Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Degree of ill–posedness of Statistical inverse problems

Peter Mathé

submitted: 19th August 2004

Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstrasse 39
10117 Berlin, Germany
E-Mail: mathe@wias-berlin.de

No. 959
Berlin 2004

2000 *Mathematics Subject Classification.* 65J20, 62G20.

Key words and phrases. Degree of ill–posedness, statistical inverse problem, variable Hilbert scale.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. We introduce the notion of the degree of ill–posedness of linear operators in operator equations between Hilbert spaces. For specific assumptions on the noise this quantity can be computed explicitely. Next it is shown that the degree of ill–posedness as introduced explains the loss of accuracy when solving inverse problems in Hilbert spaces for a variety of instances.

1. INTRODUCTION

We study the solution of operator equations $Ax = y$ under presence of noise, which means we are given an operator $A\colon X \to Y$, acting between Hilbert spaces and data

$$y_\delta = Ax + \delta\xi, \tag{1.1}$$

where δ represents the noise level and $\delta\xi$ the noise inherent in the data y_δ. Our goal is to study the lack of accuracy when reconstructing the unknown solution x based on data y_δ.

Ideally, if the noise is bounded $\|\xi\| \leq 1$ and the operator is boundedly invertible, then we would be able to reconstruct the unknown solution x by $\hat{x} := A^{-1}y_\delta$, and this would result in an error bound

$$\|x - \hat{x}\| = \|\delta A^{-1}\xi\| \leq \delta \|A^{-1}\colon Y \to X\|, \quad \text{i.e., of the order}\delta.$$

But in general, either when the noise is unbounded, in particular statistical, and/or the operator does not have a bounded inverse, then there will be a lack in accuracy when reconstructing x based on data y_δ. The question arises, whether one can describe this loss as dependent on the operator A and/or the type of noise ξ. This leads to the notion of the *degree of ill–posedness* of the operator in the presence of noise. The loss of accuracy will then be seen to depend on this degree of ill–posedness and on some *a priori smoothness* assumption on the exact solution.

In this study we assume that the underlying operator A is compact and injective and the noise is either *bounded deterministic* or centered *Gaussian white noise*, such that for any functional $a \in Y$ it holds

$$\mathbf{E}\,|\langle \xi, a\rangle|^2 = \|a\|^2.$$

In this framework, an estimate for approximating x, based on observations y_δ is given as an arbitrary (measurable) mapping, say $\hat{x}$ into X. Its error at any problem instance $x \in X$ is then given by

$$e^{\det}(x, \hat{x}, \delta) := \sup_{\|\xi\| \leq 1} \|x - \hat{x}(y_\delta)\|$$

in the deterministic case or

$$e^{stat}(x, \hat{x}, \delta) := \left(\mathbf{E}\|x - \hat{x}\|^2\right)^{1/2}$$

for Gausssian white noise. The worst-case error over a class F of problem instances is given as

$$e^{\bullet}(F, \hat{x}, \delta) := \sup_{x \in F} e^{\bullet}(x, \hat{x}, \delta),$$

with $\bullet \in \{det, stat\}$. The best possible order of accuracy it defined my minimization over all estimators, i.e.,

$$e^{\bullet}(F, \delta) := \inf_{\hat{x}} e^{\bullet}(F, \hat{x}, \delta).$$

In the present context we are interested in the asymptotic behavior of $e^{\bullet}(F, \delta)$ as $\delta \to 0$, when the class F of problem instances is given by general source conditions in the form

$$A_{\varphi}(R) := \{x \in X, \quad x = \varphi(A^*A)v, \ \|v\| \leq R\}, \tag{1.2}$$

where φ is a function on the spectrum of the operator A^*A. Further restrictions will be imposed later on.

Previous approaches. There were various attempts to formalize the notion of *degree of ill–posedness.* The first appearance of this notion probably dates back to

G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., **14** (1977), pp. 651–667.

Later this question was discussed in some talk by

M. Nussbaum, *Deterministische und stochastische Modellierung von Inversen Problemen, WIAS 1994.* This resulted in a joint study

M. Nussbaum and Sergei Pereverzev. *The degree of ill–posedness in stochastic and deterministic noise models*, Preprint 509, WIAS, Berlin, 1999.

In a similar way this was explored in the context of Hilbert scales in

P. Mathé and S. V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal. **38**(2001), pp. 1999–2021.

The distinction between the *smoothness of the solution* and the intrinsic *lack of invertibility of the operator* became transparent. Within the context of classical Hilbert scales, say H^s, $s \in \mathbb{R}$, the latter is defined as

$$\alpha := \sup \left\{s, \quad \mathbf{E}\|A^{-1}\xi\|_s^2 < \infty\right\}.$$

For operators acting along such scales, i.e. for some $a > 0$ it holds $\|Ax\|_{\lambda+a} \asymp \|x\|_{\lambda}$, this quantity was proven to be $\alpha = -(a + 1/2)$, which can be interpreted as – *step size plus one half*, the latter being the contribution due to white noise.

For variable Hilbert scales $\{X_{\psi}\}$, as introduced below, and deterministic noise this was extended to (with α and ψ being functions)

$$\alpha := \inf \left\{\psi, \quad \|A^{-1} \colon Y \to X_{\psi}\| \leq 1\right\},$$

as elaborated in [5]. It could be seen, that this minimization problem has a solution given by $\alpha(t) = 1/\sqrt{t}$. Below we shall also review this result.

In the present context of white noise the natural definition should read

$$\alpha := \inf \left\{\psi, \quad \mathbf{E}\|A^{-1}\xi\|_{\psi}^2 \leq 1\right\}.$$

The following questions arise naturally.

– Can this definition be given a precise meaning?

– Does this relate to the lack of accuracy for operator equations under statistical noise?

In Section 2 we shall provide a framework in which the above definition becomes meaningful. Moreover, we shall calculate α, as a function of the underlying operator A under both assumptions on the noise.

Then in Section 3 we shall see that for a variety of cases the degree of ill–posedness is inherent in the loss of accuracy as this has been shown for classical Hilbert scales in [3].

We close our investigations with some representative examples.

2. The degree of ill–posedness under general source conditions

In contrast to the usual approach, where smoothness is given by (finite) differentiability properties, we express smoothness in terms of the compact injective operator A from (1.1). This approach is related to *variable Hilbert scales*, as introduced by Hegland [2]. We briefly recall the basic concept.

2.1. **Variable Hilbert scales.**

We shall assume, that the scale is generated by A^*A. The singular numbers of A^*A are denoted by $(s_k)_{k=1}^{\infty}$, arranged in non-increasing order. In particular $a := s_1 = \|A^*A\|$.

Each function $\varphi : (0, a] \to [0, \infty)$ can be assigned a pre-Hilbert space as follows. First let

$$F := \left\{ x,\ x = \sum_{j=1}^{n} \langle x, u_j \rangle u_j, \quad n < \infty \right\},$$

be the linear space of finite expansions in $u_1, u_2, \dots$, the eigenbasis of A^*A. Given φ we can endow F with scalar product

$$\langle x, y \rangle_{\varphi} := \sum_{j=1}^{\infty} \frac{\langle x, u_j \rangle \langle y, u_j \rangle}{\varphi^2(s_j)}, \quad x, y \in F.$$

The completion of F in this scalar product is denoted by X_{φ}, such that $A_{\varphi}(R)$ from (1.2) is the ball of radius R in X_{φ}. It is easy to verify, that $A_{\varphi}(R) \subset X$ is relatively compact only, if $\lim_{t \to 0} \varphi(t) = 0$, such that we assume φ is non-decreasing and $\varphi(0+) = 0$. We shall call such functions *index (functions)* throughout.

Definition 1. *Let $\mathcal{I}(0, a]$ denote the class of all functions $\varphi\colon (0, a] \to [0, \infty)$, which are non-decreasing and $\lim_{t \searrow 0} \varphi(t) = 0$. We agree to call functions from $\mathcal{I}(0, a]$* index functions.

The collection $\{X_{\varphi},\ \varphi \in \mathcal{I}(0, a]\}$ is called variable Hilbert scale. A detailed account of such variable Hilbert scales can be found in [2, 10] and [5, 4]. We shall however not use much of the theory developed towards numerical analysis under general source conditions.

2.2. **Bounded deterministic noise.** Here we review one result from [5]. This is for completeness, but also to see how calculations are carried out within the present framework. In [5] the function

$$\alpha := \inf \left\{ \psi, \quad \|A^{-1} \colon Y \to X_\psi\| \le 1 \right\}$$

was defined to be the degree of ill–posedness. But the operator A is not boundedly invertible from Y into X, therefore X_ψ cannot embed into X, thus ψ cannot be an index. Therefore we require that ψ is the adjoint of an index function, i.e., $1/\psi \in \mathcal{I}(0,a]$.

Moreover, it is important to notice that the index functions are only determined on the spectrum of A^*A, thus we identify index functions which coincide on the spectrum of the operator A^*A.

Definition 2. *Let*

$$\rho := \sup \left\{ \psi \in \mathcal{I}(0,a], \quad \|A^{-1} \colon Y \to X_{1/\psi}\| \le 1 \right\} \tag{2.3}$$

be the point-wise supremum on s_j, $j = 1, 2, \ldots$. The function $\alpha = 1/\rho$ is called degree of ill–posedness *of equation* (1.1) *under bounded deterministic noise.*

The degree of ill–posedness of the operator A under bounded deterministic noise is calculated next.

Proposition 1. *For a compact injective operator A and bounded noise it holds true that*

$$\rho(t) = \sqrt{t}, \qquad 0 < t \le a. \tag{2.4}$$

Thus the degree of ill–posedness is $t \to 1/\sqrt{t}$.

Proof. Let ψ be any index function satisfying the restriction from (2.3). For an arbitrary $\|y\| \le 1$ we can argue as follows.

$$1 \ge \|A^{-1}y\|^2_{1/\psi} = \|\psi(A^*A)A^{-1}y\|^2 = \|\psi^2(A^*A)(A^*A)^{-1}y\|.$$

Thus $\|\psi^2(A^*A)(A^*A)^{-1}\| \le 1$, which translates to $\psi^2(t)/t \le 1$ and proves $\sqrt{t}$ to be an upper bound. But is is easy to see that $\sqrt{t}$ itself fulfills the restriction, so the upper bound is attained and the proof is complete. □

Remark 1. *Notice that in variable Hilbert scales this degree of ill–posedness is an invariant. Once the scale fits to the operator the degree of ill–posedness no longer depends on the operator. Therefore, the theory of regularization of linear ill–posed problems in variable Hilbert scales has nice geometric properties, see* [5].

2.3. **Gaussian white noise.** Based on Proposition 1 there is good reason to mimic the definition for white noise as above with the same modification as for deterministic noise.

Definition 3. *Let*

$$\rho := \sup \left\{ \psi \in \mathcal{I}(0,a], \quad \mathbf{E}\|A^{-1}\xi\|^2_{1/\psi} \le 1 \right\}. \tag{2.5}$$

The function $\alpha = 1/\rho$ is said to be the degree of ill–posedness *for operator equations* (1.1) *under Gaussian white noise.*

As we shall see next, this can be calculated explicitely. Recall that $s_j,\ j = 1, 2, \dots$ denotes the decreasing sequence of singular numbers of A^*A.

Theorem 1. *The degree of ill–posedness for operator equations* (1.1) *under Gaussian white noise is*

$$\alpha(t) = \left(\sum_{s_k \geq t} 1/s_k\right)^{1/2}. \tag{2.6}$$

Let us denote $\Psi(t) := \left(\sum_{s_k \geq t} 1/s_k\right)^{-1}$. This function is closely related to the harmonic mean of the singular numbers of A^*A. A more balanced variant of the degree of ill–posedness can be given in terms of $t_j,\ j = 1, 2, \dots,$ the sequence of singular numbers of A as $\alpha(t) = \left(\sum_{t_k^2 \geq t} 1/t_k^2\right)^{1/2}$.

Proof. We shall show that $1/\alpha^2(t) = \Psi(t)$. To this end we rewrite for any index ψ

$$\mathbf{E}\|A^{-1}\xi\|^2_{1/\psi} = \operatorname{tr} \psi^2(A^*A)(A^*A)^{-1} = \sum_{j=1}^{\infty} \frac{\psi^2(s_j)}{s_j}.$$

Let us consider the family $\rho_k \in \mathcal{I}(0, a]$ of piece-wise constant functions defined as

$$\rho_k(t) := \begin{cases} \sqrt{\Psi(s_k)}, & t \geq s_k \\ 0, & \text{else.} \end{cases}$$

Plainly $\sum_{j=1}^{\infty} \rho_k^2(s_j)/s_j = 1$. Note also that for any j it holds true that $\max\{\rho_k^2(s_j), k = 1, 2, \dots\} = \Psi(s_j)$, such that ρ from (2.5) obeys

$$\rho^2 \geq \sup\left\{\rho_k^2,\ k = 1, 2, \dots\right\} = \Psi.$$

Now suppose that for some k it holds true that $\rho^2(s_k) > \Psi(s_k)$. By definition, for all $\varepsilon > 0$ we can find ψ_ε with $\psi_\varepsilon^2(s_k) \geq \rho^2(s_k) - \varepsilon$ and $\mathbf{E}\|A^{-1}\xi\|^2_{1/\psi_\varepsilon} \leq 1$. But, because ψ_ε is non-decreasing we deduce

$$1 \geq \sum_{j=1}^{k} \frac{\psi_\varepsilon^2(s_j)}{s_j} \geq \frac{\psi_\varepsilon^2(s_k)}{\Psi(s_k)}.$$

This yields $\Psi(s_k) \geq \psi_\varepsilon^2(s_k) \geq \rho^2(s_k) - \varepsilon$. Letting $\varepsilon \to 0$ allows to complete the proof. $\square$

In Section 4 we shall estimate the degree of ill–posedness for important special cases.

3. Relation to best possible accuracy

Here we shall establish that the degree of ill–posedness as introduced above is inherent in the loss of accuracy when reconstructing the solution x from (1.1) under noisy data y_δ. It will be transparent that this loss is determined by both, the degree of ill–posedness of the operator and the *a priori smoothness* $\varphi \in \mathcal{I}(0, a]$ of the solution $x \in A_\varphi(R)$, see (1.2). Precisely, given the a priori smoothness φ and degree of ill–posedness α let $\Theta := \varphi/\alpha \in \mathcal{I}(0, a]$.

3.1. **Bounded deterministic noise.** Again we review one result for deterministic noise from [5]. An increasing function φ is said to obey a Δ_2–condition, if there is $1 \leq C < \infty$ for which $\varphi(2t) \leq C\varphi(t),\ 0 < t \leq a$.

In the present setup we have $\Theta(t) = \sqrt{t}\varphi(t),\ 0 < t \leq a$. Theorem 1 from [5] asserts the following.

Theorem 2. *If the function* $t \to \varphi^2((\Theta^2)^{-1}(t)$ *is concave, then*

$$e^{det}(A_\varphi(R), \delta) \leq R\varphi(\Theta^{-1}(\delta/R)), \qquad 0 < \delta \leq aR.$$

If, in addition φ *obeys a* Δ_2 *condition then*

$$e^{det}(A_\varphi(R), \delta) \asymp R\varphi(\Theta^{-1}(\delta/R)), \qquad 0 < \delta \leq aR.$$

Remark 2. *The concavity of the above composite function is crucial. However, this is not restrictive, because in many cases this is fulfilled. As was established in* [5] *this holds true if* $\log(\varphi)$ *is concave, thus for all polynomial index functions, but also for index functions of logarithmic type, as* $t \to \log^{-\mu} 1/t$*, at least if* t *is small enough.*

We will *not* discuss topics as there are discretization, adaptation.... These are interesting and important issues, but here our attention is towards best possible accuracy, which is ideal but serves as a benchmark for more practical cases.

3.2. **Gaussian white noise.** Our analysis for Gaussian white noise will be based on Pinsker's seminal paper [9], where a general result was established, which allows to draw conclusions for the present setup of general source conditions. We mention that for some special cases a similar analysis was carried out in [6].

Upper bound: Regularization. It is well known that approximate solutions to (1.1) require regularization. If the best possible accuracy shall be achieved, then the regularization must be capable to take the a priori smoothness into account. The systematic study of this issue goes back to [11]. In the context of general source conditions this was extended in [5]. However, the easiest way to retain the best possible accuracy is *spectral cut–off* (hard threshold). This is not always feasible but for our purpose sufficient. We will not formally introduce the machinery of regularization methods, but note the following. The original equation (1.1) can be rewritten, using the singular value decomposition as

$$y_\delta = \sum_{j=1}^{\infty} \sqrt{s_j}\langle x, u_j\rangle v_j + \delta\xi, \tag{3.7}$$

which leads to an infinite system

$$\langle y_\delta, v_j\rangle = \sqrt{s_j}\langle x, u_j\rangle + \delta\xi_j, \quad j = 1, 2, \ldots$$

where the ξ_j are i.i.d standard normal. As estimate $\hat{x}$ of x based on observations $y_{\delta,1}, y_{\delta,2}, \ldots$ we shall use

$$\hat{x}_n(y_\delta) := \sum_{j=1}^{n} \frac{1}{\sqrt{s_j}}\langle y_\delta, v_j\rangle u_j.$$

where the choice of $n = n(\delta)$ plays the role of a regularizing parameter. The upper bound for such type of estimators is provided in

Proposition 2. *Let α be the degree of ill–posedness of the operator equation* (1.1). *For an index function φ let*

$$\Theta(t) := \frac{\varphi(t)}{\alpha(t)}, \qquad 0 < t \le a. \tag{3.8}$$

Given $\delta > 0$ let

$$n_* := \max \{n, \quad \Theta(t_n) \ge \delta/R\}. \tag{3.9}$$

Uniformly for $x \in A_\varphi(R)$ the estimator $\hat{x}_{n_*}(y_\delta)$ provides

$$\left(\mathbf{E}\|x - \hat{x}_{n_*}\|^2\right)^{1/2} \le \sqrt{2} R \varphi(\Theta^{-1}(\delta/R)). \tag{3.10}$$

Note that Θ has jumps, and we define the inverse by $\Theta^{-1}(s) = \inf \{u, \ \Theta(u) \ge s\}$.

Proof. It is easy to see that

$$\|x - \hat{x}_{n_*}\|^2 = \sum_{j=n_*+1}^{\infty} |\langle x, u_j\rangle|^2 + \delta^2 \sum_{j=1}^{n_*} |\langle \xi, v_j\rangle|^2 / t_j.$$

Thus we can bound

$$\begin{aligned}\mathbf{E}\|x - \hat{x}_{n_*}\|^2 &= \sum_{j=n_*+1}^{\infty} |\langle x, u_j\rangle|^2 + \delta^2 \sum_{j=1}^{n_*} 1/t_j \\ &\le \varphi^2(t_{n_*}) \sum_{j=n_*+1}^{\infty} |\langle x, u_j\rangle|^2 / \varphi^2(t_j) + \delta^2/\Psi(t_{n_*}) \\ &\le R^2 \varphi^2(t_{n_*}) + \delta^2 \alpha^2(t_{n_*}).\end{aligned}$$

Therefore, if n_* is chosen as in (3.9), then

$$\mathbf{E}\|x - \hat{x}_{n_*}\|^2 \le 2R^2 \varphi^2((\Theta^2)^{-1}(\delta^2/R^2)),$$

from which the proof can be completed. □

Remark 3. *A refined analysis as carried out in* [9]*, and using a more subtle regularization provides the exact constant, but is not of such a form, which is easy to interpret.*

Lower bounds: Reduction to regression. Again we start from (3.7) to obtain an infinite system of equations. This set of equations is given a suitable form by considering

$$y_{\delta,j} = \theta_j + \delta \xi_j, \quad j = 1, 2, \dots, \tag{3.11}$$

which is the standard regression problem with independent (Gaussian) noise, having variances $\sigma_j^2 = \delta^2/s_j, \ j = 1, \dots,$ and $\theta_j := \langle x, u_j\rangle, \ j = 1, 2, \dots$.

This regression problem is only complete, after fixing assumptions on the unknown $\theta := (\theta_1, \theta_2, \dots)$. In terms of Fourier coefficients $\langle x, u_j\rangle$ with respect to the eigenbasis

$u_1, u_2, \dots$ the assumption (1.2) rewrites as

$$\theta \in B_R := \left\{ (\theta_1, \theta_2, \dots), \quad \sum_{j=1}^{\infty} \frac{\theta_j^2}{\varphi^2(s_j)} \le R^2 \right\}.$$

This is exactly the setup of the paper [9] by M. S. Pinsker. It will be convenient to rephrase Pinsker's results, who aimed at providing the exact asymptotics. Let a_j and $\sigma_j,\ j = 1, 2, \dots$ as above. Theorem 1 from [9] can be stated as follows.

Fact 1. *There is $c_1 > 0$ with the following property. If $\nu = \nu(\delta)$ is such, that*

$$\delta^2 \sum_{j;\ \varphi(s_j) > \nu} \frac{1}{s_j} \frac{1}{\nu\varphi(s_j)} \left(1 - \frac{\nu}{\varphi(s_j)}\right) = R^2, \tag{3.12}$$

then the error of the best estimator can be bounded from above and below by

$$R\nu \ge e^{stat}(A_\varphi(R), \delta) \ge c_1 R\nu.$$

Remark 4. *The derivation of this statement from Pinsker's Theorem 1 can be found in* [6]. *Under additional assumptions Pinsker is even able to show, that* $\lim_{\delta\to 0} e^{stat}(A_\varphi(R), \delta)/R\nu(\delta) = 1$.

Recall the definition of Θ from (3.8) and note that Θ is increasing and $\lim_{t\searrow 0} \Theta(t) = 0$.

In [5, Prop. 2] the lower bound for reconstruction in the general ill–posed problem for bounded deterministic noise was proved under additional geometric assumptions. In the statistical framework an analogous assumption will be made. In that paper the asymptotic error behavior is described through a concave function. The respective function in the present context turns out to be $t \to \varphi^2((\Theta^2)^{-1}(t))$. For our proof to work, slightly more is assumed. Namely, there is $0 < r < 1$ for which

$$\varphi^2((\Theta^2)^{-1}(2(1 + r^2)t)) \le 2\varphi^2((\Theta^2)^{-1}(t)), \quad 0 < t < \Theta^2(a). \tag{3.13}$$

Remark 5. *For concave functions* (3.13) *holds true with* $r = 0$.

Now we are ready to state and prove the main result in this section.

Proposition 3. *Suppose that (3.13) is fulfilled. Let t_* be chosen from*

$$n_* := \sup\left\{ n, \qquad \Theta(t_n) \ge \frac{r\delta}{2R} \right\} \tag{3.14}$$

as $t_* := t_{n_*}$. Then the error of the best estimator can be bounded from below by

$$e^{stat}(A_\varphi(R), \delta) \ge c_1 R\varphi(t_*). \tag{3.15}$$

First we provide some auxiliary estimate.

Lemma 1. *Under assumption (3.13) we have for $0 < 2s \le \varphi^2(a)$ that*

$$1 - \frac{\alpha^2((\varphi^2)^{-1}(2s))}{\alpha^2((\varphi^2)^{-1}(s))} \ge \frac{r^2}{1 + r^2}. \tag{3.16}$$

Proof. Because $t \to \varphi^2((\Theta^2)^{-1}(t))$ is increasing, the same holds true for the inverse, such that the above inequality implies by a change of variables

$$2(1+r^2)\Theta^2((\varphi^2)^{-1}(s)) \le \Theta^2((\varphi^2)^{-1}(2s)),$$

hence

$$2\frac{\Theta^2((\varphi^2)^{-1}(s))}{\Theta^2((\varphi^2)^{-1}(2s))} \le \frac{1}{1+r^2}.$$

Rewriting this in terms of α^2, using $\Theta = \varphi/\alpha$, we obtain

$$2\frac{s}{\alpha^2((\varphi^2)^{-1}(s))} \cdot \frac{\alpha^2((\varphi^2)^{-1}(2s))}{2s} = \frac{\alpha^2((\varphi^2)^{-1}(2s))}{\alpha^2((\varphi^2)^{-1}(s))} \le \frac{1}{1+r^2}.$$

from which the assertion easily follows. □

Let us notice that (3.13) is equivalent to

$$\frac{\sum_{2s>\varphi^2(s_j)\ge s} 1/s_j}{\sum_{\varphi^2(s_j)\ge 2s} 1/s_j} \ge r^2, \qquad \text{as } s \to 0, \tag{3.17}$$

as can be seen from the proof of Lemma 1. This representation is less intuitive but useful for inspecting examples, below.

Proof of Theorem 3. Given δ, let n_* be from (3.14). We shall show, that for $\bar{t}$, which is obtained from the solution ν from (3.12) via $\varphi(\bar{t}) = 2\nu$, necessarily $\bar{t} \ge t_*$. Indeed, we can conclude, using $t := (2\nu)^2$ and $t = \varphi^2(\bar{t})$ intermediately, that

$$\begin{aligned} R^2 &= \delta^2 \sum_{j;\ \varphi(s_j)>\nu} \frac{1}{s_j}\frac{1}{\nu\varphi(s_j)}\left(1 - \frac{\nu}{\varphi(s_j)}\right) \ge \frac{\delta^2}{8\nu^2} \sum_{j;\ 4\nu>\varphi(s_j)\ge 2\nu} \frac{1}{s_j} \\ &= \frac{\delta^2}{2t}\left(\alpha^2((\varphi^2)^{-1}(t) - \alpha^2((\varphi^2)^{-1}(2t))\right) \\ &\ge \frac{\delta^2}{2\Theta^2(\bar{t})}\left(1 - \frac{\alpha^2((\varphi^2)^{-1}(2t))}{\alpha^2((\varphi^2)^{-1}(t))}\right). \end{aligned}$$

Using Lemma 1 we end up with $R^2 \ge r^2\delta^2/4\Theta^2(\bar{t})$ from which we easily deduce $\bar{t} \ge t_*$. Using Fact 1 the proof can be completed. □

Comparing the upper and lower bound for the best possible accuracy we can state the following result.

Theorem 3. *Under assumption* (3.13) *the following estimates hold true.*

$$c_1 R\varphi(\Theta^{-1}(r\delta/2R) \le e^{stat}(A_\varphi(R), \delta) \le \sqrt{2}R\varphi(\Theta^{-1}(\delta/R). \tag{3.18}$$

If the function $t \to \varphi(\Theta^{-1}(t)$ *obeys a* Δ_2*–condition then*

$$e^{stat}(A_\varphi(R), \delta) \asymp R\varphi(\Theta^{-1}(\delta/R), \qquad \text{as } \delta \to 0.$$

Remark 6. *By its very construction, on the spectrum* $s_1, s_2, \ldots$ *it holds true that* $\alpha(s_j) \ge 1/\sqrt{s_j} = \alpha_{det}(s_j)$*. Therefore*

$$\varphi((\varphi/\alpha_{det})^{-1}(s_j)) \le \varphi((\varphi/\alpha)^{-1}(s_j)), \quad j = 1, 2, \ldots,$$

such that under (3.13) *and on the spectrum the best possible accuracy under statistical noise is harder than under bounded deterministic noise. For severely ill–posed* *problems, i.e., if $\alpha(t) \asymp \alpha_{det}(t)$, and if φ obeys a Δ_2–condition, then the rates coincide. This was observed in previous studies, see e.g.* [8].

4. Examples

It might be of interest to estimate the degree of ill–posedness and the related best possible accuracy in important special cases. For these examples the results for bounded deterministic noise are known, see [5], and will not be reviewed here. Therefore we entail details for statistical noise, only.

Example 1. *Let us assume that the singular numbers t_j of the operator A asymptotically behave like $t_j \asymp {}^1 j^{-a}$. Then $s_j \asymp j^{-2a}$ hence $\Psi(t) \asymp t^{\frac{2a+1}{2a}}$. The degree of ill–posedness in this case asymptotically behaves like $\alpha(t) \asymp (1/t)^{(a+1/2)/(2a)}$, $0 < t \leq a$.*

Let us additionally assume that smoothness is given by $\varphi(t) = t^{\mu/2a}$, see for instance [5]. *In this case it is easy to estimate the quotient from* (3.17) *as follows.*

$$\frac{\sum_{2s>\varphi^2(s_j)\geq s} 1/s_j}{\sum_{\varphi^2(s_j)\geq 2s} 1/s_j} \geq c\frac{1/(\varphi^2)^{-1}(s)\#\{j,\ s \leq \varphi^2(s_j) < 2s\}}{(1/(\varphi^2)^{-1}(2s))^{(2a+1)/2a}}$$
$$\geq c((\varphi^2)^{-1}(2s))^{1/2a}\#\left\{j,\ s \leq \varphi^2(s_j) < 2s\right\} \geq r^2$$

as $s \to 0$.

Thus Theorem 3 applies and we obtain $e^{stat}(A_\varphi(R),\delta) \asymp R(\delta/R)^{\frac{\mu}{\mu+a+1/2}}$.

In this particular case the analysis in [9] even yields that

$$\lim_{\delta\to 0} e^{stat}(A_\varphi(R),\delta)/(R(\delta/R)^{\frac{\mu}{\mu+a+1/2}}) = 1.$$

Example 2. *For severely ill–posed problems, i.e., when the singular numbers behave like $t_j \asymp e^{-\kappa j}$ for some $\kappa > 0$, we obtain $\Psi(t) \asymp \kappa t$, and the degree of ill–posedness asymptotically behaves like $\alpha(t) \asymp 1/\sqrt{\kappa t}$, $0 < t \leq a$. It is worthwhile to notice that asymptotically this is the same as the degree of ill–posedness for bounded deterministic noise, cf. Proposition 1.*

If for some $\mu > 0$ the a priori smoothness is $\varphi(t) = \log^{-1/\mu} 1/t^2$, then we can bound the enumerator from (3.17) *by the last summand and the denominator by using $\sum_{j=1}^{l} b^j \leq b^{l+1}/(b-1)$, $b > 1$ to obtain*

$$\frac{\sum_{2s>\varphi^2(s_j)\geq s} 1/s_j}{\sum_{\varphi^2(s_j)\geq 2s} 1/s_j} \geq (e^\kappa - 1)\frac{e^{\kappa 1/(2\kappa)(1/s)^\mu}}{e^{\kappa(1/(2\kappa)(1/s)^\mu+1)}} \geq 1 - e^{-\kappa}.$$

Again Theorem 3 applies and that the degree of ill–posedness correctly predicts the asymptotic error $e^{stat}(A_\varphi(R),\delta) \asymp R\log^{-1/\mu}\delta^2/R^2$.

In the specific situation when one wants to recover the initial distribution in the heat conduction problem from noisy data at time $T > 0$, then this problem reduces to a severely ill–posed, similar to the one from above. In this particular

[1] For sequences $a_j, b_j, j = 1,2,\ldots$ the symbol $a_j \asymp b_j$ means that there are constants $0 < c < C < \infty$ for which $ca_j \leq b_j \leq Ca_j$, $j = 1,2,\ldots$.

case, which is not covered by [9], and when the a priori smoothness is given by $\varphi(t) = (2T/\log(1/t^2))^{\beta/2}$, the authors in [1] can prove that

$$\lim_{\delta\to 0} e^{stat}(A_\varphi(R),\delta)/(R\left(\frac{2T}{\log\delta^2/R^2}\right)^{\beta/2}) = 1,$$

which again shows that the asymptotics abstained above is exact.

5. Conlusion

In this study we have formally introduced the notion of the degree α of ill–posedness for linear operator equation in Hilbert spaces. This definition extends previous approaches as found in the literature. In contrast to equations with bounded deterministic noise this degree depends on spectral properties of the involved operator by a quantity related to the harmonic mean of the singular numbers.

For ill–posed problems in variable Hilbert scales, i.e., when smoothness is measured in terms of general source conditions by some index function φ, it can be seen that the best possible accuracy for reconstructing the unknown solution is determined by the pair (α,φ) through the function $\Theta(t) = \varphi(t)/\alpha(t)$. Under some additional restriction the asymptotic error of reconstruction at noise level δ behaves like $\delta \to \varphi(\Theta^{-1}(\delta))$. This extends previous study for bounded deterministic noise.

References

[1] G. Golubev and R. Khasminskii. A statistical approach to the Cauchy problem for the Laplace equation. In *State of the art in probability and statistics (Leiden, 1999)*, volume 36 of *IMS Lecture Notes Monogr. Ser.*, pages 419–433. Inst. Math. Statist., Beachwood, OH, 2001.

[2] Markus Hegland. Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization. *Appl. Anal.*, 59(1-4):207–223, 1995.

[3] Peter Mathé and Sergei V. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38(6):1999–2021, 2001.

[4] Peter Mathé and Sergei V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(6):1263–1277, 2003.

[5] Peter Mathé and Sergei V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[6] Peter Mathé and Sergei V. Pereverzev. Regularization of some linear ill–posed problems with discretized random noisy data. submitted, 2004.

[7] M. Nussbaum and Sergei Pereverzev. The degree of ill–posedness in stochastic and deterministic noise models. Preprint 509, WIAS, Berlin, 1999.

[8] Sergei Pereverzev and Eberhard Schock. Morozov's discrepancy principle for Tikhonov regularization of severely ill-posed problems in finite-dimensional subspaces. *Numer. Funct. Anal. Optim.*, 21(7-8):901–916, 2000.

[9] M. S. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission*, 16(2):52–68, 1980.

[10] Ulrich Tautenhahn. Optimality for ill-posed problems under general source conditions. *Numer. Funct. Anal. Optim.*, 19(3-4):377–398, 1998.

[11] G. M. Vaĭnikko and A. Yu. Veretennikov. Итерационные процедуры в некорректных задачах. "Nauka", Moscow, 1986.

[12] Grace Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667, 1977.