Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Optimal discretization and Degrees of ill-posedness for inverse estimation in Hilbert scales in the presence of random noise

Peter Mathé [1] Sergei V. Pereverzev [2]

submitted: 19 Januar 1999

[1] Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D–10117 Berlin, Germany
E-mail: mathe@wias-berlin.de

[2] Ukrainian Academy of Sciences, Inst. of Mathematics, Tereshenkivska Str. 3, Kiev 4, Ukraine
E-mail: serg-p@mail.kar.net

Preprint No. 469
Berlin 1999

1991 Mathematics Subject Classification. 62G05, 65J10.

Key words and phrases. Ill-posed problems, inverse estimation, operator equations, Gaussian noise, optimal difficulty, regularized inverse estimator, histogram estimator.

The work of S. V. Pereverzev was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG).

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

Fax:	+ 49 30 2044975
E-Mail (X.400):	c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
E-Mail (Internet):	preprint@wias-berlin.de
World Wide Web:	http://www.wias-berlin.de/

ABSTRACT. The problem of minimizing the difficulty of the inverse estimation of some unknown element x_0 from noisy observations $y_\delta = Ax_0 + \delta\xi$ in dependence of the nature of the random noise ξ is considered. It is shown that a combination of a Tikhonov regularization estimator with a certain projection scheme is order optimal in the sense of difficulty for a wide class of operators A acting along Hilbert scales.

1. INTRODUCTION

Suppose we wish to recover an element x of some Hilbert space X, but we are only able to observe data near $y = Ax$, where A is a compact linear operator from X to X. Such linear inverse problems arise in scientific settings, ranging from stereological microscopy (Abel's integral equation) and physical chemistry (Fujita's equation) to satellite geodesy (gravity gradiometry equation).

Moreover, we assume that the data are noisy, so that we observe y_δ given by

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

where ξ is some stochastic process and δ is a small positive number, used for measuring the noise level. Operator equation (1.1) with random noise is an example of a statistical ill-posed inverse problem. Typically A is an integral operator of the form

$$Ax(t) = \int_0^1 a(t,\tau)x(\tau)d\tau, \tag{1.2}$$

acting from $X = L_2(0,1)$ to $L_2(0,1)$ and $x(t)$ is the probability density function of some random variable that we cannot observe directly. The statistical problem we address is to estimate $x(t)$ from noisy measurements (1.1). Such noisy integral equations (1.1), (1.2) are considered throughout Wahba's work, see, e.g. Wahba (1977). We further mention Nychka and Cox (1989), Johnstone and Silverman (1991), Donoho (1995), Mair and Ruymgaart (1996), Lukas (1998). For direct density and regression estimation, when A is the identity operator I, the reader is referred to Nussbaum (1985), Speckman (1985), Donoho and Johnstone (1991), Kerkyacharian and Picard (1992).

Suppose further that even the observations (1.1) cannot be observed exactly but they can only be observed in discretized or binned form. To be precise, assume that instead of (1.1) we have only vector $\varphi(y_\delta) = \{y_{\delta,i} = (\varphi_i, y_\delta)\}_{i=1}^n$ defined by

$$y_{\delta,i} = (\varphi_i, y_\delta) = (y, \varphi_i) + \delta\xi_i = (Ax, \varphi_i) + \delta(\xi, \varphi_i), \tag{1.3}$$

where $(\cdot,\cdot)$ is the inner product in X and $\varphi = \{\varphi_i \in X,\ i = 1, 2, \ldots, n\}$ is the so-called design of the statistical experiment consisting in obtaining the values (1.3). In the sequel we denote by Φ_n the set of all designs φ determined by collections of at most n elements $\varphi_i \in X$. If we have the possibility to choose the design φ, then typically elements of the singular value decomposition or wavelet-vaguelette decomposition of the operator A play the role of φ_i. The methods for regularized solution of integral equations (1.1), (1.2) from discrete noisy data (1.3) for such

φ_i were recently studied by Johnstone, Silverman (1991) and Donoho (1995) and Golubev, Khasiminskii (1997). But very often the design φ is fixed beforehand and does not depend on the operator A. A simple example is that of estimating a continuous probability density function $x(t)$ from binned data or histogram for $y_\delta(t)$. In the case of noisy integral equation (1.1), (1.2) we assume that the interval $[0,1]$ is partitioned into histogram bins $[u_{i-1,n}, u_{i,n})$ with bin limits $0 = u_{0,n} < u_{1,n} < \ldots < u_{n,n} = 1$ and instead of (1.1), (1.2) we have the vector with components

$$y_{\delta,i} = \frac{1}{u_{i,n} - u_{i-1,n}} \int\limits_{u_{i-1,n}}^{u_{i,n}} Ax(t)dt + \delta\xi_i, \quad i = 1,2,\ldots,n, \tag{1.4}$$

having the form (1.3) for $\varphi_i(t) = (u_{i,n} - u_{i-1,n})^{-1}\chi_{i,n}(t)$, where $\chi_{i,n}(t)$ is the characteristic function of $[u_{i-1,n}, u_{i,n})$. The approximate solution of Abel's integral equation based on histograms (1.4) was considered by Nychka and Cox (1989). It is easy to see that fixed designs of normalized characteristic functions $\varphi_i(t) = (u_{i,n} - u_{i-1,n})^{-1}\chi_{i,n}(t)$ of histogram bins are not generated by the wavelet-vaguelette decomposition of Abel's integral operator. Therefore, the powerful scheme proposed by Donoho (1995) does not apply. It is the purpose of this paper to provide a common background for a variety of discretized observations (1.3). For this reason we will only need that the design $\varphi = \{\varphi_i \in X,\ i = 1,2,\ldots,n\}$ has good approximation properties, but does not necessarily depend on the operator A. Note, that for $\varphi \in \Phi_n$ the number n defines the amount of discrete information used for recovering the unknown solution x_0 of the equation $Ax = y$.

On the other hand, as was indicated by Johnstone and Silverman (1991), there is a substantial statistical literature which is concerned with such questions; how many bins to use with a histogram estimator and a given noise intensity δ. If we concentrate on the case of linear estimators of x_0 from discretized observations of more general form (1.3) and it is a priori known that the unknown solution belongs to some set $\mathfrak{M} \subset X$ then the answer is connected with the behavior of the quantity

$$\Delta_{n,\delta}(A,\mathfrak{M},\xi)^2 := \inf_{\varphi\in\Phi_n} \inf_{S\in\mathcal{L}_n(X)} \sup_{x\in\mathfrak{M}} \mathbf{E}\|x - S\circ\varphi(Ax+\delta\xi)\|^2, \tag{1.5}$$

where $\mathcal{L}_n(X)$ denotes the set of all linear mapping from $\mathbb{R}^n$ to X and $\mathbf{E}$ denotes the expectation with respect to the noise ξ. Below we indicate designs and estimators which are order optimal in the sense of (1.5) simultaneously for all operators A from some sufficiently wide class. We note that a quantity, related to (1.5) was considered earlier by Donoho et al. (1990) in the specific case when $A = I$ and $S\circ\varphi$ is the orthogonal projector on span $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$.

The recovery of the unknown solution x from indirect measurements (1.3), blurred by random noise is usually studied under the assumption that prior knowledge regarding the smoothness of the solution is available. For greater flexibility we embed the general problem (1.1) into an abstract Hilbert scale. Regularization of ill-posed problems in Hilbert scales was introduced by Natterer (1984). Statistical inverse estimation in Hilbert scales has first been studied by Mair and Ruymgaart (1996).

But in their asymptotic consideration these authors did not consider the case of noisy discretized observations (1.3).

2. Setup

To be precise, a Hilbert scale $\{X_s\}_{s\in\mathbb{R}}$ is a family of Hilbert spaces X_s with inner product $(x,y)_s := (L^s x, L^s y)$, where L is some unbounded self-adjoint strictly positive operator in a dense domain of X, and X_s is defined as the completion of the intersection of domains of all operators $\left\{L^k\right\}_{k\in\mathbb{R}}$, endowed with the norm $\|x\|_s := (x,x)_s^{1/2}$. A Hilbert scale satisfies the following interpolation inequalities

$$\|x\|_0 \le \|x\|_{-a}^{\frac{s}{a+s}} \|x\|_s^{\frac{a}{a+s}}, \quad x \in X_s, \tag{2.1}$$

where $\|\cdot\|_0 = \|\cdot\|$ is the norm in the initial Hilbert space X. Moreover, we assume that the Hilbert scale $\{X_s\}$ is scaled as the Sobolev scale $W_2^s(0,1)$. Namely, if $J_s : X_s \to X$ is the canonical embedding, then we assume that

$$\mathrm{a}_n(J_s) := \inf\{\|J_s - U\|_{X_s\to X},\ \operatorname{rank} U < n\} \asymp n^{-s}, \tag{2.2}$$

where a_n is n-th approximation number (see Pietsch (1978)) and " $\asymp$ " means equivalent in the sense of order.

We turn to properties required of the operator A. We assume that A acts along the Hilbert scale in the following way: For some parameter $a > 0$ there exist constants $d, D > 0$ such that

$$d\|x\|_{\nu-a} \le \|Ax\|_\nu \le D\|x\|_{\nu-a} \tag{2.3}$$

holds for all $x \in X_{\nu-a}$ and $\nu \in \mathbb{R}$. The parameter a in (2.3) can be interpreted as a "degree of ill-posedness" of equations involving A, analytical in nature.

Moreover, in the sequel we will assume, that the exact solution x_0 of the equation $Ax = y$ belongs to some fixed ball

$$X_\mu^R := \{x, \quad \|x_0\|_\mu \le R\} \tag{2.4}$$

for some X_μ.

Remark 2.1. We illustrate assumption (2.3) by introducing Symm's equation

$$\int_\Gamma \log(|u-v|) z(v) dS_v = g(u), \quad u \in \Gamma, \tag{2.5}$$

arising from the Dirichlet boundary value problem for the Laplace equation in some region with boundary curve Γ. Assuming that Γ admits a C^∞–smooth 1–periodic parameterization γ: $[0,1] \to \Gamma$ we can rewrite (2.5) as

$$(Ax)(t) := \int_0^1 \log(|\gamma(t) - \gamma(\tau)|) x(\tau) d\tau = y(t),$$

where $x(t) := z(\gamma(t))\,|\gamma'(t)|$ and $y(t) := g(\gamma(t))$. It can be seen that the operator A, just defined obeys condition (2.3) with $a = 1$ within the scale $X_s := \tilde{W}_2^s(0,1)$, $s \in \mathbb{R}$, of Sobolev spaces of 1–periodic functions, see e.g. Bruckner et al. (1995) for details. We add that more generally, given an operator A that does not fit some standard

Hilbert scale, one can often construct a scale, adapted to this operator. This is the case when $A: X \to X$ acts compactly and injectively in some Hilbert space X. Then A meets condition (2.3) with $a = 1/2$ in the scale generated by $L: = (A^*A)^{-1}$, see Natterer (1984) and Hegland (1995) for further details.

We turn to assumptions made for the noise. The first model of random noise was initially considered by Bakushinskii (1969). Here ξ is supposed to be a centered X-valued random vector defined on some probability space $(\Omega, \Sigma, \mathbb{P})$ with bounded variance, i.e.,

$$\mathbf{E}(\xi) = 0, \qquad \mathbf{E}\|\xi\|_0^2 \leq 1. \tag{2.6}$$

The second model is connected with Gaussian white noise. Here ξ is a weak or generalized random element, such that for any $f \in X$ the inner product (f, ξ) is a measurable function, mapping a probability space $(\Omega, \Sigma, \mathbb{P})$ into $\mathbb{R}$ equipped with its Borel σ-field. Moreover, for any $f, g \in X$ the functions $(f, \xi), (g, \xi)$ are square-summable with respect to the probability measure $\mathbb{P}$ and

$$\mathbf{E}(f, \xi) = \mathbf{E}(g, \xi) = 0, \quad \mathbf{E}(f, \xi)(g, \xi) = (f, g). \tag{2.7}$$

We extend this to a parameterized family of noise by introducing ξ^β, where ξ^β is such that for some constant c_β and for some orthonormal basis $\{u_k\}$ of X we have $\xi_k = c_\beta k^\beta (u_k, \xi^\beta)$, $k = 1, 2, \ldots$, are i.i.d $N(0,1)$. Note that for $\beta = 0$ and $c_0 = 1$ ξ^0 is Gaussian white noise with properties (2.7). On the other hand, for arbitrary small $\varepsilon > 0$ the noise $\xi^{\frac{1}{2}+\varepsilon}$ is an X-valued random function satisfying (2.6) for appropriate c_β.

We note that noise introduces an additional degree of ill-posedness, this time of statistical nature.

The notion of "degree of ill-posedness" has been coined by Wahba (1977) to quantify the interplay between "nastiness" of operator A and "dimensionality" of the regularizing set X_μ^R. On the other hand, we can expect additional influence of the noise ξ on the degree of ill-posedness. This influence has been observed by Nussbaum (1994) for the special case when the operator A is a-fold integration and the Hilbert scale consists of Sobolev spaces, i.e. $X_s = W_2^s(0,1)$. Namely, for deterministic noise $\xi \in L_2(0,1)$, $\|\xi\| \leq 1$, the optimal order for recovering $x_0 \in W_2^\mu(0,1)$ from noisy data (1.3) is $\delta^{\frac{\mu}{\mu+a}}$. On the other hand, it might be interesting to note that in the example thus mentioned, but for Gaussian white noise $\xi(t)$ the optimal (minimax) rate of convergence for recovering $x_0 \in W_2^\mu(0,1)$ is $\delta^{\frac{\mu}{\mu+a+1/2}}$, where the error criterion is modified appropriately.

In the present paper we obtain the same rates of convergence for the general case of equations (1.1) with operators satisfying (2.3) and for different type of random noise which cover deterministic and Gaussian white noise.

3. Discretized noisy observations as Galerkin data for regularization in Hilbert scales

Below we derive an order optimal numerical scheme to solve problem (1.1).

3.1. **Description of the method.** For the sake of simplicity we assume that the $\{\varphi_1, \varphi_2, \dots, \varphi_n\}$ are orthonormal. Then

$$Q_n f := \sum_{i=1}^{n} (f, \varphi_i)\varphi_i$$

denotes the orthogonal projector onto $\operatorname{span}\{\varphi_1, \varphi_2, \dots, \varphi_n\}$. Using Q_n we can rewrite (1.3) as

$$Q_n y_\delta = Q_n(Ax + \delta\xi). \tag{3.1}$$

This is the standard form of the Galerkin projection scheme for the approximate solution of operator equation (1.1). But if (1.1) is ill-posed, regularization techniques are required for solving (3.1). The most widely used method for regularization in Hilbert scales is the Tikhonov method. In statistics this method is called regularization estimator. Statistical justification for such estimator has been given by Li (1982) and Speckman (1985). Tikhonov regularization of the Galerkin method for the approximate solution of (3.1), and hence of (1.1) is obtained by minimizing

$$\|Q_n Ax - Q_n y_\delta\|_0^2 + \alpha\|x\|_s^2 \tag{3.2}$$

over some finite-dimensional subspace V_m of X_s, where we assume that the true solution $x_0 \in X_\mu \subset X_s$ for $\mu \geq s$.

It follows from Neubauer (1988) that the unique minimizer $x^\delta_{\alpha,n,m}$ of (3.2) has the form

$$x^\delta_{\alpha,n,m} = x^\delta_{\alpha,n,m}(\xi) = G_{\alpha,n,m} y_\delta, \tag{3.3}$$

where

$$G_{\alpha,n,m} = (T^\#_{n,m} T_{n,m} + \alpha I)^{-1} T^\#_{n,m} = L^{-s}(B^*_{n,m} B_{n,m} + \alpha I)^{-1} B^*_{n,m},$$

$$T_{n,m} = Q_n A P_{m,s}, \quad B_{n,m} = Q_n A P_{m,s} L^{-s},$$

$P_{m,s}$ is the orthogonal projector form X_s onto V_m and $K^*, K^\#$ denote the adjoint operators of K with respect to the corresponding inner products $(\cdot,\cdot)$ and $(\cdot,\cdot)_s$. In particular

$$T^\#_{n,m} = P_{m,s} L^{-2s} A^* Q_n, \quad B^*_{n,m} = L^s P_{m,s} L^{-2s} A^* Q_n.$$

Note that representation (3.3) is needed only for the analysis of the rate of convergence. The construction of $x^\delta_{\alpha,n,m}$ actually reduces to solving a system of $\min\{m,n\}$ linear algebraic equations and is completely determined by the choice of design $\varphi \in \Phi_n$, parameter s, finite dimensional subspace $V_m \subset X_s$ and finally by choosing the regularization parameter α in (3.2).

To estimate the performance of the approximating $x^\delta_{\alpha,n,m}$ additional properties of the design $\varphi \in \Phi_n$ as well as of the choice of spaces V_m are required.

To be precise, we assume that for all $x \in X_{a+\nu}$

$$\inf_{g \in \mathrm{span}\{\varphi_1, \varphi_2, \dots, \varphi_n\}} \|x - g\|_0 \le c n^{-(a+\nu)} \|x\|_{a+\nu}, \quad \nu \le s, \tag{3.4}$$

where c is some positive constant (for simplicity we often use the same symbol c for possibly different constants). Note, that from (2.2) we deduce that the best possible order of approximating elements from $X_{a+\nu}$ in X_0 by linear combinations of at most n design elements is $n^{-(a+\nu)}$. Thus we assume in (3.4) that this order is achieved by the chosen design $\varphi \in \Phi_n$.

Moreover, as in Neubauer (1988) we require, that the finite-dimensional space V_m obeys

$$\inf_{g \in V_m} \|x - g\|_s \le c m^{-(a+s)} \|x\|_{a+2s}, \quad x \in X_{a+2s}. \tag{3.5}$$

Conditions (3.4) and (3.5) can be written in the form

$$\|I - Q_n\|_{X_{a+\nu} \to X} \le c\, n^{-(a+\nu)}, \quad \nu \le s, \tag{3.6}$$

and

$$\|I - P_{m,s}\|_{X_{a+2s} \to X_s} \le c\, m^{-(a+s)}. \tag{3.7}$$

If (3.7) is fulfilled and $s \ge (\mu - a)/2$ then standard interpolation techniques, we refer to Babuška and Aziz (1972) for details, yield inequality

$$\|(I - P_{m,s})x_0\|_s \le c\, m^{-(\mu - s)} \tag{3.8}$$

whenever $x_0 \in X_\mu^R$, which will be useful below.

The performance of the Tikhonov regularization estimator $x^\delta_{\alpha,n,m}$ will be based on the risk $\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|^2$. Making use of (2.6) or (2.7), respectively, this can be rewritten

$$\begin{aligned} \mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|^2 &= \mathbf{E}\|(G_{\alpha,n,m}A - I)x_0 + \delta G_{\alpha,n,m}\xi\|^2 \\ &= \mathbf{E}\|(G_{\alpha,n,m}A - I)x_0\|^2 \\ &\quad - 2\delta \mathbf{E}(G^*_{\alpha,n,m}(G_{\alpha,n,m}A - I)x_0, \xi) \\ &\quad + \delta^2 \mathbf{E}\|G_{\alpha,n,m}\xi\|^2 \\ &= \|(G_{\alpha,n,m}A - I)x_0\|^2 + \delta^2 \mathbf{E}\|G_{\alpha,n,m}\xi\|^2. \end{aligned} \tag{3.9}$$

In the sequel the terms

$$b^2_{\alpha,n,m}(x_0) := \|(G_{\alpha,n,m}A - I)x_0\|^2, \quad v_{\alpha,n,m}(\xi) := \delta^2 \mathbf{E}\|G_{\alpha,n,m}\xi\|^2$$

will be considered as bias and variance of the risk, respectively. Now we turn to estimate these quantities separately.

3.2. **Estimate of the bias.** The bias is bounded from above in

Lemma 3.1. *Let the assumptions* (2.3), (2.4), (3.7), (3.6) *be fulfilled. Assume that for some* $\varkappa < 1$

$$\|I - Q_n\|_{X_a \to X} \le \varkappa\, D^{-1}(2^a d^s)^{\frac{1}{a+s}} \alpha^{\frac{a}{2(a+s)}}, \tag{3.10}$$

where D, d *are the constants from* (2.3). *Then*

$$b_{\alpha,n,m}(x_0) \le c\Big[\alpha^{\frac{\mu}{2(a+s)}} + m^{-s}\Big(1 + m^{-a}\alpha^{-\frac{a}{2(a+s)}}\Big)\Big(\alpha^{\frac{\mu-s}{2(a+s)}} + \|(I - P_{m,s})x_0\|_s\Big)\Big].$$

Proof. Let $y_0 = Ax_0$ be the true free term of our equation. Consider the elements $x^0_{\alpha,n,m} = G_{\alpha,n,m} y_0$ and

$$x^0_{\alpha,m} = (T^\#_m T_m + \alpha I)^{-1} T^\#_m y_0,$$

where $T_m = AP_{m,s}$, $T^\#_m = P_{m,s} L^{-2s} A^*$. It follows from Lemma 2.2, Lemma 3.2 by Neubauer (1988) that

$$\|(T^\#_m T_m + \alpha I)^{-1} T^\#_m\|_{X \to X} \le (2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}}, \tag{3.11}$$

and

$$\begin{aligned} &\|x_0 - x^0_{\alpha,m}\| \\ &\le c\Big[\alpha^{\frac{\mu}{2(a+s)}} + m^{-s}\Big(1 + m^{-a}\alpha^{-\frac{a}{2(a+s)}}\Big)\Big(\alpha^{\frac{\mu-s}{2(a+s)}} + \|(I - P_{m,s})x_0\|_s\Big)\Big]. \end{aligned} \tag{3.12}$$

Note that

$$b_{\alpha,n,m}(x_0) = \|x_0 - x^0_{\alpha,n,m}\| \le \|x_0 - x^0_{\alpha,m}\| + \|x^0_{\alpha,m} - x^0_{\alpha,n,m}\|. \tag{3.13}$$

Moreover, from (3.10), (3.11) and (2.3) it follows that

$$\begin{aligned} &\|x^0_{\alpha,m} - x^0_{\alpha,n,m}\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1}[(T^\#_m - T^\#_{n,m}) y_0 - (T^\#_m T_m - T^\#_{n,m} T_{n,m}) x^0_{\alpha,n,m}]\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1} T^\#_m [(I - Q_n) y_0 - (I - Q_n) A P_{m,s} x^0_{\alpha,n,m}]\| \\ &\quad = \|(T^\#_m T_m + \alpha I)^{-1} T^\#_m (I - Q_n) A (x_0 - x^0_{\alpha,n,m})\| \\ &\quad \le (2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}} \|I - Q_n\|_{X_a \to X_0} \|A(x_0 - x^0_{\alpha,n,m})\|_a \\ &\quad \le D(2^a d^s)^{-\frac{1}{a+s}} \alpha^{-\frac{a}{2(a+s)}} \|I - Q_n\|_{X_a \to X_0} b_{\alpha,n,m}(x_0) \\ &\quad \le \varkappa b_{\alpha,n,m}(x_0). \end{aligned} \tag{3.14}$$

Combining (3.13) and (3.14) we obtain

$$b_{\alpha,n,m}(x_0) \le (1 - \varkappa)^{-1} \|x_0 - x^0_{\alpha,m}\|. \tag{3.15}$$

The assertion of the lemma follows from (3.12) and (3.15). □

3.3. **Estimate of the variance.** We turn to bounding the variance. The basic step towards this goal is given by

Lemma 3.2. *Let the assumptions* (2.3), (2.4), (3.7), (3.6) *be fulfilled. If* $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$ *then*

$$\|G_{\alpha,n,m}\|_{X\to X} \le c\alpha^{-\frac{a}{2(a+s)}},$$

where c *does not depend on* α, n, m.

Proof. It is well known that for an arbitrary compact operator B from X to X

$$\|(B^*B+\alpha I)^{-1}B^*\|_{X\to X} \le \frac{1}{2\sqrt{\alpha}}, \quad \|B(B^*B+\alpha I)^{-1}B^*\|_{X\to X} \le 1. \tag{3.16}$$

In particular we have for any $f \in X$ the bound

$$\begin{aligned} \|G_{\alpha,n,m}f\|_s &= \|L^s G_{\alpha,n,m} f\| \\ &= \|(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \le \frac{1}{2\sqrt{\alpha}}\|f\|. \end{aligned} \tag{3.17}$$

Moreover, from (2.3) and (3.16) it follows that

$$\begin{aligned} \|G_{\alpha,n,m}f\|_{-a} &\le d^{-1}\|AG_{\alpha,n,m}f\| \\ &= d^{-1}\|AL^{-s}(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\ &\le d^{-1}\|B_{n,m}(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\ &+ d^{-1}\|(AL^{-s}-B_{n,m})(B^*_{n,m}B_{n,m}+\alpha I)^{-1}B^*_{n,m}f\| \\ &\le d^{-1}\Big(1+\frac{1}{2\sqrt{\alpha}}\|AL^{-s}-B_{n,m}\|_{X\to X}\Big)\|f\|. \end{aligned} \tag{3.18}$$

Now we derive an estimate for $\|AL^{-s}-B_{n,m}\|$:

$$\begin{aligned} \|AL^{-s}-B_{n,m}\|_{X\to X} &\le \|AL^{-s}-AP_{m,s}L^{-s}\|_{X\to X} \\ &+ \|AP_{m,s}L^{-s}-Q_nAP_{m,s}L^{-s}\|_{X\to X}. \end{aligned} \tag{3.19}$$

By (2.3) and (3.7) we can continue

$$\begin{aligned} \|AL^{-s}-AP_{m,s}L^{-s}\|_{X\to X} &\asymp \|(I-P_{m,s})L^{-s}\|_{X\to X_{-a}} \\ &= \|L^{-a}(I-P_{m,s})L^{-s}\|_{X\to X} \\ &= \|L^sL^{-s-a}(I-P_{m,s})L^{-s}\|_{X\to X} \\ &= \|L^{-s-a}(I-P_{m,s})\|_{X_s\to X_s} \\ &= \|(I-P_{m,s})L^{-s-a}\|_{X_s\to X_s} \\ &= \|(I-P_{m,s})\|_{X_{2s+a}\to X_s} \\ &\le c\, m^{-(s+a)}. \end{aligned} \tag{3.20}$$

(Note that $L^{-\nu}: X_t \to X_t$ is self-adjoint for $\nu \ge t$). Further, using (2.3) and (3.6) we find

$$\begin{aligned} \|AP_{m,s}L^{-s} - Q_nAP_{m,s}L^{-s}\|_{X\to X} &\le \|I-Q_n\|_{X_{a+s}\to X}\|AP_{m,s}L^{-s}\|_{X\to X_{a+s}} \\ &\le c\, n^{-(a+s)}\|P_{m,s}L^{-s}\|_{X\to X_s} \\ &\asymp n^{-(a+s)}\|P_{m,s}\|_{X_s\to X_s} = n^{-(a+s)}. \end{aligned} \tag{3.21}$$

If $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$ then (3.18)–(3.21) imply

$$\|G_{\alpha,n,m}f\|_{-a} \le c\,\|f\|.$$

Together with (2.1) and (3.17) this yields

$$\|G_{\alpha,n,m}f\|_0 \le \|G_{\alpha,n,m}f\|_s^{\frac{a}{a+s}}\|G_{\alpha,n,m}f\|_{-a}^{\frac{s}{a+s}} \le c\,\alpha^{-\frac{a}{2(a+s)}}\|f\|.$$

The lemma is proved. □

Now we are in a position to estimate the variance $v_{\alpha,n,m}(\xi)$ for X-valued random noise ξ meeting the conditions (2.6). Namely, from Lemma 3.2 it follows that for $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$

$$\begin{aligned} v_{\alpha,n,m}(\xi) = \delta^2\mathbf{E}\|G_{\alpha,n,m}\xi\|^2 &\le \delta^2\|G_{\alpha,n,m}\|^2_{X\to X}\mathbf{E}\|\xi\|^2 \\ &\le c\,\delta^2\,\alpha^{-\frac{a}{(a+s)}}. \end{aligned} \tag{3.22}$$

But such straightforward way is unsuitable for generalized white noise ξ satisfying (2.7) because in this case $\mathbf{E}\|\xi\|^2 = \infty$. Instead we note that $G_{\alpha,n,m}\xi = G_{\alpha,n,m}Q_n\xi$, and we conclude

$$v_{\alpha,n,m}(\xi) = \delta^2\mathbf{E}\|G_{\alpha,n,m}Q_n\xi\|^2 \le \delta^2\|G_{\alpha,n,m}\|^2_{X\to X}\mathbf{E}\|Q_n\xi\|^2.$$

Using (2.7) and keeping in mind that $\{\varphi_i\}$ is an orthonormal system we have $\mathbf{E}(\varphi_i,\xi)^2 = (\varphi_i,\varphi_i) = 1, \quad i = 1,2,\ldots,n$, such that, applying Lemma 3.2 again, we arrive at

$$v_{\alpha,n,m}(\xi) \le c\,\delta^2\,\alpha^{-\frac{a}{(a+s)}}\mathbf{E}\Big(\sum_{k=1}^{n}(\varphi_i,\xi)^2\Big) = c\,\delta^2\,\alpha^{-\frac{a}{(a+s)}}n \tag{3.23}$$

for $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$.

3.4. **Parameter choice and convergence rates.** In order to optimize the rate of convergence for the global risk we will determine $\alpha = \alpha(\delta)$ in such a way that the rates of bias and variance in (3.9) are of the same order as $\delta \to 0$. This is accomplished in

Theorem 3.1. *Let the assumptions* (2.3), (2.4), (3.7), (3.6) *be fulfilled and* $s \ge (\mu - a)/2$.

If the random noise ξ *satisfies the conditions* (2.6) *then for* $\alpha \asymp \delta^{\frac{2(a+s)}{a+\mu}}$, $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}} \asymp \delta^{-\frac{1}{a+\mu}}$

$$\mathbf{E}\|x_0 - x^{\delta}_{\alpha,n,m}(\xi)\|^2 \le c\,\delta^{\frac{2\mu}{\mu+a}}. \tag{3.24}$$

In the case of generalized white noise ξ *satisfying* (2.7) *for* $\alpha \asymp \delta^{\frac{2(a+s)}{a+\mu+1/2}}$, $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}} \asymp \delta^{-\frac{1}{a+\mu+1/2}}$

$$\mathbf{E}\|x_0 - x^{\delta}_{\alpha,n,m}(\xi)\|^2 \le c\,\delta^{\frac{2\mu}{\mu+a+1/2}}. \tag{3.25}$$

In (3.24) *and* (3.25) *the constant* c *does not depend on* δ, α, n, m.

Proof. We prove only (3.25). The estimate (3.24) is established in a similar manner. From (3.6) it follows that for some $n \asymp \alpha^{-\frac{1}{2(a+s)}}$ condition (3.10) is fulfilled. Then, using Lemma 3.1 and (3.23) for $n \asymp m \asymp \alpha^{-\frac{1}{2(a+s)}}$, $\alpha \asymp \delta^{\frac{2(a+s)}{a+\mu+1/2}}$ we obtain

$$\begin{aligned}\mathbf{E}\|x_0 - x^{\delta}_{\alpha,n,m}(\xi)\|^2 &\le c\,\alpha^{\frac{\mu}{a+s}} + c\,n\,\delta^2\alpha^{-\frac{a}{a+s}}\\ &\le c\big(\alpha^{\frac{\mu}{a+s}} + \delta^2\alpha^{-\frac{2a+1}{2(a+s)}}\big) \le \delta^{\frac{2\mu}{a+\mu+1/2}},\end{aligned}$$

which yields the required result. $\square$

4. Lower bounds for $\Delta_{n,\delta}(A, \mathfrak{M}, \xi)$

In this section we obtain lower bounds in Hilbert scales under additional assumptions, relating properties of the operator A to the generator L of the Hilbert scale.

Following Mair and Ruymgaart (1996) we assume that the eigenvectors of the operator L generating the Hilbert scale $\{X_s\}$ coincide with the eigenvectors of A^*A. This means that both the operator L^{-1} and the operator A from (1.1) can be represented in the form

$$L^{-1}g = \sum_{k=1}^{\infty} l_k(g, f_k) f_k, \quad Ag = \sum_{k=1}^{\infty} \gamma_k(g, f_k) u_k, \tag{4.1}$$

where $\{f_k\}$, $\{u_k\}$ are some orthonormal bases of X. >From (4.1), (2.2) and (2.3) it follows, in particular, that

$$l_k \asymp k^{-1}, \quad \gamma_k \asymp k^{-a}. \tag{4.2}$$

Now we can state

Theorem 4.1. *Let the assumptions* (2.2) – (2.4) *and* (4.1) *be fulfilled. Then*

$$\Delta_{n,\delta}(A, X^R_\mu, \xi^\beta) \ge c\{n^{-\mu} + \delta^{\frac{\mu}{\mu+a-\beta+1/2}}\}$$

We rely on the following Lemma, originally proven in Korostelev and Tsybakov (1993), Chapt. 9.

Lemma 4.1. *Suppose we are given*

$$v_k = \theta_k + \delta\sigma_k\xi_k, \quad k = 1, 2, \dots, \tag{4.3}$$

where ξ_k *are i.i.d* $N(0,1)$, $\sigma_k \asymp k^b$ *and* $\theta = (\theta_1, \theta_2, \dots)$ *is unknown, but belongs to*

$$B^R_\mu := \{\theta : \sum_k \lambda_k^2\theta_k^2 \le R^2, \ \lambda_k \asymp k^\mu\}.$$

Then

$$\inf_{\hat{\theta}(v)} \sup_{\theta \in B_\mu^R} \mathbf{E}\|\theta - \hat{\theta}(v)\|_{l_2}^2 \asymp \delta^{\frac{2\mu}{\mu+b+1/2}},$$

where the inf *is taken over all estimators* $\hat{\theta}(v)$ *based on observations* (4.3).

Proof of Theorem 4.1. We observe that the composition $S \circ \varphi$ is a linear mapping in X with rank at most n. Then by (2.7) and arguing as in (3.9) we obtain

$$\begin{aligned} \mathbf{E}\|x - S\circ\varphi(Ax + \delta\xi^\beta)\|^2 &= \|x - S\circ\varphi(Ax)\|^2 + \delta^2\mathbf{E}\|S\circ\varphi(\xi^\beta)\|^2 \\ &\geq \|x - S\circ\varphi(Ax)\|^2. \end{aligned}$$

Uniformly for $x \in X_\mu^R$ this yields

$$\begin{aligned} \Delta_{n,\delta}(A, X_\mu^R, \xi^\beta)^2 &\geq \inf_{\varphi\in\Phi_n} \inf_{S\in\mathcal{L}_n(X)} \sup_{x\in X_\mu^R} \|x - S\circ\varphi(Ax)\|^2 \\ &\geq R^2 \mathrm{a}_{n+1}^2(J_\mu) \asymp n^{-2\mu}, \end{aligned} \tag{4.4}$$

where we used (2.2) for the last asymptotics. On the other hand, it follows from the very definition that

$$\Delta_{n,\delta}(A, X_\mu^R, \xi^\beta)^2 \geq \inf_{\hat{x}} \sup_{x\in X_\mu^R} \mathbf{E}\|x - \hat{x}\|^2, \tag{4.5}$$

where the inf is taken over all estimators based on observations

$$y_\delta = Ax + \delta\xi^\beta \tag{4.6}$$

Let $\theta_k = (x, f_k)$ and $v_k = (y_\delta, u_k)\gamma_k^{-1}$, $k = 1, 2, \ldots$. We can represent (4.6) in the equivalent form

$$v_k = \theta_k + \delta\sigma_k\xi_k, \quad \sigma_k = c_\beta k^{-\beta}\gamma_k^{-1} \asymp k^{a-\beta}, \quad k = 1, 2, \ldots \tag{4.7}$$

and any estimator $\hat{\theta}$ based on (4.7) gives the estimator $\hat{x} = \sum_k \hat{\theta}_k f_k$ for x, and conversely. Applying Lemma 4.1 with $b = a - \beta$, $\lambda_k = l_k^{-\mu} \asymp k^\mu$ we obtain

$$\inf_{\hat{x}} \sup_{x\in X_\mu^R} \mathbf{E}\|x - \hat{x}\|^2 \sim \inf_{\hat{\theta}} \sup_{\theta\in B_\mu^R} \mathbf{E}\|\theta - \hat{\theta}\|_{l_2}^2 \asymp \delta^{\frac{2\mu}{\mu+a-\beta+1/2}} \tag{4.8}$$

The assertion of the theorem follows from (4.4), (4.5), (4.8). □

Remark 4.1. As in Donoho et al. (1990) the number n can be understood as the difficulty of the estimator $S\circ\varphi(Ax+\delta\xi)$ based on the design $\varphi \in \Phi_n$. For $\beta = 1/2+\varepsilon$, $\varepsilon > 0$ arbitrarily small, the noise ξ^β is X–valued and satisfies conditions (2.6). Then Theorem 4.1 and Lemmas 3.1 and 3.2 yield for $m \asymp \alpha^{-1/(2(a+s))} \asymp \delta^{-1/(a+\mu)}$ and sufficiently large $n \geq c\delta^{-1/(a+\mu)}$ the estimate

$$c_1\delta^{\frac{2\mu}{\mu+a-\varepsilon}} \leq \Delta_{n,\delta}(A, X_\mu^R, \xi^\beta)^2 \leq \sup_{x\in X_\mu^R} \mathbf{E}\|x - x_{\alpha,n,m}^\delta(\xi^\beta)\|^2 \leq c_2\delta^{\frac{2\mu}{\mu+a}}.$$

For such noise the lower bound for the difficulty of estimation with optimal precision $\delta^{\mu/(\mu+a-\varepsilon)}$ is $\delta^{-1/(\mu+a-\varepsilon)}$. The estimator $x_{\alpha,n,m}^\delta(\xi^\beta)$ however has difficulty $\delta^{-1/(\mu+a)}$ at precision $\delta^{\mu/(\mu+a)}$, which is close to being optimal. We stress that within the framework of Tikhonov regularization (3.2) and (3.3), the precision will not change by enlarging the design beyond $\delta^{-1/(\mu+a)}$.

For Gaussian white noise ξ^0 the situation is different. Theorems 3.1 and 4.1 imply for $n \geq m \asymp \alpha^{-1/(2(a+s))}$ the bounds

$$\begin{aligned} c_1(n^{-2\mu} + \delta^{\frac{2\mu}{\mu+a+1/2}}) &\leq \Delta_{n,\delta}(A, X_\mu^R, \xi^0)^2 \\ &\leq \sup_{x \in X_\mu^R} \mathbf{E}\|x - x_{\alpha,n,m}^\delta(\xi^0)\|^2 \\ &\leq c_2(\alpha^{\mu/(a+s)} + n\delta^2\alpha^{-a/(a+s)}), \end{aligned}$$

and the corresponding estimator $x_{\alpha,n,m}^\delta(\xi^0)$ with $n \asymp \alpha^{-1/(2(a+s))} \asymp \delta^{-1/(\mu+a+1/2)}$ attains the lower bound for the difficulty $n_{opt} \asymp \delta^{-1/(\mu+a+1/2)}$ of estimation with optimal precision simultaneously for all operators A meeting conditions (2.3) and (4.1). But using designs larger than n_{opt} can spoil the precision of the Tikhonov regularization estimator.

Remark 4.2. If we allow the randomness to be degenerate then for $n \asymp m \asymp \delta^{-\frac{1}{a+\mu}}$, $\alpha \asymp \delta^{\frac{2(a+s)}{a+\mu}}$ and for deterministic noise ξ, $\|\xi\| \leq 1$, from (3.9), Lemma 3.1, Lemma 3.2 it follows that

$$\|x_0 - x_{\alpha,n,m}^\delta(\xi)\| \leq \|(G_{\alpha,n,m}A - I)x_0\| + \delta\|G_{\alpha,n,m}\xi\| \leq \delta^{\frac{\mu}{\mu+a}}. \tag{4.9}$$

Formally (4.9) coincides with the classical nonrandom result by Natterer (1984) and Neubauer (1988). But in Natterer (1984) the discretization effects in ill-posed problems were not considered. Neubauer (1988) studied only semi-discrete schemes, that is when we pass from equation $Ax = y_\delta$ to $AP_{m,s}x = y_\delta$. Therefore, it seems that estimate (4.9) is new even in the deterministic case.

5. Application to Abel's equation

Here we apply the results of the previous sections to the regularization of histograms (1.4), where A is Abel's integral operator of the form

$$Ax(t) := \frac{1}{\pi}\int_t^1 \frac{x(\tau)d\tau}{\sqrt{\tau - t}}, \quad t \in (0,1). \tag{5.1}$$

Noisy Abel's equation (1.1), with operator (5.1) arises from a diverse range of applications in the physical sciences and in stereological microscopy. Some pertinent references are Nychka and Cox (1989), Johnstone and Silverman (1991) and Donoho (1995).

In order to apply our results we let $X = L_2(0,1)$. Then as mentioned in Remark 2.1 we generate a Hilbert scale via $L = (A^*A)^{-1}$. In this case the conditions (2.3) are fulfilled for $a = 1/2$. For simplicity we let $\mu = 1$ in (2.4) . This means that $x_0(t)$ can be represented in the form

$$x_0(t) = A^*Ag_0(t), \tag{5.2}$$

where $g_0(t) \in L_2(0,1)$. Using Corollary 1 in Samko (1968) we obtain the following representations

$$Af(t) = A^*Vf(t), \quad A^*f(t) = \frac{1}{\pi}\int_0^t \frac{f(\tau)d\tau}{\sqrt{t-\tau}}, \tag{5.3}$$

and the operator

$$Vf(t) := \frac{1}{\pi\sqrt{t}}\int_0^1 \frac{\sqrt{\tau}f(\tau)d\tau}{\tau - t}$$

acts boundedly from $L_2(0,1)$ into the space $L_{2,2\varepsilon}(0,1)$ of functions that are square-summable on (0,1) with weight $t^{2\varepsilon}$, where $\varepsilon > 0$ is arbitrarily small. This means that for any $f \in L_2(0,1)$ there exists $f_\varepsilon \in L_2(0,1)$ such that

$$Vf(t) = t^{-\varepsilon}f_\varepsilon(t) \quad \text{and} \quad \|f_\varepsilon\| \le c\|f\|. \tag{5.4}$$

Moreover, from the properties of fractional integration we have

$$A^*A^*f(t) = \int_0^t f(\tau)d\tau. \tag{5.5}$$

Then (5.2)–(5.5) imply that

$$x_0(t) = A^*Ag_0(t) = A^*A^*Vg_0(t) = \int_0^t \tau^{-\varepsilon}g_{0,\varepsilon}(\tau)d\tau. \tag{5.6}$$

Thus, x_0 has derivative $x_0' \in L_{2,2\varepsilon}(0,1)$ for any small $\varepsilon > 0$. In terms of the modulus of continuity

$$\omega_2(f,h) = \left\{\sup_{0<t\le h}\int_0^{1-t} |f(t+\tau) - f(\tau)|^2 d\tau\right\}^{1/2}, \quad 0 < h < 1,$$

for functions $f \in L_2(0,1)$ we can estimate the smoothness of the solution x_0 by

$$\omega_2(x_0,h) = O\big(h^{1-\varepsilon}\big) \tag{5.7}$$

for any small $\varepsilon > 0$.

As in Nychka and Cox (1989) we assume that bin limits of histograms (1.4) obey

$$\max_{1\le i\le n}\big(u_{i,n} - u_{i-1,n}\big) \asymp \min_{1\le i\le n}\big(u_{i,n} - u_{i-1,n}\big) \asymp n^{-1}. \tag{5.8}$$

As before we represent (1.4) in the form (3.1), where Q_n is the orthogonal projector on the subspace of piecewise constant functions having discontinuities at the points $\{u_{i,n}\}$. It is well known (see, e.g. Plato (1998)) that under (5.8) we have

$$\|(I - Q_n)A\|_{L_2\to L_2} \le c\,n^{-1/2}, \quad \|(I - Q_n)A^*\|_{L_2\to L_2} \le c\,n^{-1/2}.$$

This is just the condition (3.6) for $a = 1/2$ and $\nu = 0$.

A straightforward application of Theorem 3.1 in the present case requires to take $s \ge \frac{\mu - a}{2} = \frac{1}{4}$. On the other hand, it is inconvenient to use Tikhonov estimator (3.2), (3.3) when s has the form of a fraction. But for $s = 1$ the condition (3.6) breaks down for our case because using Q_n we cannot obtain an accuracy being superior to $O(n^{-1})$. Therefore we let $s = 0$. Then condition $s \geq \frac{\mu-a}{2}$ is violated, but if we only slightly change the value m ($m = \alpha^{-\frac{1}{1-\varepsilon}}$ instead of $m = \alpha^{-1}$) then estimate (5.7) allows to obtain the same order of global risk as in Theorem 3.1.

Since $s = 0$, we let $P_{m,0} = Q_m$, the orthogonal projector like Q_n but corresponding to m bins. Estimate (5.7) implies

$$\|(I - P_{m,0})x_0\|_0 = \|(I - Q_m)x_0\|_{L_2} \leq c\omega_2(x_0, m^{-1}) \leq cm^{-1+\varepsilon}, \tag{5.9}$$

for any small $\varepsilon > 0$. Using (5.9) instead of (3.8) we arrive at

Theorem 5.1. *Let us suppose that the exact solution of Abel's equation* (1.1), (5.1) *satisfies the condition* (5.2), *and let $x^\delta_{\alpha,n,m}(\xi)$ be a regularized solution obtained from noisy histogram data* (1.4), (5.8) *within the framework of Tikhonov regularization* (3.2), (3.3) *for $P_{m,0} = Q_m$. If the random noise ξ satisfies condition* (2.6) *then for $\alpha \asymp \delta^{2/3}$, $n \asymp \delta^{-2/3}$, $m \asymp \delta^{-2/3(1-\varepsilon)}$, where $0 < \varepsilon < 1$, the following estimate holds true*

$$\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|^2 \leq c\, \delta^{4/3}.$$

In the case of generalized white noise ξ satisfying (2.7), *for $\alpha \asymp \delta^{1/2}$, $n \asymp \delta^{-1/2}$, $m \asymp \delta^{-1/2(1-\varepsilon)}$*

$$\mathbf{E}\|x_0 - x^\delta_{\alpha,n,m}(\xi)\|^2 \leq c\, \delta.$$

Remark 5.1. Under condition (5.2), which means $\mu = 1$, for Abel's integral equation (1.1), (5.1) with generalized white noise we obtained the same order of precision δ as in Donoho (1995), where wavelet-vaguelette estimator was used. But we can not apply the Theorem 4.1 for estimating the difficulty because the exact order of the eigenvalues of operator $L^{-1} = A^*A$ is unknown (see, for example, Nychka, Cox (1989)).

References

[1] I. Babuška and A. K. Aziz. Survey lectures on the mathematical foundation of the finite element method. In Aziz. A. K., editor, *The Mathematical Foundation of the Finite Element Method with Appl. to Part. Diff. Equations*, New York, 1972. Acad. Press.

[2] A. B. Bakushinskii. On the construction of regularizing algorithms under random noise. *Soviet Math. Dokl.*, 189:231–233, 1969.

[3] G. Bruckner, S. Prößdorf, and G. Vainikko. Error bounds of discretization methods for boundary integral equations with noisy data. Preprint 194, WIAS, 1995.

[4] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. Tech. Rep. Statist. Dep. Standford Univ., 1991.

[5] D. L. Donoho, R. C. Liu, and B. MacGibbon. Minimax risk over hyperrectangles, and implications. *Ann. Statist.*, 18:1416–1437, 1990.

[6] D.L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2:101–126, 1995.

[7] G. Golubev and Khasiminskii R. Statistical approach to inverse boundary value problems for partial differential equations. Preprint 370, WIAS, 1997.

[8] M. Hegland. Variable Hilbert scale and their interpolation inequalities with applications to Tikhonov regularization. *Applic. Analysis*, 59:207–223, 1995.

[9] I. M. Johnstone and B. W. Silverman. Discretization effects in statistical inverse problems. *J. Complexity*, 7:1–34, 1991.

[10] G. Kerkyacharian and D. Picard. Density estimation in Besov spaces. *Statist. Probab. Letters*, 13:15–24, 1992.

[11] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*, volume 82 of *Lect. Notes Statist.* Springer, New York, 1993.

[12] K. C. Li. Minimaxity of the method of regularization of stochastic processes. *Ann. Statist.*, 10:937–942, 1982.

[13] M.A. Lukas. Comparisons of parameter choice methods for regularization with discrete noisy data. *Inverse Problems*, 14:161–184, 1998.

[14] B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56:1424–1444, 1996.

[15] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applic. Analysis*, 18:29–37, 1984.

[16] A. Neubauer. On a posteriori parameter choice for Tikhonov regularization in Hilbert scale leading to optimal convergence rates. *SIAM J. Numer. Anal.*, 25:1313–1326, 1988.

[17] M. Nussbaum. Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.*, 13:984–997, 1985.

[18] M. Nussbaum. Degrees of ill-posedness in stochastic and deterministic noise models. in Algorithms and Complexity for Continuous Problems. Dagstuhl-Seminar-Report 101, 1994.

[19] D.W. Nychka and D.D Cox. Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.*, 17:556–572, 1989.

[20] A. Pietsch. *Operator Ideals*. Deutscher Verlag Wiss., Berlin, 1978.

[21] R. Plato. The Galerkin scheme for Lavrentiev's m-times iterated method to solve linear accretive Volterra integral equations of the first kind. *BIT*, 37:404–423, 1997.

[22] S.G. Samko. On generalized Abel equation and on operators of fractional integration. *Differantsial'nye Uravneniya*, 4:298–314, 1968.

[23] P. Speckman. Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Math.*, 13:970–984, 1985.

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