Georg-August-Universität

Göttingen

Some Considerations Concerning Regularization and Parameter Choice Algorithms

Frank Bauer

Preprint Nr. 2006-20

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

Some Considerations Concerning Regularization and Parameter Choice Algorithms

Frank Bauer†

†University of Göttingen,
Institute for Numerical and Applied Mathematics,
Lotzestr. 16-18, 37083 Göttingen, Germany
and
Institute for Mathematical Stochastics,
Maschmühlenweg 8-10, 37073 Göttingen, Germany

E-mail: `bauer@mathematik.uni-kl.de`

Abstract. Using a Bayesian type approach to inverse problems many phenomenons occurring the practice can be explained in an comparably consistent and easy way. In particular we can prove the quasi-optimality criterion to choose the regularization parameter just imposing minor additional a priori assumptions about the solution and the measurement noise.

AMS classification scheme numbers: 65J22,60G60,62H12

1. Introduction

In inverse problems one is concerned to reconstruct a quantity $x \in \mathcal{X}$ out of noisy measurements $y^\delta = y + \delta\xi$ of a quantity $y \in \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are separable Hilbert spaces and $\widetilde{A} : \mathcal{X} \to \mathcal{Y}$ is a linear compact operator

$$\widetilde{A}x = y.$$

Because $\widetilde{A}$ is not continuously invertible we need to regularize. Well known regularization techniques are Spectral Cut-Off or Tikhonov regularization [1]. All regularization procedures require a regularization parameter whose choice is crucial.

In the theory of inverse problems one has the result by Bakushinskii [2] telling that without the knowledge of the noise level δ one cannot choose a regularization operator which guarantees a result with low error, neither absolutely nor in rate. However in stochastics purely data driven methods like cross validation [1] exist which can be proven in some non-inverse problems situations to return optimal solutions.

In practical situations normally we do not know the noise level and hence we are bound to use parameter choice algorithms which are supposed to work without this information. Examples are L-curve [3], Generalized Cross-Validation [4], Quasi-Optimality [1, 5] or smoothing of the data and inverting without or with minor regularization.

When one tries to prove parameter choice procedures similar to quasi-optimality one encounters that one has a lack of inequalities for the error terms from below. These inequalities seem to fail e.g. for regularization methods like Spectral Cut-Off if the energy just concentrates in a small number of Fourier coefficients. For the L-curve method there also exist very easy counter examples [6]. On the other hand in practice a couple of these methods work rather well [3]. So the question is if one can, at least in average, guarantee some lower bounds or to show that the counter examples to these methods are rare exceptions.

The general approach we will present is in principle not new and can be considered as a Bayesian framework. We will assume that the solution has some random structure, i.e. is drawn according to some prior distribution [7]. However in contrary to [7] we will not fix this prior but just assume that there is one with certain very general properties. In some sense similar considerations have been made in the statistical community, however not in the context of inverse problems, see e.g. [8, 9].

Using this assumption we will introduce several different optimality criteria which we will prove to be equivalent under certain conditions. A particular interesting additional result is that we can prove both the quasi-optimality criterion and smoothing of the data along with the inversion of the unregularized operator to be valid and order optimal regularization methods.

2. Prior Assumptions

We will restrict ourselves for the ease of notation to self-adjoint problems right now. The considerations would work using some more careful notation also in the general setting, however by multiplying a linear operator equation $\widetilde{A}x = y$ by $\widetilde{A}^*$ we can always get an operator equation $\widetilde{A}^*\widetilde{A}x = \widetilde{A}^*y$ respectively $\sqrt{\widetilde{A}^*\widetilde{A}}x = \sqrt{\widetilde{A}^*\widetilde{A}}^{-1}\widetilde{A}^*y$ with a self adjoint operator $\widetilde{A}^*\widetilde{A}$.

2.1. Operator

Consider the bounded self-adjoint compact linear operator

$$A : \mathcal{X} \to \mathcal{X}$$

where $\mathcal{X}$ is a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\| \cdot \|$.

A bounded self-adjoint linear compact operator between separable Hilbert spaces has a singular value decomposition

$$Ax = \sum_{k=1}^{\infty} \sigma(k) \langle x, u_k \rangle u_k \tag{1}$$

where $\{u_k\}_{k \in \mathbb{N}}$ is an orthonormal basis system of $\mathcal{X}$ (see e.g. [1]). $\sigma(\cdot)$ is assumed to be a continuous monotonously decreasing function $\sigma : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ asymptotically going to 0, i.e. $\lim_{k\to\infty} \sigma(k) = 0$.

The function σ should not decrease faster than exponentially, i.e. there is a positive constant c_σ such that for all $k \in \mathbb{N}_0$ it holds

$$\sigma(k + 1) \geq c_\sigma \sigma(k)$$

2.2. Equation

We want to solve the equation

$$Ax = y. \tag{2}$$

We cannot measure y but just a perturbed version $y^\delta = y + \delta\xi$, where ξ is a normalized error element.

2.3. The Solution x

We assume to have the following stochastical prior for x. All Fourier coefficients $\langle x, u_k \rangle$ should be independently distributed with normal distribution $\mathcal{N}(0, \gamma(k)^2)$.

The function $\gamma : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is assumed to be monotonously decreasing and square sumable, i.e.

$$\sum_{k=1}^{\infty} \gamma(k)^2 < \infty$$

Hence x fulfills the following relations for all $k \neq l$

$$\mathbb{E}\langle x, u_k\rangle = 0$$

$$\mathbb{E}\langle x, u_k\rangle \langle x, u_l\rangle = 0$$

and

$$\mathbb{E}\langle x, u_k\rangle^2 = \gamma(k)^2 \tag{3}$$

The function γ should not decrease faster than exponentially, i.e. there is a positive constant c_γ such that for all k it holds

$$\gamma(k+1) \geq c_\gamma \gamma(k) \tag{4}$$

Remark

Although it seems unusual to impose such a condition on x there are several lines of argumentation which support this. First of all source conditions can be seen (asymptotically) as a bound from above such that $\langle x, u_k\rangle^2 \leq \gamma(k)^2$. What we do is on the one hand relaxing this bound and on the other hand requiring that in average $\langle x, u_k\rangle^2$ behaves like $\gamma(k)^2$. This means in particular that we do not just have something like a bound from above but also from below without being strict in this sense.

Furthermore assume the case that we already know that x is in some space along the Hilbert scale, e.g. $\|A^\mu x\| < \infty$ or written differently $x \in \mathcal{H}_{A^\mu}$. For any $\nu < \mu$ the space $\mathcal{H}_{A^\nu}$ is a meagre subset of $\mathcal{H}_{A^\mu}$. This implies if we would draw an x out of $\mathcal{H}_{A^\mu}$ with probability zero x is also an element of $\mathcal{H}_{A^\nu}$. I.e. the Fourier coefficients are not likely to decrease fast enough to support x in $\mathcal{H}_{A^\nu}$. In order to have this behavior for all $\nu < \mu$ it is a sensible assumption that the Fourier coefficients behave like described above. In a non-functional analytic but stochastic setting one can find a similar argumentation in [10, 8]

2.4. The Error element ξ

We assume that the normalized error element is randomly chosen and has the property that the (formal) Fourier coefficients $\langle \xi, v_k\rangle$ are all independently distributed according to the normal distribution $\mathcal{N}(0, \varepsilon(k)^2)$.

The function $\varepsilon : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ should fulfill the following properties:

- $\varepsilon\sigma^{-p}$, $p \in \{0, 1\}$ is a monotonous function
- $\lim_{k\to\infty} \varepsilon(k)/(\sigma(k)\gamma(k)) \to \infty$

Without the second property we do not have an ill-posed problem. It holds

$$\mathbb{E}\langle \xi, v_k\rangle = 0$$

$$\mathbb{E}\langle \xi, v_k\rangle \langle \xi, v_l\rangle = 0$$

and

$$\mathbb{E}\langle \xi, v_k\rangle^2 = \varepsilon(k)^2 \tag{5}$$

The function ε should not decrease or increase faster than exponentially, i.e. there is a positive constant c_ε such that for all k it holds

$$c_\varepsilon^{-1}\varepsilon(k) \geq \varepsilon(k+1) \geq c_\varepsilon \varepsilon(k) \tag{6}$$

2.5. Independence

In all cases we will assume that x and ξ are independent. This means in particular that for all combinations of k and l it holds

$$\mathbb{E}\left\langle x, u_k\right\rangle \left\langle \xi, v_l\right\rangle = 0$$

This allows us in particular the following consideration. Let the operators $A_1 : \mathcal{X} \to \mathcal{X}$ and $A_2 : \mathcal{X} \to \mathcal{X}$.

Then it holds

$$\mathbb{E}\|A_1 x + A_2\xi\|^2 = \mathbb{E}\|A_1 x\|^2 + \mathbb{E}\|A_2\xi\|^2 + 2\mathbb{E}\left\langle A_1 x, A_2\xi\right\rangle = \mathbb{E}\|A_1 x\|^2 + \mathbb{E}\|A_2\xi\|^2$$

The triangle inequality would just give us (apart from a factor of 2) the $\leq$ sign. Actually, the possibility to decompose (in average) the norms in this way keeping the equality was one of the reasons to introduce the above conditions on x and ξ. Note that in order to have this equality we just require that one of the two quantities x and ξ is such a random variable.

2.6. Regularization

Now we define the notion of a regularization family [11, 1]. This is a family of functions $\{g_n\}_{n\in\mathbb{N}}$ which should replace the inversion operator $^{-1} : s \mapsto s^{-1}$ by a family of bounded approximands. It fulfills the following properties: for all $n < m$ and for all $s \in \mathbb{R}^+$

- $sg_n(s)$ monotonous
- $g_n(s) \leq g_m(s)$ and $g_n \neq g_m$
- $\lim_{n\to\infty} sg_n(s) = 1$
- $\sum_{k=1}^{\infty} g_n(\sigma(k))^2\varepsilon(k)^2 < \infty$

Examples of regularization families are

- Spectral Cut-Off:

$$g_n(s) = \begin{cases} s^{-1} & s \geq \sigma(n) \\ 0 & s < \sigma(n) \end{cases}$$

- Tikhonov

$$g_n(s) = \frac{s}{s^2 + q^n}$$

for some $0 < q < 1$

3. First Considerations

Now we can do some calculations always assuming $n \leq m$. In order to reduce the number of equations and furthermore to introduce interesting additional aspects we will always consider vectors of the form $A^p x$. For $p = 0$ we have x, for $p = 1$ we have y.

3.1. Abbreviations

In order to have a simpler access to some quantities we will replace some sums which we will have later on with easier expressions. Define the functions $\overline{\varrho}_p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ and $\overline{\varphi}_p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ by

$$\overline{\varrho}_p(k)^2 = \sum_{k=1}^{\infty} \sigma(k)^{2p} \varepsilon(k)^2 g_n(\sigma(k))^2 \tag{7}$$

$$\overline{\varphi}_p(k)^2 = \sum_{k=1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \left(1 - \sigma(k) g_n(\sigma(k))\right)^2 \tag{8}$$

where for any point $k < l < k+1$ the function is the straight line between the points at k and $k+1$.

Due to the fact that none of the functions σ, γ and ε is exhibiting behavior worse than exponentially we have that neither $\overline{\varrho}_p$ nor $\overline{\varphi}_p$ is exhibiting behavior worse than exponentially. This means that there exist positive constants c_{ϱ_p} and c_{φ_p} such that

$$\overline{\varrho}_p(k+1) \leq c_{\varrho_p} \overline{\varrho}_p(k) \tag{9}$$

$$\overline{\varphi}_p(k+1) \geq c_{\varphi_p} \overline{\varphi}_p(k) \tag{10}$$

3.2. Expectations

Using the definitions of x and ξ we can easily evaluate the following expectations.

$$\mathbb{E}\|A^p x\|^2 = \sum_{k=1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 = \overline{\varphi}_p(0)^2 \tag{11}$$

$$\mathbb{E}\|A^p x_n\|^2 = \sum_{k=1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \sigma(k)^2 g_n(\sigma(k))^2 = \overline{\varphi}_p(0)^2 - \overline{\varphi}_p(n)^2 \tag{12}$$

$$\begin{aligned} \mathbb{E}\|A^p x_n^\delta\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p} \left(\gamma(k)^2 \sigma(k)^2 + \delta^2 \varepsilon(k)^2\right) g_n(\sigma(k))^2 \\ &= \overline{\varphi}_p(0)^2 - \overline{\varphi}_p(n)^2 + \delta^2 \overline{\varrho}_p(n)^2 \end{aligned} \tag{13}$$

$$\begin{aligned} \mathbb{E}\|A^p (x_m - x_n)\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \sigma(k)^2 \left(g_m(\sigma(k)) - g_n(\sigma(k))\right)^2 \\ &= \overline{\varphi}_p(n)^2 - \overline{\varphi}_p(m)^2 \end{aligned} \tag{14}$$

$$\begin{aligned}
\mathbb{E}\|A^p\left(x_m^\delta - x_n^\delta\right)\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p}\left(\gamma(k)^2\sigma(k)^2 + \delta^2\varepsilon(k)^2\right)\left(g_m(\sigma(k)) - g_n(\sigma(k))\right)^2 \\
&= \overline{\varphi}_p(n)^2 - \overline{\varphi}_p(m)^2 + \delta^2\overline{\varrho}_p(m)^2 - \delta^2\overline{\varrho}_p(n)^2 \qquad (15)\\
\mathbb{E}\|A^p\left(x_n^\delta - x_n\right)\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p}\delta^2\varepsilon(k)^2 g_n(\sigma(k))^2 = \delta^2\overline{\varrho}_p(n)^2 \qquad (16)\\
\mathbb{E}\|A^p\left(x - x_n\right)\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p}\gamma(k)^2\left(1 - \sigma(k)g_n(\sigma(k))\right)^2 = \overline{\varphi}_p(n)^2 \qquad (17)\\
\mathbb{E}\|A^p\left(x - x_n^\delta\right)\|^2 &= \sum_{k=1}^{\infty} \sigma(k)^{2p}\gamma(k)^2\left(1 - \sigma(k)g_n(\sigma(k))\right)^2 + \delta^2\varepsilon(k)^2 g_n(\sigma(k))^2 \\
&= \overline{\varphi}_p(n)^2 + \delta^2\overline{\varrho}_p(n)^2 \qquad (18)
\end{aligned}$$

4. Spectral Cut-Off

First of all we will consider the spectral cut-off regularization scheme because it has the advantage that

$$g_n(\sigma(k))\sigma(k) = \begin{cases} 1 & k \le n \\ 0 & k > n \end{cases}$$

And hence also the term $(m \ge n)$ gets

$$\left(g_m(\sigma(k)) - g_n(\sigma(k))\right)\sigma(k) = \begin{cases} 0 & k \le n \\ 1 & n+1 \le k \le m \\ 0 & k > m \end{cases}$$

Therefore the preceding sums (7) and (8) can be replaced by simpler versions, in particular:

$$\varrho_p(k)^2 = \sum_{k=1}^{n} \sigma(k)^{2p-2}\varepsilon(k)^2 \qquad (19)$$

$$\varphi_p(k)^2 = \sum_{k=n+1}^{\infty} \sigma(k)^{2p}\gamma(k)^2 \qquad (20)$$

4.1. Concentration

In order to gain more stochastical stability we will not consider all possible regularized solutions but a much smaller number which will still yield, if appropriately chosen, optimal rates. The function choosing the "right" regularized solutions will be called concentration.

Definition 4.1 (Concentration) *The function $l_p : \mathbb{N} \to \mathbb{N}$ is called concentration iff there exist positive constants $c_{l,\mathrm{up}}$ and $c_{l,\mathrm{reglow}}$ and $k_{minp} \ge 1$ such that it holds*

- $l_p(0) = 0$
- $l_p(1) = k_{minp}$
- $l_p(n+1) - l_p(n) \geq l_p(n) - l_p(n-1)$
- $\varrho_p(l_p(n+1))^2 \leq c_{l,\text{up}}^2 \varrho_p(l_p(n))^2$
- $\varphi_p(l_p(n+1))^2 \geq c_{l,\text{reglow}}^2 \varphi_p(l_p(n))^2$

Due to (9) and (10) concentrations always exist.

Definition 4.2 (Valid exponential concentrations) *A concentration additionally fulfilling for* $c_{l,\text{regup}} < 1$ *and* $c_{l,\text{low}} > 1$

$$c_{l,\text{low}}^2 \varrho_p(l_p(n))^2 \leq \varrho_p(l_p(n+1))^2 \leq c_{l,\text{up}}^2 \varrho_p(l_p(n))^2 \tag{21}$$

and

$$c_{l,\text{reglow}}^2 \varphi_p(l_p(n+1))^2 \leq \varphi_p(l_p(n))^2 \leq c_{l,\text{regup}}^2 \varphi_p(l_p(n+1))^2 \tag{22}$$

is called to be exponentially valid.

4.2. Optimality Criteria

We have four different optimality criteria. We will show that they are equivalent in certain cases and guarantee that we have an optimal rate solution.

Definition 4.3 (Norm criterion) *The parameter* $m \in \mathbb{N}$ *fulfills the* norm criterion *w.r.t.* p *if it holds*

$$\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 \leq c_{\text{other}}^2 \min_{n \in \mathbb{N}} \mathbb{E}\|A^p\left(x - x_n^\delta\right)\|^2 \tag{23}$$

Definition 4.4 (Sum criterion) *A parameter* $m \in \mathbb{R}^+$ *fulfills the* sum criterion *w.r.t.* p *if it holds*

$$c_{\min}^2 \varphi_p(m)^2 \leq \delta^2 \varrho_p(m)^2 \leq c_{\max}^2 \varphi_p(m)^2 \tag{24}$$

Definition 4.5 (Integral criterion) *A parameter* $m \in \mathbb{R}^+$ *fulfills the* integral criterion *w.r.t.* p *if it holds*

$$c_{\min}^2 \int_m^\infty \sigma(x)^{2p} \sigma(x)^2 \gamma(x)^2 dx \leq \int_1^m \sigma(x)^{2p} \delta^2 \varepsilon(x)^2 dx \leq c_{\max}^2 \int_m^\infty \sigma(x)^{2p} \sigma(x)^2 \gamma(x)^2 dx \tag{25}$$

Definition 4.6 (Intersection criterion) *A parameter* $m \in \mathbb{R}^+$ *fulfills the* intersection criterion *if it holds*

$$c_{\min}^2 \sigma(m)^2 \gamma(m)^2 \leq \delta^2 \varepsilon(m)^2 \leq c_{\max}^2 \sigma(m)^2 \gamma(m)^2 \tag{26}$$

This is the only definition which is not depending on p.

4.3. Optimal Solutions

According to the above definitions we define

Definition 4.7 (Norm optimal parameter) *The parameter* $n_{\text{norm}} \in \mathbb{N}$ *is called* norm optimal *w.r.t. p if it holds*

$$\mathbb{E}\|A^p\left(x - x^\delta_{n_{\text{norm}}}\right)\|^2 = \min_{n\in\mathbb{N}} \mathbb{E}\|A^p\left(x - x^\delta_n\right)\|^2 \tag{27}$$

This criterion is also known under the name oracle inequality.

Definition 4.8 (Sum optimal parameter) *The parameter* $r_{\text{isect}} \in \mathbb{R}^+$ *is called* continuous sum optimal *if it holds*

$$\varphi_p(r_{\text{sum}})^2 = \delta^2 \varrho_p(r_{\text{sum}})^2 \tag{28}$$

It is called sum optimal *if*

$$n_{\text{sum}} = \lceil r_{\text{sum}} \rceil \tag{29}$$

A concentrated version is

$$k_{\text{sum}} = \min_{k\in\mathbb{N}}\{l(k) \geq n_{\text{sum}}\} \tag{30}$$

Definition 4.9 (Integral criterion) *The parameter* $r_{\text{int}} \in \mathbb{R}^+$ *is called* continuous integral optimal *if it holds*

$$\int_{r_{\text{int}}}^{\infty} \sigma(x)^{2p}\sigma(x)^2\gamma(x)^2 dx = \int_1^{r_{\text{int}}} \sigma(x)^{2p}\delta^2\varepsilon(x)^2 dx \tag{31}$$

It is called integral optimal *if*

$$n_{\text{int}} = \lceil r_{\text{int}} \rceil \tag{32}$$

Definition 4.10 (Intersection optimal parameter) *The parameter* $r_{\text{isect}} \in \mathbb{R}^+$ *is called* continuous intersection optimal *if it holds*

$$\sigma(r_{\text{isect}})^2\gamma(r_{\text{isect}})^2 = \delta^2\varepsilon(r_{\text{isect}})^2 \tag{33}$$

It is called intersection optimal *if*

$$n_{\text{isect}} = \lceil r_{\text{isect}} \rceil \tag{34}$$

Lemma 4.1 *Assume m fulfills the sum criterion and that l is a concentration with $l(k) \leq m \leq l(k+1)$. Then $l(k)$ and $l(k+1)$ fulfill the sum criterion.*

Proof
Using the monotonicity of ϱ_p and φ_p together with the inequalities $\varrho_p(l_p(n+1))^2 \leq c^2_{l,\text{up}}\varrho_p(l_p(n))^2$ and $\varphi_p(l_p(n+1))^2 \geq c^2_{l,\text{reglow}}\varphi_p(l_p(n))^2$ of the definition of a concentration the result is straightforward.

4.4. Requirements

Definition 4.11 (Valid p) *The parameter p is called* valid *if*

- $\sigma(k)^{2p}\delta^2\sigma(k)^{-2}\varepsilon(k)^2$ *is increasing*
- $\sigma(k)^{2p}\gamma(k)^2$ *is decreasing.*

Remark
This means in particular that also ϱ_p is increasing and φ_p is decreasing. This property makes that the problem we consider actually behaves in the way we expect from an inverse problem.

Definition 4.12 (Lower Equal Behavior) *Two real valued functions $f(k)$ and $g(k)$ are called to fulfill the* lower equal behavior condition*, if either*

- *Both of them can be bounded from below by a positive polynomial in k*
- *Both of them can be bounded from below by a positive exponential function in k*

Definition 4.13 (Equal Behavior) *Two real valued functions $f(k)$ and $g(k)$ are called to fulfill the* equal behavior condition*, if either*

- *Both of them can be bounded from below by a positive polynomial and above by a polynomial in k*
- *Both of them can be bounded from below by a positive exponential function and above by an exponential function in k*

Remark
Please note that this behavior condition trivially transfers to $\varphi_p(k)^2$ and $\delta^2\varrho_p(k)^2$ and vice versa as long as we have a valid p.

4.5. Results without additional assumptions

Now we will show step by step a theorem which will be the key part of the later considerations.

Theorem 4.2 *Assume that p is valid. Then the following diagram holds*

$$\begin{array}{ccccc} & & \textit{Intersection criterion} & & \\ & & \Downarrow & & \\ \textit{Norm criterion}(p) & \Leftarrow & \textit{Sum criterion}(p) & \Leftrightarrow & \textit{Integral criterion}(p) \end{array}$$

Lemma 4.3 (Sum criterion(p) $\Leftrightarrow$ Integral criterion(p)) *Let p be valid. Then the sum and the integral criterion are equivalent.*

Proof
Trivial consequence of the monotonicity of $\sigma(k)^{2p}\delta^2\sigma(k)^{-2}\varepsilon(k)^2$ and $\sigma(k)^{2p}\gamma(k)^2$ and their property of not changing faster than exponentially.

Lemma 4.4 (Sum criterion(p) ⇒ Norm criterion(p)) *Assume that p is valid and for $m \in \mathbb{N}$*

$$c_{\min}^2 \varphi_p(m)^2 \leq \delta^2 \varrho_p(m)^2 \leq c_{\max}^2 \varphi_p(m)^2$$

Then it holds for n_{norm}

$$\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 \leq \left(1 + \max\{c_{\max}^2, c_{\min}^{-2}\}\right) \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2$$

Proof

There are two cases to consider

Case 1: $m < n_{\text{norm}}$

Due to ϱ_p increasing it holds

$$\begin{aligned}\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 =& \varphi_p(m)^2 + \delta^2 \varrho_p(m)^2 \\ &\leq \left(1 + c_{\max}^2\right) \delta^2 \varrho_p(m)^2 \\ &\leq \left(1 + c_{\max}^2\right) \delta^2 \varrho_p(n_{\text{norm}})^2 \\ &\leq \left(1 + c_{\max}^2\right) \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2\end{aligned}$$

Case 2: $m > n_{\text{norm}}$

It holds due to φ_p decreasing

$$\begin{aligned}\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 =& \varphi_p(m)^2 + \delta^2 \varrho_p(m)^2 \\ &\leq \left(1 + c_{\min}^{-2}\right) \varphi_p(m)^2 \\ &\leq \left(1 + c_{\min}^{-2}\right) \varphi_p(n_{\text{norm}})^2 \\ &\leq \left(1 + c_{\min}^{-2}\right) \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2\end{aligned}$$

Hence the above inequality holds.

Corollary 4.5 *Let p be valid. It holds that there is a positive constant c_{other} independent of δ such that*

$$\mathbb{E}\|A^p\left(x - x_{n_{\text{sum}}}^\delta\right)\|^2 \leq c_{\text{other}}^2 \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2$$

and

$$\mathbb{E}\|A^p\left(x - x_{n_{\text{int}}}^\delta\right)\|^2 \leq c_{\text{other}}^2 \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2$$

Proof

On the one hand we have by definition

$$\varphi_p(n_{\text{sum}})^2 \leq \delta^2 \varrho_p(n_{\text{sum}})^2$$

and on the other hand

$$\varphi_p(n_{\text{sum}} - 1)^2 > \delta^2 \varrho_p(n_{\text{sum}} - 1)^2$$

Using $\delta^2 \varrho_p(n_{\text{sum}}-1)^2 \geq c_{\varrho_p}^2 \varrho_p(n_{\text{sum}})^2$ and $\varphi_p(n_{\text{sum}})^2 \geq c_{\varphi_p}^2 \varphi_p(n_{\text{sum}}-1)^2$ we get

$$\varphi_p(n_{\text{sum}})^2 > c_{\varphi_p}^2 c_{\varrho_p}^2 \delta^2 \varrho_p(n_{\text{sum}})^2$$

This means in particular that we are in the situation of the last theorem and hence the proposition holds for the first equation. The second can be shown using the same arguments and the equivalence of the sum and the integral criterion.

Lemma 4.6 (Intersection criterion ⇒ Norm criterion(p)) *Let p be valid. Assume for $m \in \mathbb{N}$*

$$c_{\min}^2 \sigma(m)^2 \gamma(m)^2 \leq \delta^2 \varepsilon(m)^2 \leq c_{\max}^2 \sigma(m)^2 \gamma(m)^2$$

Then it holds for n_{norm}

$$\mathbb{E}\|A^p \left(x - x_m^\delta\right)\|^2 \leq \max\{c_{\max}^2, c_{\min}^{-2}\} \mathbb{E}\|A^p \left(x - x_{n_{\text{norm}}}^\delta\right)\|^2$$

Proof
Again, there are two cases to consider:
Case 1: $m \leq n_{\text{norm}}$
For all $m \leq k \leq n_{\text{norm}}$ it holds:

$$\delta^2 \varepsilon(k)^2 \geq \delta^2 \varepsilon(m)^2 \geq c_{\min}^2 \sigma(m)^2 \gamma(m)^2 \geq c_{\min}^2 \sigma(k)^2 \gamma(k)^2$$

Hence we have

$$\sum_{k=m+1}^{n_{\text{norm}}} \delta^2 \sigma(k)^{2p} \sigma(k)^{-2} \varepsilon(k)^2 \geq c_{\min}^2 \sum_{k=m+1}^{n_{\text{norm}}} \sigma(k)^{2p} \sigma(k)^2 \gamma(k)^2$$

and so

$$\begin{aligned}
\mathbb{E}\|A^p \left(x - x_m^\delta\right)\|^2 =& \sum_{k=1}^{m} \sigma(k)^{2p} \delta^2 \sigma(k)^{-2} \varepsilon(k)^2 \\
&+ \sum_{k=m+1}^{n_{\text{norm}}} \sigma(k)^{2p} \gamma(k)^2 + \sum_{k=n_{\text{norm}}+1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \\
\leq& \sum_{k=1}^{m} \sigma(k)^{2p} \delta^2 \sigma(k)^{-2} \varepsilon(k)^2 + c_{\min}^{-2} \sum_{k=m+1}^{n_{\text{norm}}} \delta^2 \sigma(k)^{2p} \sigma(k)^{-2} \varepsilon(k)^2 \\
&+ \sum_{k=n_{\text{norm}}+1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \\
\leq& c_{\min}^{-2} \mathbb{E}\|A^p \left(x - x_{n_{\text{norm}}}^\delta\right)\|^2
\end{aligned}$$

Case 2: $m \geq n_{\text{norm}}$
For all $m \geq k \geq n_{\text{norm}}$ it holds:

$$\delta^2 \varepsilon(k)^2 \leq \delta^2 \varepsilon(m)^2 \leq c_{\max}^2 \sigma(m)^2 \gamma(m)^2 \leq c_{\max}^2 \sigma(k)^2 \gamma(k)^2$$

Hence we have

$$\sum_{k=m+1}^{n_{\text{norm}}} \delta^2 \sigma(k)^{2p} \sigma(k)^{-2} \varepsilon(k)^2 \leq c_{\max}^2 \sum_{k=m+1}^{n_{\text{norm}}} \sigma(k)^{2p} \sigma(k)^2 \gamma(k)^2$$

and so

$$\begin{aligned}
\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 =& \sum_{k=1}^{m} \sigma(k)^{2p} \delta^2 \sigma(k)^{-2} \varepsilon(k)^2 + \sum_{k=m+1}^{n_{\text{norm}}} \sigma(k)^{2p} \delta^2 \sigma(k)^{-2} \varepsilon(k)^2 \\
&+ \sum_{k=n_{\text{norm}}+1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \\
\leq& \sum_{k=1}^{m} \sigma(k)^{2p} \delta^2 \sigma(k)^{-2} \varepsilon(k)^2 \\
&+ c_{\max}^2 \sum_{k=m+1}^{n_{\text{norm}}} \sigma(k)^{2p} \gamma(k)^2 + \sum_{k=n_{\text{norm}}+1}^{\infty} \sigma(k)^{2p} \gamma(k)^2 \\
\leq& c_{\max}^2 \mathbb{E}\|A^p\left(x - x_{n_{\text{norm}}}^\delta\right)\|^2
\end{aligned}$$

Hence the inequality holds.

Lemma 4.7 (Intersection criterion ⇒ Sum criterion(p)) *Let p be valid. Assume for $m \in \mathbb{N}$*

$$c_{\min}^2 \sigma(m)^2 \gamma(m)^2 \leq \delta^2 \varepsilon(m)^2 \leq c_{\max}^2 \sigma(m)^2 \gamma(m)^2$$

Then it holds

$$c_{\text{low}}^2 \varphi_p(m)^2 \leq \delta^2 \varrho_p(m)^2 \leq c_{\text{up}}^2 \varphi_p(m)^2$$

Proof
The same line of arguments as in the last theorem works in this case.

4.6. Results with behavior assumptions

Now we will use the equal behavior assumptions to show step by step

Theorem 4.8 *Assume that p is valid. Assume furthermore that for p we have the same behavior condition. Then the following diagram holds*

$$\begin{array}{ccccc}
\textit{Norm criterion}(p) & \Leftrightarrow & \textit{Sum criterion}(p) & \Leftrightarrow & \textit{Integral criterion}(p) \\
 & & \Updownarrow & & \\
 & & \textit{Intersection criterion} & &
\end{array}$$

Due to Theorem 4.2 we can restrict our attention to Norm criterion(p) ⇒ Sum criterion(p) and Integral criterion(p) ⇒ Intersection criterion. As far as we can see the only way to prove these two results is making use of the equal behavior conditions which will insure that the points described by the different optimality criteria and the real intersection point are not too far away from each other.

Lemma 4.9 (Polynomial behavior) *Assume that we have functions* $f : \mathbb{R}^+ \to \mathbb{R}^+$ *decreasing and* $g : \mathbb{R}^+ \to \mathbb{R}^+$ *increasing. Let* $f(1) = g(1) = 1$ *and for a variable* $x \geq 1$ *we know that it holds* $g(x) \leq c_{\text{other}}$*.*

Assume furthermore that f *and* g *exhibit lower polynomial behavior, i.e. there are positive constants* $c_{\min}$*,* $c_{\max}$ *such that for all* $y > 1$

$$-c_{\max} \ln y \leq \ln f(y)$$

and

$$c_{\min} \ln y \leq \ln g(y)$$

Then it holds that there is a positive constant c *just depending on* c_{other}*,* $c_{\min}$ *and* $c_{\max}$ *such that*

$$g(x)/f(x) \leq c$$

Proof

We have

$$c_{\min} \ln x \leq \ln g(x) \leq \ln c_{\text{other}}$$

and hence

$$\ln x \leq (\ln c_{\text{other}}) / c_{\min}$$

Therefore we have

$$\ln f(x) \geq -(c_{\max} \ln c_{\text{other}}) / c_{\min}$$

and so

$$f(x) \geq \exp(-(c_{\max} \ln c_{\text{other}}) / c_{\min})$$

which yields the proposition.

Lemma 4.10 (Exponential behavior) *Assume that we have functions* $f : \mathbb{R} \to \mathbb{R}^+$ *decreasing and* $g : \mathbb{R} \to \mathbb{R}^+$ *increasing. Let* $f(0) = g(0) = 1$ *and for a variable* $x \geq 0$ *we know that it holds* $g(x) \leq c_{\text{other}}$*.*

Assume furthermore that f *and* g *exhibit lower exponential behavior, i.e. there are positive constants* $c_{\min}$*,* $c_{\max}$ *such that for all* $y > 1$

$$-c_{\max} y \leq \ln f(y)$$

and

$$c_{\min} y \leq \ln g(y)$$

Then it holds that there is a positive constant c *just depending on* c_{other}*,* $c_{\min}$ *and* $c_{\max}$ *such that*

$$g(x)/f(x) \leq c$$

Proof

Exactly the same as last proof when replacing $\ln x$ by x.

Theorem 4.11 *Assume that we have functions $f : \mathbb{R}^+ \to \mathbb{R}^+$ decreasing and $g : \mathbb{R}^+ \to \mathbb{R}^+$ increasing intersecting at a point z, i.e. $f(z) = g(z)$. Assume that for a variable $x \geq z$ we know that it holds $g(x) \leq c_{\text{other}} g(z)$.*

Assume furthermore that f and g fulfill the lower equal behavior condition.

Then it holds that there is a positive constant c just depending on the behavior of f and g but not on the intersection point z or the value of $g(z)$ such that

$$g(x)/f(x) \leq c$$

Proof

This is a trivial consequence of the two preceding lemmas doing a rescaleing to z and $g(z)$.

Remark

By mirroring and exchanging f and g we also get the same results for the values smaller than the intersection point.

Theorem 4.12 (Norm criterion(p) $\Rightarrow$ Sum criterion(p)) *Assume that p is valid and that φ_p and ϱ_p fulfill the lower equal behavior condition.*

Assume that for $m \in \mathbb{N}$

$$\mathbb{E}\|A^p\left(x - x_m^\delta\right)\|^2 \leq c_{\text{other}}^2 \mathbb{E}\|A^p\left(x - x_{n_{\text{sum}}}^\delta\right)\|^2$$

Then it holds

$$c_{\text{low}}^2 \varphi_p(m)^2 \leq \delta^2 \varrho_p(m)^2 \leq c_{\text{up}}^2 \varphi_p(m)^2$$

Proof

Like beforehand, two cases:

Case 1: $m < n_{\text{sum}}$

It holds trivially

$$\delta^2 \varrho_p(m)^2 \leq \varphi_p(m)^2$$

The other side of the inequality gets

$$\varphi_p(m)^2 \leq c_{\text{other}}^2 \left(\varphi_p(n_{\text{sum}})^2 + \delta^2 \varrho_p(n_{\text{sum}})^2\right) \leq c_{\text{other}}^2 \left(1 + c_{\varphi_p}^{-2} c_{\varrho_p}^{-2}\right) \varphi_p(n_{\text{sum}})^2$$

Using Theorem 4.11 we get our result.

Case 2: $m \geq n_{\text{sum}}$

Exactly the same as the last case.

Theorem 4.13 (Integral criterion(p) $\Rightarrow$ Intersection criterion) *Assume that p is valid and that $\sigma(x)^{2p}\sigma(x)^2\gamma(x)^2$ and $\sigma(x)^{2p}\delta^2\varepsilon(x)^2$ fulfill the equal behavior condition, that they intersect and that we know:*

$$c_{\min}^2 \int_m^\infty \sigma(x)^{2p}\sigma(x)^2\gamma(x)^2 dx \leq \int_1^m \sigma(x)^{2p}\delta^2\varepsilon(x)^2 dx \leq c_{\max}^2 \int_m^\infty \sigma(x)^{2p}\sigma(x)^2\gamma(x)^2 dx$$

Then it also holds

$$c_{\text{low}}^2 \sigma(m)^2\gamma(m)^2 \leq \delta^2\varepsilon(m)^2 \leq c_{\text{up}}^2 \sigma(m)^2\gamma(m)^2$$

Proof

As the proposition is equivalent to

$$c_{\text{low}}^2 \sigma(x)^{2p}\sigma(m)^2\gamma(m)^2 \le \sigma(x)^{2p}\delta^2\varepsilon(m)^2 \le c_{\text{up}}^2\sigma(x)^{2p}\sigma(m)^2\gamma(m)^2$$

we can rewrite to the following problem:

$$c_{\min}^2 \int_m^\infty f(x)dx \le \int_1^m g(x)dx \le c_{\max}^2 \int_m^\infty f(x)dx$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is decreasing, $g : \mathbb{R}^+ \to \mathbb{R}^+$ increasing, f and g intersecting and both exhibiting the same behavior (either polynomial or exponential). We need to show that there exist c_{low} and c_{up} such that

$$c_{\text{low}}^2 f(m) \le g(m) \le c_{\text{up}}^2 f(m)$$

Obviously we have that also the primitives $F(\cdot) = \int_\cdot^\infty f(x)dx$ and $G(\cdot) = \int_1^\cdot g(x)dx$ of f and g respectively exhibit the same behavior as f and g. Furthermore F is decreasing, G increasing.

Now there is a unique point r for which

$$F(r) + G(r) = \int_r^\infty f(x)dx + \int_1^r g(x)dx$$

is minimal due to f and g intersecting and it obviously holds $f(r) = g(r)$. Due to the lower equal behavior condition we can argue along the lines of the last proof that there is a positive constant such that

$$c^{-1}F(r) \le G(r) \le cF(r)$$

We need to distinguish two cases:

Case 1: $r < m$

Obviously $f(m) \le g(m)$. Furthermore it holds

$$G(m) \le c_{\max}^2 F(m) \le c_{\max}^2 F(r) \le c_{\max}^2 cG(r)$$

Using that G, f and g exhibit the same behavior we get the assertion as in the supporting lemmas of Theorem 4.11.

Case 2: $r > m$

The same argumentation as for the last case, just f and g exchanged.

4.7. Considerations

Definition 4.14 (Qualification) *A regularization family is called* qualified *for p if there is a valid concentration l for which we have*

$$c_{\min}^2 \varrho_p(l(n))^2 \le \sum_{k=1}^\infty \sigma(k)^{2p}\delta^2\varepsilon(k)^2 g_n(\sigma(k))^2 \le c_{\max}^2 \varrho_p(l(n))^2$$

$$c_{\text{low}}^2 \varphi_p(l(n))^2 \le \sum_{k=1}^\infty \sigma(k)^{2p}\gamma(k)^2 \left(1 - \sigma(k)g_n(\sigma(k))\right)^2 \le c_{\text{up}}^2 \varphi_p(l(n))^2$$

Remark

The spectral cut-off scheme is trivially qualified.

Remark

Looking through the proofs of the last section we can remark two things. First of all they will work even if we would use concentrations which are valid and we get the same *rates. Furthermore if a regularization family is qualified for p it exhibits the same rate behavior as the associated concentrated spectral cut-off scheme.*

Therefore we can put the results together and get

Theorem 4.14 *Assume that p and q are valid. Assume furthermore that for p we have the same behavior condition and the regularization family $\{g_n\}_{n\in\mathbb{N}}$ is qualified for both p and q.*

Then the following diagram holds

$$\begin{array}{ccccc} \textit{Norm criterion}(p) & \Leftrightarrow & \textit{Sum criterion}(p) & \Leftrightarrow & \textit{Integral criterion}(p) \\ & & \Updownarrow & & \\ & & \textit{Intersection criterion} & & \\ & & \Downarrow & & \\ \textit{Norm criterion}(q) & \Leftarrow & \textit{Sum criterion}(q) & \Leftrightarrow & \textit{Integral criterion}(q) \end{array}$$

This has very interesting and important consequences. It tells that it will be sufficient to regularize in one space along the Hilbert scale which fulfills the same behavior condition and that we get a good regularization for *all* other spaces along the Hilbert scale for free.

This means in particular that if the noise is bad enough smoothing in the data space and then inverting the operator without any kind of regularization is a valid regularization method. In the other direction we can conclude that a good regularized solution in the original space also yields a rate optimal approximation to the data.

5. Stopping Strategies

Now we will analyze a stopping strategy which neither requires a tuning parameter nor the knowledge of δ.

Definition 5.1 *Assume that l is a valid exponential concentration and p is valid.*

Assume furthermore that $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ is function which fulfills that

$$\varphi_p(l(n))^2 \chi(l(n))^2$$

is exponentially decreasing and

$$\varrho_p(l(n))^2 \chi(l(n))^2$$

is exponentially increasing.

The parameter $n_{\mathrm{diff},\chi}$ is called minimum difference regularization parameter if

$$k_{\mathrm{diff},\chi} = \operatorname*{armin}_{k\in\mathbb{N}} \left\{ \mathbb{E}\|A^p \left(x^\delta_{l(k+1)} - x^\delta_{l(k)}\right)\|^2 \chi(l(k))^2 \right\} \tag{35}$$

and

$$n_{\mathrm{diff},\chi} = l(k_{\mathrm{diff},\chi})$$

Theorem 5.1 *There is a positive constant c_{other} such that*

$$\mathbb{E}\|A^p \left(x^\delta_{l(k_{\mathrm{diff},\chi})} - x\right)\|^2 \le c^2_{\mathrm{other}} \mathbb{E}\|A^p \left(x^\delta_{n_{\mathrm{isect}}} - x\right)\|^2$$

Proof
It holds

$$\mathbb{E}\|A^p \left(x^\delta_{l(k+1)} - x^\delta_{l(k)}\right)\|^2 = \varphi_p(l(k))^2 - \varphi_p(l(k+1))^2 + \delta^2 \varrho_p(l(k+1))^2 - \delta^2 \varrho_p(l(k))^2$$

and hence using the exponential behavior of both φ_p and ϱ_p there are positive constants $c_{\min}$ and $c_{\max}$ such that

$$c^2_{\min}\left(\varphi_p(l(k))^2 + \delta^2\varrho_p(l(k))^2\right) \le \mathbb{E}\|A^p \left(x^\delta_{l(k+1)} - x^\delta_{l(k)}\right)\|^2 \le c^2_{\max}\left(\varphi_p(l(k))^2 + \delta^2\varrho_p(l(k))^2\right)$$

Using the same line of arguments as in the Theorem 4.11 and its supporting lemmas we get

$$c^2_{\mathrm{low}}\varphi_p(l(k))^2 \le \delta^2\varrho_p(l(k))^2 \le c^2_{\mathrm{up}}\varphi_p(l(k))^2$$

Hence we get the above result. The multiplication with $\chi(l(n))^2$ does not change the proof because it does not change the principal behavior of φ_p and ϱ_p.

The above result just holds in expectation, i.e. in particular just with the expected regularization parameter. For real world situations we can just formulate the following conjecture right now:

Conjecture 5.1 *Define according to the last definition the regularization parameter*

$$\widetilde{k_{\mathrm{diff},\chi}} = \operatorname*{armin}_{k\in\mathbb{N}} \left\{ \|A^p \left(x^\delta_{l(k+1)} - x^\delta_{l(k)}\right)\|^2 \chi(l(n))^2 \right\}$$

$$\widetilde{n_{\mathrm{diff},\chi}} = l(\widetilde{k_{\mathrm{diff},\chi}})$$

Then it holds analogously to the last theorem

$$\mathbb{E}\|A^p \left(x^\delta_{l(\widetilde{k_{\mathrm{diff},\chi}})} - x\right)\|^2 \le c^2_{\mathrm{other}} \mathbb{E}\|A^p \left(x^\delta_{n_{\mathrm{isect}}} - x\right)\|^2$$

For the choice of $\chi(l(n))^2$ there seem to be basically two good choices:

- $\chi = 1$, $p = 0$
  In this case the parameter choice procedure is very similar two the quasi-optimality principle. The only difference is that it is formulated for concentrated Spectral Cut-Off right now and not for Tikhonov regularization [1, 12]. In experiments this choice lacks for a wrong choice of the concentration statistical stability, i.e. at least some assumptions on the concentration have to be made in order to fulfill Conjecture 5.1.
- $\chi(l(n))^2 = \varrho(l(n))^{-1}$, $p = 0$
  In this case we are at some point in an intermediate situation between the Lepskij-type balancing principle as proposed in [11] and the quasi-optimality.
  This choice has in comparison to the quasi-optimality the advantage that we can guarantee for both the increasing and decreasing part a minimum speed which results in practice in slightly worse constants but much higher stochastical stability. Methods like the one presented in [13] remind remotely to this kind of parameter choice rule.

6. Numerics

In order to test the described method we have set up the following example problem:

- A is a diagonal operator with 200 elements and Eigenvalue decay k^{-3}.
- Solution x is a Gaussian random vector with Fourier coefficients decaying as $k^{-2.5}$ along the Eigen spaces.
- The noise was chosen as a Gaussian random vector with error level $\delta = 10^{-9}$
- We generate 50 different solutions and for each of them we also generate 10 different noisy input vectors. So we were treating 500 different problems for each case.

As regularization methods we used

- Exponentially concentrated spectral-cut-off as used in the proofs; figure 1.
- Tikhonov regularization; figure 2.

As parameter choice methods we used

- A tuned version of the balancing principle with $\kappa = 0.5$, see [14]; grey bars.
- The intermediate stopping procedure between balancing principle and quasi-optimality, i.e. $\chi(l(n))^2 = \varrho(l(n))^{-1}$; black bars.
- Quasi-optimality; white bars.

As we “knew” the real solution in our experiment we could also generate the optimal possible regularized solution with respect to the chosen regularization scheme. This was used for comparison purposes.

In the left half of the diagrams we displayed a bar plot showing the chosen regularization parameter minus the optimal possible one. This means in particular that numbers smaller than 0 mean oversmoothing. In the right half of the diagrams we displayed a barplot showing the ratio between chosen solution and the norm optimal one. This means in particular that the ratio cannot get smaller than 1.

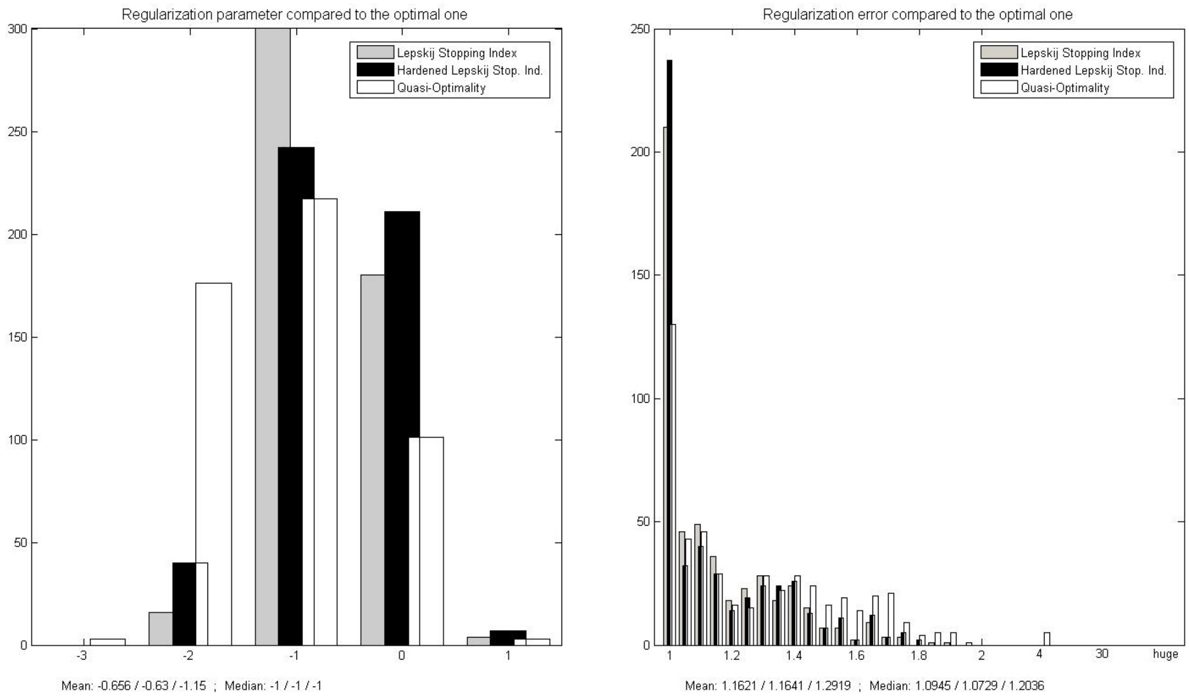

Figure 1. Experiment for expoentially concentrated cut-off

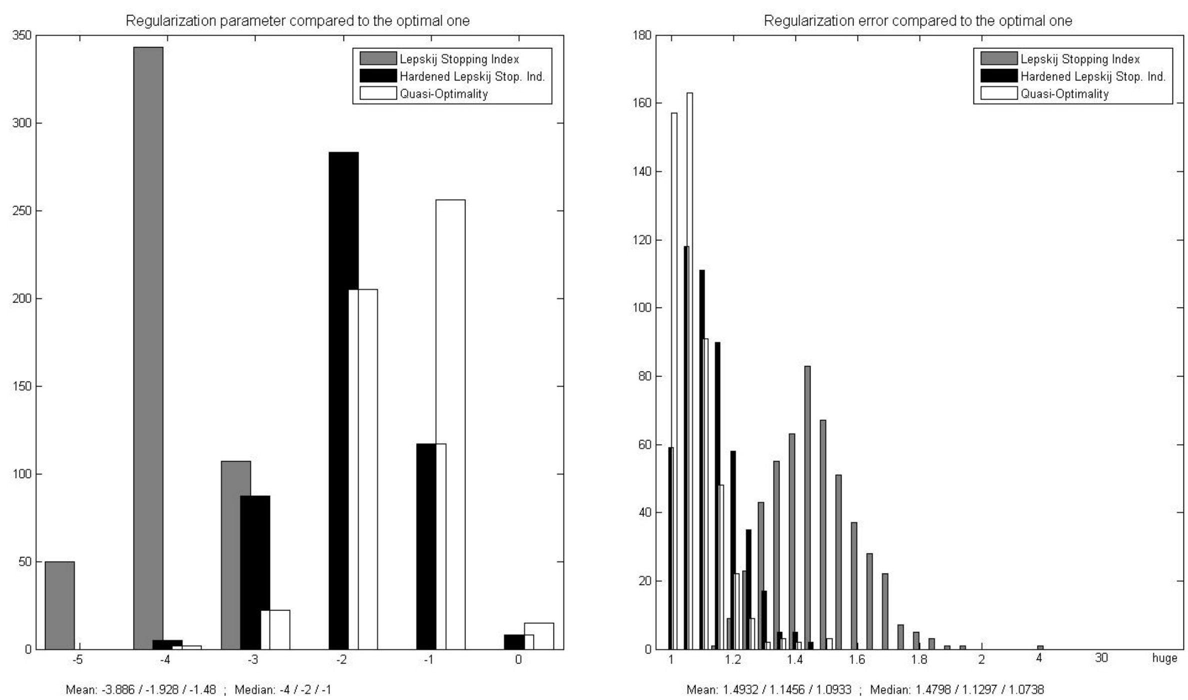

Figure 2. Experiment for Tikhonov

We observe that the proposed methods perform statistically seen very stable and are in average better than the balancing principle. In fact we performed this type of experiment with very many different setups for the operator behavior, smoothness of the solution, type (Gaussian or heavy tails) of the noise and color of the noise.

In all situations covered by the proof the quasi-optimality performed very well, in the ones which are not covered it sometimes worked very well and sometimes not at all. In comparison the mixture between balancing and quasi-optimality performed very well in all situations we tested, in particular it coped very well with heavy tails in the noise and situations where the equal behavior condition was violated.

7. Conclusion

It is clearly questionable if it is sensible to impose a prior condition on the quantity we search for. There are not very many situations where one can definitely say that such or an equivalent condition holds. However, as we have more or less complete freedom for the function describing the smoothness of x and the fact that we just use it formally but do not require the knowledge in methods like the one presented this approach has merely the property that it qualifies the situations for which one can construct counterexamples along the Bakushinskii veto [2] as very rare and hence possible to ignore in reality.

In practice quite a lot of methods which lack of thorough foundation are in use, e.g. quasi-optimality. We hope that the proposed approach can help to explain why these parameter choice heuristics work in certain situations and not in others. Furthermore one can use these tools to design new heuristics which can be proven to work at least in a very special case.

Acknowledgments

I would like to thank N. Bissantz (Univ. of Göttingen), B. Hofmann (Univ. of Chemnitz), B. Mair (Univ. of Florida), P. Mathé (WIAS Berlin), S. Pereverzev (RICAM Linz), M. Pricop (Univ. of Göttingen), M. Reiss (Univ. of Heidelberg) and L. Rosasco (Univ. of Genova) for very enlightening discussions on this and related topics.

Furthermore I would like to acknowledge gratefully the financial support of the Graduiertenkolleg 1023, University of Göttingen.

References

[1] Engl H, Hanke M, Neubauer A. Regularization of Inverse Problems. Dordrecht, Boston, London: Kluwer Academic Publisher; 1996.

[2] Bakushinsky A. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. Comput Maths Math Phys 1984;24(4):181–182.

[3] Hansen P. Rank-deficient and discrete ill-posed problems. SIAM Monographs on Mathematical Modeling and Computation. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); 1998. Numerical aspects of linear inversion.

[4] Wahba G. Practical approximate solutions to linear operator equations when the data are noisy. SIAM Journal on Numerical Analysis 1977;14(4):651–667.

[5] Hofmann B. Regularization of Applied Inverse and Ill-Posed Problems. Leipzig: Teubner; 1986.

[6] Yagola A, Leonov A, Titarenko V. Data Errors and an Error Estimation for Ill-Posed Problems. Inverse Problems in Engineering 2002;10(2):117 – 129.

[7] Kaipio J, Somersalo E. Statistical and computational inverse problems. vol. 160 of Applied Mathematical Sciences. New York: Springer-Verlag; 2005.

[8] Belitser E, Ghosal S. Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. The Annals of Statistics 2003;31(2):536–559.

[9] Johnstone I, Silverman B. Empirical Bayes selection of wavelet thresholds. The Annals of Statistics 2005;33(4):1700–1752.

[10] Zhao L. Bayesian aspects of some nonparametric problems. The Annals of Statistics 2000;28(2):532–552.

[11] Mathé P, Pereverzev S. Geometry of Linear Ill-Posed Problems in Variable Hilbert Spaces. Inverse Problems 2003;19(3):789–803.

[12] Tikhonov A, Arsenin V. Solutions of Ill-Posed Problems. New York: Wiley; 1977.

[13] Raus T. About regularization parameter choice in case of approximately given error bounds of data. Tartu Ül Toimetised 1992;(937):77–89.

[14] Bauer F, Pereverzev S. Regularization without Preliminary Knowledge of Smoothness and Error Behavior. European Journal of Applied Mathematics 2005;16(3):303–317.

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:

2006-01	R. Schaback	Limit Problems for Interpolation by Analytic Radial Basis Function
2006-02	N. Bissantz, T. Hohage, A. Munk, F. Ruymgaart	Convergence rates of general regularization methods for statistical inverse problems and applications
2006-03	J. Brimberg, H. Juel, A. Schöbel	Locating a circle on a sphere
2006-04	J. Brimberg, H. Juel, A. Schöbel	Locating a circle on the plane using the minimax criterion
2006-05	L. Ling, R. Opfer, R. Schaback	Results on Meshless Collocation Techniques
2006-06	G. Lube, T. Knopp et.al.	Domain decomposition methods for indoor air flow simulation
2006-07	T. Apel, T. Knopp, G. Lube	Stabilized finite element methods with anisotropic mesh refinement for the Oseen problem
2006-08	R. Schaback	Recovery of Functions From Weak Data Using Unsymmetric Meshless Kernel-Based Methods
2006-09	H.W. Hannacher, S. Horn, A. Schöbel	Stop Location Design in Public Transportation Networks: Covering and Accessibility Objective
2006-10	Q.T. Le Gia, F.J. Narcowich, J.D. Ward, H. Wendland	Continuous and Discrete Least-Squares Approximation by Radial Basis Functions on Spheres
2006-11	R. Schaback	Unsymmetric Meshless Methods for Operator Equations
2006-12	P. Giesl, H. Wendland	Meshless Collocation: Error Estimates with Application to Dynamical Systems
2006-13	H. Wendland	On the stability of meshless symmetric collocation for boundary value problems
2006-14	G. Lube	Stabilized FEM for incompressible flow. Critical review and new trends

2006-15	J. Puerto, A. Schöbel, S. Schwarze	A Class of Infinite Potential Games
2006-16	S. Hein, T. Hohage, W. Koch, J. Schöberl	Acoustic Resonances in High Lift Configuration
2006-17	T. Hohage, M.L. Rapún, F.J. Sayas	Detecting corrosion using thermal measurements
2006-18	H. Wendland	Spatial coupling in aeroelasticity by meshless kernel-based methods
2006-19	F. Bauer, O. Ivanyshyn	Optimal Regularization with Two Interdependent Regularization Parameters
2006-20	F. Bauer	Some Considerations Concerning Regularization and Parameter Choice Algorithms