Georg-August-Universität

Göttingen

Local Solutions to Inverse Problems in Geodesy: The Impact of the Noise Covariance Structure upon the Accuracy of Estimation

Frank Bauer, Peter Mathé, Sergei Pereverzev

Preprint Nr. 2005-14

Preprint-Serie des
Instituts für Numerische und Angewandte Mathematik
Lotzestr. 16-18
D - 37083 Göttingen

Local Solutions to Inverse Problems in Geodesy: The Impact of the Noise Covariance Structure upon the Accuracy of Estimation

Frank Bauer [1] , Peter Mathé [2] , Sergei Pereverzev [3]

[1] Institute for Mathematical Stochastics, University of Göttingen, Maschmühlenweg 8-10, 37073 Göttingen, Germany
Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany
Tel.: ++49 551 39 13 523, Fax: ++49 551 39 13 505, Email: bauer@mathematik.uni-kl.de

[2] Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany
Tel.: ++49 30 20372 550, Fax: ++49 30 2044975, Email: mathe@wias-berlin.de

[3] Johann Radon Institute for Computational and Applied Mathematics (RICAM), Altenbergerstrae 69, 4040 Linz, Austria
Tel.: ++43 732 2468 5215, Fax: ++43 732 2468 5212, Email: sergei.pereverzyev@oeaw.ac.at

Abstract

In many geoscientific applications one needs to recover the quantities of interest from indirect observations blurred by colored noise. Such quantities of interest often corresponds to the values of bounded linear functionals defined on the solution of some observation equation.

For example, various potential quantities are derived from harmonic coefficients of the Earth's gravity potential. Each such coefficient is the value of corresponding linear functional.

The goal of the paper is to discuss a new way to use information about noise covariance structure which allows to estimate the functionals of interest with order optimal risk and does not involve a covariance operator directly in the estimation process. It is done on the base of a balancing principle for the choice of regularization parameter which is new in geoscientific applications. A number of tests demonstrate its applicability.

MSC Classification: 47A52, 65J22, 49J55, 93E14

Keywords: Regularization by local data, ill-posed inverse problems, Gaussian random noise, satellite gravity gradiometry

1 Introduction

The question addressed in the present paper is:

How can colored observation noise be taken into account properly without direct access to a noise covariance operator?

This question is prompted by satellite missions such as GRACE(1998) and GOCE [Eur99], launched or planned, respectively, aiming to develop a model of the Earth's gravity field from satellite gravity observations.

These observations are used to determine the gravity field, which in mathematical terms leads to the observation equations written as a standard Gauss-Markov model

$$y^\delta = Ax + \delta\xi \tag{1}$$

Here x is the unknown gravity potential at the Earth's surface which should be recovered from noisy observations y^δ where ξ is a Gaussian random variable with zero expectation $\mathbb{E}\xi = 0$ and covariance $\operatorname{cov}(\xi) = \mathbf{K}_\xi^2$; the a priori variance is denoted by δ.

Due to the huge number of observations and unknowns it is reasonable to consider (1) as an operator equation in Hilbert spaces with compact operator A mapping from the solution space $\mathcal{X}$ into the observation space $\mathcal{Y}$ equipped with the corresponding inner products $\langle \cdot, \cdot \rangle$ and norms $||\cdot||$, respectively. In this context the covariance operator is a bounded self-adjoint non-negative operator $\mathbf{K}_\xi : \mathcal{Y} \to \mathcal{Y}$ such that for all $f, g \in \mathcal{Y}$

$$\mathbb{E}\langle f, \xi\rangle \langle g, \xi\rangle = \langle \mathbf{K}_\xi^2 f, g\rangle = \langle \mathbf{K}_\xi f, \mathbf{K}_\xi g\rangle . \tag{2}$$

A much more detailed description of the underlying physical and mathematical problem in the case of satellite-to-satellite tracking (GRACE) and satellite gravity gradiometry (GOCE), respectively, can be found e.g., in [Fre99].

As A is compact it is not continuously invertible (see e.g., [EHN96]) and we are facing an ill-posed problem which in this context is normally referred to as "downward-continuation".

1.1 Noise models and regularization

In any discretized version of (1) the ill-posedness is reflected in the ill-condition of the corresponding normal equation matrix. Ill-conditioned normal equations are not new in Geodesy, especially when satellite observations are used for gravity field estimation, see e.g., [KK02].

So far, various algorithms for computing a regularized solution have been proposed in connection with the determination of the gravity filed on the Earth's surface from space-borne geodata. Truncated singular value decomposition, Tikhonov regularization and least-squares estimation (LS) with stochastic prior information are among the widely used regularization methods.

LS-estimators for the gravity field recovery have been discussed recently by [PP02] and by [KDB03]. At the same time the authors of the latter paper point out that this approach can be used only if the covariance operator is known exactly.

Tikhonov regularization and truncated singular value decomposition are more robust with respect to misspecification of the noise. The results from [LKB01] indicate that the knowledge of just δ will not guarantee to obtain a good solution, and the choice of the regularization parameter for these methods is a severe topic.

At the same time, for new satellite missions, such as GOCE, one cannot expect to have a good description of the noise covariance. Indeed, as it has been indicated by [KK02] the measurement equipment on board of the GOCE satellite has never been validated in the orbit and it is likely that it will be contaminated with an aliasing signal caused by the unmodeled high frequencies of the gravitational field.

Furthermore numerical experiments as reported by [PP02] show that the use of a rough approximation of the covariance operator in the Tikhonov regularization scheme leads to a very poor performance.

Therefore one needs algorithms, which are capable to deal with different noise models. In this study we shall present one, which intrinsically separates computation and adaptation to unknown noise as well as to unknown smoothness of the solution to (1).

1.2 Data functional strategy

In some cases one is not interested in completely knowing x, but of some derived quantities of it, only. As it has been pointed out by [And86] such derived quantities often correspond to bounded linear functionals $L(x)$ of the solution x. For example, for the success of the GOCE mission it will be sufficient to obtain an accurate estimation of the Fourier coefficients of the gravity potential up to degree 300. Each of these coefficients is the value of a corresponding linear functional. Another example is the gravity potential at some fixed point of some region of interest.

We present a solution to this problem using the *data-functional strategy*, as introduced by [And86]

The idea is as follows. Instead of computing an approximate solution x^δ for any set of data y^δ and then evaluate $L(x^\delta) = \langle l, x^\delta \rangle$ where l is the Ritz representer of L, we find a suitable Ritz representer of the functional on the data, say z such that $\langle l, x^\delta \rangle \sim \langle z, y^\delta \rangle$. The advantage is obvious. Once z has been determined, the evaluation of $\langle z, y^\delta \rangle$ is fast and stable.

The original idea by [And86] was to find z from the equation

$$l = A^* z. \tag{3}$$

However such a z need not exist. Therefore we construct $z = z_\alpha$ regularizing (3) by some reasonable regularization method.

At any parameter α the error can be computed exactly as

$$\mathbb{E} \left| \langle l, x \rangle - \langle z_\alpha, y^\delta \rangle \right|^2 = |\langle l, x \rangle - \langle z_\alpha, Ax \rangle|^2 + \delta^2 \|\mathbf{K}_\xi z_\alpha\|^2. \tag{4}$$

It is well-known and will be crucial in our analysis below, that the bias $|\langle l, x \rangle - \langle z_\alpha, Ax \rangle| \to 0$ as $\alpha \to 0$, under natural assumptions on the smoothness of l and the solution x, even when $A^* z = l$ itself has no solution.

1.3 Outline

The outline of the study is as follows.

- We first present the algorithm in section 2 and indicate its behavior, which crucially depends on some known bound on the covariance, but adapts to the underlying smoothness.
- In section 3 we shall exhibit how the error behaves when smoothness is given in terms of general source conditions.
- In section 4 we compare the impact of different kind of error models on the achievable convergence rates.
- In section 5 we will describe how one can estimate the necessary information about the noise behavior and the impact of estimation errors.
- We present computational results in section 6 and conclude with some discussion in section 7.

2 The adaptive algorithm

2.1 Description of the algorithm

The design of the algorithm starts from the bias-variance decomposition (4).

The algorithm is based on some ordered set

$$\Delta_M = \{\alpha_i : 0 < \alpha_0 < \alpha_1 < \ldots < \alpha_M\}$$

of regularization parameters, say α for Tikhonov regularization, or iteration number, say n for some iterative schemes.

To be specific we focus on Tikhonov regularization, thus we let $z_\alpha := (\alpha I + AA^*)^{-1} Al$ for the data functional strategy and compute successively z_α along $\alpha \in \Delta_M$. It remains to select α.

We are going to present a method (balancing principle) for the adaptive choice of $\alpha_+ \in \Delta_M$. As we will see such an α_+ can be chosen *without* any a priori information concerning the smoothness of the unknown solution.

The idea of our adaptive principle has its origin in the paper by [Lep90] devoted to the statistical estimation from direct observation blurred by Gaussian white noise that corresponds to (1) with A and $\mathbf{K}_\xi$ being equal to the identity. In the context of general statistical estimation this idea was realized in [GP00, Tsy00, Bau04] and [BP05]. But only the Gaussian white noise model, corresponding to the case $\mathbf{K}_\xi = \mathrm{I}$ has been discussed there.

The algorithm requires partial knowledge on the covariance, specifically a valid bound

$$\|\mathbf{K}_\xi z_\alpha\| \le 1/\Omega(\alpha), \qquad \alpha \in \Delta_M, \tag{5}$$

for an increasing function Ω which obeys $\Omega(0) = 0$.

Remark 2.1 *Notice that z_α is independent of any data, and the required bound can be estimated accurately for any underlying model assumption on the noise as it will be discussed in section 5.*

We assume that $0 < \delta < 1$ is sufficiently small. Choose $q > 1$, let $\alpha_0 = \Omega^{-1}(\delta)$, $M \sim \ln(1/\Omega^{-1}(\delta))$ and consider $\Delta_M = \{\alpha_i = \alpha_0 q^i,\ i = 0, 1, \ldots, M\}$.

Let $\Delta_M^{\Omega,\delta}$ be the set of all $\alpha_j \in \Delta_M$ such that

$$\max_{\alpha_j \ge \alpha_i \in \Delta_M} \left\{ \left| \langle z_{\alpha_j}, y^\delta \rangle - \langle z_{\alpha_i}, y^\delta \rangle \right| \frac{\Omega(\alpha_i)}{\delta} \right\} \le 4\kappa, \tag{6}$$

where κ is a tuning parameter which will be specified later on.

The regularization parameter α_+ is chosen as

$$\alpha_+ := \max \left\{ \alpha_j, \qquad \alpha_j \in \Delta_M^{\Omega,\delta} \right\}. \tag{7}$$

2.2 Optimality properties of the algorithm

The choice of the parameter α_+ according to (7) has the following remarkable properties, which will be described in terms of *admissible functions*, introduced as follows. We start with the bias-variance decomposition from (4). As discussed there it is natural to assume that there is a function $\Phi(\alpha) = \Phi(\alpha, x, A, l)$, which is non-decreasing, continuous and obeys $0 = \Phi(0) \le \Phi(\alpha) \le 1$ and

$$|\langle l - A^* z_\alpha, x \rangle| \le \Phi(\alpha), \qquad \alpha > 0. \tag{8}$$

In view of the error representation (4) the quantity

$$\mathbf{e}_\delta^2(x, z_\alpha, \Delta_M) := \inf \left\{ \Phi(\alpha)^2 + \frac{\delta^2}{\Omega(\alpha)^2},\ \Phi \text{ admis.}, \alpha \in \Delta_M \right\}, \tag{9}$$

is the (square of the) best possible accuracy for estimating $\langle l, x \rangle$ by the data functional strategy z_α, $\alpha \in \Delta_M$ using the bound (5).

To establish the error behavior we need the following technical assumption on $\Omega(t)$, valid in a vast majority of cases. It increases not faster than with a power rate, i.e. $t^{\beta_2} \le \Omega(t) \le t^{\beta_1}$. This means in particular that there is a constant $p = \max_{\alpha_i \in \Delta_M} \Omega(\alpha_{i+1})/\Omega(\alpha_i)$.

Theorem 2.1 *Under the above assumptions on $t \mapsto \Omega(t)$ we have*

$$\mathbb{E} \left| \langle l, x \rangle - \langle z_{\alpha_+}, y^\delta \rangle \right|^2 \le 36 p^2 \kappa^2 \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M) + c_0 \ln[1/\Omega^{-1}(\delta)] \exp\left(-\kappa^2/8\right),$$

where c_0 is an absolute constant.

Proof
Let Φ be an admissible function. Consider

$$\alpha_{j_0} = \alpha_{j_0}(\Phi) = \max\{\alpha_i \in \Delta_M : \Phi(\alpha_i) \le \delta/\Omega(\alpha_i)\},$$

and

$$\alpha_{j_1} = \alpha_{j_1}(\Phi) = \operatorname{argmin} \left\{ \Phi(\alpha_i)^2 + \delta^2/\Omega(\alpha_i)^2, \alpha_i \in \Delta_M \right\}.$$

As in [BP05, MP05] this leads to

$$\mathbb{E} \left| \langle l, x \rangle - \langle z_{\alpha_+}, y^\delta \rangle \right|^2 \le \frac{36 \delta^2 \kappa^2}{\Omega(\alpha_{j_0})^2} + c_0 \ln[1/\Omega^{-1}(\delta)] \exp\left(-\frac{\kappa^2}{8}\right),$$

and it is enough to prove that

$$\delta^2/\Omega(\alpha_{j_0})^2 \le p^2 \left(\Phi(\alpha_{j_1})^2 + \delta^2/\Omega(\alpha_{j_1})^2 \right), \tag{10}$$

because then the statement of the theorem will follow from the fact that Φ is an arbitrary admissible function.

If $\alpha_{j_1} \le \alpha_{j_0}$ then $\delta/\Omega(\alpha_{j_0}) \le \delta/\Omega(\alpha_{j_1})$ and (10) is obviously satisfied. Otherwise $\alpha_{j_0} < \alpha_{j_0+1} \le \alpha_{j_1}$ and by definition of α_{j_0} it holds true that $\delta/\Omega(\alpha_{j_0+1}) \le \Phi(\alpha_{j_0+1})$ and

$$\frac{\delta^2}{\Omega(\alpha_{j_0})^2} \le p^2 \frac{\delta^2}{\Omega(\alpha_{j_0+1})^2} \le p^2 \Phi(\alpha_{j_0+1})^2 \le p^2 \Phi(\alpha_{j_1})^2 \le p^2 \left(\Phi(\alpha_{j_1})^2 + \frac{\delta^2}{\Omega(\alpha_{j_1})^2} \right),$$

which completes the proof.

In order to obtain the best possible rate we need to balance $\kappa^2 \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M)$ and $\ln[1/\Omega^{-1}(\delta)] \exp\left(-\frac{\kappa^2}{8}\right)$. This balance will be achieved for different values of κ, depending on different kind of ill-posedness of the problem.

Arguments as in [Bau04, BP05] give the following result.

Corollary 2.2 *Suppose (1) is a* moderately ill-posed problem, *i.e., $\mathbf{e}_\delta(x, z_\alpha, \Delta_M) = O(\delta^\mu)$ for some $0 <$*

$\mu < 1$. *Then, for a problem dependent constant d, and setting* $\kappa = 4\sqrt{d \ln 1/\delta}$, *we obtain*

$$\mathbb{E}\left|\langle l, x\rangle - \left\langle z_{\alpha+}, y^\delta \right\rangle\right|^2 \le c_1 (\ln 1/\delta)\, \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M),$$

where c_1 *is independent of* δ.

This means in particular that in the power scale the accuracy of our adaptive data-functional strategy is of the same order as the best possible one.

Corollary 2.3 *If (1) describes a* severely ill-posed problem, *i.e.,* $\mathbf{e}_\delta(x, z_\alpha, \Delta_M) = O\left(\ln^{-\mu} 1/\delta\right)$ *for some* $\mu > 0$. *Then, for a problem dependent constant d, and with* $\kappa = 4\sqrt{d \ln\ln 1/\delta}$ *we obtain*

$$\mathbb{E}\left|\langle l, x\rangle - \left\langle z_{\alpha+}, y^\delta \right\rangle\right|^2 \le c_1 (\ln\ln 1/\delta)\, \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M),$$

where c_1 *is independent of* δ.

Again, measured in the logarithmic scale the accuracy of our adaptive data-functional strategy is of the same order as the best possible one.

3 Error analysis under general source conditions

In both cases as discussed in Corollaries 2.2 and 2.3, up to a lower order factor the quantity $\mathbf{e}_\delta(x, z_\alpha, \Delta_M)$ from (9) determines the rate of approximating the unknown value $L(x)$ based on observations y^δ. Hence, it is interesting to discuss this in more detail in some mathematical framework.

Apart form the noise properties the achievable accuracy for estimating $\langle l, x\rangle$ is essentially determined by

- the smoothness of the unknown solution x,
- the smoothness of the Ritz representer l and
- the degree of ill-posedness of the operator in (1).

The benchmark for the smoothness of x is provided by the Picard-Criterion, which is based on the singular value decomposition of A from (1) as

$$Ax = \sum_{j=1}^{\infty} a_j \langle v_j, x\rangle\, u_j, \quad x \in \mathcal{X}, \tag{11}$$

For the sake of simplicity we will assume in the sequel that the orthonormal systems $\{u_j\}$ and $\{v_j\}$ form bases in the spaces $\mathcal{Y}$ and $\mathcal{X}$, respectively, such that A is injective and has dense range. This assumption is satisfied for the satellite gravity gradiometry case, see e.g. [FP01] and [PS99].

3.1 Measuring smoothness through general source conditions

Picard's criterion asserts that for an operator A with singular value decomposition (11) and $\ker(A) = \{0\}$ the equation $Ax = y$ has a unique solution $x \in \mathcal{X}$ if and only if

$$\sum_{j=1}^{\infty} \left|\langle u_j, y\rangle\right|^2 a_j^{-2} < \infty. \tag{12}$$

In this case the solution is given by

$$x := \sum_{j=1}^{\infty} \langle u_j, y\rangle\, a_j^{-1} v_j.$$

One can impose additional smoothness of x by enforcing additional summability properties of $\left|\langle u_j, y\rangle\right|^2$.

Precisely, let $\varphi : [0, ||A^*A||] \to \mathbb{R}_0^+$ be a non-decreasing function with $\varphi(0) = 0$. If

$$\sum_{j=1}^{\infty} \left|\langle u_j, y\rangle\right|^2 / [a_j^2 \varphi(a_j^2)^2] < \infty,$$

then the following expression is finite and gives rise to

$$||x||_\varphi^2 := \sum_{j=1}^{\infty} \left|\frac{\langle v_j, x\rangle}{\varphi(a_j^2)}\right|^2 = \sum_{j=1}^{\infty} \left|\frac{\langle u_j, y\rangle}{a_j \varphi(a_j^2)}\right|^2 < \infty, \tag{13}$$

a norm on $\mathcal{X}$. In particular, $x = A^{-1}y$ can (formally) be represented as $x = \sum_{j=1}^{\infty} \varphi(a_j^2)\, \langle v_j, r\rangle\, v_j = \varphi(A^*A)r$, for some

$$r = \sum_{j=1}^{\infty} \frac{\langle u_j, y\rangle}{a_j \varphi(a_j^2)} v_j \in \mathcal{X}, \quad \text{with } ||r|| = ||x||_\varphi < \infty.$$

Thus the additional smoothness of x can be expressed as

$$x \in A_\varphi(R) := \{v \in \mathcal{X} : v = \varphi(A^*A)u, ||u|| \le R\},$$

which is called *source condition.*

Please note that the set $A_\varphi(R)$ is just the ball of radius R in a Hilbert space $\mathcal{X}_\varphi = \{v : ||v||_\varphi < \infty\}$. The variety of spaces constructed in this way has been studied frequently, see e.g., [MP03]). In particular we mention that the dual space of $\mathcal{X}_\varphi$ is given by $\mathcal{X}_{1/\varphi}$, and $\mathcal{X}_\psi$ is embedded in $\mathcal{X}_{1/\varphi}$ whenever $0 < \psi \le 1/\varphi$. Therefore we assume that the solution functional l obeys $l \in A_\psi(R_1)$ for some $0 < \psi \le 1/\varphi$, in order to ensure that the linear functional $\langle l, x\rangle$ is well-defined for $x \in A_\varphi$.

Remark 3.1 *The fact that x belongs to a specific Sobolev space can be reformulated in terms of general source conditions, see e.g. [PS99].*

3.2 Finding admissible functions Φ

If both the solution x and the solution functional l are given in terms of source conditions as in section 3.1, then we can find admissible functions, bounding (8). Precisely, let $z_\alpha = (\alpha I + AA^*)^{-1} Al = A(\alpha I + A^*A)^{-1} l$.

Proposition 3.1 *Suppose that $x \in A_\varphi(R)$ and $l \in A_\psi(R_1)$. If $\lambda \mapsto \varphi(\lambda)\psi(\lambda)/\lambda$ is non-increasing then the function $\alpha \mapsto RR_1\varphi(\alpha)\psi(\alpha)$ is admissible.*

Proof
Using spectral calculus we have for $x = \varphi(A^*A)v$ and $l = \psi(A^*A)u$ the bound

$$\begin{aligned}
&|\langle l - A^*z_\alpha, x\rangle| \\
&\quad = \left|\left\langle (I - A^*A(\alpha I + A^*A)^{-1})\psi(A^*A)u, \varphi(A^*A)v\right\rangle\right| \\
&\quad = \alpha \left|\left\langle \varphi(A^*A)(\alpha I + A^*A)^{-1}\psi(A^*A)u, v\right\rangle\right| \\
&\quad \leq \alpha ||u||\, ||v||\, ||\varphi(A^*A)(\alpha I + A^*A)^{-1}\psi(A^*A)|| \\
&\quad \leq \alpha RR_1 \sup_{\lambda \in [0, ||A^*A||]} \left|(\alpha + \lambda)^{-1}\varphi(\lambda)\psi(\lambda)\right|.
\end{aligned}$$

For a non-increasing function $\lambda \mapsto \varphi(\lambda)\psi(\lambda)/\lambda$ it has been shown in [MP03] that

$$\alpha \sup_{\lambda \in [0, ||A^*A||]} \left|(\alpha + \lambda)^{-1}\varphi(\lambda)\psi(\lambda)\right| \leq \varphi(\alpha)\psi(\alpha),$$

and the proof is complete.

Remark 3.2 *If in contrast the function $\lambda \mapsto \varphi(\lambda)\psi(\lambda)/\lambda$ is non-decreasing there is a constant $c = c(||A||)$ such that*

$$\alpha \sup_{\lambda \in [0, ||A^*A||]} \left|(\alpha + \lambda)^{-1}\varphi(\lambda)\psi(\lambda)\right| \leq c\alpha,$$

where we refer to [MP03], again. Hence in both cases (8) is satisfied, although with different rates.

Similar results hold for other regularization methods such as truncated singular value decomposition.

3.3 Finding Ω: Bounding the variance

It was mentioned in remark 2.1 that bounds for $||\mathbf{K}_\xi z_\alpha||$ can be obtained in many cases. Here we highlight some particular case, suited to the present setup.

Let $\mathbf{K}_\xi$ and A be related in such a way that for some non-decreasing function, known a priori it holds true that

$$||\mathbf{K}_\xi f|| \leq ||\varrho(AA^*)f||, \qquad \text{for all } f \in \mathcal{Y}. \tag{14}$$

Assumption (14) is fulfilled in several important cases of interest. For $\rho(\lambda) = 1,\ \lambda \in [0, \|AA^*\|]$, this reflects the boundedness of $\mathbf{K}_\xi$, as e.g., for Gaussian white noise.

We denote the set of self-adjoint non-negative operators $\mathbf{K}_\xi$ which fulfill (14) by $\mathbb{K}_\varrho(A)$.

Remark 3.3 *It has been observed by [Böt05] that the assumption (14) is equivalent to*

$$\text{range}(\mathbf{K}_\xi) \subset \text{range}\left(\varrho\left(AA^*\right)\right).$$

We discuss condition (14) in some particular case when the (random) Fourier coefficients $\langle u_j, \xi\rangle$ are independent random variables. In this case, for an operator A with singular value decomposition (11) and for any $i \neq j$ it holds true that

$$\left\langle \mathbf{K}_\xi^2 u_i, u_j\right\rangle = \mathbb{E}\langle u_i, \xi\rangle \langle u_j, \xi\rangle = 0.$$

Thus, in the basis $\{u_j\}$ the operator $\mathbf{K}_\xi$ is diagonal, and its singular values k_j are given by

$$k_j^2 = \mathbb{E}|\langle u_j, \xi\rangle|^2, \qquad j = 1, 2, \ldots$$

Then assumption (14) imposes bounds

$$k_j \leq \varrho(a_j^2), \qquad j = 1, 2, \ldots \tag{15}$$

in terms of the non-decreasing function ϱ. Of course the above reasoning extends to the situation, when both $\mathbf{K}_\xi$ and AA^* can be diagonalized in a common basis.

Furthermore, assuming the independence of the (random) noise Fourier coefficients for the basis that diagonalizes some design matrix it is reasonable to construct a covariance prior using the same basis as this has been done in [FM02]. Then assumption (14) is automatically satisfied for some ϱ.

We shall see next, that (14) induces a function Ω, required in (5), in a natural way.

Let z_α be obtained using Tikhonov regularization.

Proposition 3.2 *Let $l \in A_\psi(R_1)$ and $\mathbf{K}_\xi$ obey (14). If $\lambda \mapsto \varrho(\lambda)\psi(\lambda)/\sqrt{\lambda}$ is non-increasing and $\sqrt{\lambda}\rho(\lambda)\psi(\lambda)$ is non-decreasing then assumption (5) holds for*

$$\Omega(\alpha) = \sqrt{\alpha}/\left(2R_1\varrho(\alpha)\psi(\alpha)\right).$$

Proof
For l from above property (14) yields

$$\begin{aligned}||\mathbf{K}_\xi z_\alpha|| \leq& ||\varrho(AA^*)(\alpha I + AA^*)^{-1}Al|| \\ =& ||(A^*A)^{1/2}\varrho(A^*A)(\alpha I + A^*A)^{-1}\psi(A^*A)u|| \\ \leq& R_1 \sup_{\lambda\in[0,||A^*A||]} \left|\sqrt{\lambda}\varrho(\lambda)(\alpha+\lambda)^{-1}\psi(\lambda)\right|.\end{aligned}$$

Here we use the fact that for any continuous function g it holds $g(AA^*)A = Ag(A^*A)$ and for any $f \in \mathcal{X}$ we have $||Af|| = ||(A^*A)^{1/2}f||$.

If we now assume that $\lambda \mapsto \varrho(\lambda)\psi(\lambda)/\sqrt{\lambda}$ is non-increasing, then from [MP03] it follows that

$$\sup_{\lambda\in[0,||A^*A||]} \left|\sqrt{\lambda}\varrho(\lambda)(\alpha+\lambda)^{-1}\psi(\lambda)\right| \leq 2\varrho(\alpha)\psi(\alpha)/\sqrt{\alpha},$$

which completes the proof.

3.4 Bounding the error from above

The descriptions of admissible functions given in proposition 3.1 and for Ω from proposition 3.2 allow for an error estimate.

The following function turns out to be important. Let

$$\theta_\varrho(\lambda) := \sqrt{\lambda}\varphi(\lambda)/\varrho(\lambda), \quad \lambda > 0. \tag{16}$$

Precisely we state

Theorem 3.3 *Let $\alpha = \alpha_\delta$ be the solution to the equation $\varphi(\alpha) = \delta\varrho(\alpha)/\sqrt{\alpha}$, i.e. $\alpha_\delta = \theta_\varrho^{-1}(\delta)$. Under the assumptions of propositions 3.1 and 3.2 there is a constant C such that as $\delta \to 0$ we have*

$$\mathbb{E}\left|\langle l, x\rangle - \langle z_{\alpha_\delta}, y^\delta\rangle\right|^2 \leq C\varphi^2((\theta_\varrho)^{-1}(\delta))\psi^2((\theta_\varrho)^{-1}(\delta)).$$

Proof
Under the conditions on φ, ψ and ϱ indicated in propositions 3.1 and 3.2 for any $x \in A_\varphi(R)$, $l \in A_\psi(R_1)$, and $\mathbf{K}_\xi \in \mathbb{K}_\varrho(A)$ we have

$$\begin{aligned}&\sqrt{\mathbb{E}|\langle l, x\rangle - \langle z_\alpha, Ax + \delta\xi\rangle|^2} \\ &\qquad \leq R_1 R\varphi(\alpha)\psi(\alpha) + 2R_1\delta\varrho(\alpha)\psi(\alpha)/\sqrt{\alpha} \\ &\qquad \leq c\psi(\alpha)\left(\varphi(\alpha) + \delta\varrho(\alpha)/\sqrt{\alpha}\right), \quad \alpha > 0,\end{aligned}$$

where z_α is the data-functional strategy based on Tikhonov regularization. The choice of α from above provides us with the desired upper bound.

This results exhibits the same error behavior as obtained by the Lepskiĭ-type balancing principle, although that works *without* a priori knowledge of the smoothness.

3.5 Bounding the error from below

It is interesting to see, that the upper bound provided in theorem 3.3 cannot be improved, at least asymptotically. To avoid degeneracies we have to compute the supremum over covariances and solution functionals within the classes described above. Precisely we introduce the modulus of continuity by letting

$$\tau_\delta^2(\varphi, \psi, \varrho) := \sup_{\mathbf{K}_\xi\in\mathbb{K}_\varrho(A)} \sup_{l\in A_\psi(R_1)} \inf_z \sup_{x\in A_\varphi(R)} \mathbb{E}\left|\langle l, x\rangle - \langle z, y^\delta\rangle\right|^2. \tag{17}$$

The same arguments as in [Don94] lead to

$$\tau_\delta(\varphi, \psi, \varrho) \geq \frac{R_1}{2}\sup\left\{||\psi(A^*A)x|| : x \in A_\varphi(R), ||\frac{1}{\varrho(A^*A)}Ax|| \leq \delta\right\}.$$

The latter sup can be estimated in the same way as in [MP03] and provides us with the following lower bound. Recalling the function $\theta_\varrho(\lambda)$ from (16) we get:

Proposition 3.4 *Assume that*

- *there is a constant $0 < \sigma < 1$ such that for the singular values a_i of A as described in (11) we have $a_{i+1}/a_i \geq \sigma,\ i = 1, 2 \ldots,$*
- *the functions φ and ψ fulfill $\varphi(2\lambda) \sim \varphi(\lambda)$ and $\psi(2\lambda) \sim \psi(\lambda)$*
- *the function $\varphi^2((\theta_\varrho^2)^{-1}(\lambda))\psi^2((\theta_\varrho^2)^{-1}(\lambda))$ is concave.*

Then there is c_2 for which

$$\tau_\delta(\varphi, \psi, \varrho) \geq c_2\varphi((\theta_\varrho)^{-1}(\delta))\psi((\theta_\varrho)^{-1}(\delta)), \quad \text{as } \delta \to 0.$$

We summarize our analysis in the following

Corollary 3.5 *Let the assumptions of propositions 3.1–3.4 be satisfied. Then*

$$\tau_\delta(\varphi, \psi, \varrho) \asymp \varphi((\theta_\varrho)^{-1}(\delta))\psi((\theta_\varrho)^{-1}(\delta)), \quad \text{as } \delta \to 0.$$

The order of the minimax risk $\tau_\delta(\varphi, \psi, \varrho)$ is realized by the data-functional strategy z_α with the regularization parameter $\alpha = \theta_\varrho^{-1}(\delta)$ chosen independently of ψ and $l \in A_\psi(R_1)$.

The same order is attained when choosing the regularization parameter from (7), adaptively.

We close this section with a specific example, corresponding to a moderately ill-posed problem.

Example

Assume that the singular values of A, cf. (11), are decaying with a power rate, i.e. $a_j \sim j^{-r}$. Assume also that $\varphi(\lambda) \sim \lambda^\mu$, $\psi(\mu) \sim \lambda^\nu$ and $\varrho(\lambda) \sim \lambda^\beta$. In this context $x \in A_\varphi$ is equivalent to x belonging to some Sobolev space.

Now we have that $\theta_\varrho(\lambda) \sim \lambda^{\mu-\beta+1/2}$ and, under the condition that $\beta - 1/2 < \mu$, $\nu + \beta < 1/2$ and $-\mu \le \nu$, corollary 3.5 provides us with the order of the minimax risk as

$$\tau_\delta(\varphi, \psi, \varrho) \asymp \delta^{\frac{2\mu+2\nu}{2\mu-2\beta+1}}, \qquad \text{as } \delta \to 0. \tag{18}$$

4 Comparison of Different Noise Models

4.1 Statistical noise vs. deterministic one

It is interesting to compare the accuracy of reconstructing $L(x) = \langle l, x\rangle$, both under statistical noise, as this was done here, and under bounded deterministic noise, i.e., when $||\xi|| \le 1$, a usual setup for studying ill-posed problems.

For statistical noise it follows from (15) that

$$\mathbb{E}||\xi||^2 = \sum_{j=1}^{\infty} \mathbb{E}|\langle u_j, \xi\rangle|^2 \le \sum_{j=1}^{\infty} \varrho^2(a_j^2) \asymp \sum_{j=1}^{\infty} j^{-4\beta r}. \tag{19}$$

Thus, for $\beta > 1/4r$ the random element ξ has a strong second moment $\mathbb{E}||\xi||^2 < \infty$ and can be regarded as an element from the observation space $\mathcal{Y}$, thus suits the deterministic noise model. For such a noise model the analogue of the minimax risk at any functional l is defined as

$$\tau_\delta^{\det}(l, \varphi, \psi) = \inf_z \sup_{x \in A_\varphi(R)} \sup_{||\xi|| \le \delta} \left|\langle l, x\rangle - \langle z, y^\delta\rangle\right|,$$

and the quantity $\tau_\delta^{\det}(\varphi, \psi) := \sup_{l \in A_\psi(R_1)} \tau_\delta^{\det}(l, \varphi, \psi)$ must be compared to the modulus from (17). We know from [EN94, MP02] that

$$\tau_\delta^{\det}(\varphi, \psi) \asymp \delta^{(2\mu+2\nu)/(2\mu+1)}.$$

Again, from (19) and for $\beta \in [0, 1/4r]$ the noise generating random process ξ does not take values in the observation space $\mathcal{Y}$, a.s.

From this point of view such a noise is stronger than a deterministic one. Nevertheless, for the Gaussian white noise corresponding to $\beta = 0$ the order of the minimax risk (18) coincides with the order of its deterministic counterpart.

This result is not new. [Don94] observes that the minimax risk of statistical estimation from observations blurred by a Gaussian white noise has the same asymptotic behavior as its deterministic counterpart.

It looks surprisingly enough that for colored noise ξ with $\beta \in]0, 1/4r[$ it is possible to obtain a better order of accuracy than for the deterministic noise model. This is because in this case it is very improbable that the error concentrates in the high frequencies, actually it is expected to get smaller there.

So information about the noise covariance structure, as given by ϱ in (14) is important and can essentially improve the performance of the data processing.

4.2 Local vs. global solutions

One more important fact should be pointed out. Discussing inverse estimation from noisy observation it is reasonable to distinguish global and local regularization. Within the global regularization one tries to recover the whole solution x as an element of some space $\mathcal{X}$. In the local regularization, as discussed here, the goal is to specify only some details such as a point value or some wavelet or Fourier coefficient, respectively.

For colored observation noise local regularization allows to obtain better accuracy than for deterministic noise in contrast to global regularization. In [MP05] it has been shown that under the same assumptions as above the best possible (global) accuracy for the deterministic noise model has the order $\delta^{\frac{2\mu}{2\mu+1}}$ while for colored noise it is of order $\delta^{\frac{2\mu}{2\mu+1+1/r-4\beta}}$ for $0 < \beta < 1/4r$.

This means in particular that colored noise will not improve recovering the whole solution but it will enhance recovering some local details.

5 Noise Behavior

For the adaptive choice of the regularization parameter α_+ according to (6) the quantity δ/Ω is of high importance. It roughly describes the error behavior with respect to regularization. First we will give a couple of possibilities of obtaining δ/Ω in practice and then discuss the impact of estimation errors.

5.1 Estimation

As (5) indicates we need $\delta||\mathbf{K}_\xi z_\alpha|| \le \delta/\Omega(\alpha)$ for all $\alpha \in \Delta_M$. Therefore we will try to estimate $\delta||\mathbf{K}_\xi z_\alpha||$

instead of $\delta/\Omega(\alpha)$.

5.1.1 Knowing $\mathbf{K}_\xi$

If we already know $\mathbf{K}_\xi$ by some other means, then we can directly compute $\delta||\mathbf{K}_\xi z_\alpha||$ and hence $\delta/\Omega(\alpha)$.

Please note that due to just considering $||\mathbf{K}_\xi z_\alpha||$ this is much more stable towards misspecification of $\mathbf{K}_\xi$ than taking $\mathbf{K}_\xi$ itself as in other regularization procedures.

5.1.2 $ARMA(p,q)$ Models

The quantity $\delta||\mathbf{K}_\xi z_\alpha||$ can be estimated for $ARMA(p,q)$ models in a number of cases which have been extensively discussed in [KDB03].

5.1.3 Several Data Sets

Any method for estimating $\mathbb{E}|\langle \xi, z_\alpha\rangle|^2$ can be used in order to get the necessary information about $||\mathbf{K}_\xi z_\alpha||$. A particular easy and feasible possibility is the following one:

When we regard satellite missions we are often in the situation that we have a large number of data points which can be parted into several independent data sets. If we assume that ξ_1 and ξ_2 are independent and generated by the same random process, then the respective data sets y_1^δ and y_2^δ obey

$$\mathbb{E}|\langle z_\alpha, y_1^\delta\rangle - \langle z_\alpha, y_2^\delta\rangle|^2 \\ = \delta^2\mathbb{E}\langle z_\alpha, \xi_1 - \xi_2\rangle^2 = 2\delta^2||\mathbf{K}_\xi z_\alpha||^2.$$

5.1.4 Locally known Data

Sometimes we are in the situation that we have a very accurate knowledge of small parts of x but no information about the other parts of x, e.g. looking at terrestrial measurements. Assuming that the noise situation in this small part Γ_{part} is the same as on the whole of Γ and remembering that

$$\langle f, g\rangle = \int_\Gamma fg\; d\omega$$

we can define a subset inner product by

$$\langle f, g\rangle_{\text{part}} = \int_{\Gamma_{\text{part}}} fg\; d\omega$$

This directly leads to

$$\mathbb{E}|\langle z_\alpha, Ax + \delta\xi\rangle_{\text{part}} - \langle z_\alpha, Ax\rangle_{\text{part}}|^2 \\ = \delta^2\mathbb{E}\left|\langle z_\alpha, \xi\rangle_{\text{part}}\right|^2 \approx \left(\frac{\int_{\Gamma_{\text{part}}} 1 d\omega}{\int_\Gamma 1 d\omega}\right)^2 \delta^2||\mathbf{K}_\xi z_\alpha||^2.$$

5.2 Estimation Errors

Denote the estimated version of δ/Ω by $\widetilde{\delta}/\widetilde{\Omega}$. Note that κ is also dependent on δ and hence we additionally get an estimated version $\widetilde{\kappa}$. By setting $\widetilde{\chi} = (\widetilde{\kappa}\widetilde{\delta}/\widetilde{\Omega})/(\kappa\delta/\Omega)$ we get a modification of the balancing principle (6). Let $\widetilde{\Delta}_M^{\widetilde{\Omega},\widetilde{\delta}}$ be the set of all $\alpha_j \in \Delta_M$ such that

$$\max_{\alpha_j \geq \alpha_i \in \Delta_M} \left\{ \left|\langle z_{\alpha_j}, y^\delta\rangle - \langle z_{\alpha_i}, y^\delta\rangle\right| \frac{\widetilde{\Omega}(\alpha_i)}{\widetilde{\delta}} \right\} \leq 4\widetilde{\kappa},$$

which is equivalent to

$$\max_{\alpha_j \geq \alpha_i \in \Delta_M} \left\{ \left|\langle z_{\alpha_j}, y^\delta\rangle - \langle z_{\alpha_i}, y^\delta\rangle\right| \frac{\Omega(\alpha_i)}{\delta} \right\} \leq 4\left(\kappa\widetilde{\chi}(\alpha_j)\right), \tag{20}$$

and $\alpha_{\widetilde{+}} := \max\left\{\alpha_j, \quad \alpha_j \in \widetilde{\Delta}_M^{\widetilde{\Omega},\widetilde{\delta}}\right\}$.

This means in particular that we can see the randomness in the estimation of δ/Ω as an additional factor close to 1 which enters our equations as modification of κ. In particular this means that with the "tuning" parameter $\kappa\widetilde{\chi}$ instead of κ theorem 2.1 still holds.

Now define a maximum $\chi_{max} = \max_{\alpha\in\Delta_M} \widetilde{\chi}(\alpha)$ and a minimum $\chi_{min} = \min_{\alpha\in\Delta_M} \widetilde{\chi}(\alpha)$ and assume that 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:

$$\mathbb{P}\{\chi_{max} > \tau\} \leq C_1 \exp\left(-C_2\left(\tau - 1\right)\right) \tag{21}$$

and

$$\mathbb{P}\{\chi_{min}^{-1} > \tau\} \leq C_1 \exp\left(-C_2\left(\tau - 1\right)\right) \tag{22}$$

Furthermore $\chi = (\chi_{max}, \chi_{min})$ shall be uncorrelated to the noise element ξ.

Then we have

Lemma 5.1 *If (1) describes a severely ill-posed problem, which means*

$$\mathbf{e}_\delta(x, z_\alpha, \Delta_M) = O\left(\ln^{-\mu} 1/\delta\right)$$

for some $0 < \mu$*, then for a problem dependent constant* d *(which can be precomputed) and*

$$\kappa = 4\sqrt{d \ln\ln\delta^{-1}}$$

we have for δ *small enough*

$$\mathbb{E}_\chi \mathbb{E}_\xi \left| \langle l, x\rangle - \left\langle z_{\alpha_{\mp}}, y^\delta \right\rangle \right|^2 \le c_2 \left(\ln\ln 1/\delta\right)^2 \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M),$$

where c_2 *is independent of* δ*.*

Proof

As χ was assumed to be uncorrelated to ξ we have

$$\mathbb{E}_\chi \left(\mathbb{E}_\xi \left| \langle l, x\rangle - \left\langle z_{\alpha_{\mp}}, y^\delta \right\rangle \right|^2 \right) \le \mathbb{E}_\chi \left(\Pi(\chi)\right),$$

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

$$\begin{aligned}\Pi(\chi) =& C \left(\ln\delta^{-1}\right)^{\mu_0 - \chi^2_{min} \ln\ln\delta^{-1}} + \\ & C\chi^2_{max} \left(\ln\ln\delta^{-1}\right)^2 \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M)\end{aligned}$$

for some constants C and μ_0 independent of δ and χ as proved in theorem 2.1.

Now we introduce functions Π_1 and Π_2 as

$$\Pi_1(\chi_{min}) := \left(\ln\delta^{-1}\right)^{2\mu + \mu_0 - \chi^2_{min}\ln\ln\delta^{-1}}$$

and

$$\Pi_2(\chi_{max}) := \chi^2_{max},$$

hence

$$\begin{aligned}\Pi(\chi) = & \, C\left(\ln 1/\delta\right)^{-2\mu} \Pi_1(\chi_{min}) \\ & + C\left(\ln\ln 1/\delta\right)^2 \mathbf{e}_\delta^2(x, z_\alpha, \Delta_M)\, \Pi_2(\chi_{max}).\end{aligned}$$

Using (22) and that $\delta > 0$ obeys

$$2\mu + \mu_0 - (1/4)\ln\ln 1/\delta < 0,$$

we conclude as in [Bau04] that

$$\mathbb{E}_\chi \Pi_1(\chi_{min}) \le \tilde{c} < \infty.$$

Similarly, using (21) we get $\mathbb{E}_\chi \Pi_2(\chi_{max}) \le \tilde{c} < \infty$. Taking into account that $\mathbf{e}_\delta(x, z_\alpha, \Delta_M) = O\left(\ln^{-\mu} 1/\delta\right)$ and putting things together the lemma is proved.

This means in particular that our adaptive regularization strategy even works when we estimate the behavior of $\delta||\mathbf{K}_\xi||$ in contrast to e.g. [KDB03].

We just loose the factor $\ln\ln[1/\delta]$ which is negligible in the logarithmic scale.

Similar results also hold for moderately ill-posed problems as has been shown in [Bau04].

6 Numerics

6.1 Remarks on the Balancing Principle

One of the most delicate points in practice seems to be the choice of the tuning parameter κ. On the one hand some estimations in the proof seem to be too conservative and we just need κ to be a function which is ascending in δ much slower than proposed if at all.

On the other hand in practice we are not really interested in rates but we would like to have the best possible result where the tuning parameter κ is stable over a large range of error levels δ.

Extensive trials with model problems yield the following qualitative results where the tuning parameter κ was chosen problem dependent (mostly around 0.2 – 0.7) for one instance of the problem and then kept constant for the whole experiment.

The balancing principle

- is rather insensitive to misspecified noise models;
- seems to be superior to other parameter choice regimes, like e.g., the L-curve method, we refer to [Bau04] for a discussion.

We also observed that Tikhonov regularization is slightly more stable than truncated singular value decomposition but in general yields slightly worse results.

As one can see it from theorem 3.3 and proposition 3.4, the regularization parameter realizing an accuracy of optimal order for functional estimation does not depend on the functional of interest. It gives a heuristic reason to use the parameter given by balancing principle also for the whole solution recovery, especially for the reconstruction of the solutions of severely ill-posed problems such as satellite gravity gradiometry. Numerical tests supporting this heuristic reason are described below.

6.2 Case study

As an example we consider a similar case as for the GOCE mission. We assume to have gravity data at an orbit height of about $400km$ and tried to reconstruct the gravitational field at the height of the Earth's surface.

As described in subsection 5.1.4 we use a small known part of the Earth to generate δ/Ω which represents our knowledge about the noise behavior.

In particular we want to show that the balancing principle exhibits stable behavior under small errors obtained from estimating δ/Ω. Therefore we considered the *same* noisy input data set and determined a regularization parameter chosen by the balancing principle with respect to a number of different given small reference data sets.

In particular we carried out the following 28 experiments. The first time we partitioned the Earth's surface into four, the second time into eight and the third time into 16 different parts. Later on we will display a table where we indicate for which of the experiments which regularization parameter was chosen. We will denote by q_i^k the regularization parameter chosen on the base of estimating δ/Ω corresponding to the $1/k$ part with number i used as reference set; in the table it will be put in the row corresponding to its value.

6.2.1 Requirements

The final simulation was designed under the following requirements:

- noise model with correlated noise,
- possibility to judge the goodness of solutions,
- enough computations to show the reliability of the method,
- very limited computer resources.

Therefore we decided to use the following setup

- simulated satellite data on an integration grid,
- uncorrelated noise and additionally space correlated noise,
- truncated singular value decomposition.

6.2.2 Technical Remarks

We used a Driscoll-Healy grid (as e.g., [May01]) as data location at an orbit height of 6% of the Earth radius which roughly corresponds to an average satellite height of 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 (cf. e.g. [Dea98]). 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 which was added on the points of the integration grid.

Table 1: Results

Degree	$\frac{\text{Bias}}{\text{Variance}}$	Reg. Params.
...	...	
57	0.1726	
58	0.2119	
59	0.2124	
60	0.2500	q_0^{16}
61	0.2808	q_0^4 , q_0^8
62	0.3016	q_1^8 , q_6^8 , q_5^{16} , q_{10}^{16} , q_{15}^{16}
63	0.3348	q_3^4 , q_2^8 , q_1^{16} , q_8^{16}
64	0.3940	q_1^4 , q_2^4 , q_5^8 , q_7^8 , q_4^{16} , q_{11}^{16} , q_{13}^{16}
65	0.4376	q_2^{16}
66	0.4363	q_3^8 , q_{14}^{16}
67	0.4727	q_4^8 , q_3^{16} , q_6^{16}
68	0.5192	
69	0.5430	
70	0.6805	
71	0.7270	
72	0.6788	
73	0.8126	q_7^{16} , q_{12}^{16}
74	0.7578	
75	0.9144	q_9^{16}
76	**1.0299**	
77	1.0720	
78	1.1522	
79	1.2194	
80	1.3589	
81	1.3088	
...	...	

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

For our purposes we observed that reliable results were obtained for $\kappa \approx 0.25$.

After having chosen this parameter we proceeded with our experiment. Note that this parameter seems to be strongly dependent from the chosen underlying grid and other model parameters. However, once adapted to the particular problem it seems to handle different data situations.

6.2.3 Discussion

We observe that the optimal cut-off point is chosen reliably and rather near to the optimal one (degree 76 where the bias to variance ratio gets bigger than 1). Some more results, e.g. regarding the choice by an L-curve method (yields a cut-off around degree 40) can be found in [Bau04].

Slightly better results (in average) were produced by a tuning parameter $\kappa = 0.225$, however this was paid by accepting some outliers.

7 Conclusion

The presented adaptive parameter choice is an easy to use heuristics which allows to choose the regularization parameter robustly even when we have a very limited knowledge about the underlying noise structure. It relies on a problem depending tuning parameter which can be chosen in advance because the behavior of the chosen regularization parameter does not change considerably even when we change the noise level over several orders of magnitude or change the type of error.

We could show that in theory we can achieve (almost) optimal convergence rates and for a data-functional strategy we can even use the color of the noise as an advantage which helps to improve the solution. For this result it suffices to estimate the noise behavior; this does not spoil the optimal rate result.

From the practical point of view the Lepskiĭ-type balancing principle works reliably, even when we only use about 6% of the Earth's surface as calibration data, only. The robustness has been shown not just in the gravity gradiometry case but also for numerous other inverse problems.

Acknowledgements

The support of the Graduiertenkolleg "Mathematics and Practice", University of Kaiserslautern and the Graduiertenkolleg "Identification in mathematical models", University of Göttingen is kindly acknowledged.

Part of this work was written during the visit of the third author to the Weierstrass Institute Berlin (May 2005). The support of this institute is gratefully acknowledged.

References

[And86] R.S. Anderssen, *The linear functional strategy for improperly posed problems*, Inverse problems, Proc. Conf., Oberwolfach/Ger. 1986 (Basel, Boston, Stuttgart) (J. Cannon and U. Hornung, eds.), International Series of Numerical Mathematics, vol. 77, Birkhäuser Verlag, 1986, pp. 11–30.

[Bau04] Frank Bauer, *An alternate approach to the oblique derivative problem in potential theory*, Ph.D. thesis, Universtiy of Kaiserslautern, Department of Mathematics, Geomathematics Group, 2004.

[Böt05] A. Böttcher, Personal Communications, 2005.

[BP05] F. Bauer and S. Pereverzev, *Regularization without preliminary knowledge of smoothness and error behavior (accepted)*, European J. Appl. Math. (2005).

[Dea98] R. Deakin, *Derivatives of the earth's potential*, Geomatics Research Australasia **68** (1998), 31–60.

[Don94] D. Donoho, *Statistical estimation and optimal recovery*, Ann. Stat. **22** (1994), no. 1, 238–270.

[EHN96] H. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.

[EN94] H. Engl and A. Neubauer, *A parameter choice strategy for (iterated) tikhonov regularization of ill-posed problems leading to superconvergence with optimal rates*, Appl. Anal. **27** (1994), no. 1–3, 5–18.

[Eur99] EuropeanSpaceAgency, *Gravity field and steady-state ocean circulation mission. report for mission selection – the four candidate earth explorer core missions.*, ESA SP-1233 (1999).

[FM02] W. Freeden and T. Mayer, *On multiscale denoising of spherical functions: Basic theory and numerical aspects*, ETNA **14** (2002), 56–78.

[FP01] W. Freeden and S. Pereverzev, *Spherical tikhonov regularization wavelets in satellite gravity gradiometry with random noise*, Journal of Geodesy **74** (2001), 730–736.

[Fre99] W. Freeden, *Multiscale modelling of spaceborne geodata*, European Consortium for Mathematics in Industry, B. G. Teubner, Stuttgart, Leipzig, 1999.

[GP00] Alexander Goldenshluger and Sergei Pereverzev, *Adaptive estimation of linear functionals in hilbert scales from indirect white noise observations*, Probability Theory and Related Fields **118** (2000), no. 2, 169–186.

[KDB03] R. Klees, P. Ditmar, and P. Broersen, *How to handle colored observation noise in large least-squares problems*, J. Geod. **76** (2003), 629–640.

[KK02] J. Kusche and R. Klees, *Regularization of gravity field estimation from satellite gravity gradients*, J. Geod. **76** (2002), 359–368.

[Lep90] O.V. Lepskiĭ, *On a problem of adaptive estimation in Gaussian white noise.*, Theory of Probability and its Applications **35** (1990), no. 3, 454–466.

[LKB01] M. Lonkhuyzen, R. Klees, and J. Bouman, *Regularization for the gravity field recovery from goce observations*, Proc. IAG Int. Symp. Gravity, Geoid Geodynamics 2000, Banff, 2001.

[May01] Carsten Mayer, *Potential operators, jump relations and wavelets*, 2001.

[MP02] Peter Mathé and Sergei Pereverzev, *Direct estimation of linear functionals from indirect noisy observations*, Journal of Complexity **18** (2002), no. 2, 501–516.

[MP03] Peter Mathé and Sergei Pereverzev, *Geometry of linear ill-posed problems in variable hilbert spaces*, Inverse Problems **19** (2003), no. 3, 789–803.

[MP05] Peter Mathé and Sergei Pereverzev, *Regularization of some linear inverse problems with discretized random noise data (submitted).*

[PP02] R. Pail and G Plank, *Assessment of three numerical solution strategies for gravity field recovery from goce satellite gravity gradiometry implemented on a parallel platform.*, Journal of Geodesy **76** (2002), 462–474.

[PS99] S. Pereverzev and E. Schock, *Error estimates for band-limited spherical regularization wavelets in an inverse problem of satellite geodesy*, Inverse Problems **15** (1999), no. 4, 881–890.

[Tsy00] A. Tsybakov, *On the best rate of adaptive estimation in some inverse problems*, C. R. Acad. Sci., Paris, Sér. I, Math. **330** (2000), no. 9, 835–840.

Institut für Numerische und Angewandte Mathematik
Universität Göttingen
Lotzestr. 16-18
D - 37083 Göttingen

Telefon:	0551/394512
Telefax:	0551/393944

Email: trapp@math.uni-goettingen.de URL: http://www.num.math.uni-goettingen.de

Verzeichnis der erschienenen Preprints:

2005-01	A. Schöbel, S. Scholl:	Line Planning with Minimal Traveling Time
2005-02	A. Schöbel	Integer programming approaches for solving the delay management problem
2005-03	A. Schöbel	Set covering problems with consecutive ones property
2005-04	S. Mecke, A. Schöbel, D. Wagner	Station location - Complex issues
2005-05	R. Schaback	Convergence Analysis of Methods for Solving General Equations
2005-06	M. Bozzini, L. Lenarduzzi, R. Schaback	Kernel B–Splines and Interpolation
2005-07	R. Schaback	Convergence of Unsymmetric Kernel–Based Meshless Collocation Methods
2005-08	T. Hohage, F.J. Sayas	Numerical solution of a heat diffusion problem by boundary element methods using the Laplace transform
2005-09	T. Hohage	An iterative method for inverse medium scattering problems based on factorization of the far field operator
2005-10	F. Bauer, T. Hohage	A Lepskij-type stopping rule for regularized Newton methods
2005-11	V. Dolean, F. Nataf, G. Rapin	New constructions of domain decomposition methods for systems of PDEs Nouvelles constructions de méthodes de décomposition de domaine pour des systèmes d'équations aux dérivées partielles
2005-12	F. Nataf, G. Rapin	Construction of a New Domain Decomposition Method for the Stokes Equations
2005-13	Hon, Y.C., Schaback, R.	Solvability of Partial Differential Equations by Meshless Kernel Methods
2005-14	Frank Bauer, Peter Mathé, Sergei Pereverzev	Local Solutions to Inverse Problems in Geodesy: The Impact of the Noise Covariance Structure upon the Accuracy of Estimation