B02

Proceedings of the 5th International Conference on Inverse Problems in Engineering: Theory and Practice, Cambridge, UK, 11-15th July 2005

AUTOMATIC REGULARIZATION FOR ILL-POSED PROBLEMS WITH STOCHASTICAL NOISE ESTIMATE

F. BAUER

University of Göttingen, Department of Mathematics, Institute for Numerical and Applied Mathematics, Lotzestr. 16-18, 37083 Göttingen, Germany

e-mail: bauer@mathematik.uni-kl.de

Abstract - We consider the compact operator $A : \mathcal{X} \longrightarrow \mathcal{Y}$ for the separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. The problem $Ax = y$ is called (exponentially) ill-posed when the singular values σ_n of the operator A tend to zero (with exponential rate).

Classically one assumes that y is biased with "deterministic noise", i.e. that we have $y^\delta = y + \xi$ with $||\xi|| < \delta$. Instead we will assume to have "stochastical noise", i.e. $y^\delta = y + \xi$ where for all $f \in \mathcal{Y}$ we have that $\xi_f =< f, \xi >_{\mathcal{Y}}$ is a Gaussian random variable fulfilling $\mathbb{E} \langle f, \xi \rangle_{\mathcal{Y}} = 0$ and $\mathbb{E} \left(\langle f, \xi \rangle_{\mathcal{Y}} \right)^2 = \delta^2 (||f||_{\mathcal{Y}})^2$. This means that in terms of a Fourier expansion each Fourier coefficient is disturbed with Gaussian white noise. Regularization in this case is harder to perform than for the classical case.

For asymptotically optimal regularization with respect to $\delta \longrightarrow 0$ we have two important parameters: the smoothness of x and the error behavior of ξ under the application of the regularization operators. It is well-known that optimal regularization is possible when both smoothness and error (behavior) are known; however this situation normally does not occur in practice.

We will show that both for the "deterministic noise" case as well for the "stochastical noise" case we can regularize in an (asymptotically) optimal way (i.e. choosing a regularization parameter) just using the error behavior. We will furthermore show that we actually do not need complete knowledge on the error but can either use information obtained from several measurements or a small very accurately known part of the space domain. This means, in particular, that a repetition of measurements can be used not only for error reduction, but also for determining the regularization parameter in an optimal way. Another possibility is obtaining partial very accurate information (e.g. a test sample in some part) to optimally determine the regularization parameter.

A practical application to the "downward-continuation" problem for satellite observed gravitational data in geosciences will be shown.

1. INTRODUCTION

The primary objective of modern satellite missions like GOCE and CHAMP is determining the geopotential field precisely with high spatial resolution. In particular its knowledge is the base of further geoscientifical investigations like prospecting, exploration, solid earth physics or physical oceanography. For further information on this topic the reader is referred to [4, 5, 13] and the references therein.

A number of interesting mathematical problems is associated to this theme. We will concentrate on the following one [6]. When we have approximated the geopotential field v at the height r of the satellite orbit $x \in \Omega_r$ (for reasons of simplicity now assumed as sphere) by

$$v(x) = \sum_{n=0}^{\infty} \sigma_n \sum_{k=-n}^{n} v^\wedge(n,k) Y_n^k \left(\frac{x}{r} \right) \tag{1}$$

then it reads on the height R of the earth's surface $x \in \Omega_R$ (also assumed as sphere)

$$v(x) = \sum_{n=0}^{\infty} \sum_{k=-n}^{n} v^\wedge(n,k) Y_n^k \left(\frac{x}{R} \right) \tag{2}$$

where $\sigma_n = \left(\frac{R}{r}\right)^n$ and Y_n^k denote the standard spherical harmonics. So the downward-continuation (i.e. determination of the geopotential field out of satellite data) is a severely ill-posed problems because the eigenvalues σ_n of the downward-continuation operator $\Lambda_{R/r}$ decrease with exponential rate.

In order to solve this problem we need to regularize it. Because we are dealing with measurements it is sensible to assume that our data are biased with random noise [5]. In this text we will concentrate on the spectral cut-off scheme as regularization procedure, i.e. our regularized solution v_N reads

$$v_N(x) = \sum_{n=0}^{N} \sum_{k=-n}^{n} v^\wedge(n,k) Y_n^k \left(\frac{x}{R} \right) \tag{3}$$

The important question is how big one should choose the regularization parameter N; if it is too low we are far away from the real solution even if there would be no noise at all, if it is too high the noise completely conceals the data. If we know the smoothness of the solution (i.e. in which Sobolev space it lives) and the behavior of the noise w.r.t. the regularization operators we can regularize optimally. However, in practice we normally do neither know the smoothness nor the exact behavior of the error. Therefore we need to employ methods which are not straight forward.

We will show that some knowledge concerning the error behavior is sufficient in order to regularize in an asymptotically near to optimal way. Furthermore we will present how to use approximations to the error behavior in order to regularize optimally and show it to be working using a particular example from satellite gradiometry.

Please note that most results in the first part are based on the work of S. Pereverzev. He has shown that in the error regions we are interested in, the satellite problem is much more behaving like an ordinarily ill-posed problem [12]. Therefore we will mainly concentrate on this case. Theoretical results on the exponentially ill-posed case are available in [2, 1].

2. PARAMETER SELECTION FOR REGULARIZATION

2.1. Preliminaries and Notation

From now on, let $\mathcal{X}$ and $\mathcal{Y}$ be separable Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$ and with basis $\{u_k\}_{k \in \mathbb{N}}$ and $\{v_k\}_{k \in \mathbb{N}}$, respectively. If no confusion is likely to arise we will denote the inner product just by $\langle \cdot, \cdot \rangle$. Additionally assume $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ if not stated otherwise.

Furthermore assume A is a map $A : \mathcal{X} \to \mathcal{Y}$ which is a continuous linear operator with infinite rank. A shall admit a singular value decomposition

$$Ax = \sum_{k=1}^{\infty} s_k v_k \langle u_k, x \rangle \tag{4}$$

where $s_k \geq s_{k+1} > 0$ for all $k \in \mathbb{N}$, the positive natural numbers.

We will consider two different kinds of noise, namely the classical deterministic noise and the physically seen more sensible, but also more difficult to treat, stochastic noise case.

Definition 2.1 (Deterministic Noise)
The data y_δ are biased with deterministic noise (in comparison to y) if $||y - y^\delta|| \leq \delta$, i.e. there exists a $\xi \in \mathcal{Y}$ with $||\xi|| \leq 1$ such that $y^\delta = y + \delta\xi$.

Definition 2.2 (Stochastic Gaussian White Noise)
Let $(\Omega, \Sigma, \mathbb{P})$ be the ordinary probability space. Furthermore let $y^\delta = y + \delta\xi$, where ξ is a random vector fulfilling

- *For all $y \in \mathcal{Y}$ we have that $\xi_y(\omega) = \langle y, \xi \rangle$, where $\xi_y(\omega) : \Omega \to \mathbb{R}$ is a random variable. Assume furthermore $\forall t : \{\omega \,|\, \omega \in \Omega, \xi_y(\omega) \leq t\} \in \Sigma$*
- $\mathbb{E}_\xi \langle y, \xi \rangle = 0$
- $\mathbb{E}_\xi \langle y, \xi \rangle^2 = ||y||^2$
- *ξ_y is normally distributed around 0.*

Then y^δ is called to be biased with stochastic Gaussian white noise.

2.2. Regularization with all Information

Assume the sequence of regularization operators $\{A_n\}_{n\in\mathbb{N}}$. We will consider the following noisy solutions of our operator equation (noise element $\delta\xi$ with the standard formulation for a stochastical noise element ξ):

$$x_n^\delta = A_n^+(Ax+\delta\xi) = (A_n^*A_n)^{-1}A_n^*(Ax+\delta\xi) = x_n^0 + \delta\eta_n^\xi \tag{5}$$

where

$$\eta_n^\xi = A_n^+\xi = (A_n^*A_n)^{-1}A_n^*\xi \tag{6}$$

is a Gaussian random element. (The spectral cut-off scheme fulfills this property, e.g.). From now on we assume that there exist functions ρ and ψ fulfilling the following assumptions, which of course depends on the used regularization method:

Assumption 2.1

Assume that there exist decreasing functions $\rho, \psi : [1,\infty[\to [0,a]$, $\lim_{n\to\infty}\rho(n) = \lim_{n\to\infty}\psi(n) = 0$ which fulfill

- $\rho(n+1) \geq c\rho(n)$ *for a constant* c
- $\mathbb{E}_\xi ||\eta_n^\xi||^2 \leq \frac{1}{\rho(n)^2}$
- $||x - x_n^0|| \leq \psi(n)$.

Remark

The function ρ may be associated with how the operator A spreads the error over the various frequencies, whereas the function ψ is determined by the smoothness of the solution x (i.e. in which Sobolev space it lives, e.g.). Since $||\eta_n^\xi|| \leq 1/\rho(n)$ implies $\mathbb{E}_\xi||\eta_n^\xi||^2 \leq 1/\rho(n)^2$ we can restrict our attention to the stochastical noise case.

Please note that in principle $\psi(\cdot)$ can depend on x. As we assume that we do not know $\psi(\cdot)$, later on, this does not pose a problem. It is just important in order to formulate rates [2].

Now we can determine the optimal regularization parameter via the following result which still needs the input of smoothness and error level but works for the stochastical noise case:

Lemma 2.1

When choosing

$$n_{opt} = \min\left\{n : \ \psi(n) \leq \frac{\delta}{\rho(n)}\right\} \tag{7}$$

we have

$$\sqrt{\mathbb{E}_\xi||x - x_{n_{opt}}^\delta||^2} \leq \frac{\sqrt{2}}{c}\psi\left((\psi\rho)^{-1}(\delta)\right) \tag{8}$$

A more detailed proof can be found [2].

Proof

We have:

$$\mathbb{E}_\xi||x - x_n^\delta||^2 = ||x - x_n^0||^2 - 2\delta\mathbb{E}_\xi\left\langle x - x_n^0, (A_n^*A_n)^{-1}A_n^*\xi\right\rangle + \delta^2\mathbb{E}_\xi||\eta_n^\xi||^2 \leq \psi(n)^2 + \frac{\delta^2}{\rho(n)^2} \tag{9}$$

which holds because of Definition 2.2. Balancing for the best possible order of accuracy yields that we need a n_0 fulfilling $\psi(n_0)\rho(n_0) = \delta$. We have:

$$\psi(n_{opt})\rho(n_{opt}) \leq \delta = \psi(n_0)\rho(n_0) < \psi(n_{opt}-1)\rho(n_{opt}-1) \tag{10}$$

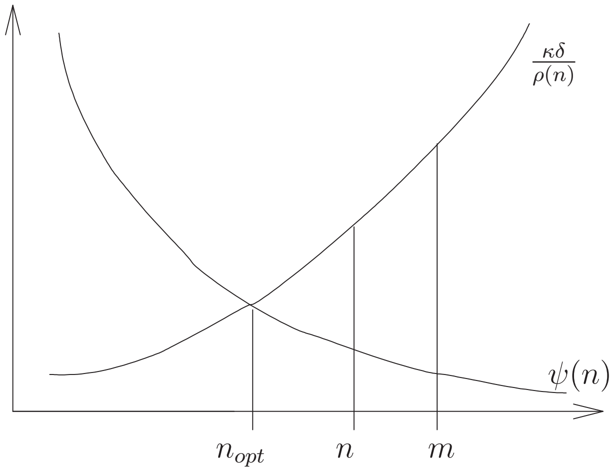
This yields

$$\mathbb{E}_\xi||x - x_{n_{opt}}^\delta||^2 \leq \psi(n_{opt})^2 + \frac{\delta^2}{\rho(n_{opt})^2} \leq \frac{2\psi(n_0)^2\rho(n_0)^2}{\rho(n_0+1)^2} \leq \frac{2}{c^2}\psi\left((\psi\rho)^{-1}(\delta)\right)^2 \tag{11}$$

q.e.d.

2.3. Regularization without Known Smoothness

The above result cannot be used in the case when we do not know the smoothness of our solution. Therefore consider for $n < m$ and $n, m \in \{k : \psi(k) \leq \frac{\kappa\delta}{\rho(k)}\}$ the following picture where we have, in view of Assumption 2.1, the "expected" behavior of the random variable $||\eta_n^\xi||$. Please note that the computation is not mathematically rigorous and is just meant as motivation. The parameter κ will help us to control the tail behavior of the stochastical noise vector, later on.



$$
\begin{aligned}
||x_n^\delta - x_m^\delta|| &\leq ||x - x_n^\delta|| + ||x - x_m^\delta|| \\
&\leq \psi(n) + \delta||\eta_n^\xi|| + \psi(m) + \delta||\eta_m^\xi|| \\
&\leq \frac{2\kappa\delta}{\rho(n)} + \frac{2\kappa\delta}{\rho(m)} \\
&\leq \frac{4\kappa\delta}{\rho(m)}
\end{aligned}
\tag{12}
$$

Now we use an idea by Lepskij [9] and take

$$
n_* = \min\left\{n : ||x_n^\delta - x_m^\delta|| \leq \frac{4\kappa\delta}{\rho(m)},\ N = \rho^{-1}(\delta) > m > n\right\} \tag{13}
$$

Remark
In real applications it might be better to choose

$$
n_* = \min\left\{n : ||x_n^\delta - x_m^\delta|| \leq \frac{2\kappa\delta}{\rho(n)} + \frac{2\kappa\delta}{\rho(m)},\ N = \rho^{-1}(\delta) > m > n\right\} \tag{14}
$$

However, for the subsequent proofs we will use the simpler version.

Before starting with the main results we need the following supporting lemma which can be easily proven [2] using [8]:

Lemma 2.2
The following probability estimate holds:

$$
\mathbb{P}_\xi\left\{||\eta_n^\xi||\rho(n) > \tau\right\} \leq 4\exp\left(-\frac{\tau^2}{8}\right). \tag{15}
$$

Theorem 2.3
Let n_* be chosen as above with $\kappa \geq 1$. Then we have:

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

where C_1 and C_2 are constants.
In the deterministic noise case we additionally have $C_1 = 0$ from a certain δ onward.

Right now we just want to give a sketch of the proof, a much more detailed version can be found in [2].

Proof
We will employ the following strategy: We estimate the behavior of the error in a limited region which is bounded by the expected error behavior times κ; this is by far the most probable part for the error vector. For the rest we will use that the tail behavior of the error vector is exponentially decreasing. Note that this tail behavior is 0 for the deterministic noise part (therefore $C_1 = 0$).
Define

$$\Xi_\rho(\omega) = \max_{1\leq n\leq N} ||\eta_n^\xi||\rho(n) \tag{17}$$

and divide the probability space in two subspaces:

$$\Omega_\kappa = \{\omega : \Xi_\rho(\omega) \leq \kappa\} \quad \text{and} \quad \overline{\Omega_\kappa} = \Omega \setminus \Omega_\kappa \tag{18}$$

Part 1: ($\omega \in \Omega_\kappa$)
First we show $n_{opt} \geq n_*$:

$$||x_n^\delta - x_{n_{opt}}^\delta|| \leq ||x - x_n^\delta|| + ||x - x_{n_{opt}}^\delta|| \leq \psi(n) + \frac{\kappa\delta}{\rho(n)} + \psi(n_{opt}) + \frac{\delta}{\rho(n_{opt})} \leq \frac{2\kappa\delta}{\rho(n)} + \frac{2\kappa\delta}{\rho(n_{opt})} \leq \frac{4\kappa\delta}{\rho(n)} \tag{19}$$

which tells that

$$n_* = \min\left\{n : ||x_n^\delta - x_m^\delta|| \leq \frac{4\kappa\delta}{\rho(n)},\ N = \rho^{-1}(\delta) > m > n\right\} \leq n_{opt} \tag{20}$$

Then using $n_{opt} \geq n_*$ we have for all $\omega \in \Omega_\kappa$

$$||x - x_{n_*}^\delta|| \leq ||x - x_{n_{opt}}^\delta|| + ||x_{n_{opt}}^\delta - x_{n_*}^\delta|| \leq \frac{2\delta}{\rho(n_{opt})} + \frac{4\kappa\delta}{\rho(n_{opt})} \leq 6\frac{\kappa}{c}\left(c\frac{\delta}{\rho(n_{opt})}\right) \leq 6\frac{\kappa}{c}\psi\left((\psi\rho)^{-1}(\delta)\right) \tag{21}$$

Hence we get

$$\int_{\Omega_\kappa} ||x - x_{n_*}^\delta||^2 d\mathbb{P}_\xi(\omega) \leq |\Omega_\kappa|\ ||x - x_{n_*}^\delta||^2 \leq 36\frac{\kappa^2}{c^2}\psi\left((\psi\rho)^{-1}(\delta)\right)^2 \tag{22}$$

Part 2: ($\omega \in \overline{\Omega_\kappa}$)
Remember that we defined $n_{opt} \leq N = \rho^{-1}(\delta)$. Hence we get $\frac{\delta}{\rho(N)} = 1$ and $\psi(N) \leq \delta||\eta_N^\xi||$ and thus

$$||x - x_{n_*}^\delta|| \leq ||x - x_N^\delta|| + ||x_N^\delta - x_{n_*}^\delta|| \leq 2\delta||\eta_N^\xi|| + \frac{4\kappa\delta}{\rho(N)} \leq 2\frac{\delta||\eta_N^\xi||\rho(N)}{\rho(N)} + 4\kappa \leq 2\Xi_\rho + 4\Xi_\rho = 6\Xi_\rho \tag{23}$$

Using this result we obtain:

$$\int_{\overline{\Omega_\kappa}} ||x - x_{n_*}^\delta||^2 d\mathbb{P}_\xi(\omega) \leq 36\int_{\overline{\Omega_\kappa}} \Xi_\rho(\omega)^2 d\mathbb{P}_\xi(\omega) \leq 36\sqrt{\int_{\overline{\Omega_\kappa}} \Xi_\rho(\omega)^4 d\mathbb{P}_\xi(\omega)}\sqrt{\int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}_\xi(\omega)} \tag{24}$$

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

$$G(\tau) = 1 - F(\tau) = \mathbb{P}_\xi\left\{\Xi_\rho(\omega) > \tau\right\} \leq \sum_{n=1}^{N} \mathbb{P}_\xi\left\{||\eta_n^\xi||\rho(n) > \tau\right\} \leq 4N\exp\left(-\frac{\tau^2}{8}\right) \tag{25}$$

So we get:

$$\begin{aligned}\int_{\overline{\Omega_\kappa}} \Xi_\rho^4 d\mathbb{P}_\xi(\omega) &= -\int_\kappa^\infty \tau^4 d(1 - F(\tau)) \leq -\tau^4 G(\tau)|_0^\infty + 4\int_0^\infty \tau^3 G(\tau)d\tau \\ &= 4\int_0^\infty \tau^3 G(\tau)d\tau \leq 4N\int_0^\infty \tau^3\exp\left(-\frac{\tau^2}{8}\right)d\tau = 2^9 N\int_0^\infty u\exp(-u)du = 2^9 N\end{aligned} \tag{26}$$

The other part gets:

$$\int_{\overline{\Omega_\kappa}} 1 d\mathbb{P}_\xi(\omega) \leq 4 \exp\left(-\frac{\kappa^2}{8}\right). \tag{27}$$

Hence we get

$$\int_{\overline{\Omega_\kappa}} \Xi_\rho^2 d\mathbb{P}_\xi(\omega) \leq 2^{11/2} N \exp\left(-\frac{\kappa^2}{16}\right) \leq 2^{11/2} \rho^{-1}(\delta) \exp\left(-\frac{\kappa^2}{16}\right). \tag{28}$$

This yields

$$\mathbb{E}_\xi ||x - x_{n_*}^\xi||^2 \leq 36 \cdot 2^{11/2} \rho^{-1}(\delta) \exp\left(-\frac{\kappa^2}{16}\right) + 36 \frac{\kappa^2}{c^2} \psi\left(\left(\psi\rho\right)^{-1}(\delta)\right)^2. \tag{29}$$

This is exactly the proposition. q.e.d.

Now the main task will be choosing an appropriate κ.

2.4. Remarks on Smoothness and Error Spread

Now we want to give some short remarks what the terms ρ^{-1} and $\psi((\rho\psi)^{-1})$ actually mean in practice. Assume that we have $\psi(n) = n^{-r}$ which means a finite smoothness of the solution.

Considering an ordinarily ill-posed problem we have $\rho(n) \asymp n^{-a}$, i.e. there are constants c_1 and c_2 such that $c_1 n^{-a} \leq \rho(n) \leq c_2 n^{-a}$ from a certain n onwards. For reasons of simplicity we therefore assume $\rho(n) = n^{-a}$. Then we have:

$$\rho^{-1}(\delta) = \delta^{-\frac{1}{a}} \tag{30}$$

$$\psi\left(\left(\psi\rho\right)^{-1}(\delta)\right) = \delta^{\frac{r}{r+a}}. \tag{31}$$

2.5. Relaxations

Now we may choose κ according to our needs. The factor $\chi > 0$ below should always be close to 1. We will do some balancing process:

Lemma 2.4
Assume that our problem is ordinarily ill-posed with stochastical noise and polynomial behavior of the smoothness index function ψ. Now choose $\kappa = \chi 4 \ln \delta^{-1}$.
Then we have

$$\mathbb{E}_\xi ||x - x_{n_*}^\delta||^2 \leq C_1 \delta^{-\frac{1}{a} + \chi^2 \ln \delta^{-1}} + 16 C_2 \chi^2 \left(\ln \delta^{-1}\right)^2 \psi\left(\left(\psi\rho\right)^{-1}(\delta)\right)^2 \tag{32}$$

The proof just consist out of inserting ρ^{-1} and κ.

Corollary 2.5
Assume that our problem is ordinarily ill-posed with stochastical noise and polynomial behavior of the smoothness index function ψ. Now choose $\kappa = 4 \ln \delta^{-1}$.
Then we have if δ small enough for some constant C

$$\sqrt{\mathbb{E}_\xi ||x - x_{n_*}^\delta||^2} \leq C \left(\ln \delta^{-1}\right) \psi\left(\left(\psi\rho\right)^{-1}(\delta)\right) \tag{33}$$

Remark
This result is in principle the same as in [7, 10]. *In contrast we still have the possibility to work with the unspecified parameter χ which will turn out to be important, now.*

The Theorem 2.3 and its corollary tell that under certain conditions we just need the error (and error behavior) and can obtain an almost order optimal regularization procedure even under the hard assumption of stochastic rather than deterministic error.

2.6. Regularization without Known Smoothness and Error Behavior

As we have seen the above results still hold even if we introduce an additional parameter χ. Of course, no-one would like to obstruct our optimal κ by purpose. In practice we do not know the error behavior ρ. However by some means we can sometimes get hold of an estimation, e.g. by using additional local knowledge or two independent but equally disturbed input data sets.

If we denote the estimated version of $\frac{\delta}{\rho}$ by $\widetilde{\frac{\delta}{\rho}}$ and the corresponding estimated κ by $\widetilde{\kappa}$ we see that we get a slightly modified choice rule: (denote $N := \rho^{-1}(\delta)$ and choose $\kappa = 4\ln\delta^{-1}$)

$$n_* = \min\left\{n: \; ||x_n^\delta - x_m^\delta|| \le \frac{4\widetilde{\kappa}\widetilde{\delta}}{\widetilde{\rho}(m)}, \; N > m > n\right\} = \min\left\{n: \; ||x_n^\delta - x_m^\delta|| \le \frac{4\kappa\chi(m)\delta}{\rho(m)}, \; N > m > n\right\}$$
$$\in \left\{\min\left\{n: \; ||x_n^\delta - x_m^\delta|| \le \frac{4\kappa\chi_s\delta}{\rho(m)}, \; N > m > n\right\}, \ldots, \min\left\{n: \; ||x_n^\delta - x_m^\delta|| \le \frac{4\kappa\chi_h\delta}{\rho(m)}, \; N > m > n\right\}\right\} \tag{34}$$

where

$$\chi(m) = \frac{4\widetilde{\kappa}\widetilde{\delta}}{\widetilde{\rho}(m)} \Bigg/ \frac{4\kappa\delta}{\rho(m)} = \frac{\widetilde{\kappa}\widetilde{\delta}\rho(m)}{\kappa\delta\widetilde{\rho}(m)} \tag{35}$$

and the worst cases are

$$\chi_s = \min_{m<N}\{\chi(m)\} \quad \text{and} \quad \chi_h = \max_{m<N}\{\chi(m)\}. \tag{36}$$

Due to the different possibilities for the noise it is very hard to give exact estimates on χ_s and χ_h respectively [2]. However we can show a different very important result, namely that if χ is near enough at 1, then we can regularize in an optimal way.

Assumption 2.2
There exist constants c_1 and c_2 where the second one can be chosen big enough independent of δ such that the following properties hold:

$$\mathbb{P}_\chi\{\chi > \tau\} \le c_1 \exp\left(-c_2\left(\tau - 1\right)\right) \tag{37}$$

and

$$\mathbb{P}_\chi\{\chi^{-1} > \tau\} \le c_1 \exp\left(-c_2\left(\tau - 1\right)\right) \tag{38}$$

Furthermore χ shall be uncorrelated to η_n^ξ for all n.

Then the following theorems holds.

Theorem 2.6
Assume that our problem is ordinarily ill-posed with stochastical noise and polynomial behavior of the smoothness index function ψ and that the above assumption holds. Now choose $\kappa = 4\ln\widetilde{\delta}^{-1}$.
Then, if δ small enough for some constant C we have

$$\sqrt{\mathbb{E}_\xi||x - x_{n_*}^\delta||^2} \le C\left(\ln\delta^{-1}\right)\psi\left(\left(\psi\rho\right)^{-1}(\delta)\right). \tag{39}$$

Proof
As χ was assumed to be uncorrelated to η_n^ξ for all n we have

$$\mathbb{E}_\chi\left(\mathbb{E}_\xi||x - x_{n_*}^\delta||^2\right) \le \mathbb{E}_\chi\left(\Pi(\chi)\right) \tag{40}$$

where the function $\Pi(\chi)$ is defined as

$$\Pi(\chi) = C_1\delta^{-\frac{1}{a}+\chi^2\ln\delta^{-1}} + 16C_2\chi^2\left(\ln\delta^{-1}\right)^2\psi\left(\left(\psi\rho\right)^{-1}(\delta)\right)^2 \tag{41}$$

as in the lemma 2.4.
We have that $\delta^{\frac{r}{r+a}}$ is descending at least as fast as $\psi((\psi\rho)^{-1}(\delta))$. We introduce functions Π_1 and Π_2:

$$\Pi_1(\chi) = \delta^{-\frac{1}{a}-\frac{r}{r+a}+\chi^2\ln\delta^{-1}} \quad (42)$$
$$\Pi_2(\chi) = \chi^2 \quad (43)$$

and hence

$$\Pi(\chi) = C_1\delta^{\frac{r}{r+a}}\Pi_1(\chi) + 16C_2\left(\ln\delta^{-1}\right)^2\psi\left(\left(\psi\rho\right)^{-1}(\delta)\right)^2\Pi_2(\chi) \quad (44)$$

So it suffices to show that $\mathbb{E}_\chi\Pi_1(\chi)$ and $\mathbb{E}_\chi\Pi_2(\chi)$ can be bounded independent of δ.

$$\begin{aligned}
\mathbb{E}_\chi\Pi_1(\chi) &= \int_0^\infty \Pi_1(\tau)d\mathbb{P}_\chi\{\chi<\tau\} = \int_\infty^0 \Pi_1(\tau^{-1})d\mathbb{P}_\chi\{\chi^{-1}>\tau\} \\
&= \Pi_1\left(\tau^{-1}\right)\mathbb{P}_\chi\{\chi^{-1}>\tau\}\big|_\infty^0 - \int_\infty^0 \mathbb{P}_\chi\{\chi^{-1}>\tau\}d\Pi_1(\tau^{-1}) \\
&= 0-0+\int_0^\infty \mathbb{P}_\chi\{\chi^{-1}>\tau\}d\Pi_1(\tau^{-1}) = \int_0^\infty \mathbb{P}_\chi\{\chi^{-1}>\tau\}\frac{\partial}{\partial\tau}\Pi_1(\tau^{-1})d\tau \\
&\leq \int_0^2 \frac{\partial}{\partial\tau}\Pi_1(\tau^{-1})d\tau + \int_2^\infty \mathbb{P}_\chi\{\chi^{-1}>\tau\}2\tau^{-3}(\ln\delta^{-1})^2\delta^{-\frac{1}{a}-\frac{r}{r+a}+\tau^{-2}\ln\delta^{-1}}d\tau \\
&\leq \Pi_1\left(\frac{1}{2}\right) - 0 + 2(\ln\delta^{-1})^2\delta^{-\frac{1}{a}-\frac{r}{r+a}+0\ln\delta^{-1}}\int_2^\infty \mathbb{P}_\chi\{\chi^{-1}>\tau\}\tau^{-3}d\tau \\
&\leq \delta^{-\frac{1}{a}-\frac{r}{r+a}+4\ln\delta^{-1}} + 2(\ln\delta^{-1})^2\delta^{-\frac{1}{a}-\frac{r}{r+a}}\int_2^\infty c_1\exp\left(-c_2\left(\tau-1\right)\right)\tau^{-3}d\tau \\
&\leq \delta^{-\frac{1}{a}-\frac{r}{r+a}+4\ln\delta^{-1}} + 2(\ln\delta^{-1})^2\delta^{-\frac{1}{a}-\frac{r}{r+a}}c_1c_2^{-1}\exp(-c_2) \leq \widetilde{C}
\end{aligned} \quad (45)$$

where $\widetilde{C}$ is a constant which can be bounded independent of δ because of our assumption regarding c_2 and δ small enough which implies $\frac{1}{a}+\frac{r}{r+a}<4\ln\delta^{-1}$.
The second term can be evaluated in a similar way:

$$\begin{aligned}
\mathbb{E}_\chi\Pi_2(\chi) &= \int_0^\infty \Pi_2(\tau)d\mathbb{P}_\chi\{\chi<\tau\} = \int_0^\infty \tau^2 d\left(1-\mathbb{P}_\chi\{\chi>\tau\}\right) = -\int_0^\infty \tau^2 d\mathbb{P}_\chi\{\chi>\tau\} \\
&= -\tau^2\mathbb{P}_\chi\{\chi>\tau\}\big|_0^\infty + \int_0^\infty \mathbb{P}_\chi\{\chi>\tau\}d\tau^2 = 0-0+\int_0^\infty 2\tau\mathbb{P}_\chi\{\chi>\tau\}d\tau \\
&= \int_0^2 2\tau\mathbb{P}_\chi\{\chi>\tau\}d\tau + \int_2^\infty 2\tau\mathbb{P}_\chi\{\chi>\tau\}d\tau \leq \int_0^2 2\tau d\tau + \int_2^\infty 2\tau c_1\exp\left(-c_2\left(\tau-1\right)\right)d\tau \\
&= 4+2c_1c_2^{-2}(1+2c_2)\exp\left(-c_2\right) \leq \widetilde{C}
\end{aligned} \quad (46)$$

which can be assured when c_2 is chosen big enough.
Using that $\delta^{\frac{r}{r+a}}$ is descending at least as fast as $\psi((\psi\rho)^{-1}(\delta))$ we immediately get our result. q.e.d.

This means in particular that we are having the same rate of convergence even when we are having inexact input data.

3. NUMERICS

We have tested our method using simulated satellite data. This has several particular advantages. It is a severely ill-posed problem where we exactly know the degree of ill-posedness as required in our theorems. 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 consequently evades 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 [11] 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.

For the noise estimation we generated a small second data set of degrees 8–32 (i.e., about 900 actual data) and compared it with the biased approximation of our noisy data. Note that one could have also used a second noisy approximation, but this would have just increased computation time without giving any mathematical valuable information. We only need to consider more Fourier coefficients to obtain the same accuracy in the estimation (degrees 8–36).

For our purposes we observed that a $\kappa = 0.25$ seems to be a good choice which corresponds (roughly) to an accepted Variance/Bias ratio of 1:1 at the cutting point. After having chosen this parameter we proceeded with our experiment.

3.2. Data

We present two different representations of the data: the L-curves corresponding to our two input data sets and bias/variance behavior where we chose a log scale for better observability.

We displayed 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 auto-regularization method by $*$.

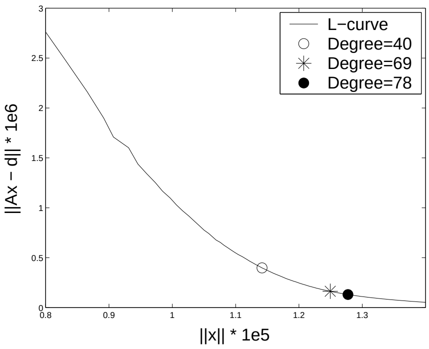


Figure 1: Data set at 200km.

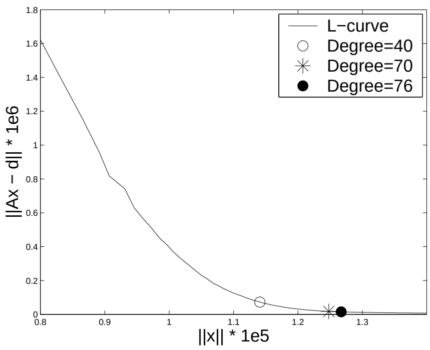


Figure 2: Data set at 400km.

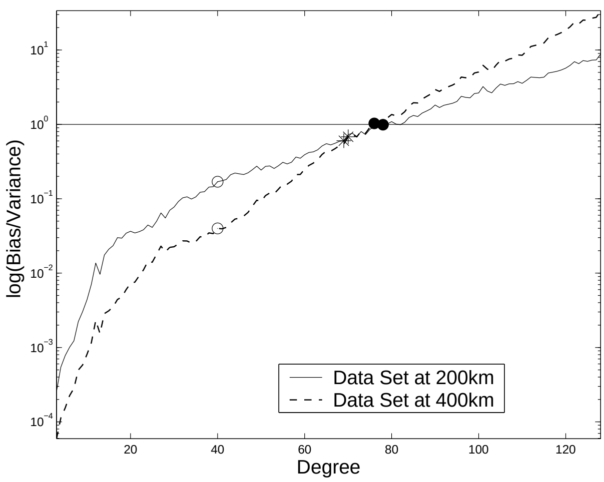


Figure 3: Bias/Variance ratio with respect to the degrees.

4. DISCUSSION AND OUTLOOK

Although widely used as a rule of thumb in practice the L-curve method cannot really solve an inverse problem (see e.g. [14]). In contrast our method is proven to have almost optimal convergence rates, performs well in practice and also does not require inaccessible knowledge.

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 although we are working with an exponentially ill-posed problem.

Another experiment where we used locally known exact data of the size of one sixteenth of the Earth's surface also returned very promising results.

Please note that we did *not* choose κ according to the proposed choice rule, and not even greater 1, but adapted to the problem. This difference is in our opinion due to the fact that we are for a specific problem in practice not just searching for an order optimal regularization parameter (which we get for the κ proposed in the proof) but for the best possible solution. This means we rely on a parameter which is not chosen theoretically correct but yields outstanding numerical results.

Numerical experiments indicate that once chosen an appropriate κ the results get very stable towards changes in the error level (even over an order of magnitude) and the smoothness of the data.

REFERENCES

1. F. Bauer and S. Pereverzev, Regularization without preliminary knowledge of smoothness and error behavior,. *European J. App. Math. (submitted)*, 2004.

2. F. Bauer, *An Alternate Approach to the Oblique Derivative Problem in Potential Theory.* PhD thesis, Universtiy of Kaiserslautern, Dept. of Mathematics, Geomathematics Group, 2004.

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

4. W. Freeden, *Multiscale Modelling of Spaceborne Geodata.* European Consortium for Mathematics in Industry. B. G. Teubner, Stuttgart, Leipzig, 1999.

5. W. Freeden and S. Pereverzev, Spherical tikhonov regularization wavelets in satellite gravity gradiometry with random noise. *J. Geodesy* (2001) **74**, 730–736.

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

7. H. Harbrecht, S. Pereverzev and R. Schneider, Self-regularization by projection for noisy pseudodifferential equations of negative order. *Num. Math.* (2002) **95**, 123–143.

8. M. Ledoux and M. Talagrand, *Probability in Banach spaces : Isoperimetry and Processes.* Springer, New York, 1991.

9. O.V. Lepskij, On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications* (1990) **35**(3), 454–466.

10. P. Mathé and S. Pereverzev, Direct estimation of linear functionals from indirect noisy observations. *J. Complexity* (2002) **18**(2), 501–516.

11. C. Mayer, *Potential operators, jump relations and wavelets*, Diploma thesis, Universtiy of Kaiserslautern, Dept. of Mathematics, Geomathematics Group, 2001.

12. S. Perevezev. On adaptive inverse estimation of linear functionals from random noisy data – case study: GOCE data processing. In *Oberwolfach Workshop Report 27*, p. 1380, 2004.

13. P. Schwintzer. Earth gravity field recovery from satellite orbit perturbations. *Geowissenschaften* (1997) **15**, 85–89.

14. A. Yagola, A. Leonov, and V. Titarenko. Data errors and an error estimation for ill-posed problems. *Inverse Problems in Engineering* (2002) **10**(2), 117 – 129.