$A\mathbf{f} = \sum_{n=1}^{\infty} \sigma_n \langle \mathbf{f}, \mathbf{v}_n \rangle \mathbf{u}_n \tag{B.8}$$

of the operator A is called a *singular value decomposition* (SVD). The equation of the first kind $A\mathbf{f} = \mathbf{g}$ has a solution if $\mathbf{g} \in \overline{R(A)}$ and

$$\sum_{n=1}^{\infty} \sigma_n^{-2} |\langle \mathbf{g}, \mathbf{u}_n \rangle|^2 < \infty. \tag{B.9}$$

This is called the *Picard condition.* ●

The Picard condition is a 'smoothness condition' for the right-hand side $\mathbf{g}$. Since $\mathbf{g} \in \overline{R(A)}$ one can write $\mathbf{g} = \sum_n g_n \mathbf{u}_n$, see equation (B.7). Because $\sigma_n^{-2} \to \infty$ for $n \to \infty$ the coefficients g_n of $\sum_n \sigma_n^{-2} g_n^2$ have to decay fast enough with respect to the singular values in order to fulfil (B.9). The solution

$$\mathbf{f} = \sum_{n=1}^{\infty} \frac{\langle \mathbf{g}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n$$

is not unique since any solution $\mathbf{f}_s = \mathbf{f} + \mathbf{h}$ where $\mathbf{h} \in N(A)$ is also a solution of $A\mathbf{f} = \mathbf{g}$.

For a degenerate or finite dimensional operator the sums become finite.

Singular value decomposition with finite dimensions. Let $A \in \mathbb{R}^{m\times n}$, with $m \geq n$. The singular value decomposition of A is then $A = U\Sigma V^T$ (Figure B.1).

The matrices U and V are orthogonal, which means $U^TU = UU^T = I_m$ and $V^TV = VV^T = I_n$ respectively. Since the last $m - n$ rows of Σ only contain zeros, the last $m - n$ columns of U could be cancelled. This is called the thin singular value decomposition, Golub and van Loan (1996). The resulting smaller matrix U becomes semi-orthogonal, $U^TU = I_n, UU^T \neq I_m$, compare also Lanczos (1961). The range of A is spanned by

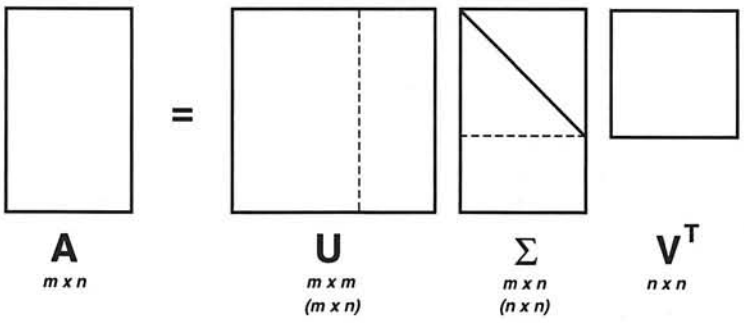

Figure B.1: *Singular value decomposition, the thin singular value decomposition is denoted by the dashed lines.*

the first n columns of U, provided that all singular values are non-zero. The domain of A is spanned by the columns of V. Therefore, the large singular values denote the combination of unknowns x_i that is well represented by a combination of measurements y_j. In contrast, the small singular values reveal which linear combination of unknowns are hardly recoverable from the measurements.

Generalized singular value decomposition

For later use it is necessary to define the generalized singular value decomposition (GSVD). Here Hansen (1997) and Golub and van Loan (1996) are followed, compare also Hansen (1989). Firstly, the generalized eigenvalues for a pair of symmetric (positive definite) matrices (S, T) are defined, after which the generalized singular values are discussed.

Generalized eigenvalues. Given a symmetric matrix $S \in \mathbb{R}^{n \times n}$ and a symmetric positive definite matrix $T \in \mathbb{R}^{n \times n}$. The *symmetric-definite generalized eigenvalue problem* is to find a nonzero vector $\mathbf{x}$ and a scalar λ such that $S\mathbf{x} = \lambda T\mathbf{x}$, λ is a *generalized eigenvalue* Golub and van Loan (1996). The set of generalized eigenvalues $\lambda(S, T)$ is determined by

$$\lambda(S, T) = \{\lambda | \det(S - \lambda T) = 0\}.$$

Remark: The prove of the existence of the generalized eigenvalue decomposition involves the condition that the (weighted) sum of the matrices S and T must be non-negative definite (= positive semi-definite). When at least one of these matrices is positive definite this condition is fulfilled. Here we are concerned with matrices $A^T A$ and $L^T L$ of which the first is assumed to be positive definite (it has full column rank). More on the mathematical background can be found in (Golub and van Loan, 1996, Sec. 8.7).

Generalized singular values. The GSVD of the matrix pair (A, L) is a generalization of the SVD of A in the sense that the generalized singular values of (A, L) are the square roots of the generalized eigenvalues of the matrix pair $(A^T A, L^T L)$, Hansen (1997). Thus, going from generalized eigenvalues to generalized singular values resembles the step from eigenvalues to singular values.

Let $A \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{p\times n}$ with $m \geq n \geq p$. Then the GSVD is a decomposition of A and L in the form

$$A = U \begin{pmatrix} \Sigma & 0_{p\times n-p} \\ 0_{n-p\times p} & I_{n-p} \end{pmatrix} X^{-1}, \quad L = V \begin{pmatrix} M & 0_{p\times n-p} \end{pmatrix} X^{-1}$$

where $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{p\times p}$ and $U^TU = I_n$, $V^TV = I_p$. $X \in \mathbb{R}^{n\times n}$ is nonsingular, and Σ and M are $p \times p$ diagonal matrices with elements:

$$0 \leq \sigma_1 \leq \ldots \leq \sigma_p \leq 1, \quad 1 \geq \mu_1 \geq \ldots \geq \mu_p > 0$$

which are normalized such that

$$\Sigma^T\Sigma + M^TM = I_p.$$

Then the *generalized singular values* γ_i of (A, L) are defined as

$$\gamma_i = \sigma_i/\mu_i, \quad i = 1, \ldots, p$$

and they appear in non-decreasing order (opposite to the singular value ordering for historical reasons, Hansen (1997)).

The first p columns of $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ satisfy

$$\mu_i^2 A^TA\mathbf{x}_i = \sigma_i^2 L^TL\mathbf{x}_i, \quad i = 1, \ldots, p$$

hence $A^TA\mathbf{x}_i = \gamma_i^2 L^TL\mathbf{x}_i$. Thus, the $\mathbf{x}_i$ are called the *generalized singular vectors* of the pair (A, L). For $p < n$ the matrix $L \in \mathbb{R}^{p\times n}$ always has a nontrivial null-space $N(L)$, Hansen (1997). The last $n - p$ columns $\mathbf{x}_i$ of X satisfy

$$L\mathbf{x}_i = 0, \quad i = p+1, \ldots, n$$

and they are therefore basis vectors for the null-space $N(L)$.

Relation with SVD. Only when L is the indentity matrix I_n, the matrices U, Σ and V in the GSVD of (A, L) are identical to U, Σ and V of the SVD, except for the ordering of the singular values and vectors, since $p = n$, $X^{-1} = M^{-1}V^T$ and $A = U\Sigma M^{-1}V^T$. In general there is no connection between the singular values and vectors of SVD and GSVD. However, when L is well-conditioned (has a 'small' condition number, the smallest possible number is one) it can be shown that the matrix X is also well-conditioned, Hansen (1997). The diagonal matrix Σ displays therefore, the ill-conditioning of A.

Discrete Picard condition. Hansen (1990) introduces the Picard condition for finite dimension. The unperturbed $\mathbf{y}$ in a discrete ill-posed problem with regularization matrix L satisfies the discrete Picard condition if the Fourier coefficients $|\mathbf{u}_i^T\mathbf{y}|$ on the average decay faster than the generalized singular values γ_i.[2] A visual inspection of a plot of the

[2]The (discrete) Picard condition is usually not satisfied because the spectrum of the measurement errors, or perturbation of $\mathbf{y}$, does not decrease fast enough. Hence, the least-squares solution, which involves σ_n^{-1}, 'blows up'. The idea of regularization then might be to impose a constraint on the unknown signal: $\|L\mathbf{x}\|_2$ has to be finite, where L is the regularization matrix. See Chapters 2 and 3 for further details.

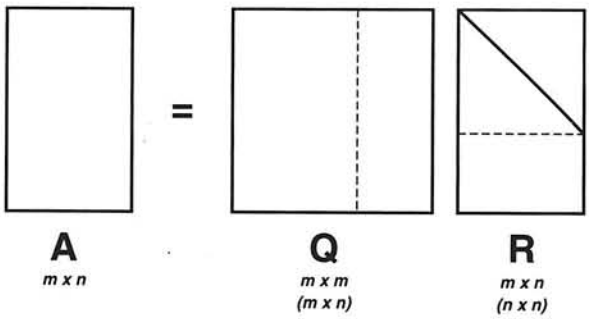

Figure B.2: *QR factorization, the thin factorization is denoted by the dashed lines.*

Fourier coefficients $|\mathbf{u}_i^T\mathbf{y}|$ and the generalized singular values γ_i could reveal the faster decay. Alternatively, one may define the ratio

$$\rho_i \equiv \gamma_i^{-1}\left(\prod_{j=i-q}^{i+q}|\mathbf{u}_i^T\mathbf{y}|\right)^{1/(2q+1)}, \quad i = q+1,\ldots,n-q$$

which is the moving geometric mean, with q a small integer, Hansen (1990). This ratio should decay monotonically to zero.

QR factorization

Although the QR factorization is not a spectral decomposition we need to mention it briefly, since it is used in the *transformation to standard form*, Chapter 4. The QR factorization of an m-by-n matrix A is given by

$$A = QR$$

where $Q \in \mathbb{R}^{m\times m}$ is orthogonal and $R \in \mathbb{R}^{m\times n}$ is upper triangular. If A has full column rank then the first n columns of Q form an orthonormal basis for $R(A)$. Also in this case a thin version exists, Golub and van Loan (1996); Strang (1988). See Figure B.2 for a visualization of the QR factorization.

B.3 Summary

The subject of this study are integral equations of the first kind, represented by a linear, compact operator A mapping a function $\mathbf{f}$ from a Hilbert space F to a function $\mathbf{g}$ from a Hilbert space G, as well as their discrete counterparts $A, \mathbf{x}$ and $\mathbf{y}$, where A is a matrix and $\mathbf{x}$ and $\mathbf{y}$ are vectors. The model, A, is assumed to be exact. The measurements $\mathbf{y}^\epsilon$ and $\mathbf{g}^\epsilon$, however, are not exact, leading to the (weighted) least-squares minimization of the error. It can be shown that the generalized inverse A^+ of A gives the same solution.

A generalized singular value decomposition exists of the matrix pair (A, L), which can be derived from the generalized eigenvalue problem of the symmetric (positive definite) matrix pair (S, T). If L is the identity matrix then the usual singular value decomposition is obtained, which in turn can be derived from the eigenvalue problem of the symmetric

semi-positive definite matrix (or operator) T. The singular values, together with the eigenvectors, completely describe the operator A. The singular values form the spectrum or, in other words, they are the coefficients with respect to the basis (the eigenvectors).

References

Akhiezer, N. and Glazman, I. (1981). *Theory of linear operators in Hilbert space*, volume 1. Pitman, third edition.

Anger, G., Gorenflo, R., Jochmann, H., Moritz, H., and Webers, W., editors (1993). *Inverse problems: principles and applications in geophysics, technology, and medicine*, volume 74 of *Mathematical Research*. Akademie Verlag. Proceedings of the International Conference held in Potsdam, August 30 - September 3, 1993.

Backus, G. and Gilbert, J. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophys. J. R. astr. Soc.*, **13**, 247–276.

Backus, G. and Gilbert, J. (1968). The resolving power of gross earth data. *Geophys. J. R. astr. Soc.*, **16**, 169–205.

Blaser, J., Cornelisse, J., Cruise, A., Damour, T., Hechler, F., Hechler, M., Jafry, Y., Kent, B., Lockerbie, N., Paik, H., Ravex, A., Reinhard, R., Rummel, R., Speake, C., Sumner, T., Touboul, P., and Vitale, S. (1996). STEP Satellite Test of the Equivalence Principle. Report on the Phase A study. ESA SCI(96)5.

Bouman, J. (1993). *The normal matrix in gravity field determination with satellite methods; its stabilization, its information content and its use in error propagation.* Master's thesis, Delft University of Technology.

Bouman, J. and Koop, R. (1997). Quality differences between Tikhonov regularization and generalized biased estimation in gradiometric analysis. *DEOS Progress Letters*, **97.1**, 42–48.

Bouman, J. and Koop, R. (1998). Regularization in gradiometric analysis. *Physics and Chemistry of the Earth*, **23**(1), 41–46.

Eldén, L. (1977). Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, **17**, 134–145.

Eldén, L. (1982). A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT*, **22**, 487–502.

Engl, H. (1997). *Integralgleichungen.* Springer-Verlag.

Engl, H., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems.* Kluwer Academic Publishers.

ESA (1996). Gravity Field and Steady-State Ocean Circulation Mission. Report for assessment. ESA SP-1196(1).

Gelderen, M. van (1992). The geodetic boundary value problem in two dimensions and its iterative solution. Publications on geodesy. New series no. 35, Netherlands Geodetic Commission.

Gelderen, M. van and Koop, R. (1997). The use of degree variances in satellite gradiometry. *Journal of Geodesy*, **71**, 337–343.

Gerstl, M. and Rummel, R. (1981). Stability investigations of various representations of the gravity field. *Reviews of geophysics and space physics*, **19**(3), 415–420.

Golub, G. and van Loan, C. (1996). *Matrix computations.* The Johns Hopkins University Press, third edition.

Golub, G., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**(2), 215–223.

Groetsch, C. (1980). *Elements of applicable functional analysis.* Marcel Dekker.

Groetsch, C. (1984). *The theory of Tikhonov regularization for Fredholm integral equations of the first kind.* Pitman.

Groetsch, C. (1993). *Inverse problems in the mathematical sciences.* Vieweg Verlag.

Gruber, T., Anzenhofer, M., and Rentsch, M. (1995). The 1995 GFZ high resolution gravity model. In R. Rapp, A. Cazenave, and R. Nerem, editors, *IAG Symposia 116 - Global gravity field and its temporal variations.* Springer-Verlag.

Haagmans, R. (1988). *Detailed gravity anomalies derived from SEASAT altimeter data; A comparison of two alternative approaches: least squares collocation and a method based on FFT.* Master's thesis, Delft University of Technology.

Hansen, P. (1989). Regularization, GSVD and truncated GSVD. *BIT*, **29**, 491–504.

Hansen, P. (1990). The discrete Picard condition for discrete ill-posed problems. *BIT*, **30**, 658–672.

Hansen, P. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, **34**(4), 561–580.

Hansen, P. (1997). *Regularization Tools, A Matlab package for analysis and solution of discrete ill-posed problems, Version 2.1 for Matlab 5.0.* Department of Mathematical Modelling, Technical University of Denmark. http://www.imm.dtu.dk/~pch.

Hansen, P. and O'Leary, D. (1993). The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, **14**(6), 1487–1503.

Heck, B. (1990). An evaluation of some systematic error sources affecting terrestrial gravity anomalies. *Bulletin Géodésique*, **64**, 88–108.

Heiskanen, W. and Moritz, H. (1967). *Physical geodesy.* W.H. Freeman and Co.

Hemmerle, W. (1975). An explicit solution for generalized ridge regression. *Technometrics*, **17**, 309–314.

Hemmerle, W. and Brantle, T. (1978). Explicit and constrained generalized ridge estimation. *Technometrics*, **20**(2), 109–120.

Hirsch, M. (1996). Analyse und Numerik überbestimmter Randwertprobleme in der Physikalischen Geodäsie. Reihe C No. 453, Deutsche Geodätische Kommission.

Hoerl, A. and Kennard, R. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, **12**(1), 55–67.

Ivanov, V. (1962). Integral equations of the first kind and an approximate solution for the inverse problem of potential. *Soviet Math. Doklady*, **3**, 210–212.

Jekeli, C. (1978). An investigation of two models for the degree variances of global covariance functions. Report No. 275, Ohio State University.

Kan, J. van and Segal, A. (1993). *Numerieke methoden voor partiële differentiaalvergelijkingen.* DUM.

Kaula, W. (1966). *Theory of satellite geodesy.* Blaisdell Pub. Co.

Kitagawa, T. (1987). A deterministic approach to optimal regularization - the finite dimensional case. *Japan J. Appl. Math.*, **4**, 371–391.

Kress, R. (1989). *Linear integral equations.* Springer-Verlag.

Kreyszig, E. (1988). *Advanced engineering mathematics.* John Wiley and Sons, sixth edition.

Kreyszig, E. (1989). *Introductory functional analysis with applications.* John Wiley and Sons.

Lanczos, C. (1961). *Linear differential operators.* Van Nostrand Company Ltd.

Lawson, C. and Hanson, R. (1974). *Solving least squares problems.* Prentice-Hall.

Lerch, F., Iz, H., and Chan, J. (1993). Gravity model solution based upon SLR data using eigenvalue analysis: Alternative methodology. In D. Smith and D. Turcotte, editors, *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, volume 24 of *Geodynamics Series*, pages 213–219. American Geophysical Union.

Louis, A. (1989). *Inverse und schlecht gestellte Probleme.* Teubner.

Marsh, J., Lerch, F., Putney, B., Felsentreger, T., Sanchez, B., Klosko, S., Patel, G., Robbins, J., Williamson, R., Engelis, T., Eddy, W., Chandler, N., Chinn, D., Kapoor, S., Rachlin, K., Braatz, L., and Pavlis, E. (1989). The GEM-T2 gravitational model. TM 100746, NASA.

Martensen, E. and Ritter, S. (1997). Potential theory. In F. Sansò and R. Rummel, editors, *Geodetic Boundary Value Problems in View of the One Centimeter Geoid*, volume 65 of *Lecture notes in earth sciences*, pages 19–66. Springer-Verlag.

Meissl, P. (1975). Elements of functional analysis. In B. Brosowski and E. Martensen, editors, *Mathematical Geodesy, part I*, volume 12 of *Methoden und Verfahren de mathematischen Physik*, pages 19–78. Bibliographisches Institut.

Meissl, P. (1976). Hilbert spaces and their application to geodetic least squares problems. *Bollettino di Geodesia e Scienze Affini*, **35**(1), 49–80.

Moritz, H. (1980). *Advanced physical geodesy.* Wichmann.

Morozov, V. (1984). *Methods for solving incorrectly posed problems.* Springer-Verlag.

Nashed, M. (1976). Aspects of generalized inverses in analysis and regularization. In M. Nashed, editor, *Generalized inverses and applications*, pages 193–244.

Nerem, R., Lerch, F., Marshall, J., Pavlis, E., Putney, B., Tapley, B., Eanes, R., Ries, J., Schutz, B., Shum, C., Watkins, M., Klosko, S., Chan, J., Luthcke, S., Patel, G., Pavlis, N., Williamson, R., Rapp, R., Biancale, R., and Nouel, F. (1994). Gravity model development for Topex/Poseidon: Joint Gravity Models 1 and 2. *Journal of Geophysical Research*, **99**(C12), 24421–24447.

Neyman, Y. (1985). Improperly posed problems in geodesy and methods of their solution. In K. Schwarz, editor, *Local Gravity Field Approximation*, pages 499–566.

Oppenheim, A., Willsky, A., and Young, I. (1983). *Signals and Systems.* Prentice-Hall.

Parker, R. (1994). *Geophysical inverse theory.* Princeton University Press.

Phillips, D. (1962). A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery*, **9**, 84–97.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical recipes in C: the art of scientific computing.* Cambridge University Press, second edition.

Rapp, R. (1972). Geopotential coefficient behavior to high degree and geoid information by wavelength. Report No. 180, Ohio State University.

Rapp, R. (1979). Potential coefficient and anomaly degree variance modelling revisited. Report No. 293, Ohio State University.

Rapp, R., Wang, Y., and Pavlis, N. (1991). The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models. Report No. 410, Ohio State University.

Rauhut, A. (1992). *Regularization methods for the solution of the inverse Stokes problem.* Ph.D. thesis, The University of Calgary.

Regińska, T. (1996). A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, **17**(3), 740–749.

Rummel, R. (1992). Physical geodesy I. Lecture notes, Delft University of Technology, Faculty of Geodesy, Delft.

Rummel, R. (1997). Spherical spectral properties of the earth's gravitational potential and its first and second derivatives. In F. Sansò and R. Rummel, editors, *Geodetic Boundary Value Problems in View of the One Centimeter Geoid*, volume 65 of *Lecture notes in earth sciences*, pages 359–404. Springer-Verlag.

Rummel, R., Schwarz, K., and Gerstl, M. (1979). Least squares collocation and regularization. *Bulletin Géodésique*, **53**, 343–361.

Sansò, F. (1989). On the foundations of various approaches to improperly posed problems. In F. Sansò, editor, *Workshop on theory and practice of inverse problems*, Ricerche di geodesia topografia e fotogrammetria, pages 85–153. Clup.

Schuh, W. (1996). Tailored numerical solution strategies for the global determination of the earth's gravity field. Technical Report Folge 81, Mitteilungen der geodätischen Institute der Technischen Universität Graz.

Schwarz, K. (1973). Investigations on the downward continuation of aerial gravity data. Report No. 204, Ohio State University.

Schwarz, K. (1979). Geodetic improperly posed problems and their regularization. *Bollettino di Geodesia e Scienze Affini*, **38**(3), 389–416.

Schwintzer, P. (1990). Sensitivity analysis in least squares gravity field modelling by means of redundancy decomposition of stochastic a priori information. Deutsches Geodätiches Forschungs-Institut, internal report.

Schwintzer, P., Reigber, C., Bode, A., Kang, Z., Zhu, S., Massmann, F., Raimondo, J., Biancale, R., Balmino, G., Lemoine, J., Moynot, B., Marty, J., Barlier, F., and Boudon, Y. (1997). Long-wavelength global gravity field models: GRIM4-S4, GRIM4-C4. *Journal of Geodesy*, **71**, 189–208.

Snieder, R. (1998). The role of nonlinearity in inverse problems. *Inverse Problems*, **14**, 387–404.

Strang, G. (1986). *Introduction to applied mathematics.* Wellesley-Cambridge press.

Strang, G. (1988). *Linear algebra and its applications.* Harcourt Brace Jovanovich, third edition.

Tapley, B. (1996). Gravity recovery and climate experiment (GRACE); Proposal to NASA's earth system science pathfinder program. Technical report, JPL.

Tikhonov, A. (1963a). Regularization of incorrectly posed problems. *Soviet Math. Dokl.*, **4**, 1624–1627.

Tikhonov, A. (1963b). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, **4**, 1035–1038.

Tikhonov, A. and Arsenin, V. (1977). *Solutions of ill-posed problems.* Winston and Sons.

Trampert, J. and Snieder, R. (1996). Model estimations biased by truncated expansions: possible artifacts in seismic tomography. *Science*, **271**, 1257–1260.

Tscherning, C. (1978). Introduction to functional analysis with a view to its applications in approximation theory. In H. Moritz and H. Sünkel, editors, *Approximation methods in Geodesy*, pages 157–191. Herbert Wichmann Verlag.

Tscherning, C. (1986). Functional methods for gravity field approximation. Fourth int. summer school in the mountains Admont, Austria.

Tscherning, C. and Rapp, R. (1974). Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models. Report No. 208, Ohio State University.

Vinod, H. and Ullah, A. (1981). *Recent advances in regression methods.* Marcel Dekker.

Vogel, C. (1996). Non-convergence of the L-curve regularization parameter selection method. *Inverse Problems*, **12**, 535–547.

Wahba, G. (1990). *Spline models for observational data.* SIAM.

Wenzel, H. (1985). Hochauflösende Kugelfunktionsmodelle für das Gravitationspotential der Erde. Technical Report 137, Wissenschaftliche Arbeiten der Fachrichtung Vermussungswesen der Universität Hannover.

Wing, G. (1991). *A primer on integral equations of the first kind: the problem of deconvolution and unfolding.* SIAM.

Xu, P. (1992a). Determination of surface gravity anomalies using gradiometric observables. *Geophysical Journal International*, **110**, 321–332.

Xu, P. (1992b). The value of minimum norm estimation of geopotential fields. *Geophysical Journal International*, **111**, 170–178.

Xu, P. (1997). Truncated SVD estimators for discrete linear ill-posed problems, with applications in recovery of regional gravity fields from space gradiometric observables. Submitted to *Geophysical Research Letters.*

Xu, P. and Rummel, R. (1994a). Generalized ridge regression with applications in determination of potential fields. *Manuscripta Geodetica*, **20**, 8–20.

Xu, P. and Rummel, R. (1994b). A simulation study of smoothness methods in recovery of regional gravity fields. *Geophysical Journal International*, **117**, 472–486.

Publications of the Delft Institute for Earth-Oriented Space Research:

97.1	Bouman, J.	A survey of global gravity models.
97.2	Bruijne, A. de	Wavelet and Radon analysis for detection of elongated structures in profile measurements.
97.3	Onselen, K. van	Quality investigation of vertical datum connection.
98.1	Hanssen, R.	Atmospheric heterogeneities in ERS tandem SAR interferometry

Request additional copies: Ms. W.G. Coops-Luijten, Delft University of Technology, Faculty of Civil Engineering and Geosciences, Thijsseweg 11, 2629 JA, Delft, The Netherlands, e-mail: deos@geo.tudelft.nl, costs: NLG 15,– (excl. postal charges).

3031526

ISBN 90-407-1798-2 Delft University Press