

## REGULARIZATION OF SOME LINEAR ILL-POSED PROBLEMS WITH DISCRETIZED RANDOM NOISY DATA

PETER MATHÉ AND SERGEI V. PEREVERZEV

ABSTRACT. For linear statistical ill-posed problems in Hilbert spaces we introduce an adaptive procedure to recover the unknown solution from indirect discrete and noisy data. This procedure is shown to be order optimal for a large class of problems. Smoothness of the solution is measured in terms of general source conditions. The concept of operator monotone functions turns out to be an important tool for the analysis.

### 1. INTRODUCTION

In this study we analyze the numerical solution of operator equations  $Ax = y$  under the presence of noise, which means we are given

$$(1) \quad y_\delta = Ax + \delta\xi,$$

where  $A: X \rightarrow Y$  acts compactively and injective between Hilbert spaces  $X$  and  $Y$ . The noise  $\xi$  is assumed to be centered and (weak) Gaussian in  $Y$ . Further restrictions will be imposed later on.

When processing data, several decisions have to be made. First one has to agree upon the method of reconstruction. Our analysis will be carried out for *Tikhonov's regularization*, because this is most often used in practice. Then, and this is the heart of the present study, several parameters have to be chosen.

Typically the a priori information on the solution is not precise, so decisions must be made on the basis of some *adaptive procedure*. Here we shall analyze a Lepskiĭ-type procedure. We stress that this procedure is only suited for one-parameter families of estimators, as this is the case for kernel-type statistical estimators with varying bandwidth. Within the present context, the approximation obtained by discretized regularization depends on at least two parameters; the regularization parameter, and the amount of discrete data to be used. To apply the Lepskiĭ strategy it is thus necessary to relate both parameters, independent of the underlying true smoothness. This is one of the accomplishments of the present paper, and is exhibited in Theorem 3 by estimating the impact of each parameter upon accuracy.

The question arises as to which conditions of the described procedure may achieve the best possible (order of) reconstruction of the true solution. To this end we shall present an analysis for the case when the smoothness is measured in terms of *general source conditions*, an approach which recently became attractive. It is more flexible

---

Received by the editor February 2, 2005 and, in revised form, August 26, 2005.

2000 *Mathematics Subject Classification*. Primary 62G05; Secondary 62G20, 65J20.

*Key words and phrases*. Statistical ill-posed problem, general source condition, operator monotone function.

©2006 American Mathematical Society  
Reverts to public domain 28 years from publication

to describe smoothness than just the usual scales of Hilbert spaces. In particular severely ill-posed problems can be well described and analyzed within this framework. To achieve tight upper bounds a new concept is introduced into the analysis of statistical ill-posed problems, the notion of *operator monotone functions*. This class of functions allows us to analyze discretization particularly well. The upper bounds are accompanied with lower bounds, and are presented as a specification of Pinsker's results.

The outline of the paper is as follows. First we shall provide the setup under which the analysis is carried out. Then we shall present the adaptive procedure to be analyzed in this study. The theoretical properties of it will be given in an Appendix.

In Section 4 we introduce smoothness in terms of general source conditions and prove uniform error bounds for Tikhonov regularization for smoothness classes given through operator monotone functions. We close this section with proving lower bounds. Finally we prove the optimality (up to a logarithmic factor) of the adaptive procedure for such smoothness classes.

## 2. SETUP

The description of the problem as given by (1) must be completed with information on the available data and the nature of the noise.

**2.1. Discrete data.** Our focus shall be on the regularization of problem (1), based on discrete data, given through some orthonormal system. Precisely, instead of (1) we shall limit ourselves to a *one-sided* discretization of the form  $B := QA$  with a finite rank orthogonal projection  $Q$  in  $Y$ . The interpretation is natural. Turning from (1) to

$$Qy_\delta = QAx + \delta Q\xi$$

means that instead of complete data, generically expressed as  $y_\delta$ , we turn to a finite set  $\{y_{\delta,1}, \dots, y_{\delta,n}\}$  of data given by

$$(2) \quad y_{\delta,i} = \langle Ax, \psi_i \rangle + \delta \xi_i, \quad i = 1, \dots, n,$$

where  $Q$  is the orthogonal projection onto the linear span of  $\{\psi_1, \dots, \psi_n\}$ .

At this point it is important to note that we assume to have observations without repetitions, which means that each of the functionals  $y_{\delta,i} = \langle y_\delta, \psi_i \rangle$  is observed only once.

*Remark 1.* Let us briefly discuss a similar, although different, model which has been studied elsewhere, and which is related to repeated observations. Indeed, if for each  $i = 1, \dots, n$  we are given repetitions

$$y_{\delta,i}^k = \langle Ax, \psi_i \rangle + \delta \xi_i^k, \quad i = 1, \dots, n, \quad k = 1, \dots, M,$$

where for each  $i$  the noise  $\xi_i^k$ ,  $k = 1, \dots, M$  is zero-mean, independent and identically distributed (i.i.d.), then we may decrease the variance at functionals  $\psi_i$  by letting

$$(3) \quad \hat{y}_{\delta,i} := \frac{1}{M} \sum_{k=1}^M y_{\delta,i}^k = \langle Ax, \psi_i \rangle + \frac{\delta}{M} \sum_{k=1}^M \xi_i^k.$$

Observations that are modified in such a way now correspond to the model (2) studied here, but with noise level  $\mathbf{E} |\langle Ax, \psi_i \rangle - \hat{y}_{\delta,i}|^2 \sim \delta^2/M$ .

Under these circumstances and asymptotically one can even achieve arbitrary accuracy. Observation models including repetitions have been discussed in [6, 17]. For example, the observation model from [17, Chapt. 7] can be interpreted as the model with  $M = \sqrt{n}$  repetitions, and the analog of (2) appears as

$$y_{\delta,i} = \sqrt{n}\langle Ax, \psi_i \rangle + \delta\xi_i, \quad i = 1, \dots, n.$$

It can thus be considered as an observation model without repetitions, as studied here, but with the noise level  $\delta/\sqrt{n}$ .

**2.2. The noise.** As mentioned in the Introduction, we will further impose restrictions on the nature of the random noise  $\xi$ . Most importantly we assume that it is centered and has second weak moments, i.e., for all (functionals)  $a \in Y$  the expectations  $\mathbf{E}|\langle \xi, a \rangle|^2 < \infty$  are finite. Thus we can introduce the *covariance operator*  $K$  by  $\langle Ka, b \rangle := \mathbf{E}\langle \xi, a \rangle \langle \xi, b \rangle$ ,  $a, b \in Y$ .

We mention the following extreme cases; see [16, Chapt. 3.2]. Each *bounded* symmetric nonnegative operator  $K$  induces a random element with weak second moment. For Gaussian random elements and  $K := I$  we thus obtain the *Gaussian white noise model*. If the operator  $K$  has a finite trace, then the random element possesses a strong second moment, i.e.,  $\mathbf{E}\|\xi\|^2 < \infty$ . In this case, the theory of statistical ill-posed problems does not distinguish from the classical ill-posed ones with bounded deterministic noise, as can be seen from the results below in Section 4.3.

Here we shall study classes of statistical ill-posed problems, for noise with covariance  $K$  which has the following property.

**Assumption 1.** *There are a parameter  $1 \leq p \leq \infty$  and a constant  $C_p$ , such that for any discretization projection  $Q$  it holds true that*

$$(4) \quad \text{trace } QKQ \leq C_p(\text{rank}(Q))^{1/p}.$$

This class of covariance operators is denoted by  $\mathcal{K}_p$ . Note that the noisy perturbations  $\{\xi_1, \dots, \xi_n\}$  form a random vector having covariance  $QKQ$ , if  $K$  was the original covariance of the random element  $\xi$ , thus  $\mathbf{E}\|Q\xi\|^2 = \text{trace } QKQ$ .

*Remark 2.* If the covariance operator  $K$  belongs to some *Schatten class*  $\mathcal{S}_q$  of operators containing all operators with  $q$ -summable singular numbers, i.e.,  $\|K\|_{\mathcal{S}_q}^q := \sum_{j=1}^{\infty} s_j^q(K) < \infty$ , then (4) is fulfilled for arbitrary projections  $Q$  and with dual index  $p$ , i.e.,  $1/p + 1/q = 1$ . (We agree to let  $\mathcal{S}_{\infty}$  denote the space of all bounded linear operators in  $Y$ .) We refer to [12, 2.11.23] for more details.

Summarizing, in our study the reconstruction  $x_{\alpha,\delta}$  of the solution  $x$  will be based on noisy data  $y_{\delta,1}, y_{\delta,2}, \dots, y_{\delta,n}$ , with the assumptions made above. It will be obtained from Tikhonov regularization as

$$x_{\alpha,\delta} = (\alpha I + B^*B)^{-1}B^*y_{\delta}$$

with parameter  $\alpha$  to be chosen adaptively. If the observation functionals  $\{\psi_i\}$  form an orthonormal system in  $Y$ , then the regularized approximation  $x_{\alpha,\delta}$  can be represented in the form  $x_{\alpha,\delta} = \sum_{k=1}^n \zeta_k A^* \psi_k$ , where the vector  $(\zeta_1, \zeta_2, \dots, \zeta_n)$  constitutes the solution to the system of linear equations

$$\alpha\zeta_i + \sum_{j=1}^n \langle A^* \psi_j, A^* \psi_i \rangle \zeta_j = y_{\delta,i}, \quad i = 1, \dots, n.$$

Thus, to construct  $x_{\alpha,\delta}$  one needs to know the observations  $y_{\delta,i}$  and the inner products  $\langle A^*\psi_j, A^*\psi_i \rangle$ . The covariance operator itself is not used in the regularization procedure. As we will see in the sequel, to accomplish the regularization one needs to know only two numbers  $p$  and  $C_p$  from Assumption 1, or reliable estimates for them. This allows us to apply Tikhonov regularization in situations when the covariance structure of the noise is not completely known.

### 3. THE ADAPTIVE PROCEDURE

Within the present approach we shall use an adaptation procedure similar to the one introduced by Lepskiĭ [5]. It is based on a finite number of approximations at different levels of the regularization parameter. Since 1990 Lepskiĭ's approach has found many applications, there are many results, either generalizing the scope of applicability or modifying it for specific applications. We only mention [2, 15], where the same approach has been applied to inverse problems. Within our approach the variance term may be quite general and a suitable reference is not available. Therefore we outline the procedure in some general form, providing details and proofs in the Appendix.

**3.1. An abstract oracle principle.** Suppose we are given a finite set  $\{x_1, \dots, x_m\}$  of random elements in some metric space  $(M, d)$ , given on some probability space  $(\Omega, \mathcal{F}, P)$  and a decreasing function  $\Psi : \{1, \dots, m\} \rightarrow \mathbb{R}^+$ . Let  $x \in M$  be any (deterministic) element.

**Definition 1.** A nondecreasing function  $\Phi : \{1, \dots, m\} \rightarrow \mathbb{R}^+$  is called *admissible* for  $x$  if there exists a family of nonnegative random variables  $\rho(j)$ ,  $j = 1, 2, \dots, m$ , for which

- (i)  $d(x, x_j) \leq \Phi(j) + \rho(j)$ ,  $j = 1, 2, \dots, m$ ,
- (ii)  $\mathbf{E}\rho^2(j) \leq \Psi^2(j)$ ,  $j = 1, 2, \dots, m$ ,

and

- (iii)  $\Phi(1) \leq \Psi(1)$ .

Furthermore we assume the following *concentration inequality* for  $\rho$ : there are constants  $0 < b < 1 \leq a$  for which uniformly in  $j = 1, 2, \dots, m$  it holds true that

$$(5) \quad P(\rho(j) \geq t\Psi(j)) \leq a \exp(-bt^2), \quad t > 0.$$

*Remark 3.* As can be drawn from item (i), the functions  $\Phi$  and  $\rho$  correspond to an error estimate with  $\rho$  being the noise term, the second moment of which is controlled by  $\Psi^2$ .

The adaptive procedure as well as their properties are stated in

**Theorem 1** (Statistical oracle inequality). *Let  $\{x_1, \dots, x_m\}$  be a finite set of random elements in  $(M, d)$  and let  $\Psi : \{1, \dots, m\} \rightarrow \mathbb{R}^+$  be decreasing. Furthermore we assume that there is  $D \geq 1$  for which  $\Psi(i) \leq D\Psi(i+1)$ ,  $i = 1, \dots, m-1$ .*

*Given some  $\kappa \geq 1$ , let us consider the (random) index*

$$\bar{j} := \max \{j, \quad d(x_i, x_j) \leq 4\kappa\Psi(i) \text{ for all } i \leq j\}.$$

For every admissible  $\Phi$ , with concentration (5), the following estimate holds true:

$$(6) \quad (\mathbf{E}d^2(x, x_{\bar{j}}))^{1/2} \leq 6D\kappa \min \{ \Phi(j) + \Psi(j), j = 1, 2, \dots \} + \frac{15}{b} \Psi(1) \sqrt{ma} \exp(-b/4\kappa^2).$$

The proof is carried out in the Appendix.

**3.2. Description of the procedure.** In the application below, the adaptive procedure is based on Tikhonov regularization of some projection scheme. Precisely, for a discretization  $Q$ , corresponding  $B := QA$ , and parameter  $\alpha$ , we let

$$x_{\alpha, \delta} := (\alpha I + B^* B)^{-1} B^* Q y_{\delta}.$$

Due to the nature of the noise, the elements  $x_{\alpha, \delta}$  are random in  $X$ . The elements for the abstract oracle principle will be obtained along a sequence  $\alpha_1, \alpha_2, \dots, \alpha_m$  of regularization parameters.

Precisely, for a given  $q > 1$  and  $\alpha_1 = \delta^2$ , let  $\alpha_j := q^j \alpha_1, j = 2, \dots, m$ , where  $m$  is determined from  $q^{m-1} \alpha_1 \leq 1 < q^m \alpha_1$ . In particular  $2 \log_q 1/\delta \leq m \leq 1 + 2 \log_q 1/\delta$ .

Corresponding to  $\alpha = \alpha_j$  we choose  $n = n(\alpha)$  and a projection  $Q = Q_{n(\alpha)}$ , hence  $B = Q_{n(\alpha)} A$ , and use

$$x_j := x_{\alpha_j, \delta} := (\alpha_j I + B^* B)^{-1} B^* y_{\delta}, \quad j = 1, 2, \dots, m.$$

*Remark 4.* As will be seen below, the interplay between  $\alpha$  and  $Q = Q_n$  is important. Two scenarios are natural.

First, we may successively choose  $\alpha$  and then let  $n = n(\alpha)$  be chosen appropriately. This presumes that potentially we are able to use an arbitrary amount of data.

Second, we may encounter the situation when the amount of data is given. Then one may wish to successively reduce the amount of data used and adapt  $\alpha$  accordingly to finally arrive at the  $\alpha$  which is best for the data at hand.

The choice of the best parameters is dependent on the actual smoothness of the true solution. This is well known and within the framework of general source conditions; it has recently been studied by the authors in [9].

As mentioned above the decreasing function  $\Psi$  is obtained from a valid bound on the variance term for Tikhonov regularization.

**Lemma 1.** *Let  $B := QA$ . Under Assumption 1 it holds that*

$$\mathbf{E} \| (\alpha I + B^* B)^{-1} B^* \xi \|^2 \leq C_p \frac{(\text{rank } Q)^{1/p}}{4\alpha}.$$

*Proof.* Because  $B^* \xi = B^* Q \xi$  we can bound the expectation

$$\mathbf{E} \| (\alpha I + B^* B)^{-1} B^* \xi \|^2 \leq \| (\alpha I + B^* B)^{-1} B^* \|^2 \mathbf{E} \| Q \xi \|^2 \leq C_p \frac{(\text{rank } Q)^{1/p}}{4\alpha}$$

by Assumption 1 and a simple norm bound. □

For the procedure to work we shall assume that the amount of used data  $n = n(\alpha)$  is a decreasing function of  $\alpha$ , which is very natural and not restrictive. Moreover, we assume that the bound  $C_p$  for the covariance from (4) is available. If  $n = n(\alpha) \geq 1$  is decreasing, then the function

$$(7) \quad \Psi_p(\alpha) := C_p \frac{(n(\alpha))^{1/p}}{4\alpha}, \quad \alpha > 0,$$

is decreasing.

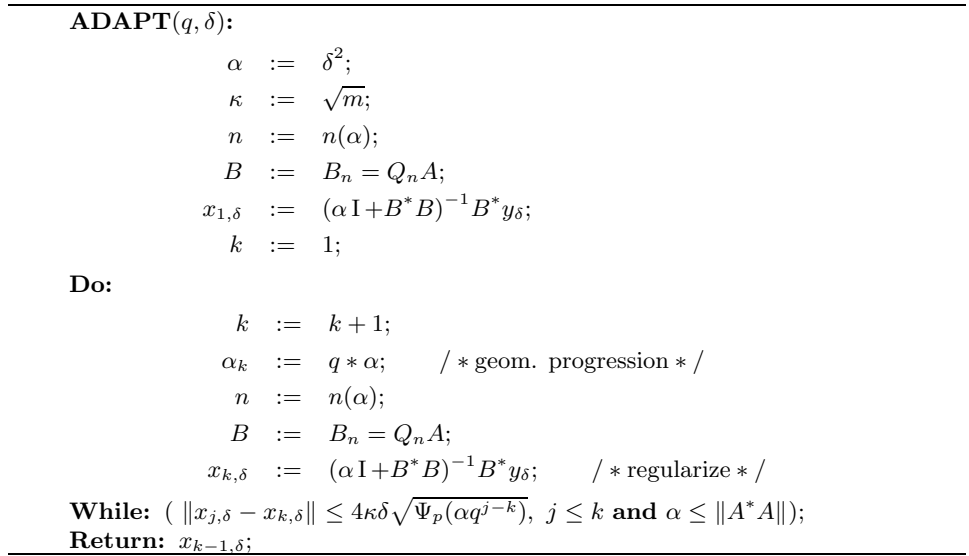


FIGURE 1. The adaptive strategy

Finally we let  $\Psi(j) := \delta\sqrt{\Psi_p(\alpha_j)}$ ,  $j = 1, 2, \dots, m$ , which is decreasing as  $j$  increases from 1 to  $m$ . To apply Theorem 1 we shall assume that  $n$  does not decrease too fast, which is reasonable if one wants to retain the best possible accuracy, which will be seen from Theorem 3. Precisely, we suppose that  $n$  decreases such that for some  $\bar{D}$  it holds true that

$$(8) \quad n(t) \leq \bar{D}n(qt), \quad t > 0.$$

Under (8) the function  $\Psi$  used in the adaptive procedure obeys  $\Psi(j) \leq D\Psi(j + 1)$  for  $D := \sqrt{q(\bar{D})^{1/p}}$ .

The numerical procedure is shown schematically in Figure 1. We let  $\kappa := \sqrt{m} \sim \log_q^{1/2}(1/\delta)$ , which can be seen to be best. Finally we observe that for the present setup of  $\alpha_1, \alpha_2, \dots$ , there are  $\delta_0 > 0$  and  $C = C(D, q) > 6D$  such that for  $\delta \leq \delta_0$  it holds true that

$$\frac{15}{b}\Psi(1)\sqrt{ma} \exp(-b/4m) \leq (C - 6D)\Psi(m).$$

Applying Theorem 1 with the functions and parameters above we obtain

**Theorem 2.** *Suppose we have chosen the parameter  $q > 1$  and that  $n = n(\alpha)$  obeys (8). For any  $x \in X$  the resulting approximation  $x_{\bar{\alpha},\delta}$  of the adaptive strategy yields*

$$\begin{aligned} (\mathbf{E}\|x - x_{\bar{\alpha},\delta}\|^2)^{1/2} &\leq C\sqrt{\lceil 2\log_q(1/\delta) \rceil} \\ &\quad \times \min \left\{ \Phi(j) + \delta\sqrt{\Psi_p(\alpha_j)}, j = 1, 2, \dots, \Phi \text{ admiss. for } x \right\} \end{aligned}$$

for  $\delta \leq \delta_0$  and for some constant  $C = C(D, q)$ .

*Remark 5.* The above strategy has only one parameter to choose. If  $q$  is large, then there are only a few comparisons to be made, but the approximation to the true value of  $\alpha$  can only be rough. A choice of  $q > 1$  close to 1 will result in a large

number of possible terminal values, thus allowing us to recover the best one more accurately, but the additional log-factor in the estimate will be large.

As seen from Theorem 2, the adaptive strategy provides the best possible choice among the approximations  $x_{\alpha,\delta}$  and all functions for admissible  $x$  in the sense of Definition 1, where  $\Psi(j) = \delta\sqrt{\Psi_p(\alpha_j)}$ ,  $j = 1, 2, \dots, m$ . Two questions arise:

- (1) Given  $x$ , can we infer something about admissible functions for  $x$  and the given  $\Psi$ ?
- (2) How does the error bound from Theorem 2 relate to the best possible accuracy?

#### 4. GENERAL SOURCE CONDITIONS

In this section we shall describe a large class of functions  $\varphi$  and corresponding sets  $A_\varphi(R) \subset X$  which possess the following property. If  $x \in A_\varphi(R)$ , then the functions  $\varphi$  provide  $\Phi$ , which are admissible.

Because we want to bound the distance of  $x_{\hat{\alpha},\delta}$  to the true solution  $x$ , this amounts to proving error estimates for Tikhonov regularization  $x_{\alpha,\delta} := (\alpha I + B^*B)^{-1}B^*y_\delta$  in which  $\delta^2\Psi_p(\alpha)$  is an estimate for the variance.

**4.1. The error criterion.** For any covariance structure  $K \in \mathcal{K}_p$  of the noise and for any estimator  $\hat{x}$  of the true solution  $x$  based on data  $y_{\delta,1}, y_{\delta,2}, \dots, y_{\delta,n}$  the error is given by

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

where expectation is with respect to the noise. The worst-case error over a class  $F$  of problem instances is given as

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

The best possible order of accuracy is defined by minimization over all estimators, i.e.,

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

Lower bounds are of interest only for noise which is not degenerate, so we shall study

$$e(F, \mathcal{K}_p, \delta) := \sup_{K \in \mathcal{K}_p} \inf_{\hat{x}} e(F, K, \hat{x}, \delta).$$

**4.2. Smoothness in terms of general source conditions.** Here we are going to specify the class  $F$  of problem instances. In contrast to the usual approach, where smoothness is given by (finite) differentiability properties, we want to keep the class of admissible functions for the adaptive strategy as large as possible.

To this end we start with the description of the Moore–Penrose generalized solution of the equation  $Ax = y$  for a compact injective operator  $A: X \rightarrow Y$ . Under this assumption there is a singular value decomposition  $Ax = \sum \sqrt{s_k} \langle x, u_k \rangle v_k$ , for orthonormal systems  $u_1, u_2, \dots$  in  $X$  and  $v_1, v_2, \dots$  in  $Y$ , and a (decreasing) sequence of singular numbers  $s_j = s_j(A^*A) > 0$ ,  $j = 1, 2, \dots$ . Throughout we shall denote  $a := s_1(A^*A) = \|A^*A\|$ . The Moore–Penrose solution is then given through

$$x^+ := \sum_{j=1}^{\infty} \frac{1}{\sqrt{s_j}} \langle y, v_j \rangle u_j.$$

Picard's criterion asserts that  $x^+ \in X$  iff  $\sum |\langle y, v_j \rangle|^2 / s_j < \infty$ . This implies a minimal decay of the Fourier coefficients  $\langle y, v_j \rangle$ . Therefore it seems natural to measure smoothness of  $x$  by enforcing some faster decay. This is achieved through an increasing function  $\varphi: (0, a] \rightarrow \mathbb{R}_+$ ,  $\lim_{t \rightarrow 0} \varphi(t) = 0$ , by requiring  $\sum |\langle y, v_j \rangle|^2 / (s_j \varphi^2(s_j)) < \infty$ . Then

$$v := \sum_{j=1}^{\infty} \frac{1}{\sqrt{s_j} \varphi(s_j)} \langle y, v_j \rangle u_j \in X$$

and

$$x^+ = \sum_{j=1}^{\infty} \varphi(s_j) \langle v, u_j \rangle v_j = \varphi(A^*A)v \in X.$$

Therefore we shall seek solutions to (1) in terms of general source conditions, given by

$$(9) \quad A_\varphi(R) := \{x \in X, \quad x = \varphi(A^*A)v, \quad \|v\| \leq R\},$$

where the symbol  $A$  refers to the underlying operator. The function  $\varphi$  is called an *index function*. In such a form inverse problems have been studied by several authors; we mention [3, 14, 11] and previous study by the present authors [10, 9].

There is good reason to further restrict the class of admissible index functions. In general, the smoothness expressed through general source conditions is not stable with respect to perturbations in the involved operator  $A^*A$ . But regularization is carried out for a nearby operator  $B$ , and it is desirable that the class  $A_\varphi(R)$  is robust with respect to this type of discretization. This can be achieved by requiring  $\varphi$  to be *operator monotone*. For a detailed exposition of this concept we refer the reader to [1]. In the context of numerical analysis for ill-posed problems this was introduced by the authors in [8].

As in [8] we introduce the partial ordering  $B_1 \leq B_2$  for self-adjoint operators  $B_1$  and  $B_2$  on some Hilbert space  $X$ , by  $\langle B_1 u, u \rangle \leq \langle B_2 u, u \rangle$  for any  $u \in X$ .

**Definition 2.** A function  $\varphi: (0, b) \rightarrow \mathbb{R}$  is operator monotone on  $(0, b)$ , if for any pair of self-adjoint operators  $B_1, B_2: X \rightarrow X$  with spectra in  $(0, b)$ , the relation  $B_1 \leq B_2$  implies  $\varphi(B_1) \leq \varphi(B_2)$ .

The important implication of the concept of operator monotonicity in the context of discretization is as follows.

**Lemma 2** (see [8]). *Suppose  $\varphi$  is an operator monotone index function on  $(0, b)$ ,  $b > a$ . Then there is  $M < \infty$ , depending on  $b - a$ , such that for any pair  $A, B, \|A\|, \|B\| \leq a$  of nonnegative self-adjoint operators on some Hilbert space  $X$ , it holds that*

$$\|\varphi(A) - \varphi(B)\| \leq M\varphi(\|A - B\|).$$

**Assumption 2.** *We assume that the source condition is given by an increasing operator monotone on  $(0, b)$  index function  $\varphi$ ,  $\varphi(0) = 0$ , for some  $b > a = \|A^*A\|$ .*

The following monotonicity property will be used below.

**Lemma 3.** *Under Assumption 2 there is  $T \geq 1$  such that*

$$\varphi(t)/t \leq T\varphi(s)/s, \quad \text{whenever } 0 < s < t \leq a < b.$$



*Proof.* As used in [8], such  $\varphi$  admit a decomposition  $\varphi = \varphi_0 + \varphi_1$  into a concave part  $\varphi_0$  and a Lipschitz part  $\varphi_1$  with Lipschitz constant, say  $c_1$ . Moreover,  $\varphi_0(0) = \varphi_1(0) = 0$  such that  $\varphi_0(s)/s \geq \varphi_0(t)/t$  whenever  $0 < s < t \leq a$ . Thus for  $t > s$  we estimate

$$\varphi(t)/t = (\varphi_0(t) + \varphi_1(t))/t \leq \varphi_0(t)/t + c_1.$$

Now, for  $T := c_1 a / \varphi_0(a) + 1$  we conclude  $\varphi(t)/t \leq T \varphi(s)/s$ . □

**4.3. The error of Tikhonov regularization.** We start with the following obvious bias-variance decomposition at the true solution  $x$ :

$$(10) \quad \mathbf{E} \|x - (\alpha I + B^* B)^{-1} B^* Q y_\delta\|^2 = \|x - (\alpha I + B^* B)^{-1} B^* Q A x\|^2 + \delta^2 \mathbf{E} \|(\alpha I + B^* B)^{-1} B^* Q \xi\|^2.$$

The noise term was bounded in Lemma 1 as

$$\delta^2 \mathbf{E} \|(\alpha I + B^* B)^{-1} B^* Q \xi\|^2 \leq C_p \delta^2 (\text{rank } Q)^{1/p} / 4\alpha.$$

Note that this is exactly the function  $\delta^2 \Psi_p(\alpha)$  with  $\Psi_p$  as in (7).

We turn to bounding the noise-free term. To this end the following result is used.

**Proposition 1.** *Let  $\alpha \rightarrow (\alpha I + B^* B)^{-1} B^*$  be the family of operators from Tikhonov regularization. The following assertions hold true:*

- (1)  $\|(\alpha I + B^* B)^{-1} B^* : Y \rightarrow X\| \leq 1 / (2\sqrt{\alpha})$ .
- (2)  $\|(\alpha I + B^* B)^{-1} B^* B : X \rightarrow X\| \leq 1$ .
- (3) *If  $\varphi$  obeys Assumption 2, then*

$$\| (I - (\alpha I + B^* B)^{-1} B^* B) \varphi(B^* B) : X \rightarrow X \| \leq T \varphi(\alpha),$$

*with  $T$  from Lemma 3.*

*Proof.* The first two assertions follow from spectral calculus. To prove item (3), we mention that the estimate may be rewritten as

$$(11) \quad \sup_{0 < t \leq a} \left| \frac{\alpha}{\alpha + t} \varphi(t) \right| \leq T \varphi(\alpha).$$

For  $t \leq \alpha$ , (11) holds trivially with  $T = 1$ . Otherwise we use Lemma 3 to derive

$$\sup_{\alpha < t \leq a} \left| \frac{\alpha}{\alpha + t} \varphi(t) \right| \leq \sup_{\alpha < t \leq a} \left| \frac{\alpha t}{\alpha + t} \varphi(t) / t \right| \leq T \sup_{\alpha < t \leq a} \left| \frac{\alpha t}{\alpha + t} \varphi(\alpha) / \alpha \right| \leq T \varphi(\alpha),$$

which completes the proof. □

Furthermore, the following remark is important. Because we are given a finite set of data in (2), corresponding to a projection  $Q$ , the error of any reconstruction based on (2) must depend on the quality with which  $QA$  is close to the original operator  $A$ , i.e., it will depend on

$$(12) \quad \eta := \|(I - Q)A : X \rightarrow Y\|.$$

Under Assumption 2, in particular using the bound from Lemma 2, we obtain uniformly in  $x \in A_\varphi(R)$  the estimate

$$\begin{aligned} \|x - (\alpha I + B^*B)^{-1}B^*QA x\| &= \|(I - (\alpha I + B^*B)^{-1}B^*B)\varphi(A^*A)v\| \\ &\leq RT\varphi(\alpha) + R\|\varphi(A^*A) - \varphi(B^*B)\| \\ &\leq R(T + M)(\varphi(\alpha) + \varphi(\eta^2)). \end{aligned}$$

To obtain best possible rates of reconstruction with Tikhonov regularization, we further impose restrictions on the operator and on the power of the design, given by the projection  $Q$ .

**Assumption 3.** *There exists  $r > 0$  such that the singular values  $s_j(A)$  of  $A$  obey*

$$(13) \quad s_j(A) \asymp j^{-r}, \quad j = 1, 2, \dots$$

As a consequence  $s_j(A^*A) \asymp j^{-2r}$ ,  $j = 1, 2, \dots$ . Furthermore, we restrict our analysis to projections  $Q$  which at least by order realize the best possible approximation.

**Assumption 4.** *There is a constant  $C \geq 1$  such that*

$$\|(I - Q)A: X \rightarrow Y\| \leq C(\text{rank } Q)^{-r}.$$

*Remark 6.* In many applications the design may be chosen independent of the operator. This is for instance the case for  $Y = Y^0$  belonging to some scale  $\{Y^r\}$  of Sobolev Hilbert spaces and  $\|A: X \rightarrow Y^r\| \leq C_1$ , where the  $n$ -widths  $a_n(Y^r, Y)$  of  $Y^r$  in  $Y$  obey  $a_n(Y^r, Y) \leq C_2 n^{-r}$ ,  $n = 1, 2, \dots$ . Then any family of projections  $Q_n$ ,  $n = 1, 2, \dots$ , with

$$\|I - Q: Y^r \rightarrow Y\| \leq C(\text{rank } Q)^{-r}$$

obeys Assumption 4 with  $C := C_1 C_2$ . This is known for many spline approximations, finite element schemes and wavelet expansions.

We mention the study [7], where such analysis was carried out for Abel’s integral equation and piecewise constant (histogram) design.

We can formulate the main result on error estimation under a known source condition. Let  $\Theta_p(t) := t^{\frac{2rp+1}{4rp}}\varphi(t)$ ,  $0 < t \leq a$ .

**Theorem 3.** *Suppose that the reconstruction  $x_{\alpha,\delta}$  of the solution  $x$  based on data  $y_{\delta,1}, y_{\delta,2}, \dots, y_{\delta,n}$  will be given through*

$$x_{\alpha,\delta} = (\alpha I + B^*B)^{-1}B^*y_\delta,$$

where  $B := QA$  is a projection method with accuracy  $\eta$  from (12). Under Assumptions 1–4 the following assertions hold true:

(1) *For any choice of  $\alpha$  and  $\eta$  we have uniformly for  $K \in \mathcal{K}_p$*

$$e(A_\varphi(R), K, x_{\bar{\alpha},\delta}, \delta) \leq R(M + T)(\varphi(\alpha) + \varphi(\eta^2)) + \delta\sqrt{\Psi_p(\alpha)}.$$

(2) *There is a constant  $C < \infty$  such that for*

$$(14) \quad \alpha^* := \inf \left\{ \alpha, \quad \Theta_p(\alpha) \geq \frac{\delta}{R} \right\}$$

and  $n \asymp \lceil (\alpha^*)^{-1/2r} \rceil$ , it holds uniformly for  $K \in \mathcal{K}_p$ ,

$$e(A_\varphi(R), K, x_{\alpha^*,\delta}, \delta) \leq CR\varphi(\Theta_p^{-1}(\delta/R)).$$

*Proof.* The first statement follows immediately from the bias-variance decomposition above and the definition of  $\eta$ .

Note that  $\Theta_p$  is increasing and  $\lim_{t \rightarrow 0} \Theta_p(t) = 0$ , such that there is a unique choice for  $\alpha^*$ . Furthermore, from  $n \asymp \lceil \alpha^{-1/2r} \rceil$  we deduce that for some  $C_1, C_2 \geq 1$  it holds true that  $\Psi_p(t) \leq C_1 (1/t)^{(2rp+1)/(2rp)}$  and  $\eta^2 \leq C_2 \alpha$ . Furthermore, by the choice from (14) the variance term is dominated by a multiple of the bias. Also, by Lemma 3 it holds true that  $\varphi(C_2 t) \leq C_2 T \varphi(t)$ ,  $t > 0$ , such that uniformly for  $K \in \mathcal{K}_p$  we can bound

$$e(A_\varphi(R), K, x_{\alpha^*, \delta}, \delta) \leq R\tilde{C} (\varphi(\alpha^*) + M\varphi(C_2\alpha^*)) \leq CR\varphi(\alpha^*),$$

which proves assertion (2). □

*Remark 7.* The above rate of reconstruction will be seen to be optimal in Section 4.4, and it is worthwhile to note that this best possible rate for reconstruction under random noise is always worse than for deterministic noise. Indeed, as can be seen in [10], for bounded deterministic noise the respective function replacing  $\Theta_p$  is  $\tilde{\Theta}(t) := \sqrt{t}\varphi(t)$ ,  $t > 0$ , and it is an easy exercise to see that then  $\varphi(\tilde{\Theta}^{-1}(t)) \leq \varphi(\Theta_p^{-1}(t))$ ,  $t > 0$ . Only for  $p = \infty$ , i.e., for  $K$  of finite trace, do these rates formally coincide.

**4.4. Lower bound: Reduction to regression.** It will be convenient to rewrite (1) (see [10] for details) by using the expansion as in Section 4.2. Precisely, we obtain from (1) the system of equations

$$\tilde{y}_{\delta,k} = \sqrt{s_k}\theta_k + \delta\tilde{\xi}_k, \quad k = 1, 2, \dots,$$

where  $\theta_1, \theta_2, \dots$  are the corresponding Fourier coefficients  $\theta_k := \langle x, u_k \rangle$  and where  $\tilde{\xi}_1, \tilde{\xi}_2, \dots$  are centered. The error criterion is uniform for covariances which obey Assumption 1, and we shall establish a lower bound for statistical noise induced by

$$Kx := C_p \sum_{j=1}^{\infty} j^{-1/q} \langle x, u_j \rangle u_j, \quad x \in X.$$

Because for any projection  $Q$  we can bound  $s_j(QKQ) \leq s_j(K)$ ,  $j \leq \text{rank}(Q)$ , and  $s_j(QKQ) = 0$ ,  $j > \text{rank}(Q)$ , we can bound  $\text{trace}(QKQ) \leq C_p \sum_{j=1}^n j^{-1/q} \leq C_p n^{1/p}$ , such that this defines an admissible covariance operator.

To give this set of equations a final form we will actually consider

$$z_{\delta,k} = \theta_k + \xi_k, \quad k = 1, 2, \dots,$$

which is the standard regression problem under covariance  $K$  with independent (Gaussian) noise and variances  $\sigma_k^2 = \delta^2 \mathbf{E}|\tilde{\xi}_k|^2/s_k$ ,  $k = 1, \dots$ .

This regression problem is only complete after fixing assumptions on the unknown  $\theta := (\theta_1, \theta_2, \dots)$ . In terms of Fourier coefficients the assumption (9) rewrites as

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

This is exactly the setup of the seminal paper [13] by M. S. Pinsker. It will be convenient to recall Pinsker’s results, which aimed at providing the exact asymptotics.

**The Pinsker result** (see [13, Thm. 1]). Let  $\mu$  be the solution to

$$(15) \quad \delta^2 \sum_{j; \varphi(s_j) > \sqrt{\mu}} \frac{\mathbf{E}|\tilde{\xi}_j|^2}{s_j} \frac{1}{\sqrt{\mu}\varphi(s_j)} \left(1 - \frac{\sqrt{\mu}}{\varphi(s_j)}\right) = R^2,$$

and put

$$\nu^2 := \delta^2 \sum_{j; \varphi(s_j) > \sqrt{\mu}} \frac{\mathbf{E}|\tilde{\xi}_j|^2}{s_j} \left(1 - \frac{\sqrt{\mu}}{\varphi(s_j)}\right).$$

There is a universal constant  $c_1 > 0$  for which the error  $e(A_\varphi(R), K, \delta)$  of the best estimator can be bounded from above and below by

$$\nu \geq e(A_\varphi(R), K, \delta) \geq c_1 \nu.$$

*Remark 8.* Under additional assumptions Pinsker is even able to show that

$$\lim_{\delta \rightarrow 0} e(A_\varphi(P), K, \delta) / \nu(\delta) = 1.$$

For our purposes we may argue as follows. Let  $\mu$  be any solution to (15). A simple estimate yields

$$\begin{aligned} R^2 &= \delta^2 \sum_{j; \varphi(s_j) > \sqrt{\mu}} \frac{\mathbf{E}|\tilde{\xi}_j|^2}{s_j} \frac{1}{\sqrt{\mu}\varphi(s_j)} \left(1 - \frac{\sqrt{\mu}}{\varphi(s_j)}\right) \\ &\leq \nu^2 \max_{j; \varphi(s_j) > \sqrt{\mu}} \frac{1}{\sqrt{\mu}\varphi(s_j)} \leq \frac{\nu^2}{\mu}. \end{aligned}$$

This may be rephrased: if  $\mu$  solves (15), then  $e(A_\varphi(R), K, \delta) \geq c_1 R \sqrt{\mu}$ . Therefore, any lower bound for solutions to (15) will provide an estimate for the best possible error from below. This will be used in the proof of Theorem 4.

Before turning to the main lower bound we emphasize that by the asymptoticity assumption (13) on the singular numbers there is a constant  $C_0$  for which

$$\sum_{s_j > \alpha} \frac{j^{-1/q}}{s_j^2} \geq C_0^2 \alpha^{-(4rp+1)/(2rp)}.$$

Recall that  $s_j = s_j(A^*A) \asymp j^{-2r}$  and  $\Theta_p(t) := t^{\frac{2rp+1}{4rp}} \varphi(t)$ ,  $t > 0$ .

**Theorem 4.** Suppose the index function  $\varphi$  is operator monotone on  $(0, b)$ . Let  $\alpha_*$  be chosen as

$$(16) \quad \alpha_* := \inf \left\{ \alpha, \quad \Theta_p(\alpha) \geq \frac{\delta C_0}{\sqrt{TR}} \right\}.$$

Under assumptions (1)–(3) there is  $c_2 > 0$  such that the error of the best estimator for any reconstruction can be bounded from below by

$$e(A_\varphi(R), \mathcal{K}_p, \delta) \geq c_1 R \varphi(\alpha_*) \geq c_2 R \varphi(\Theta_p^{-1}(\delta/R))$$

( $c_1$  is the constant from Pinsker’s result).

*Proof.* Given  $\delta$ , let  $\alpha_*$  be from (16).

We shall show that for  $\bar{\alpha}$ , which is obtained from the solution  $\mu$  from (15) via  $\varphi^2(\bar{\alpha}) = 4\mu$ , necessarily  $\bar{\alpha} \geq \alpha_*$ .

By Lemma 3 we have for any  $0 < s < t \leq a$  that  $\varphi(t)/t \leq T\varphi(s)/s$ . Thus if  $\varphi(s_j) > 2\sqrt{\mu}$ , then  $s_j \geq \bar{\alpha}$  and we conclude that

$$\begin{aligned} R^2 &= \delta^2 \sum_{j; \varphi(s_j) > \sqrt{\mu}} \frac{j^{-1/q}}{s_j} \frac{1}{\sqrt{\mu}\varphi(s_j)} \left(1 - \frac{\sqrt{\mu}}{\varphi(s_j)}\right) \\ &\geq \frac{\delta^2 \bar{\alpha}}{4T\mu} \sum_{j; \varphi(s_j) > 2\sqrt{\mu}} \frac{j^{-1/q}}{s_j^2} \geq \frac{C_0^2 \delta^2 \bar{\alpha}^{-(2rp+1)/(2rp)}}{4T\mu}. \end{aligned}$$

Because  $4\mu = \varphi^2(\bar{\alpha})$  we can finally rewrite this estimate as

$$\Theta_p^2(\bar{\alpha}) = \bar{\alpha}^{(2rp+1)/(2rp)} \varphi^2(\bar{\alpha}) \geq \frac{C_0^2 \delta^2}{TR^2},$$

from which we easily deduce  $\bar{\alpha} \geq \alpha_*$ , completing the proof. □

**Example 1.** Let us briefly discuss the case when smoothness is given by a monomial of an operator  $A$  having polynomial  $\varphi(t) := t^{\nu/2r}$ , which is operator monotone for  $\nu \leq 2r$ . Moreover assume the noise to be Gaussian white noise, i.e., covariance is the identity, hence  $p = 1$ . By Theorems 3 and 4 the error can be estimated by  $e(A_\varphi, I, \delta) \asymp CR\varphi(\Theta_1^{-1}(\delta/R))$ , which amounts to  $e(A_\varphi, I, \delta) \asymp CR(\delta/R)^{\frac{\nu}{\nu+r+1/2}}$ .

### 5. OPTIMALITY OF THE ADAPTIVE PROCEDURE

In this section we will show that up to a logarithmic factor the bound from Theorem 2 provides the optimal rate of reconstruction. First note that Theorem 3 has an important implication on the use of the adaptive procedure. As can be seen from the error bound, the rate depends on the way the regularization parameter controls the discretization level. In fact the relation  $\eta^2(\alpha) \asymp \alpha$  turns out to be appropriate, independent of smoothness assumptions and the nature of the noise. For more details on the control of the discretization accuracy  $\eta$  along the regularization parameter  $\alpha$ , we refer the reader to [9].

If we incorporate this into the adaptive procedure, then we obtain the following result.

**Theorem 5.** *Suppose  $n = n(\alpha) \asymp \lceil \alpha^{-1/2r} \rceil$ ,  $\alpha > 0$ . Under Assumptions 1–4 the following holds true:*

- (1) *There is  $C \geq 1$  for which  $\eta^2 \leq C\alpha$  and  $n(\alpha)$  obeys (8) with  $\bar{D} := q^{1/2r}$ .*
- (2) *There is  $\delta_0 > 0$  such that for  $\delta \leq \delta_0$  the function*

$$\Phi(j) := 2R(M + T)\varphi(C\alpha_j), \quad j = 1, 2, \dots,$$

*is admissible for  $x \in A_\varphi(R)$ .*

- (3) *There is  $C < \infty$  such that for the solution  $x_{\bar{\alpha}, \delta}$  obtained from the adaptive procedure and uniformly for  $K \in \mathcal{K}_p$  we have*

$$e(A_\varphi(R), K, x_{\bar{\alpha}, \delta}, \delta) \leq C \sqrt{\lceil 2 \log_q(1/\delta) \rceil} \varphi(\Theta_p^{-1}(\delta/R)), \quad \delta \leq \delta_0.$$

*Proof.* The first assertion follows from Assumption 4. We turn to the second one. The initial error decomposition (10) and assertion (1) in Theorem 3 yield that for  $\rho(j) := \delta \|(\alpha_j I + B^*B)^{-1} B^* \xi\|$  it holds true that  $\|x - x_{\alpha_j, \delta}\| \leq \Phi(j) + \rho(j)$ .

Furthermore, Lemma 1 provides  $\mathbf{E}\rho^2(j) \leq \Psi^2(j)$ . For  $\delta > 0$  small enough it also holds true that  $\Phi(1) \leq \Psi(1)$ . It remains to establish a concentration bound. To

this end note that  $\rho(j) = \|z_j\|$  and  $z_j$  is centered Gaussian. By [4, p. 59] we can bound uniformly in  $j = 1, 2, \dots, m$

$$P(\rho(j) > t\Psi(j)) \leq P(\|z_j\| > t(\mathbf{E}\|z_j\|^2)^{1/2}) \leq 4\exp(-t^2/8),$$

such that the concentration bound (5) holds true with  $a = 4$  and  $b = 1/8$ . This establishes assertion (2).

We turn to the last assertion. Observe that for  $\alpha_* := \alpha_{j_*}$  with index  $j_*$  from (17) below we can bound

$$\min \{\Phi(j) + \Psi(j), j = 1, 2, \dots\} \leq 2\Psi(\alpha_*).$$

Because  $j \rightarrow 2R(M+T)\varphi(C\alpha_j)$  is admissible for  $\delta \leq \delta_0$  we derive

$$2R(M+T)\varphi(Cq\alpha_*) \geq \Psi(q\alpha_*).$$

Applying Theorem 1 we obtain for  $\delta \leq \delta_0$  the uniform bound at the true solution  $x \in A_\varphi(R)$  in the form

$$\begin{aligned} (\mathbf{E}\|x - x_{\bar{\alpha}, \delta}\|)^{1/2} &\leq 2C(D, q)\sqrt{[2\log_q(1/\delta)]}\Psi(\alpha_*) \\ &\leq 2DC(D, q)\sqrt{[2\log_q(1/\delta)]}\Psi(q\alpha_*) \\ &\leq 4R(M+T)DC(D, q)\sqrt{[2\log_q(1/\delta)]}\varphi(Cq\alpha_*) \\ &\leq 4R(M+T)DC(D, q)Cq\sqrt{[2\log_q(1/\delta)]}\varphi(\alpha_*) \\ &\leq 4R(M+T)DC(D, q)Cq\sqrt{[2\log_q(1/\delta)]}\varphi(\alpha^*), \end{aligned}$$

where we used the fact that  $\alpha_* \leq \alpha^* := \Theta_p^{-1}(\delta/R)$ . The proof is complete.  $\square$

Thus up to constants and a logarithmic factor, and for (unknown) smoothness measured in terms of general source conditions which obey Assumption 2, the adaptive procedure provides the same accuracy as if the smoothness were known. This logarithmic factor can be considered as payment for the lack of knowledge of  $\varphi$  under the stochastic nature of noise. It is a common belief that this payment is necessary under Gaussian white noise. In the case of deterministic noise, which formally corresponds to Assumption 1 with  $p = \infty$ , the additional logarithmic factor does not appear; see [10].

We mention that additional information about the underlying ill-posed problem allows us to further reduce such a payment. For example, severely ill-posed problems are characterized by index functions  $\varphi$  such that  $\varphi(t) \geq \ln^{-d} 1/t$ , where  $d > 0$  is known. Then it is easy to see that choosing  $\kappa = \sqrt{C \ln \log_q 1/\delta}$ ,  $C > (4d+2)/b$  in (6), one can obtain an analog of Theorem 5 for severely ill-posed problems with an additional factor of order  $\sqrt{\ln \log_q 1/\delta}$  instead of  $\sqrt{\log_q 1/\delta}$ .

*Remark 9.* In practice the true solution might be smoother than expected. In this case Assumption 2 might be violated. If instead the index function for the true source condition satisfies a Lipschitz property for nonnegative operators  $B_1$  and  $B_2$ , say

$$\|\varphi(B_1) - \varphi(B_2)\| \leq M\|B_1 - B_2\|,$$

then Tikhonov regularization will still work. However there will be saturation. In the present framework it can be seen that for  $\delta \leq \delta_0$  the function

$$\Phi(j) := 2R(M + T)C\alpha_j, \quad j = 1, 2, \dots,$$

is admissible for the true solution  $x$  if  $\delta \leq \delta_0$ . Therefore, if the remaining assumptions hold true, then the adaptive procedure will provide the rate  $\delta^{(4rp)/(6rp+1)}$ , which is worse than the saturation rate  $\delta^{2/3}$  for Tikhonov regularization under bounded deterministic noise. However, it will be attained for operators with singular numbers decreasing exponentially fast.

#### APPENDIX: AN ABSTRACT ORACLE PRINCIPLE

In this Appendix we shall prove Theorem 1. To this end we gather some useful results concerning Lepskii's adaptation procedure. Recall that within the present general framework the variance term could be quite arbitrary as long as it was decreasing with the regularization parameter.

For the statistical context the basic ingredients were already introduced in Section 3. However, the proof of the statistical estimate uses a deterministic one, therefore we shall start with the deterministic oracle principle.

Let  $\{x_1, \dots, x_m\}$  be a finite and deterministic set of elements in the metric space  $(M, d)$  and let  $\Psi: \{1, \dots, m\} \rightarrow \mathbb{R}^+$  be a decreasing function. Then we may take  $\rho(j) \equiv \Psi(j)$  in Definition 1, and a nondecreasing function  $\Phi: \{1, \dots, m\} \rightarrow \mathbb{R}^+$  is admissible for  $x$ , if and only if

$$d(x, x_j) \leq \Phi(j) + \Psi(j), \quad j = 1, 2, \dots,$$

and  $\Phi(1) \leq \Psi(1)$ .

**Lemma 4** (The Lepskii principle). *Let  $\{x_1, \dots, x_m\}$  be a finite set of elements in a metric space  $(M, d)$  and let  $\Psi: \{1, \dots, m\} \rightarrow \mathbb{R}^+$  be a decreasing function. Define*

$$(17) \quad j_* := \max \{j, \text{ there is admiss. } \Phi \text{ for which } \Phi(j) \leq \Psi(j)\}.$$

*For  $\bar{j} := \max \{j \leq m, \quad d(x_i, x_j) \leq 4\Psi(i), \text{ for all } i \leq j\}$ , we have*

$$d(x, x_{\bar{j}}) \leq 6\Psi(j_*).$$

*Proof.* We first show that  $j_* \leq \bar{j}$ . If  $\Phi$  is admissible with  $\Phi(j_*) \leq \Psi(j_*)$ , then for  $i < j_*$  it holds true that

$$\Phi(i) \leq \Phi(j_*) \leq \Psi(j_*) \leq \Psi(i),$$

such that for such admissible  $\Phi$  we obtain

$$\begin{aligned} d(x_i, x_{j_*}) &\leq d(x, x_i) + d(x, x_{j_*}) \leq \Phi(i) + \Psi(i) + \Phi(j_*) + \Psi(j_*) \\ &\leq 2\Psi(i) + 2\Psi(j_*) \leq 4\Psi(i), \end{aligned}$$

hence  $j_* \leq \bar{j}$ . Using this the proof can be completed as follows:

$$d(x, x_{\bar{j}}) \leq d(x, x_{j_*}) + d(x_{j_*}, x_{\bar{j}}) \leq \Phi(j_*) + \Psi(j_*) + 4\Psi(j_*) \leq 6\Psi(j_*). \quad \square$$

**Corollary 1** (Deterministic oracle inequality). *Under the assumptions of Lemma 4 the following assertion is true. If  $D < \infty$  is such that the function  $\Psi$  in addition obeys  $\Psi(i) \leq D\Psi(i+1)$ , then*

$$d(x, x_{\bar{j}}) \leq 6D \min \{\Phi(j) + \Psi(j), \quad j = 1, \dots, m, \Phi \text{ admiss.}\}.$$

*Proof.* Let  $\Phi$  be any admissible function and let us temporarily introduce the following index:

$$\underline{j} := \arg \min \{ \Phi(j) + \Psi(j), \quad j = 1, 2, \dots, m \}.$$

The assertion of the corollary is an immediate consequence of

$$\Psi(j_*) \leq D(\Phi(\underline{j}) + \Psi(\underline{j})).$$

But this is evident, because either  $j_* \geq \underline{j}$ , in which case  $\Psi(j_*) \leq \Psi(\underline{j}) \leq \Phi(\underline{j}) + \Psi(\underline{j})$ , or  $j_* < j_* + 1 \leq \underline{j}$ . But then, by the definition of  $j_*$  it holds true that  $\Psi(j_* + 1) \leq \Phi(j_* + 1)$ , thus  $\Psi(j_*) \leq D\Psi(j_* + 1) \leq D\Phi(j_* + 1) \leq D\Phi(\underline{j})$ , from which the estimate follows.  $\square$

We turn to the proof of Theorem 1, the main estimate in the statistical context.

*Proof of Theorem 1.* Let  $\Phi$  be any admissible function and introduce the random number

$$\Pi(\omega) := \max_{j=1, \dots, m} \frac{\rho(j)}{\Psi(j)}.$$

For  $\kappa$  from above we let  $\Omega_\kappa := \{ \omega, \quad \Pi(\omega) \leq \kappa \}$ . For  $\omega \in \Omega_\kappa$  the assumptions of Corollary 1 are fulfilled, such that for such realizations it holds true that

$$d(x, x_{\bar{j}}) \leq 6D \min \{ \Phi(j) + \kappa\Psi(j), \quad j = 1, 2, \dots, m \}.$$

On the complement  $\Omega_\kappa^c$  we have  $\Pi(\omega) > \kappa$ , such that we can bound

$$\Phi(1) \leq \kappa\Psi(1) \leq \Psi(1)\Pi(\omega).$$

Using this estimate we conclude for  $\omega \in \Omega_\kappa^c$  that

$$\begin{aligned} d(x, x_{\bar{j}}) &\leq d(x, x_1) + d(x_1, x_{\bar{j}}) \leq \Phi(1) + \rho(1) + 4\kappa\Psi(1) \\ &\leq 5\Psi(1)\Pi(\omega) + \frac{\rho(1)}{\Psi(1)}\Psi(1) \leq 6\Psi(1)\Pi(\omega). \end{aligned}$$

Therefore estimate (6) will follow from

$$(18) \quad \int_{\Omega_\kappa^c} \Pi^2(\omega) P(d\omega) \leq 5 \frac{a}{b^2} m \exp(-b/2\kappa^2).$$

Indeed, we start with

$$\int_{\Omega_\kappa^c} \Pi^2(\omega) P(d\omega) \leq \left( \int \Pi^4(\omega) P(d\omega) \right)^{1/2} (P(\Omega_\kappa^c))^{1/2}.$$

The second term can easily be estimated as

$$P(\Pi(\omega) > \kappa) \leq m \max_{j=1, 2, \dots, m} P(\rho(j) \geq \kappa\Psi(j)) \leq am \exp(-b\kappa^2).$$

In the same spirit and making use of the Leibniz formula

$$\int |f(\omega)|^4 P(d\omega) = 4 \int t^3 P(|f| > t) dt,$$

we deduce that

$$\int \Pi^4(\omega) P(d\omega) \leq 24 \frac{a}{b^4} m,$$

resulting in (18).  $\square$



## REFERENCES

- [1] Rajendra Bhatia. *Matrix analysis*. Springer-Verlag, New York, 1997. MR1477662 (98i:15003)
- [2] Alexander Goldenshluger and Sergei V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields*, 118(2):169–186, 2000. MR1790080 (2001h:62055)
- [3] Markus Hegland. Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization. *Appl. Anal.*, 59(1-4):207–223, 1995. MR1378036 (97a:65060)
- [4] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes. MR1102015 (93c:60001)
- [5] O. V. Lepskii. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990. MR1091202 (93j:62212)
- [6] Bernard A. Mair and Frits H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996. MR1409127 (97k:62095)
- [7] 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. MR1856240 (2002g:62063)
- [8] Peter Mathé and Sergei V. Pereverzev. Moduli of continuity for operator valued functions. *Numer. Funct. Anal. Optim.*, 23(5-6):623–631, 2002. MR1923828 (2003g:47029)
- [9] Peter Mathé and Sergei V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(6):1263–1277, 2003. MR2036530 (2004k:65097)
- [10] Peter Mathé and Sergei V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003. MR1984890 (2004i:47021)
- [11] M. T. Nair, E. Scheck, and U. Tautenhahn. Morozov’s discrepancy principle under general source conditions. *Z. Anal. Anwendungen*, 22(1):199–214, 2003. MR1962084 (2004a:65069)
- [12] A. Pietsch. *Eigenvalues and s-Numbers*, volume 43 of *Math. und ihre Anw. in Phys. und Technik*. Geest & Portig, Leipzig, 1987. MR0917067 (88j:47022a)
- [13] M. S. Pinsker. Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission*, 16(2):52–68, 1980. MR0624591 (82j:93048)
- [14] Ulrich Tautenhahn. Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.*, 33(6):2120–2130, 1996. MR1427456 (97k:65148)
- [15] Alexandre Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(9):835–840, 2000. MR1769957 (2001c:62058)
- [16] N. N. Vakhania, V. I. Tarieladze, and S. A. Chobanjan. *Probability Distributions on Banach Spaces*. D. Reidel, Dordrecht, Boston, Lancaster, Tokyo, 1987. MR1435288 (97k:60007)
- [17] Curtis R. Vogel. *Computational methods for inverse problems*, volume 23 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. MR1928831 (2003i:65004)

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTRASSE 39, D-10117 BERLIN, GERMANY

*E-mail address:* `mathe@wias-berlin.de`

JOHANN-RADON-INSTITUTE (RICAM), ALTENBERGER STRASSE 69, A-4040 LINZ, AUSTRIA

*E-mail address:* `sergei.pereverzyev@oeaw.ac.at`