Euro. Jnl of Applied Mathematics (2005), *vol.* 16, *pp.* 303–317. © 2005 Cambridge University Press
doi:10.1017/S0956792505006236 Printed in the United Kingdom

Regularization without preliminary knowledge of smoothness and error behaviour

FRANK BAUER[1] and SERGEI PEREVERZEV[2]

[1] *Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany*
email: bauer@mathematik.uni-kl.de
[2] *Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*
email: sergei.pereverzyev@oeaw.ac.at

(*Received 7 October 2004; revised 28 April 2005*)

The mathematical formulation of many physical problems results in the task of inverting a compact operator. The only known sensible solution technique is regularization which poses a severe problem in itself. Classically one dealt with deterministic noise models and required the knowledge of smoothness of the solution or the overall error behaviour. We will show that we can guarantee an asymptotically almost optimal regularization for a physically motivated noise model under no assumptions for the smoothness and rather weak assumptions on the noise behaviour. An application to the determination of the gravitational field out of satellite data will be shown.

1 Introduction

In this article we analyse the numerical solution of so-called inverse problems where one wishes to recover some quantity of interest from indirect measurements, often blurred by noise. It is well-known that inverse problems typically lead to mathematical models represented in the form of ill-posed operator equations. This means especially that their solutions are unstable under data perturbations. Such equations arise in scientific context, ranging from stereological microscopy (Abel's integral equation) to satellite geodesy (downward-continuation problem discussed below).

Numerical methods that can cope with ill-posed problems are the so-called regularization methods. They are usually defined in an infinite-dimensional setting and have to be discretized for an implementation. On the other hand, the discretized ill-posed problems may not require further regularization. Instead it may happen that regularization is achieved by just choosing the discretization level properly. How do we choose it a posteriori? This issue is addressed in the present paper.

Let $\mathscr{X}$ and $\mathscr{Y}$ be separable Hilbert spaces over the reals. As it will be always clear which space is concerned, the inner product and corresponding norm on each of the spaces will be denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$. In this article we want to recover an element $x \in \mathscr{X}$ from observations near

$$y = Ax, \tag{1.1}$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

where $A : \mathscr{X} \to \mathscr{Y}$ is a linear compact operator with infinite rank admitting a singular value decomposition

$$Ax = \sum_{k=1}^{\infty} s_k v_k \langle u_k, x \rangle, \tag{1.2}$$

where $s_k \geqslant s_{k+1} > 0$ for all $k \in \mathbb{N}$ and $\{u_k\}_{k\in\mathbb{N}}$ and $\{v_k\}_{k\in\mathbb{N}}$ are orthonormal systems in $\mathscr{X}$ and $\mathscr{Y}$, respectively.

If not mentioned otherwise the notions used throughout this paper can be found in any introductionary book on inverse problems, such as Freeden *et al.* [4]. Theory tells that the minimal norm element minimizing the residual $||Ax - y||$ is given by $x = A^+ y$, where A^+ denotes the Moore-Penrose generalized inverse of A. $A^+ y$ is defined as the solution of the normal equation

$$A^* Ax = A^* y, \tag{1.3}$$

which is solvable in $\mathscr{X}$ iff y satisfies the Picard criterion, namely

$$\sum_{k=1}^{\infty} s_k^{-2} \langle v_k, y \rangle^2 < \infty. \tag{1.4}$$

As the singular values tend to zero as $k \to \infty$, even small perturbations y^δ of the data y may dramatically change the problem since $A^* Ax = A^* y^\delta$ is not necessarily solvable even if (1.3) is. However, if we replace A by a finite dimensional operator A_n with $\operatorname{rank}(A_n) = n$, then the equation

$$A_n^* A_n x = A_n^* y^\delta \tag{1.5}$$

is always solvable and can be reduced to a system of n linear algebraic equations.

Thus the approximation of A by a finite dimensional operator A_n with properly chosen rank n allows one to obtain a regularization for the unstable ill-posed problem (1.1), (1.3). This is sometimes called regularization by finite-dimensional approximation, or self-regularization, because no additional regularization of the approximate problem (1.5) is needed. It has been extensively discussed in the literature (e.g. see Harbrecht *et al.* [9], and the references therein).

However, the previous study of self-regularization was restricted to the case of data blurred by so-called deterministic noise, when $\xi^\delta = y^\delta - y \in \mathscr{Y}$ with $||\xi^\delta|| \leqslant \delta$. At the same time, dealing with geoscientific inverse problems one usually assumes that the data perturbations are caused by random noise. Therefore, in this paper we assume that ξ^δ is the Gaussian white noise of intensity δ [19].

Another way of putting it is that $y^\delta = y + \delta\xi$ where ξ is a zero-mean weak Gaussian random element. This specifically means that for every element $g \in \mathscr{Y}$ we can observe $\langle g, y^\delta \rangle = \langle g, Ax \rangle + \delta \langle g, \xi \rangle$, where $\langle g, \xi \rangle$ is a centered Gaussian random variable on a probability space $(\Omega, \Sigma, \mathbb{P}_\xi)$ with variance $||g||^2$. Denoting by $\mathbb{E}_\xi$ the expectation with respect to the probability measure $\mathbb{P}_\xi$ we have in addition $\mathbb{E}_\xi(\langle g_1, \xi \rangle \langle g_2, \xi \rangle) = \langle g_1, g_2 \rangle$ for all $g_1, g_2 \in \mathscr{Y}$. This means that the white noise element ξ is generated by a stochastic process with the identity covariance operator.

It should be mentioned that in reality these simplifying assumptions will not be strictly satisfied, because the accuracy of the observations depends on several limiting

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

factors which lead to coloured noise behaviour and result in correlations between the observations. However, in practice this coloured noise is usually treated by a proper filter strategy which can be considered as a whitening process, i.e. a decorrelation of the observations (e.g. see Pail & Plank [17]). This whitening is nothing but a kind of filter operator F applied to both the observations $\overline{y^\delta} = \overline{A}x + \delta\overline{\xi}$ and the design operator $\overline{A}$. As a result a noise component $\xi = F\overline{\xi}$ of the transformed (filtered) observations $y^\delta = F\overline{y^\delta} = F\overline{A}x + \delta F\overline{\xi}$ has the identity covariance operator and can be interpreted as white noise.

Thus the white noise model assumed in the present paper can be used to adequately describe the self-regularization of ill-posed problem $y^\delta = Ax + \delta\xi$, $A = F\overline{A}$ obtained after whitening.

The main difference in analyzing self-regularization for deterministic noise and for white noise models is caused by the fact that the weak random element ξ cannot be considered as an element of the observation space $\mathcal{Y}$, because for any orthonormal basis $(e_k)_{k\in\mathbb{N}} \subset \mathcal{Y}$

$$\mathbb{E}_\xi ||\xi||^2 = \sum_{k=1}^{\infty} \mathbb{E}_\xi \langle \xi, e_k \rangle^2 = \sum_{k=1}^{\infty} 1 = \infty.$$

Therefore, in contrast to the deterministic noise model one cannot characterize the overall error behaviour in terms of the estimate of the norm $||y - y^\delta||$. (For this reason the title of the article hints at the lack of knowledge of the error behaviour). As a result deterministic regularization principles such as Morozov's discrepancy principle cannot be used within the framework of a white noise model and one usually uses heuristic principles such as L-curve, because they do not depend on a priori bounds on the measurement error $||y - y^\delta||$. There is, however, a negative result by Bakushinskii [1], which tells that no convergence theory and error estimates cannot exist for such heuristic principles.

The main accomplishment of this paper is a strategy for a proper choice of the rank in (1.5) which requires only the knowledge of the noise intensity δ. For fixed data one has to compute the solutions x_n^δ of (1.5) for a number of subsequent ranks n, then from this sequence $\{x_n^\delta\}$ our algorithm will select one with an asymptotically almost optimal order of accuracy with respect to δ.

2 Parameter selection for self-regularization

2.1 Illuminating example

In the case of a known singular value decomposition a finite dimensional operator A_n can be chosen as n-th partial sum of the series (1.2), i.e.

$$A_n = \sum_{k=1}^{n} s_k v_k \langle u_k, \cdot \rangle.$$

Then a minimal norm solution $x_n^\delta = A_n^+ y^\delta$ of (1.5) admits the representation

$$x_n^\delta = \sum_{k=1}^{n} s_k^{-1} u_k \langle v_k, y^\delta \rangle \tag{2.1}$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

and the stability of the self-regularization can be characterized by the expectation $\mathbb{E}_\xi||x_n^0 - x_n^\delta||^2$ where $x_n^0 = A_n^+ y$:

$$\mathbb{E}_\xi||x_n^0 - x_n^\delta||^2 = \delta^2 \sum_{k=1}^{n} s_k^{-2} \mathbb{E}_\xi \langle v_k, \xi\rangle^2 = \delta^2 \sum_{k=1}^{n} s_k^{-2}$$

Keeping in mind that

$$x = A^+ y = \sum_{k=1}^{\infty} s_k^{-1} u_k \langle v_k, y\rangle$$

we can represent the mean squared error of the approximation x_n^δ as

$$\mathbb{E}_\xi||x - x_n^\delta||^2 = ||x - x_n^0||^2 + \mathbb{E}_\xi||x_n^0 - x_n^\delta||^2 = \sum_{k=n+1}^{\infty} s_k^{-2}\langle v_k, y\rangle^2 + \delta^2 \sum_{k=1}^{n} s_k^{-2}.$$

In view of the Picard criterion (1.4) the term $||x - x_n^0||^2$ converges to zero for $n \to \infty$, but the convergence rate is usually unknown because it depends on the exact y while only noisy data y^δ are available. At the same time the stability term $\mathbb{E}_\xi||x_n^0 - x_n^\delta||^2$ tends to infinity as $n \to \infty$ and its rate is known or can be estimated.

The important question is how big one should choose the rank n playing the role of the regularization parameter. The optimal n should balance $\mathbb{E}_\xi||x_n^0 - x_n^\delta||^2$ with the (unknown) value of $||x - x_n^0||^2$. In the present paper we discuss this question in a much more general situation, i.e. when A_n is not necessarily a partial sum of (1.2).

On the other hand the regularization method (2.1), known as a spectral cut-off scheme, is not just of academic interest. There are several practical interesting situations where a singular value decomposition of the operator A is known. For example, the downward-continuation problem [5] arising in satellite geodesy can be formulated as an equation (1.1), where x is a gravitational potential on the Earth's surface Ω_R (for reasons of simplicity now assumed as sphere with radius R), y is the same potential but observed on the satellite orbit Ω_r (also assumed to be a sphere with radius $r > R$) and y^δ as its noisy counterpart.

Indeed, it is well-known [7, 18] that in the closed outer space $\overline{\Omega}_R^{ext}$ of Ω_R a gravitational potential x can be represented as an expansion

$$x(t) = \sum_{m=0}^{\infty} \sum_{j=1}^{2m+1} H_m^j(R,t)\langle x, H_m^j(R,\cdot)\rangle, \qquad t \in \overline{\Omega}_R^{ext}$$

in terms of outer harmonics

$$H_m^j(R,t) = \left(\frac{R}{|t|}\right)^m \frac{1}{|t|} Y_m^j\left(\frac{t}{|t|}\right),$$

where $\{Y_m^j\}$ is a system of spherical harmonics L_2 orthonormalized with respect to the unit sphere in $\mathbb{R}^3$, $\langle\cdot,\cdot\rangle$ and $|\cdot|$ are the standard inner product and norm on $L_2(\Omega_R)$ and $\mathbb{R}^3$ respectively.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

Then on the satellite orbit Ω_r where $x(t) = y(t)$ with $|t| = r$ by definition, it reads as follows:

$$y(t) = \sum_{m=0}^{\infty} \left(\frac{R}{r}\right)^m \sum_{j=1}^{2m+1} \frac{1}{r} Y_m^j\left(\frac{t}{r}\right) \left\langle x, \frac{1}{R} Y_m^j\left(\frac{\cdot}{R}\right)\right\rangle.$$

This relation is nothing but the equation (1.1) with operator $A : L_2(\Omega_R) \to L_2(\Omega_r)$ admitting a singular value decomposition (1.2) where $v_k = \frac{1}{r} Y_m^j(\frac{\cdot}{r})$, $u_k = \frac{1}{R} Y_m^j(\frac{\cdot}{R})$ and $s_k = (\frac{R}{r})^m$ with $k = j + m^2$, $j \in \{1, \ldots, 2m+1\}$ and $m \in \mathbb{N}$.

It is common belief that the perturbations in the satellite data are caused by random noise. A white noise model has been discussed in this context in [6]. In § 3 we will use the above mentioned downward-continuation problem for testing our parameter choice strategy.

2.2 Regularization with complete information

To discuss the regularization of the problem (1.1) by its finite dimensional approximation (self-regularization) we assume that there is a sequence of operators $A_n : \mathcal{X} \to \mathcal{Y}$, $\operatorname{rank}(A_n) = n$ converging to A in the operator norm. Then for each n one can construct a minimal norm solution

$$x_n^\delta = A_n^+ y^\delta = \sum_{k=1}^{n} s_{k,n}^{-1} u_{k,n} \langle v_{k,n}, y^\delta\rangle$$

of the normal equation (1.5), where $\{s_{k,n}, u_{k,n}, v_{k,n}\}$ are the terms from the singular value decomposition of the operator A_n. Of course, to construct x_n^δ there is no need to know these terms.

If $\{f_{j,n}\}$ is any basis of $\operatorname{Range}(A_n)$ then A_n can be represented as

$$A_n = \sum_{j=1}^{n} f_{j,n} \langle h_{j,n}, \cdot\rangle$$

where $\{h_{j,n}\}$ forms a basis in $\operatorname{Range}(A_n^*)$. Using $\{h_{j,n}\}$ one can seek for x_n^δ in the form

$$x_n^\delta = \sum_{k=1}^{n} x_{k,n} h_{k,n} \tag{2.2}$$

where the unknown coefficients $\{x_{k,n}\}$ are defined from the following system of linear algebraic equations:

$$\sum_{k=1}^{n} a_{i,k} x_{k,n} = \langle f_{i,n}, y_\delta\rangle, \qquad i \in \{1, \ldots, n\}$$

with matrix $\mathbb{A}_n = \{a_{i,k} = \sum_{j=1}^{n} \langle f_{i,n}, f_{j,n}\rangle \langle h_{j,n}, h_{k,n}\rangle\}_{i\in\{1,\ldots,n\}, k\in\{1,\ldots,n\}}$.

This system is a discrete analogue of (1.5). It always has at least one solution and we pick out one that gives an element (2.2) with a minimal norm. Let $x_n^0 = A_n^+ y$ be a minimal norm solution of (1.5), where y^δ is replaced by the exact y. Then

$$x_n^\delta = A_n^+ (y + \delta\xi) = x_n^0 + \delta A_n^+ \xi.$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

Note that under our assumptions

$$\eta_n^\xi = A_n^+ \xi = \sum_{k=1}^{n} s_{k,n}^{-1} \langle v_{k,n}, \xi \rangle u_{k,n}$$

is a zero mean Gaussian random process which takes the values in $\mathcal{X}$ because it is a finite linear combination of orthonormal elements with Gaussian random coefficients $\gamma_k = \langle v_{k,n}, \xi \rangle$, $\mathbb{E}_\xi \gamma_k = 0$, $\mathbb{E}_\xi \gamma_k^2 = 1$ for $k \in \{1, \dots, n\}$.

In view of this example the following assumption seems to be natural:

Assumption 2.1 *Assume that there exist continuous decreasing functions* $\rho, \psi : [1, \infty[\to [0, \infty[$, $\lim_{n\to\infty} \rho(n) = \lim_{n\to\infty} \psi(n) = 0$ *which satisfy, for all* $n \in \mathbb{N}$,

- $\mathbb{E}_\xi ||\eta_n^\xi||^2 \leqslant \frac{1}{\rho^2(n)}$;
- $||x - x_n^0|| \leqslant \psi(n)$.

Moreover, assume that for some $0 < c < 1$ *and all* $t \in [1, \infty[$

- $\rho(t+1) \geqslant c\rho(t)$.

We will not assume that $\psi(n)$ is known, because it depends on unavailable exact data y. At the same time it will be shown that our algorithm adapts to unknown ψ and it is almost order optimal in the sense of accuracy expressed in terms of ψ. However, the function ρ is assumed to be known. One reason for this is that

$$\mathbb{E}_\xi ||\eta_n^\xi||^2 = \sum_{k=1}^{n} s_{k,n}^{-2},$$

and as in our illuminating example the singular values of A_n can be, in principle, calculated or estimated for each fixed n. At the same time, if for any n the operator $(A_n^* A_n)^{-1}$ exists, i.e. the matrix $\mathbb{A}_n$ is not degenerate (set $\mathbb{A}_n^{-1} = \{a_{i,k}^{(-1)}\}_{i\in\{1,\dots,n\}, k\in\{1,\dots,n\}}$) then the trace

$$\operatorname{tr} \left(A_n^* A_n\right)^{-1} = \sum_{k=1}^{n} s_{k,n}^{-2} = \sum_{k=1}^{n} a_{k,k}^{(-1)}$$

can be effectively estimated by Monte-Carlo methods in the same way as in Hutchinson [10].

Now we can determine the optimal regularization parameter via the following result which still needs the input of the unknown ψ.

Lemma 2.1 *Assume* $\delta < \psi(1)\rho(1)$. *When choosing*

$$n_{opt} = \min \left\{ n \in \mathbb{N} : \ \psi(n) \leqslant \frac{\delta}{\rho(n)} \right\}$$

we have

$$\sqrt{\mathbb{E}_\xi ||x - x_{n_{opt}}^\delta||^2} \leqslant \frac{\sqrt{2}}{c} \psi((\psi\rho)^{-1}(\delta)).$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

For a more detailed version of the proof, see Bauer [2]:

Proof 1 Using Assumption 2.1 we get

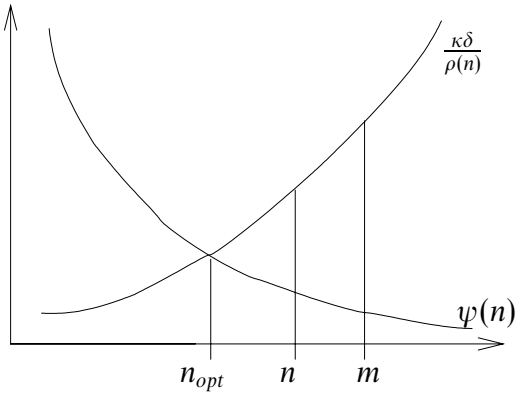
$$\mathbb{E}_\xi ||x - x_n^\delta||^2 \leqslant \psi^2(n) + \frac{\delta^2}{\rho^2(n)}.$$

Balancing for the best possible order of accuracy yields that we need a n_0 satisfying $\psi(n_0)\rho(n_0) = \delta$. Moreover owing to the monotonicity we have $\psi(n_{opt})\rho(n_{opt}) \leqslant \delta = \psi(n_0)\rho(n_0)$. However, we have $\psi(n_{opt} - 1)\rho(n_{opt} - 1) > \delta = \psi(n_0)\rho(n_0)$, which yields

$$\begin{aligned}\mathbb{E}_\xi ||x - x_{n_{opt}}^\delta||^2 &\leqslant \psi^2(n_{opt}) + \frac{\delta^2}{\rho^2(n_{opt})} \leqslant \frac{2\delta^2}{\rho^2(n_{opt})} \\ &\leqslant \frac{2\psi^2(n_0)\rho^2(n_0)}{\rho^2(n_0+1)} \leqslant \frac{2}{c^2}\psi^2(n_0) = \frac{2}{c^2}\psi^2((\psi\rho)^{-1}(\delta)).\end{aligned}$$

2.3 Regularization without knowledge of ψ

The above result cannot be used in the case when we do not know ψ. Therefore consider for $n < m$ and $n, m \in \{k : \psi(k) \leqslant \frac{\kappa\delta}{\rho(k)}\}$ the following picture; κ will be a tuning parameter:



Then we have in view of Assumption 2.1 for the "expected" behaviour of the random variable $||\eta_n^\xi||$:

$$\begin{aligned}||x_n^\delta - x_m^\delta|| &\leqslant ||x - x_n^\delta|| + ||x - x_m^\delta|| \leqslant \psi(n) + \delta||\eta_n^\xi|| + \psi(m) + \delta||\eta_m^\xi|| \\ &\leqslant \frac{2\kappa\delta}{\rho(n)} + \frac{2\kappa\delta}{\rho(m)} \leqslant \frac{4\kappa\delta}{\rho(m)}.\end{aligned}$$

Now we use an idea by Lepskij [12] and take, using $N = \lceil \rho^{-1}(\delta) \rceil$,

$$n_* = \min\left\{ n \in \mathbb{N} : \ ||x_n^\delta - x_m^\delta|| \leqslant \frac{4\kappa\delta}{\rho(m)}, \ N > m > n \right\}$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

Before starting with the main results we need the following supporting lemma:

Lemma 2.2 *The following probability estimate holds for all* $\tau \geqslant 1$*:*

$$\mathbb{P}_\xi \left\{ ||\eta_n^\xi|| \rho(n) > \tau \right\} \leqslant 4 \exp\left(-\frac{\tau^2}{8} \right).$$

Proof 2 We have, using $\mathbb{E}_\xi ||\eta_n^\xi||^2 \leqslant \frac{1}{\rho^2(n)}$ and the probability estimate for Gaussian random vectors [13],

$$\begin{aligned} \mathbb{P}_\xi \left\{ ||\eta_n^\xi|| \rho(n) > \tau \right\} &= \mathbb{P}_\xi \left\{ ||\eta_n^\xi|| > \frac{\tau}{\rho(n)} \right\} \\ &\leqslant 4 \exp\left(-\frac{\tau^2}{8\rho(n)^2 \mathbb{E}_\xi ||\eta_n^\xi||^2} \right) \leqslant 4 \exp\left(-\frac{\tau^2}{8} \right). \end{aligned}$$

Theorem 2.3 *Let* $\delta < \psi(1)\rho(1)$ *and* n_* *be chosen as above with* $\kappa \geqslant 1$*. Then we have:*

$$\mathbb{E}_\xi ||x - x_{n_*}^\delta||^2 \leqslant C_1 (\rho^{-1}(\delta)) \exp\left(-\frac{\kappa^2}{16} \right) + C_2 \kappa^2 \psi^2((\psi\rho)^{-1}(\delta)),$$

where C_1 *and* C_2 *are constants independent of the other quantities.*

Proof 3 Define

$$\Xi_\rho(\omega) = \max_{1 \leqslant n \leqslant N} ||\eta_n^\xi|| \rho(n)$$

and divide the probability space into two subspaces:

$$\Omega_\kappa = \{ \omega : \Xi_\rho(\omega) \leqslant \kappa \} \quad \text{and} \quad \overline{\Omega_\kappa} = \Omega \setminus \Omega_\kappa$$

The proof mainly consists of two different parts. The first one estimates the behaviour for all "nice" cases Ω_κ. The second one deals with the "bad" cases, where the stochastic noise property produces results far away from the average. Therefore the second part has a strong emphasis on the probability that this case actually occurs.

Part 1 ("good" event $\omega \in \Omega_\kappa$). We want to show that $n_{opt} \geqslant n_*$. For all $n \geqslant n_{opt}$ we have using $\frac{\kappa\delta}{\rho(n)} \geqslant \psi(n)$ and $\rho(n) \leqslant \rho(n_{opt})$ the following:

$$\begin{aligned} ||x_n^\delta - x_{n_{opt}}^\delta|| &\leqslant ||x - x_n^\delta|| + ||x - x_{n_{opt}}^\delta|| \leqslant \psi(n) + \delta ||\eta_n^\xi|| + \psi(n_{opt}) + \delta ||\eta_{n_{opt}}^\xi|| \\ &\leqslant \psi(n) + \frac{\kappa\delta}{\rho(n)} + \psi(n_{opt}) + \frac{\kappa\delta}{\rho(n_{opt})} \leqslant \frac{2\kappa\delta}{\rho(n)} + \frac{2\kappa\delta}{\rho(n_{opt})} \leqslant \frac{4\kappa\delta}{\rho(n)}, \end{aligned}$$

which tells us that $n_* = \min\{ n \in \mathbb{N} : ||x_n^\delta - x_m^\delta|| \leqslant \frac{4\kappa\delta}{\rho(n)},\ N > m > n \} \leqslant n_{opt}$. Then using

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

$n_{opt} \geqslant n_*$ we have, for all $\omega \in \Omega_\kappa$,

$$||x - x^\delta_{n_*}|| \leqslant ||x - x^\delta_{n_{opt}}|| + ||x^\delta_{n_{opt}} - x^\delta_{n_*}|| \leqslant \psi(n_{opt}) + \frac{\kappa\delta}{\rho(n_{opt})} + \frac{4\kappa\delta}{\rho(n_{opt})}$$
$$\leqslant \frac{2\kappa\delta}{\rho(n_{opt})} + \frac{4\kappa\delta}{\rho(n_{opt})} \leqslant 6\frac{\kappa}{c}\left(c\frac{\delta}{\rho(n_{opt})}\right) \leqslant 6\frac{\kappa}{c}\psi((\psi\rho)^{-1}(\delta)).$$

Hence we get

$$\int_{\Omega_\kappa} ||x - x^\delta_{n_*}||^2 d\mathbb{P}_\xi(\omega) \leqslant |\Omega_\kappa| \; ||x - x^\delta_{n_*}||^2 \leqslant 36\frac{\kappa^2}{c^2}\psi^2((\psi\rho)^{-1}(\delta)).$$

Part 2 ("bad" event $\omega \in \overline{\Omega_\kappa}$). Remember that we defined $n_{opt} \leqslant N = \rho^{-1}(\delta)$. Hence we get $\psi(N) \leqslant \frac{\delta}{\rho(N)} = 1$. Then

$$||x - x^\delta_{n_*}|| \leqslant ||x - x^\delta_N|| + ||x^\delta_N - x^\delta_{n_*}|| \leqslant \psi(N) + \delta||\eta^\xi_N|| + \frac{4\kappa\delta}{\rho(N)}$$
$$\leqslant \delta||\eta^\xi_N|| + \frac{5\kappa\delta}{\rho(N)} \leqslant \frac{\delta||\eta^\xi_N||\rho(N)}{\rho(N)} + 5\kappa \leqslant 6\Xi_\rho.$$

Using this result we obtain

$$\int_{\overline{\Omega_\kappa}} ||x - x^\delta_{n_*}||^2 d\mathbb{P}_\xi(\omega) \leqslant 36\int_{\overline{\Omega_\kappa}} \Xi^2_\rho(\omega) d\mathbb{P}_\xi(\omega) \leqslant 36\sqrt{\int_{\overline{\Omega_\kappa}} \Xi^4_\rho(\omega) d\mathbb{P}_\xi(\omega)}\sqrt{\int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}_\xi(\omega)}.$$

Now we estimate the two parts separately. Consider $F(\tau) = \mathbb{P}_\xi\{\Xi_\rho(\omega) \leqslant \tau\}$ for $\tau > \kappa$. Then

$$G(\tau) = 1 - F(\tau) = \mathbb{P}_\xi\{\Xi_\rho(\omega) > \tau\} \leqslant \sum_{n=1}^{N} \mathbb{P}_\xi\left\{||\eta^\xi_n||\rho(n) > \tau\right\} \leqslant 4N\exp\left(-\frac{\tau^2}{8}\right).$$

So we get

$$\int_{\overline{\Omega_\kappa}} \Xi^4_\rho d\mathbb{P}_\xi(\omega) = \int_\kappa^\infty \tau^4 dF(\tau) \leqslant \int_0^\infty \tau^4 \, dF(\tau) = -\int_0^\infty \tau^4 dG(\tau)$$
$$= -\tau^4 G(\tau)|_0^\infty + 4\int_0^\infty \tau^3 G(\tau)\, d\tau = 4\int_0^\infty \tau^3 G(\tau)\, d\tau$$
$$\leqslant 16N\int_0^\infty \tau^3 \exp\left(-\frac{\tau^2}{8}\right) d\tau = 2^9 N\int_0^\infty u\exp(-u)\, du = 2^9 N.$$

The other part gives:

$$\int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}_\xi(\omega) \leqslant 4N\exp\left(-\frac{\kappa^2}{8}\right).$$

Hence we obtain

$$\int_{\overline{\Omega_\kappa}} \Xi^2_\rho d\mathbb{P}_\xi(\omega) \leqslant 2^{11/2} N\exp\left(-\frac{\kappa^2}{16}\right) \leqslant 2^{11/2}(\rho^{-1}(\delta))\exp\left(-\frac{\kappa^2}{16}\right).$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

This yields

$$\mathbb{E}_\xi ||x - x_{n_*}^\xi||^2 \leqslant 36 \cdot 2^{11/2}(\rho^{-1}(\delta)) \exp\left(-\frac{\kappa^2}{16}\right) + 36\frac{\kappa^2}{c^2}\psi^2((\psi\rho)^{-1}(\delta))$$

which proves the theorem.

As we have seen, the inequality for the overall error consists of two parts. The second part $\psi^2((\psi\rho)^{-1}(\delta))$ is just the best order of accuracy we can reach. So we want that the first part $(\rho^{-1}(\delta))\exp(-\frac{\kappa^2}{16})$ is negligible in comparison to the second one. Therefore our main task now is to choose an appropriate κ.

We will consider two scenarios which are usually discussed in the theory of ill-posed problems [4]. One of them corresponds to the case when both functions $\psi(n)$ and $\rho(n)$ exhibit a polynomial behaviour. This means that there are positive constants μ_1, μ_2, ν_1 and ν_2 such that for all n

$$n^{-\mu_1} \leqslant \psi(n) \leqslant n^{-\mu_2}, \qquad n^{-\nu_1} \leqslant \rho(n) \leqslant n^{-\nu_2}. \tag{2.3}$$

This scenario is known as moderate ill-posedness.

The second scenario goes under the name of severe ill-posedness and corresponds to the case of exponential behaviour of $\rho(n)$ and polynomial behaviour of $\psi(n)$, i.e.

$$n^{-\mu_1} \leqslant \psi(n) \leqslant n^{-\mu_2}, \qquad \exp(-n^{\nu_1}) \leqslant \rho(n) \leqslant \exp(-n^{\nu_2}). \tag{2.4}$$

It is well known that the downward-continuation problem discussed in § 2.1 is severely ill-posed [6].

Corollary 2.4 *Assume that the conditions of Theorem 2.3 and* (2.4) *hold. Then for sufficiently small* δ *there are constants* c_3 *and* c_4 *such that*

$$c_3(\ln \delta^{-1})^{-\frac{\mu_1}{\nu_2}} \leqslant \psi((\psi\rho)^{-1}(\delta)) \leqslant c_4(\ln \delta^{-1})^{-\frac{\mu_2}{\nu_1}}$$

and, for $\kappa = 4\sqrt{c_5 \ln\ln \delta^{-1}}$ *where* $c_5 \geqslant \frac{2\mu_1+1}{\nu_2}$,

$$\mathbb{E}_\xi ||x - x_{n_*}^\xi||^2 \leqslant c \ln\ln(\delta^{-1})\psi^2((\psi\rho)^{-1}(\delta)),$$

where c *does not depend on* δ.

Proof 4 Let $n_0 = (\psi\rho)^{-1}(\delta)$. Then assuming that δ is sufficiently small that $((\psi\rho)^{-1}(\delta))^{\nu_1} \geqslant \ln(\psi\rho)^{-1}(\delta)$ we have

$$n_0^{\nu_2} \leqslant \ln\frac{1}{\psi(n_0)} + \ln\frac{1}{\rho(n_0)} = \ln(\delta^{-1}) \leqslant n_0^{\nu_1} + \mu_1 \ln n_0 \leqslant (1+\mu_1)n_0^{\nu_1}.$$

This means in particular that

$$(1+\mu_1)^{-\frac{1}{\nu_1}}(\ln \delta^{-1})^{\frac{1}{\nu_1}} \leqslant n_0 \leqslant (\ln \delta^{-1})^{\frac{1}{\nu_2}}$$

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

and

$$(\ln \delta^{-1})^{-\frac{\mu_1}{\nu_2}} \leqslant n_0^{-\mu_1} \leqslant \psi((\psi\rho)^{-1}(\delta)) \leqslant n_0^{-\mu_2} \leqslant (1+\mu_1)^{\frac{\mu_2}{\mu_1}} (\ln \delta^{-1})^{-\frac{\mu_2}{\nu_1}}.$$

However, $\rho^{-1}(\delta) \leqslant (\ln \delta^{-1})^{\frac{1}{\nu_2}}$. Then for $\kappa = 4\sqrt{c_5 \ln \ln \delta^{-1}}$, where $c_5 \geqslant \frac{2\mu_1+1}{\nu_2}$, we get

$$(\rho^{-1}(\delta)) \exp\left(-\frac{\kappa^2}{16}\right) \leqslant (\ln \delta^{-1})^{\frac{1}{\nu_2}} (\ln \delta^{-1})^{-c_5} \leqslant (\ln \delta^{-1})^{-\frac{2\mu_1}{\nu_2}} \leqslant \psi^2((\psi\rho)^{-1}(\delta))$$

and the required estimate follows from Theorem 2.3.

Corollary 2.5 *Assume that the conditions of Theorem 2.3 and* (2.3) *hold. Then*

$$\delta^{\frac{\mu_1}{\nu_2+\mu_2}} \leqslant \psi((\psi\rho)^{-1}(\delta)) \leqslant \delta^{\frac{\mu_2}{\nu_1+\mu_1}}$$

and, for $\kappa = 4\sqrt{c_6 \ln \delta^{-1}}$, $c_6 \geqslant 2\frac{\nu_2\mu_1+\nu_2+\mu_2}{\nu_2^2+\mu_2\nu_2}$,

$$\mathbb{E}_\xi ||x - x_{n_*}^\xi||^2 \leqslant c \ln(\delta^{-1}) \psi^2((\psi\rho)^{-1}(\delta)).$$

The proof is analogous to the previous one. These corollaries show that the adaptive parameter choice $n = n_*$ allows to reach almost optimal order of accuracy for both types of ill-posedness. An additional loglog (respectively log) factor does not essentially increase the error compared to the optimal one and can be considered as a payment for the lack of knowledge of ψ.

3 Numerics

We have tested our method using simulated satellite data. This has several particular advantages. Its a severely ill-posed problem where we exactly know the singular values and can estimate $\rho(n)$ easily (see illuminating example). Furthermore, the use of simulated data allows to compute signal to error ratios which will enable us to compare and evaluate the method more easily.

We assumed our data to be given on an integration grid on a sphere. This has the advantage that we do not have to bother about the (ill-posed) problem of transferring data from a satellite track to such a grid and we consequently avoid several sources of additional error. Furthermore this enables us to study our new methods in an unbiased environment.

3.1 Technical remarks

As data location we used a Driscoll-Healy grid [14] at an orbit height of 3% and 6% of the Earth radius which roughly corresponds to an average satellite height of 200 km (respectively 400 km). For approximation we used spherical harmonics up to degree 128 and we generated the data globally on a grid which allows exact integration up to degree 180 with a stable Clenshaw algorithm [3]. The model EGM96 was always used as input and reference data. The noise level was chosen in a way such that theoretically the bias to variance ratio had to pass 1.0 around the degree of 80; we used a combination of correlated and uncorrelated noise in the space domain.

As regularization method we chose the spectral cut-off scheme cutting at each degree.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

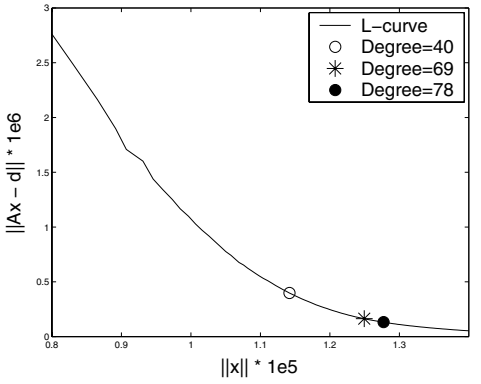


FIGURE 1. Data set at 200 km.

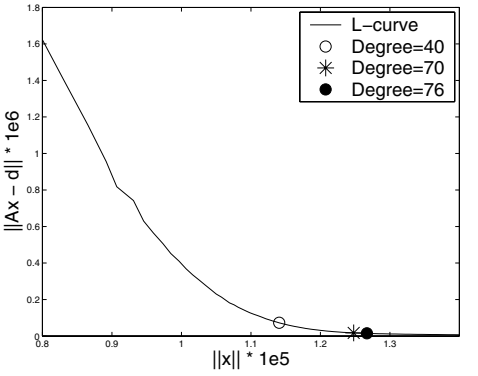


FIGURE 2. Data set at 400 km.

3.2 Some notation and the L-curve method

The given data at satellite height (projected to the space of spherical harmonics to the maximal degree of 128) is called d, our regularized solution x and the upward continuation operator A. Hence the size of x is denoted $||x||_2$ and the error occurring when choosing x is $||Ax - d||_2$.

The L-curve method now tries to estimate the point, where the curvature of $(||x||_2, ||Ax - d||_2)$ is maximal. As we will see in the next pictures this point is rather hard to obtain, especially because there are a big number of possible points nearby. Within these limitations we want to see our guess for an appropriate regularization parameter using the L-curve method. The point $\circ$ marked in the Figures 1 and 2 could also mark another degree in the range of 35–45 without really changing too much. But this also signified that the choice of the regularization parameter via this method is sometimes made at random. Note that the same (or even worse) problems occurred when we tried to use a log-log scale for determining the optimal regularization point via the L-curve method.

For a better readability of both the Table 1 and the Figures 1–3 we rescaled the occurring values by a factor of $(10^5, 10^6)$. The $\frac{\text{Bias}}{\text{Variance}}$ ratio is displayed for each degree of the solution. The optimal regularization point is where this ratio first exceeds 1.

3.3 Data

We present three different representations of the data: a table of data values where we chose the most interesting region due to space restrictions, the L-curves corresponding to our two input data sets and bias/variance behaviour where we chose a log scale for better observability.

In Figures 1–3 we display the optimal regularization point (i.e., $\frac{bias}{variance} = 1$) by $\bullet$, the regularization point proposed by the L-curve method by $\circ$ and the regularization point found by the self-regularization method by $*$.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

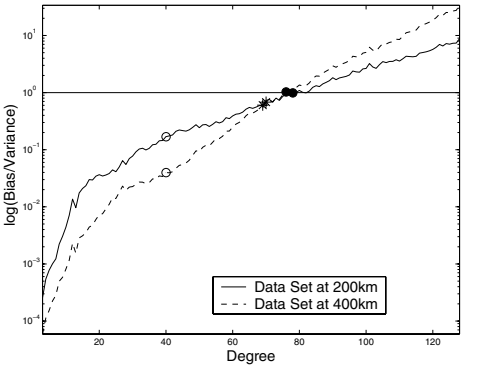


FIGURE 3. Bias/Variance ratio with respect to the degrees.

Table 1. *Numerical results*

Degree	Data Set 1 $\|x\|_2$	Data Set 1 $\|Ax-d\|_2$	Data Set 1 $\frac{\text{Bias}}{\text{Variance}}$	Data Set 2 $\|x\|_2$	Data Set 2 $\|Ax-d\|_2$	Data Set 2 $\frac{\text{Bias}}{\text{Variance}}$
3	0.2970	9.3820	0.0003	0.2970	7.4766	0.0001
...	...	...	...	...	...	...
39	1.1372	0.4137	0.1452	1.1364	0.0771	0.0340
40	1.1417	0.3977	**0.1695**	1.1408	0.0721	**0.0399**
41	1.1462	0.3840	0.1747	1.1452	0.0679	0.0393
...	...	...	...	...	...	...
67	1.2429	0.1719	0.5576	1.2388	0.0191	0.4727
68	1.2460	0.1676	0.5925	1.2418	0.0185	0.5192
69	1.2495	0.1634	**0.6020**	1.2451	0.0179	0.5430
70	1.2523	0.1589	0.7311	1.2479	0.0173	**0.6805**
71	1.2553	0.1553	0.7592	1.2510	0.0169	0.7270
72	1.2586	0.1516	0.6886	1.2540	0.0164	0.6788
73	1.2616	0.1477	0.8026	1.2571	0.0159	0.8126
74	1.2648	0.1442	0.7290	1.2607	0.0155	0.7578
75	1.2681	0.1406	0.8573	1.2640	0.0150	0.9144
76	1.2711	0.1371	0.9374	1.2671	0.0146	**1.0299**
77	1.2742	0.1339	0.9501	1.2705	0.0142	1.0720
78	1.2774	0.1307	**0.9901**	1.2739	0.0138	1.1522
79	1.2806	0.1274	1.0123	1.2775	0.0135	1.2194
80	1.2838	0.1244	1.0903	1.2811	0.0131	1.3589
81	1.2872	0.1214	1.0126	1.2849	0.0128	1.3088
...	...	...	...	...	...	...
128	1.5666	0.0024	8.8348	1.9936	0.0002	33.8520

3.4 Discussion

The data above indicate that the self-regularization method is at least not worse and perhaps even superior to the L-curve method, which itself has been proven to be reliable in a wide number of cases.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

Furthermore, we have observed that in almost all computations the method was reliable and we got a cutting point in the range of $\frac{bias}{variance} \in [0.5, 1.5]$ for the satellite case. This means that our adaptive parameter choice is worse by only a constant factor in comparison to the ideal one.

Note, that ill-conditioned systems of equations (originated by a discretisation of operator equations) are not new in geodesy, in particular when satellite observations are used for gravity field estimation. Usually Kaula stabilisation is used; that is, we add some regularization matrix to the matrix of initial system. Such stabilization scheme can be viewed as a particular case of Tikhonov regularization method. It has been discussed recently in Kusche & Klees [11].

At the same time, in Pail & Plank [17] it has been observed that an inappropriate choice of the regularization matrix for Tikhonov regularization can destroy a block-diagonal structure of the original normal equation system. Therefore, it would be desirable to treat the normal equations as they are. However, as it has been reported in Kusche & Klees [11], the straightforward solution of the full normal equation system does not provide physically meaningful gravity field models. Instead, regularization is indispensable. In this situation self-regularization combined with an adaptive parameter selection strategy could be the method of choice, and our numerical tests show that it is rather robust against correlated and uncorrelated noise.

References

[1] Bakushinskij, A. (1984) Remarks on the choice of the regularization parameter using quasi-optimality and a quotient criterion. *Computational Math. & Math. Phys.* **4**, 181–182.

[2] Bauer, F. (2004) An Alternate Approach to the Oblique Derivative Problem in Potential Theory. PhD thesis, Universtiy of Kaiserslautern, Department of Mathematics, Geomathematics Group.

[3] Deakin, R. (1998) Derivatives of the Earth's potential. *Geomatics Res. Australasia*, **68**, 31–60.

[4] Engl, H., Hanke, M. & Neubauer, A. (1996) *Regularization of Inverse Problems*. Mathematics and Its Applications, Kluwer.

[5] Freeden, W., Gervens, T. & Schreiner, M. (1998) *Constructive Approximation on the Sphere* (*With Applications to Geomathematics*). Numerical Mathematics and Scientific Computation, Oxford University Press.

[6] Freeden, W. & Pereverzev, S. (2001) Spherical Tikhonov regularization wWavelets in satellite gravity gradiometry with random noise. *J. Geodesy*, **74**, 730–736.

[7] Freeden, W. (1999) *Multiscale Modelling of Spaceborne Geodata*. European Consortium for Mathematics in Industry. Teubner.

[8] Goldenshluger, A. & Pereverzev, S. (2000) Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory & Related Fields*, **118**(2), 169–186.

[9] Harbrecht, H., Pereverzev, S. & Schneider, R. (2002) Self-regularization by projection for noisy pseudodifferential equations of negative order. *Numerische Mathematik*, **95**, 123–143.

[10] Hutchinson, M. F. (1990) A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Stat., Simulation Comput*. **19**.

[11] Kusche, J. & Klees, R. (2002) Regularization of gravity field estimation from satellite gravity gradients. *J. Geodesy*, **76**, 359–368.

[12] Lepskij, O. (1990) On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & its Applic*. **35**(3), 454–466.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press

[13] Ledoux, M. & Talagrand, M. (1991) *Probability in Banach spaces: Isoperimetry and Processes.* Springer.
[14] Mayer, C. (2001) Potential Operators, Jump Relations and Wavelets. Diploma thesis, University of Kaiserslautern.
[15] Mathé, P. & Pereverzev, S. (2002) Direct estimation of linear functionals from indirect noisy observations. *J. Complexity*, **18**(2), 501–516.
[16] Pereverzev, S. (2003) *Regularization Methods for Ill-Posed Problems of Analysis and Statistics.* Lecture Notes.
[17] Pail, R. & Plank, G. (2002) Assessment of three numerical solution strategies for gravity field recovery from GOCE satellite gravity gradiometry implemented on a parallel platform. *J. Geodesy*, **76**, 462–474.
[18] Schwintzer, P. (1997) Earth gravity field recovery from satellite orbit perturbations. *Geowissenschaften*, **15**, 85–89.
[19] Skorokhod, A. V. (1974) *Integration in Hilbert Space.* Springer.

https://doi.org/10.1017/S0956792505006236 Published online by Cambridge University Press