Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations

Alexander Goldenshluger[1], Sergei V. Pereverzev[2]

submitted: 9th July 1999

[1] Department of Statistics
University of Haifa
Mount Carmel, Haifa 31905, Israel
E-Mail: goldensh@rstat.haifa.ac.il

[2] Weierstrass Institute for Applied
Mathematics and Statistics
and
Kiev Institute of Mathematics
Tereshenkivska str. 3
252601 Kiev-4, Ukraine
E-Mail:serg-p@mail.kar.net

Preprint No. 506
Berlin 1999

1991 Mathematics Subject Classification. 62G05, 65J10, 41A25.

Key words and phrases. Adaptive estimation, discretization, Hilbert scales, inverse problems, linear functionals, regularization, minimax risk.

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

We consider adaptive estimating the value of a linear functional from indirect white noise observations. For a flexible approach, the problem is embedded in an abstract Hilbert scale. We develop an adaptive estimator that is rate optimal within a logarithmic factor simultaneously over a wide collection of balls in the Hilbert scale. It is shown that the proposed estimator has the best possible adaptive properties for a wide range of linear functionals. The case of discretized indirect white noise observations is studied, and the adaptive estimator in this setting is developed.

1 Introduction

In this paper we consider adaptive estimating linear functionals from indirect white noise observations. Let X be a separable Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$. Consider an operator equation

$$Ax = y\,, \tag{1}$$

where A is a linear injective operator from X into a dense subset Range$(A) \subset X$. Suppose that the right hand side y of (1) is observed in the presence of a Gaussian white noise of the intensity ε. This specifically means that for every element $\phi \in X$ we can observe

$$y_\varepsilon(\phi) = \langle Ax, \phi \rangle + \varepsilon \xi(\phi)\,, \tag{2}$$

where $\xi(\phi)$ is a Gaussian random variable on a probability space $\{\Omega, \mathcal{A}, \mathbb{P}\}$ with zero mean and variance $\|\phi\|^2$. In addition,

$$\mathbb{E}[\xi(\phi)\xi(\psi)] = \langle \phi, \psi \rangle, \quad \forall \phi, \psi \in X, \tag{3}$$

where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$. We are interested in estimating the value of a linear functional $\ell_f(x) = \langle f, x \rangle$ from the indirect noisy observations (2).

Statistical approach to inverse problems has been proposed by a number of authors, including Sudakov and Khalfin (1964), Bakushinskii (1969), Wahba (1977). For more recent work on this topic see, e.g., O'Sullivan (1986), Nychka and Cox (1989), Johnstone and Silverman (1991), Nussbaum (1994), Donoho (1995), Mair and Ruymgaart (1996), Golubev and Khasiminskii (1997), Chow, Ibragimov and

Khasminskii (1999), Mathe and Pereverzev (1999) and references therein. Typically in these papers A is an integral operator

$$Ax(t) = \int_T a(t,\tau)x(\tau)d\tau\,, \tag{4}$$

acting from $X = L_2(T)$ to $L_2(T)$, where T is an interval in $\mathbb{R}$, and $x(\cdot)$ is the function to be estimated. There is also a considerable literature on the optimal recovery problem, where it is assumed that the right hand side y of (1) is observed with a deterministic noise. A few selected references on the classical deterministic approach to inverse problems are Tikhonov and Arsenin (1977), Traub, Wasilkowski and Wozniakowski (1988), Engl, Hanke and Neubauer (1996).

When estimating a linear functional $\ell_f(x) = \langle f, x\rangle$, it is usually assumed that some a priori information on the unknown solution x is available. This information typically reflects prior knowledge on smoothness of x, and is stated in the form $x \in W$, where W is a prespecified subset of X. Let $\hat{\ell}^\varepsilon(x) = \hat{\ell}(x; y_\varepsilon)$ be an estimate of $\ell_f(x)$ based on the observations (2). In the framework of the minimax approach accuracy of an estimate $\hat{\ell}^\varepsilon$ is measured by its uniform with respect to W risk

$$\mathcal{R}[\hat{\ell}^\varepsilon; W] := \sup_{x\in W} \mathbb{E}|\hat{\ell}^\varepsilon(x) - \ell_f(x)|^2\,.$$

The minimax risk is defined by

$$\mathcal{R}^*[\varepsilon; W] := \inf_{\hat{\ell}^\varepsilon} \mathcal{R}[\hat{\ell}^\varepsilon; W] = \mathcal{R}[\ell^*; W],$$

where inf is taken over all possible estimates $\hat{\ell}^\varepsilon$. The main purpose is to construct *asymptotically optimal*, or in *optimal in order* estimates $\hat{\ell}^\varepsilon$ satisfying

$$\begin{aligned} \mathcal{R}[\hat{\ell}^\varepsilon; W] &= \mathcal{R}^*[\varepsilon; W](1+o(1)), \quad \varepsilon \to 0, \\ \mathcal{R}[\hat{\ell}^\varepsilon; W] &\le C(\epsilon)\mathcal{R}^*[\varepsilon; W], \quad \sup_\varepsilon C(\epsilon) < \infty \end{aligned}$$

respectively.

The outlined problem of estimating linear functionals from white noise observations is a subject of considerable literature under various assumptions on the operator A, the linear functional $\ell_f(x)$ and the solution set W. It has been extensively studied for the case of direct observations, where A is the identity operator (Speckman (1979), Li (1982), Ibragimov and Has'minskii (1984)). For models with indirect observations see, e.g., Donoho and Low (1992), Donoho (1994), Stander and Silverman (1995) and references therein. In these papers a variety of optimal in the minimax sense estimators has been developed. Typically, such estimators are highly specialized in the sense that their construction depends heavily on the solution set W. The crucial step of the construction involves selecting a smoothing parameter; to choose it optimally one should have a priori information on the solution set W. In practice, however, specifying the set W of possible solutions can present severe difficulties. Therefore, developing estimators that are optimal in the minimax sense simultaneously over a collection of solution sets W is of interest.

Let $\mathcal{W}_\varepsilon$ denote a collection of solution sets W, possibly growing as $\varepsilon \to 0$. We say that an estimate $\hat{\ell}^\varepsilon$ is *adaptive with respect to* $\mathcal{W}_\varepsilon$ if

$$\sup_{W \in \mathcal{W}_\varepsilon} \{\mathcal{R}[\hat{\ell}^\varepsilon; W]/\mathcal{R}^*[\varepsilon; W]\} \leq C(\varepsilon), \tag{5}$$

where $\sup_\varepsilon C(\varepsilon) < \infty$, or $C(\varepsilon)$ grows *slowly* as ε goes to 0 (we say that $C(\varepsilon)$ grows slowly as $\varepsilon \to 0$ if $\lim_{\varepsilon \to 0}[C(r\varepsilon)/C(\varepsilon)] = 1$ for every $r > 0$). Recently much attention has been concentrated on developing adaptive nonparametric estimators both for direct and indirect observations (Lepskii(1991), Donoho and Johnstone (1994), Barron, Birge, Massart (1999), Donoho (1995), Abramovich and Silverman (1998), Johnstone (1999), Cavalier and Tsybakov (1998)). For adaptive estimation of linear functionals from direct white noise observations we refer to Lepskii (1990), Efromovich and Low (1994) and Tsybakov (1998). Adaptive estimates that are within a logarithmic factor optimal simultaneously over a collection of the solutions sets have been proposed there. It has been shown also that the extra logarithmic factor is often unavoidable when estimating linear functionals. In particular, this fact has been established by Lepskii (1990, 1992) and Brown and Low(1996) for estimating a function (or its derivative) at a single given point from direct white noise observations. This similar result holds for indirect observations involving certain convolution operators (Goldenshluger (1998)). It should be noticed that there is a vast literature on data–driven selection of smoothing parameters in inverse problems (see, e.g., Wahba (1977), Lukas (1998) and references therein); however, the minimax properties of the related estimation methods are not usually analyzed.

The goal of the present paper is to develop an adaptive estimator of linear functionals in the general Hilbert space framework. For a flexible approach, we embed the problem in a Hilbert scale, and propose the estimator that is adaptive over a collection of balls in the Hilbert scale. Our construction exploits deterministic regularization methods along with the general adaptation scheme developed by Lepskii (1990, 1991) for estimation from direct white noise observations. We show that the accuracy of our adaptive estimator is only by a logarithmic factor worse than the one we could achieve in the case when the solution set W is known exactly. We argue also that in many important cases this extra logarithmic factor cannot be reduced; here our estimator possesses the best possible abilities for adaptation. Furthermore, we consider the case of discretized observations, where the data (2) are available only for a finite number of "probe" functions $\phi_i \in X$, $i = 1, \dots, n$. This case corresponds to grouped or binned data which are typical in statistical practice (Johnstone and Silverman (1991), Bickel and Ritov (1995)). In the context of indirect estimation in the Hilbert scales the case of discretized observations has been studied by Mathe and Pereverzev (1999). Both the "probe" functions and the number of observations n taken are important parameters of the estimation method. We consider the problem of optimal discretization, and show that our estimate associated with a data–driven choice of the design set $\Phi_n = \{\phi_i \in X, i = 1, \dots, n\}$

possesses the same adaptive minimax properties, as the adaptive estimate based on the complete observations (2).

The rest of the article is organized as follows. Section 2 introduces our notation and assumptions. In Section 3 we consider the regularized inverse estimator and show that it is optimal in order under a proper choice of the regularization parameter. In Section 4 our adaptive estimator is defined, and its accuracy is analyzed. Adaptive estimation of linear functionals from indirect discretized observations is studied in Section 5.

2 Formulation and assumptions

Recall that a Hilbert scale $\{X_r\}_{r\in\mathbb{R}}$ is a family of Hilbert spaces X_r with the inner products $\langle u,v\rangle_r := \langle L^r u, L^r v\rangle$, where L is an unbounded self-adjoint strictly positive operator in a dense domain of X. More precisely, X_r is defined as the completion of the intersection of domains of all operators $\{L^s\}_{s\in\mathbb{R}}$, endowed with the norm $\|u\|_r := \langle u,u\rangle_r^{1/2}$, $\|\cdot\|_0 = \|\cdot\|$. The first investigation of inverse problems with deterministic noise in Hilbert scales dates back to Natterer (1984). Statistical inverse estimation in Hilbert scales has been studied by Mair and Ruymgaart (1996), and Mathe and Pereverzev (1999). Usually $\{X_r\}$ are the Sobolev spaces of various kinds; in this case r is the index characterizing smoothness.

Example 1 (a). Let $X = L_2(0,1)$, and

$$X_r = \left\{ x \in L_2(0,1) : \sum_{k=1}^{\infty} k^{2r} |\langle x, \varphi_k\rangle|^2 < \infty \right\},$$

where $\varphi_1, \varphi_2, \ldots$ be an orthonormal basis of $L_2(0,1)$. In this case X_r is the domain of the operator L^r, where $L : X_1 \to L_2(0,1)$ is defined by $Lx = \sum_{k=1}^{\infty} k\langle x, \varphi_k\rangle \varphi_k$.

(b). Let $X = L_2(\mathbb{R})$, and

$$X_r = \left\{ x \in L_2(\mathbb{R}) : \int_{\mathbb{R}} (1+s^2)^r |(\mathcal{F}x)(s)|^2 ds < \infty \right\},$$

where $\mathcal{F}$ denotes the Fourier transform from $L_2(\mathbb{R})$ into itself. Then X_r is the domain of the operator L^r, where $L = S^{1/2}$ and $S : X_2 \to L_2(\mathbb{R})$ is given by $Sx = x - x''$.

The following factors determine essentially behavior of the minimax risks in estimating linear functionals $\ell_f(x) = \langle f, x\rangle$: (a) degree of ill–posedness of the operator A; (b) smoothness of the representer f; (c) smoothness of the solution x. We introduce the main assumptions on these ingredients of the problem in the Hilbert scale framework.

Throughout the paper we assume that the operator A is adapted to the Hilbert scale $\{X_r\}$ is the following sense.

Assumption 1 *The operator A acts along the Hilbert scale $\{X_r\}$: for some parameter $a \geq 0$ there exist constants $d, D > 0$ such that*

$$d\|u\|_{r-a} \leq \|Au\|_r \leq D\|u\|_{r-a}, \quad \forall u \in X_{r-a}, \quad r \in \mathbb{R}. \tag{6}$$

Examples of integral operators (4) satisfying (6) can be found in Neubauer (1988), Mair and Ruymgaart (1996). Condition (6) describes the degree of ill–posedness of the operator A relative to the Hilbert scale $\{X_r\}$. We note that even if the operator A does not fit some standard Hilbert scale, one can often construct a scale adapted to A. This is the case when $A: \; X \to X$ acts compactly and injectively in some Hilbert space X. Then A meets condition (6) with $a = 1/2$ in the scale generated by $L := (A^*A)^{-1}$; see Natterer (1984) and Hegland (1995) for further details.

The following assumption on the linear functional $\ell_f(x)$ will be used in the sequel.

Assumption 2 *The representer f of the linear functional $\ell_f(x) = \langle f, x\rangle$ belongs to the Hilbert space X_ν, and either (i) $\nu \leq a$, or (ii) $\nu < a$.*

The condition (i) is quite usual in estimating linear functionals (see Tautenhahn (1996)); it includes linear functionals that can theoretically be estimated both at the *parametric* $O(\varepsilon^2)$ and *nonparametric* rates. The condition (ii) corresponds to estimating *nonparametric* (*singular*) linear functionals, where the representer f is a generalized function relative to the Hilbert space X.

As for the unknown solution x, we suppose that x belongs to the ball $W_\mu(M) \subset X_\mu$

$$W_\mu(M) := \{x \in X_\mu : \|x\|_\mu \leq M\}$$

for some index $\mu > 0$ and constant $M > 0$. Since the dual space of X_μ is $X_{-\mu}$ (see, for example, Krein et al. (1982), p. 237), and X_r is embedded in X_s for $r > s$, we need also the condition $\nu \geq -\mu$ to ensure that the linear functional $\ell_f(x) = \langle f, x\rangle$ is well–defined.

Example 2 Let $X = L_2(0,1)$, and A be a compact integral operator given by (4). Let the Hilbert scale $\{X_r\}$ be generated by an operator L, and let $\{\varphi_k\}$ be a complete orthonormal system of eigenfunctions of the operator L, i.e. $L\varphi_k(t) = \lambda_k\varphi_k(t), \; k = 1, 2, \ldots$. Thus, $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ with $\lambda_k \to \infty$ as $k \to \infty$. Then the solution x of integral equation (1) with the integral operator (4) satisfying (6) belongs to X_μ if and only if

$$\sum_{k=1}^{\infty} \lambda_k^{2\mu} |\langle \varphi_k, x\rangle|^2 < \infty.$$

Suppose we are interested in estimating the value of $\ell_f(x) = x(t_0)$, where $t_0 \in [0,1]$. If for some $\nu \geq -\mu$

$$\sum_{k=1}^{\infty} \lambda_k^{2\nu} |\varphi_k(t_0)|^2 < \infty$$

then for $x \in X_\mu \subset X_{-\nu}$

$$x(t_0) = \langle f_{t_0}, x\rangle := \int_0^1 f_{t_0}(t)x(t)dt = \sum_{k=1}^{\infty} \varphi_k(t_0)\langle\varphi_k, x\rangle.$$

Here the representer f_{t_0} of estimated linear functional is a generalized function determined by the series

$$f_{t_0}(t) = \sum_{k=1}^{\infty} \varphi_k(t_0)\varphi_k(t)$$

converging in X_ν. In particular, if L is as in Example 1(a) and $\mu > 1/2$, then $f_{t_0} \in X_{-1/2}$.

3 Regularized inverse estimator

The inverse operator A^{-1} is not necessarily bounded in the X-topology; therefore some kind of regularization is required for estimating the value $\langle f, x\rangle$. In the context of the deterministic approach to inverse problems it was shown in Bakushinskii (1967), Groetsch (1977), Vainikko and Veretennikov (1986), and Tautenhahn (1996) that a wide variety of regularization methods can be constructed in the following way. Let $g_\alpha(\cdot)$ be a piecewise continuous function on $[0, D^2]$ depending on a *regularization parameter* $\alpha > 0$ and satisfying the following conditions:

$$\sup_{\lambda\in[0,D^2]} |\lambda^\gamma g_\alpha(\lambda)| \le c_\gamma \alpha^{\gamma-1}, \quad 0 \le \gamma \le 1, \tag{7}$$

$$\sup_{\lambda\in[0,D^2]} |\lambda^\beta[1-\lambda g_\alpha(\lambda)]| \le c_\beta \alpha^\beta, \quad 0 \le \beta \le 1, \tag{8}$$

where D is given in (6), and c_γ, c_β are positive constants. Fix a non-negative number $s \ge -\nu$ and define the regularized estimate $\hat{\ell}^\varepsilon_{\alpha,s}(x)$ of $\ell_f(x) = \langle f, x\rangle$ by

$$\hat{\ell}^\varepsilon_{\alpha,s}(x) = y_\varepsilon(AL^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}f), \tag{9}$$

where A^* is the adjoint of the operator A in X. Observe also that the condition $s \ge -\nu$ ensures that $AL^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}f \in X$, so that the estimate is well–defined. The well-known Tikhonov regularization method is characterized by (9) with $g_\alpha(\lambda) = (\lambda+\alpha)^{-1}$. In the statistical context this method has been applied to estimating the value of a linear functional by Li (1982) with the special choice $\alpha = O(\varepsilon^2)$.

Accuracy of the regularization methods depends crucially on the choice of the regularization parameter α. Let $\hat{\ell}^\varepsilon_{\alpha,s}(x)$ be the estimate (9) associated with some regularization parameter α and $s \ge \max\{0, -\nu\}$. It follows immediately from (2) and (9) that

$$\langle f, x\rangle - \hat{\ell}^\varepsilon_{\alpha,s}(x) = b_{\alpha,s}(f, x) + \varepsilon v_{\alpha,s}(f, \xi),$$

where

$$\begin{aligned} b_{\alpha,s}(f,x) &= \langle f,(I-L^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}A^*A)x\rangle, \\ v_{\alpha,s}(f,\xi) &= -\xi(AL^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}f). \end{aligned}$$

Since $v_{\alpha,s}(f)$ is a zero mean random variable, we obtain

$$\mathbb{E}|\langle f,x\rangle-\hat{\ell}^\varepsilon_{\alpha,s}(x)|^2=b^2_{\alpha,s}(f,x)+\varepsilon^2\mathbb{E}v^2_{\alpha,s}(f,\xi). \tag{10}$$

Now we establish upper bounds on the bias and variance in the right hand side of (10).

Lemma 3.1 *Let Assumptions 1, 2(i) hold, and $\hat{\ell}^\varepsilon_{\alpha,s}(x)$ be associated with $s\geq\max\{0,-\nu\}$. Then for every $\mu\in(-\nu,2s+a]$ one has*

$$\sup_{x\in W_\mu(M)}|b_{\alpha,s}(f,x)|\leq cM\|f\|_\nu\,\alpha^{\frac{\mu+\nu}{2(a+s)}}\,, \tag{11}$$

where $c=c(\nu,a,s,d,D)$ depends on ν,a,s,d,D only.

Proof Let

$$x^0_{\alpha,s}=L^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}A^*Ax \tag{12}$$

then

$$|b_{\alpha,s}(f,x)|=|\langle f,x-x^0_{\alpha,s}\rangle|\leq\|f\|_\nu\|x-x^0_{\alpha,s}\|_{-\nu}\,,$$

and the proof follows immediately from (7) and Proposition 2.2 in Tautenhahn (1996). □

Lemma 3.2 *Let Assumptions 1, 2(i) hold, and $s\geq\max\{0,-\nu\}$. Then*

$$\mathbb{E}v^2_{\alpha,s}(f,\xi)\leq c\,\alpha^{\frac{\nu-a}{a+s}},$$

where $c=c(\nu,a,s,f)$ depends on ν, a, s and f only.

Proof In order to prove the statement of the lemma, we should establish the upper bound on $\|AL^{-s}g_\alpha(L^{-s}A^*AL^{-s})L^{-s}f\|^2$. Let us introduce the operator $H=AL^{-s}$. We need the following assertion that can be found in Natterer (1984):

$$\mathrm{Range}\{(H^*H)^{r/2}\}=X_{r(a+s)},\quad |r|\leq 1. \tag{13}$$

Then for $f\in X_\nu$, $\nu\leq a$ and $s\geq\max\{0,-\nu\}$ we obtain

$$L^{-s}f\in X_{s+\nu}=\mathrm{Range}\{(H^*H)^{\frac{s+\nu}{2(a+s)}}\}.$$

This guarantees existence of $v_f \in X$ such that

$$L^{-s} f = (H^* H)^{\frac{s+\nu}{2(a+s)}} v_f. \tag{14}$$

Then (3), (13), (14) and (7) imply

$$\begin{aligned} \mathbb{E} v^2_{\alpha,s}(f,\xi) &= \|AL^{-s} g_\alpha(L^{-s}A^*AL^{-s})L^{-s}f\|^2 \\ &= \|H g_\alpha(H^*H)(H^*H)^{\frac{s+\nu}{2(a+s)}} v_f\|^2 \\ &\le \sup_{\lambda\in[0,D^2]} \left| g_\alpha(\lambda)\lambda^{\frac{s+\nu}{2(a+s)}+\frac{1}{2}} \right|^2 \|v_f\|^2 \le c\alpha^{\frac{\nu-a}{a+s}}, \end{aligned}$$

where the constant $c = c(\nu, a, s, f)$ depends on ν, a, s and f only. □

Combining Lemmas 3.1, 3.2, we obtain that under Assumptions 1, 2 the uniform risk of the estimate $\hat{\ell}^{\varepsilon}_{\alpha,s}(x)$ associated with $s \ge \max\{0, -\nu\}$ and $\alpha > 0$ admits the following upper bound

$$\mathcal{R}[\hat{\ell}^{\varepsilon}_{\alpha,s}; W_\mu(M)] \le c\left(M^2\alpha^{\frac{\mu+\nu}{a+s}} + \varepsilon^2\alpha^{\frac{\nu-a}{a+s}}\right), \quad \forall\mu \in (-\nu, 2s+a], \tag{15}$$

where $c = c(\nu, s, a, f, d, D)$. Thus, α controls trade-off between the bias and the variance of the risk. As usual in nonparametric estimation, the optimal choice of the regularization parameter minimizes the upper bound (15). We obtain that with the optimal choice $\alpha \asymp (M^{-1}\varepsilon)^{\frac{2(a+s)}{\mu+a}}$ one has

$$\mathcal{R}[\hat{\ell}^{\varepsilon}_{\alpha,s}; W_\mu(M)] \le cM^{-\frac{2(\nu-a)}{\mu+a}}\varepsilon^{\frac{2(\mu+\nu)}{\mu+a}}, \quad \forall\mu \in (-\nu, 2s+a], \tag{16}$$

where "$\asymp$" means equivalent in the sense of the order.

One can argue that the rate of convergence given in (16) cannot be improved for estimating linear functionals. Indeed, it follows from Donoho and Low (1992) that

$$\frac{1}{4}\omega^2(\varepsilon) \le \mathcal{R}^*[\varepsilon; W_\mu(M)] \le \omega^2(\varepsilon), \tag{17}$$

where the modulus of continuity $\omega(\varepsilon)$ is given by

$$\omega(\varepsilon) = \sup\{2\ell_f(x) : \|Ax\| \le \varepsilon/2, \ \ \|x\|_\mu \le M\}.$$

Since $\nu > -\mu$, we have $X_\mu \subset X_{-\nu}$, and $\langle f, x\rangle \le \|f\|_\nu \|x\|_{-\nu}$. Condition (6) implies that the constraint $\|Ax\| \le \varepsilon/2$ is equivalent to $\|x\|_{-a} \le \tilde{d}\varepsilon/2$ with some constant $\tilde{d} \in [d, D]$. Taking into account the embedding $X_{-a} \supset X_{-\nu} \supset X_\mu$, and the strict interpolation property of the Hilbert scales, we obtain

$$\sup\{\|x\|_{-\nu} : \|x\|_{-a} \le \tilde{d}\varepsilon/2, \ \|x\|_\mu \le M\} = (\tilde{d}\varepsilon/2)^{\frac{\mu+\nu}{\mu+a}} M^{-\frac{\nu-a}{\mu+a}}. \tag{18}$$

Thus, the estimate $\hat{\ell}^{\varepsilon}_{\alpha,s}(x)$ is optimal in order for every ball $W_\mu(M)$ with $\mu \in (-\nu, 2s+a]$.

It is interesting to note also the order of the risk indicated in (16) coincides with the optimal order of accuracy obtained by Engl and Neubauer (1988) and Tautenhahn (1996) in the problem of optimal recovery of $\langle f, x\rangle$.

4 Adaptive estimator

The optimal choice of the regularization parameter requires a priori information on the parameters μ and M of the solution set $W_\mu(M)$, and by this reason is not practical. In this section we introduce our adaptive estimator which is near optimal simultaneously over a wide collection of the balls $W_\mu(M)$, not just over a single one.

Let $\hat{\ell}^\varepsilon_{\alpha,s}(x)$ be the regularized inverse estimator defined in (9). Denote $\underline{\alpha} = \varepsilon^{2(a+s)/(a-\nu)}$, $\overline{\alpha} = 1$, $r_\alpha = \alpha^{(\nu-a)/(2(a+s))}$, and for a fixed real number $q > 1$ define

$$\Delta_q := \{\alpha \in [\underline{\alpha}, \overline{\alpha}] : \alpha = \alpha_j = q^j \underline{\alpha}, \;\; j = 0, 1, \dots\}.$$

Consider the family of the regularized inverse estimates $\{\hat{\ell}^\varepsilon_{\alpha,s}(x)\}$ associated with the regularization parameter α from the finite ordered set Δ_q. Let $\varkappa \geq 1$; then we define our adaptive estimate as $\hat{\ell}^\varepsilon_{\alpha_+,s}(x)$, where

$$\alpha_+ := \max\{\alpha \in \Delta_q : \; |\hat{\ell}^\varepsilon_{\alpha,s}(x) - \hat{\ell}^\varepsilon_{\eta,s}(x)| \leq 4\varkappa\varepsilon r_\eta, \;\; \forall \eta \leq \alpha, \eta \in \Delta_q\}. \tag{19}$$

Note that α_+ is well–defined; in particular, the minimal $\alpha_+ = \underline{\alpha}$ is a feasible solution to (19). Observe also that α_+ depends on the random noisy data (2), on the ill–posedness index a, on ε, on smoothness of the representer f of estimated linear functional, and on three design parameters s, q and $\varkappa$. In the sequel $\varkappa$ will be chosen as function of ε, s and q, so that actually α_+ depends on the two design parameters s and q. We would like to stress that the parameters μ and M of the solution set $W_\mu(M)$ are not involved in our construction.

Theorem 4.1 *Let Assumptions 1, 2(ii) hold, and ε be small enough such that for some constant $c_1 = c_1(\nu, a, s, f)$ one has*

$$\varepsilon\sqrt{\ln \varepsilon^{-1}} \leq c_1 \min\left\{M, \;\; M^{-\frac{\nu-a}{\mu+a}}\right\}. \tag{20}$$

Assume also that

$$\mu \in (-\nu, 2s + a]. \tag{21}$$

Then there exists a constant $c_2 = c_2(\nu, a, s, f, d, D, q)$ such that for the estimate $\hat{\ell}^\varepsilon_{\alpha_+,s}(x)$ associated with $\varkappa = c_2\sqrt{\ln \varepsilon^{-1}}$ and $s \geq \max\{0, -\nu\}$ one has

$$\mathcal{R}[\hat{\ell}^\varepsilon_{\alpha_+,s}; W_\mu(M)] \leq c_3 \left[M^{-\frac{\nu-a}{\mu+a}} (\varepsilon^2 \ln \varepsilon^{-1})^{\frac{\mu+\nu}{\mu+a}} + \varepsilon^2 \ln \varepsilon^{-1}\right], \tag{22}$$

where $c_3 = c_3(\nu, a, s, f, d, D, q)$.

Proof In the below proof $c_1, c_2, \dots$ stand for the constants depending on ν, a, s, f, d, D and q only. For brevity, we will write $\hat{\ell}_\alpha(x)$ for $\hat{\ell}^\varepsilon_{\alpha,s}(x)$, $v_\alpha(f)$ for $v_{\alpha,s}(f, \xi)$, and $b_\alpha(x)$ for $b_{\alpha,s}(f, x)$.

Denote $B_\alpha(x) = c_1\|x\|_\mu\|f\|_\nu\alpha^{(\mu+\nu)/(2(a+s))}$, where c_1 is the constant appearing in the right hand side of (11). For a fixed $\varkappa \geq 1$, define

$$\alpha_* = \max\{\alpha \in \Delta_q : B_\alpha(x) \leq \varkappa\varepsilon r_\alpha\}.$$

It follows immediately from (11) that for $q\alpha_* \in \Delta_q$

$$\varkappa\varepsilon r_{q\alpha_*} < B_{q\alpha_*}(x) \leq cM\|f\|_\nu(q\alpha_*)^{(\mu+\nu)/(2(a+s))},$$

and then

$$\alpha_* \geq (c_2\varkappa\varepsilon M^{-1})^{\frac{2(a+s)}{\mu+a}} \tag{23}$$

for some constant c_2. Condition (20) implies that the quantity in the right hand side of (23) belongs to the interval $[\underline{\alpha}, \overline{\alpha}]$.

Consider the event

$$\Omega_\varkappa = \Big\{\omega \in \Omega : \max_{\alpha\in\Delta_q}\Big(r_\alpha^{-1}|v_\alpha(f)|\Big) \leq \varkappa\Big\}.$$

Assume that $\Omega_\varkappa$ holds; then for every $\eta \in \Delta_q$ satisfying $\eta \leq \alpha_*$ we have

$$\begin{aligned} |\hat\ell_\eta(x) - \hat\ell_{\alpha_*}(x)| &\leq |\langle f, x\rangle - \hat\ell_\eta(x)| + |\langle f, x\rangle - \hat\ell_{\alpha_*}(x)| \\ &\leq |b_\eta(x)| + \varepsilon|v_\eta(f)| + |b_{\alpha_*}(x)| + \varepsilon|v_{\alpha_*}(f)| \\ &\leq 2B_{\alpha_*}(x) + \varkappa\varepsilon r_\eta + \varkappa\varepsilon r_{\alpha_*} \leq 4\varkappa\varepsilon r_\eta\,. \end{aligned}$$

This means that on the set $\Omega_\varkappa$ our adaptive rule (19) always chooses the regularization parameter α_+ greater than α_*. Thus, taking into account that $\alpha_+ \geq \alpha_*$ on the set $\Omega_\varkappa$, and (23) we obtain

$$\begin{aligned} |\langle f, x\rangle - \hat\ell_{\alpha_+}(x)| &\leq |\langle f, x\rangle - \hat\ell_{\alpha_*}(x)| + |\hat\ell_{\alpha_*}(x) - \hat\ell_{\alpha_+}(x)| \\ &\leq B_{\alpha_*}(x) + \varepsilon|v_{\alpha_*}(f)| + 4\varkappa\varepsilon r_{\alpha_*} \\ &\leq 6\varkappa\varepsilon r_{\alpha_*} \leq c_3(\varkappa\varepsilon)^{\frac{\mu+\nu}{\mu+a}}M^{-\frac{\nu-a}{\mu+a}}. \end{aligned} \tag{24}$$

Now consider the case $\omega \in \Omega_\varkappa^c = \Omega\setminus\Omega_\varkappa$. By Lemma 3.2 for $\underline{\alpha} = \varepsilon^{2(a+s)/(a-\nu)} \leq \alpha_+$ one has independently of the event $\Omega_\varkappa$

$$\begin{aligned} |\langle f, x\rangle - \hat\ell_{\alpha_+}(x)| &\leq |\hat\ell_{\underline{\alpha}}(x) - \hat\ell_{\alpha_*}(x)| + |\langle f, x\rangle - \hat\ell_{\underline{\alpha}}(x)| \\ &\leq 4\varkappa\varepsilon r_{\underline{\alpha}} + B_{\underline{\alpha}}(x) + \varepsilon|v_{\underline{\alpha}}(f)| \\ &\leq 4\varkappa\varepsilon^{-1} + \varepsilon[\mathbb{E}v_{\underline{\alpha}}^2(f)]^{1/2}\max_{\alpha\in\Delta_q}\Big(|v_\alpha(f)|[\mathbb{E}v_\alpha^2(f)]^{-1/2}\Big) \\ &\leq c_4\varkappa\varepsilon^{-1}\Theta(\xi)\,, \end{aligned} \tag{25}$$

where

$$\Theta(\xi) := \max_{\alpha\in\Delta_q}\Big(|v_\alpha(f)|[\mathbb{E}v_\alpha^2(f)]^{-1/2}\Big).$$

Since $v_\alpha(f)[\mathbb{E}v_\alpha^2(f)]^{-1/2}$ is the standard Gaussian random variable, and the cardinality of the set Δ_q does not exceed $N = c_5 \ln \varepsilon^{-1}$, we can write

$$\mathbb{P}\{\Theta(\xi) > \tau\} \le N \int_\tau^\infty \exp(-t^2/2)dt, \quad \tau > 0. \tag{26}$$

Integrating by parts we easily obtain from (26) that

$$\mathbb{E}[\Theta(\xi)]^4 \le c_0 (\ln N)^2, \tag{27}$$

where c_0 is an absolute constant. Further, Lemma 3.2 and (26) imply that

$$\begin{aligned}
\mathbb{P}\{\Omega_\varkappa^c\} &= \mathbb{P}\left\{\omega \in \Omega : \max_{\alpha \in \Delta_q} \alpha^{\frac{a-\nu}{2(a+s)}} |v_\alpha(f)| > \varkappa\right\} \\
&\le \mathbb{P}\left\{\omega \in \Omega : \max_{\alpha \in \Delta_q} |v_\alpha(f)| (\mathbb{E}v_\alpha^2(f))^{-1/2} > c_6^{-1}\varkappa\right\} \\
&= \mathbb{P}\{\Theta(\xi) > c_6^{-1}\varkappa\} \le N \int\limits_{c_6^{-1}\varkappa}^{\infty} \exp(-t^2/2)dt\,.
\end{aligned} \tag{28}$$

Using (25), (26), (27), and (28) we obtain

$$\begin{aligned}
\mathbb{E}\Big(|\langle f, x\rangle - \hat{\ell}_{\alpha_+}(x)|^2 \mathbf{1}\{\Omega_\varkappa^c\}\Big) &\le c_4 \varkappa \varepsilon^{-1} \int\limits_{\Omega_\varkappa^c} |\Theta(\xi)|^2 d\mathbb{P}(\omega) \\
&\le c_4 \varkappa \varepsilon^{-1} (\mathbb{E}|\Theta(\xi)|^4)^{1/2} [\mathbb{P}(\Omega_\varkappa^c)]^{1/2} \\
&\le c_7 \varkappa \varepsilon^{-1} \sqrt{N} \ln N \left(\int\limits_{c_6^{-1}\varkappa}^{\infty} \exp(-t^2/2)dt\right)^{1/2}.
\end{aligned}$$

Now it is evident from the above upper bound that one can choose a constant c_8 such that for $\varkappa = c_8\sqrt{\ln \varepsilon^{-1}}$ one has

$$\mathbb{E}\Big[|\langle f, x\rangle - \hat{\ell}_{\alpha_+}(x)|^2 \mathbf{1}\{\Omega_\varkappa^c\}\Big] \le \varepsilon^2 \ln \frac{1}{\varepsilon}. \tag{29}$$

With this choice of $\varkappa$, combining (29) and (24) we finally obtain

$$\mathbb{E}|\langle f, x\rangle - \hat{\ell}_{\alpha_+}(x)|^2 \le c_9 \left[M^{-\frac{\nu-a}{\mu+a}} (\varepsilon^2 \ln \varepsilon^{-1})^{\frac{\mu+\nu}{\mu+a}} + \varepsilon^2 \ln \varepsilon^{-1}\right].$$

□

If we knew in advance the parameters μ and M of the solution set $W_\mu(M)$, we could achieve the rate of convergence given in (16). The arguments of Donoho and Low (1992) show that this is the minimax rate of convergence. Therefore accuracy of our adaptive estimator coincides, up to a logarithmic in ε^{-1} factor, with the best achievable rate of convergence for the case, where the parameters of the solution set $W_\mu(M)$ are known exactly. We stress, however, that the upper bound (22) holds simultaneously for all balls $W_\mu(M)$ from the collection $\mathcal{W}_\varepsilon$ defined by (20) and (21). Comparing the upper bound (22) with the order of the minimax risk given by (17) and (18), we conclude that the estimate $\ell^\varepsilon_{\alpha_+,s}(x)$ is adaptive with respect to $\mathcal{W}_\varepsilon$ in the sense of (5).

We can argue also that in many important cases the estimate $\hat{\ell}^\varepsilon_{\alpha_+,s}(x)$ possesses the best possible abilities for adaptation; i.e. the $\ln\varepsilon^{-1}$ factor cannot be eliminated if one is interested in adaptive estimation over a collection of solution sets. In particular, if $X = L_2(0,1)$, A is the identity operator, $\ell_f(x)$ is the singular linear functional, and $\mathcal{W}_\varepsilon$ contains at least two balls $W_{\mu_1}(\cdot)$, $W_{\mu_2}(\cdot)$ with $\mu_1 \neq \mu_2$, then the extra $\ln\varepsilon^{-1}$ cannot be avoided (see Lepskii (1990), Brown and Low (1996), Efromovich and Low (1994) and Tsybakov (1998)). The same is true for some convolution operators (Goldenshluger (1998)). In these cases our estimator has the best possible adaptive properties.

5 Discretization

In this section we consider the problem of estimating a linear functional $\ell_f(x)$ from discretized indirect white noise observations. In other words, we assume that only a finite number of observations is available

$$y_\varepsilon(\phi_i) = \langle Ax, \phi_i\rangle + \varepsilon\xi(\phi_i), \quad i = 1, \dots, n, \tag{30}$$

where the set of the elements $\Phi_n := \{\phi_i \in X,\, i = 1, \dots, n\}$ is called the *design*. From now on we assume that both the design set Φ_n, and the number of observations n can be chosen. This assumption has a practical meaning, because it concerns with the important question of how many observations to use for a given noise intensity ε (cf. Johnstone and Silverman (1991)). The goal is to estimate a linear functional $\ell_f(x) = \langle f, x\rangle$ from such discretized indirect white noise observations.

Our construction is based on the Tikhonov regularized inverse estimator characterized by $g_\alpha(\lambda) = (\lambda+\alpha)^{-1}$. Suppose that Φ_n is an orthonormal system in X, and let Q_n denote the orthogonal projector onto the span$\{\phi_1, \phi_2, \dots, \phi_n\}$

$$Q_n = \sum_{i=1}^n \langle\phi_i, \cdot\rangle\phi_i.$$

Let $s \geq \max\{0, -\nu\}$, and define the regularized estimate $\hat{\ell}^\varepsilon_{\alpha,n,s}(x)$ of the linear functional $\ell_f(x) = \langle f, x\rangle$ by

$$\hat{\ell}^\varepsilon_{\alpha,n,s}(x) = y_\varepsilon(Q_nAL^{-s}(\alpha I + L^{-s}A^*Q_nAL^{-s})^{-1}L^{-s}f) \tag{31}$$

(cf. (9)). Since $Q_nAL^{-s}(\alpha I + L^{-s}A^*Q_nAL^{-s})^{-1}L^{-s}f \in \text{span}\{\phi_1, \phi_2, \dots, \phi_n\}$, the estimate is well–defined.

Another representation for $\hat{\ell}^{\varepsilon}_{\alpha,n,s}(x)$ can be obtained from the variational characterization of the Tikhonov method. Denote

$$Q_n y_\varepsilon = \sum_{i=1}^{n} y_\varepsilon(\phi_i)\phi_i, \quad Q_n\xi = \sum_{i=1}^{n} \xi(\phi_i)\phi_i.$$

Then the observations (30) can be written as

$$Q_n y_\varepsilon = Q_n(Ax + \varepsilon\xi). \tag{32}$$

Note that (32) is the standard form of the projection scheme for the approximate solution of the operator equation (1) with random noise. Let $x^{\varepsilon}_{\alpha,n,s}$ be the solution to the following minimization problem

$$\min_{u\in X_s}\{\|Q_nAu - Q_ny_\varepsilon\|^2 + \alpha\|u\|_s^2\}.$$

Equivalently, $x^{\varepsilon}_{\alpha,n,s}$ is the solution to the Euler equation

$$\alpha u + L^{-2s}A^*Q_nAu = L^{-2s}A^*Q_ny_\varepsilon, \tag{33}$$

which is, in fact, a finite-dimensional operator equation in $\text{span}\{L^{-2s}A^*\phi_i,\ i = 1, 2, \ldots, n\}$. With this notation, $\hat{\ell}^{\varepsilon}_{\alpha,n,s}(x) = \langle f, x^{\varepsilon}_{a,n,s}\rangle$.

In what follows we assume that the design sets $\Phi_1 \subset \Phi_2 \subset \cdots \subset \Phi_n \subset \cdots$ have good approximation properties in the following sense.

Assumption 3 *For every n*

$$\|I - Q_n\|_{X_r\to X_0} \le \tilde{c}\, n^{-r}, \quad \forall r \in [0, s+a], \tag{34}$$

where $\tilde{c}$ is a constant depending on s and a only.

This assumption is standard for discretization of inverse problems in Hilbert scales (see, for example, Neubauer (1988)). If $\{X_r\}$ is a scale of Sobolev spaces then (34) is valid for a wide variety of design sets, like splines, wavelets, trigonometric functions.

It follows from (30), (31) and (33) that

$$\langle f, x\rangle - \hat{\ell}^{\varepsilon}_{\alpha,n,s}(x) = b_{\alpha,n,s}(f, x) + v_{\alpha,n,s}(f, \xi),$$

where

$$\begin{aligned} b_{\alpha,n,s}(f,x) &= \left\langle f, (I - (\alpha I + L^{-2s}A^*Q_nA)^{-1}L^{-2s}A^*Q_nA)x\right\rangle, \\ v_{\alpha,n,s}(f,\xi) &= -\xi(Q_nAL^{-s}(\alpha I + L^{-s}A^*Q_nAL^{-s})^{-1}L^{-s}f). \end{aligned}$$

Now we establish upper bounds on the bias and variance of the estimate $\hat{\ell}^{\varepsilon}_{\alpha,n,s}(x)$.

Lemma 5.1 *Let Assumptions 1, 2(ii) and 3 hold, and* $s \geq \max\{0, -\nu\}$. *Then there exists a constant* $\bar{c}_1 = \bar{c}_1(a, \nu, s, d, D)$ *such that for* $n = n(\alpha) = \bar{c}_1 \alpha^{-1/(2(a+s))}$ *one has*

$$\sup_{x \in W_\mu(M)} |b_{\alpha,n(\alpha),s}(f,x)| \leq \bar{c}_2 M \|f\|_\nu \, \alpha^{\frac{\mu+\nu}{2(a+s)}}, \quad \forall \mu \in (-\nu, 2s+a],$$

where $\bar{c}_2 = \bar{c}_2(a, \nu, s, d, D)$.

Proof In the below proof $c_1, c_2, \ldots$ stand for positive constants depending on a, ν, s, d and D only.

Let $x^0_{\alpha,n,s} = (\alpha I + L^{-2s} A^* Q_n A)^{-1} L^{-2s} A^* Q_n A x$; then

$$|b_{\alpha,n(\alpha),s}(f,x)| = |\langle f, x - x^0_{\alpha,n,s}\rangle| \leq \|f\|_\nu \|x - x^0_{\alpha,n,s}\|_{-\nu},$$

and it is sufficient to bound from above the norm $\|x - x^0_{\alpha,n,s}\|_{-\nu}$. It follows from (11) that

$$\begin{aligned} \|x - x^0_{\alpha,n,s}\|_{-\nu} &\leq \|x - x^0_{\alpha,s}\|_{-\nu} + \|x^0_{\alpha,s} - x^0_{\alpha,n,s}\|_{-\nu} \\ &\leq c_1 M \|f\|_\nu \, \alpha^{\frac{\mu+\nu}{2(a+s)}} + \|x^0_{\alpha,s} - x^0_{\alpha,n,s}\|_{-\nu}, \end{aligned} \tag{35}$$

where $x^0_{\alpha,s}$ is given in (12). Let us evaluate the second term in (35). Using the formula

$$g_\alpha(L^{-2s} A^* A) = L^{-s} g_\alpha(H^* H) L^s, \quad H = A L^{-s},$$

(see, e.g., Tautenhahn (1996)) for $g_\alpha(\lambda) = (\lambda + \alpha)^{-1}$ and $H_n = Q_n H = Q_n A L^{-s}$ we have

$$\begin{aligned} x^0_{\alpha,s} \;-\; & x^0_{\alpha,n,s} = L^{-s}[g_\alpha(H^*H)H^* - g_\alpha(H_n^* H_n) H_n^*] A x \\ =\;& L^{-s}[(\alpha I + H^*H)^{-1} H^* - (\alpha I + H_n^* H_n)^{-1} H_n^*] A x \\ =\;& L^{-s}(\alpha I + H^*H)^{-1}[(H^* - H_n^*) - (H^*H - H_n^* H_n)(\alpha I + H_n^* H_n)^{-1} H_n^*] A x \\ =\;& L^{-s}(\alpha I + H^*H)^{-1} H^*(I - Q_n)[I - A L^{-s}(\alpha I + H_n^* H_n)^{-1} H_n^*] A x \\ =\;& L^{-s}(\alpha I + H^*H)^{-1} H^*(I - Q_n) A (x - x^0_{\alpha,n,s}). \end{aligned} \tag{36}$$

Further, it follows from Proposition 1 by Natterer (1984) that for any $u \in X$

$$\|u\|_{-r(a+s)} \leq d_0 \, \|(H^*H)^{r/2} u\|, \quad |r| \leq 1, \tag{37}$$

where $d_0 = [\min_{|r| \leq 1} \min\{d^r, D^r\}]^{-1}$, and d, D are the constants from (6). Combining (6), (7), (34), (36) and (37) we obtain

$$\begin{aligned} \|x^0_{\alpha,s} - x^0_{\alpha,n,s}\|_{-\nu} &= \|L^{-\nu-s}(\alpha I + H^*H)^{-1} H^*(I - Q_n) A (x - x^0_{\alpha,n,s})\| \\ &= \|(\alpha I + H^*H)^{-1} H^*(I - Q_n) A (x - x^0_{\alpha,n,s})\|_{-\nu-s} \\ &\leq d_0 \|(H^*H)^{\frac{\nu+s}{2(a+s)}} (\alpha I + H^*H)^{-1} H^*(I - Q_n) A (x - x^0_{\alpha,n,s})\| \\ &\leq d_0 \sup_\lambda |g_\alpha(\lambda) \lambda^{\frac{\nu+s}{2(a+s)} + \frac{1}{2}}| \|I - Q_n\|_{X_{a-\nu} \to X} \|A(x - x^0_{\alpha,n,s})\|_{a-\nu} \\ &\leq c_2 \, \alpha^{\frac{\nu+s}{2(a+s)} - \frac{1}{2}} n^{-(a-\nu)} \, D \, \|x - x^0_{\alpha,n,s}\|_{-\nu}. \end{aligned}$$

Now it is easy to see that there exists a constant c_3 such that for $n = c_3\,\alpha^{-1/(2(a+s))}$

$$\|x^0_{\alpha,s} - x^0_{\alpha,n,s}\|_{-\nu} \le \frac{1}{2}\,\|x - x^0_{\alpha,n,s}\|_{-\nu}. \tag{38}$$

The assertion of the lemma follows from (35) and (38). $\square$

Remark 5.1 *Reconsidering the proof of Lemma 5.1 one can see that the constant $\bar{c}_1$ can be chosen as*

$$\bar{c}_1 = \left[c d_0 D(a+s)^{-1}\left[(\nu+a+2s)^{\nu+a+2s}(a-\nu)^{a-\nu}\right]^{\frac{1}{2(a+s)}}\right]^{\frac{1}{a-\nu}},$$

where c is a constant from (34).

Lemma 5.2 *Let Assumptions 1, 2(ii) and 3 hold, and $s \ge \max\{0,-\nu\}$. Assume that $n = n(\alpha) = \bar{c}_1\alpha^{-1/(2(a+s))}$, where $\bar{c}_1$ is as in Lemma 5.1. Then*

$$\mathbb{E}[v^2_{\alpha,n(\alpha),s}(f,\xi)] \le \bar{c}_2\alpha^{\frac{\nu-a}{a+s}}, \tag{39}$$

where $\bar{c}_2 = \bar{c}_2(\nu,a,s,f)$.

Proof We have

$$\mathbb{E}v^2_{\alpha,n(\alpha),s}(f,\xi) = \|H_{n(\alpha)}g_\alpha(H^*_{n(\alpha)}H_{n(\alpha)})L^{-s}f\|^2.$$

By (14) and (7) we obtain

$$\begin{aligned}
\mathbb{E}v^2_{\alpha,n(\alpha),s}(f,\xi) &\le \|H_{n(\alpha)}g_\alpha(H^*_{n(\alpha)}H_{n(\alpha)})(H^*H)^{\frac{\nu+s}{2(a+s)}}v_f\|^2 \\
&\le c_1\Big\{\|H_{n(\alpha)}g_\alpha(H^*_{n(\alpha)}H_{n(\alpha)})(H^*_{n(\alpha)}H_{n(\alpha)})^{\frac{\nu+s}{2(a+s)}}\|_{X\to X} \\
&\quad +\|H_{n(\alpha)}g_\alpha(H^*_{n(\alpha)}H_{n(\alpha)})[(H^*H)^{\frac{\nu+s}{2(a+s)}} - (H^*_{n(\alpha)}H_{n(\alpha)})^{\frac{\nu+s}{2(a+s)}}]\|_{X\to X}\Big\}^2 \\
&\le c_2\Big\{\alpha^{\frac{\nu+s}{2(a+s)}-\frac{1}{2}} + \alpha^{-\frac{1}{2}}\|(H^*H)^{\frac{\nu+s}{2(a+s)}} - (H^*_{n(\alpha)}H_{n(\alpha)})^{\frac{\nu+s}{2(a+s)}}\|_{X\to X}\Big\}^2
\end{aligned} \tag{40}$$

Using (34) and Corollary 4.2 from Plato and Vainikko (1990) we finally obtain

$$\begin{aligned}
\|(H^*H)^{\frac{\nu+s}{2(a+s)}} - (H^*_{n(\alpha)}H_{n(\alpha)})^{\frac{\nu+s}{2(a+s)}}\|_{X\to X} &\le \|(I-Q_{n(\alpha)})AL^{-s}\|^{\frac{\nu+s}{a+s}}_{X\to X} \\
&\le \Big\{c_3\,[n(\alpha)]^{-(a+s)}\|AL^{-s}\|_{X\to X_{a+s}}\Big\}^{\frac{\nu+s}{a+s}} \\
&\le [c_4\,n(\alpha)]^{-(\nu+s)}\left(D\|L^{-s}\|_{X\to X_s}\right)^{\frac{\nu+s}{a+s}} \\
&\le c_5\,\alpha^{\frac{\nu+s}{2(a+s)}}.
\end{aligned}$$

Together with (40) this yields (39). $\square$

Now we are ready to establish an analog of Theorem 4.1 for the case of discretized observations. Let $n = n(\alpha) = \bar{c}_1\alpha^{-1/(2(a+s))}$, where $\bar{c}_1$ is defined in Lemma 5.1 (see also Remark 5.1). Then the estimate $\hat{\ell}^{\varepsilon}_{\alpha,n(\alpha),s}$ depends only on two design parameters α and s. Let α_+ be given by (19) and $n_+ = n(\alpha_+)$. Consider the estimate $\hat{\ell}^{\varepsilon}_{\alpha_+,n_+,s}(x)$ associated with the choice $\varkappa = \bar{c}_2\sqrt{\ln \varepsilon^{-1}}$, where $\bar{c}_2$ depends on ν, a, s, f, d, D and q. We stress here that $\hat{\ell}^{\varepsilon}_{\alpha_+,n_+,s}(x)$ is based on discretized observations (30), the number of which $n_+ = n(\alpha_+)$ depends on the random regularization parameter α_+. Then the following statement holds.

Theorem 5.1 *Let the conditions of Theorem 4.1 holds, and Assumption 3 is satisfied. Then*

$$\mathcal{R}[\hat{\ell}^{\varepsilon}_{\alpha_+,n_+,s}; W_\mu(M)] \le \tilde{c}\left[M^{-\frac{\nu-a}{\mu+a}}(\varepsilon^2 \ln \varepsilon^{-1})^{\frac{\mu+\nu}{\mu+a}} + \varepsilon^2 \ln \varepsilon^{-1}\right],$$

where $\tilde{c} = \tilde{c}(\nu, a, s, f, d, D, q)$.

Proof follows from Lemmas 5.1, 5.2 using the same arguments as in the proof of Theorem 4.1. □

Theorem 5.1 shows that the same rate of convergence as in (22) can be achieved even in the case where only a finite number of observations n is available. Thus, the estimate $\hat{\ell}^{\varepsilon}_{\alpha_+,n_+,s}(x)$ is adaptive over the collection of the balls $W_\mu(M)$ defined by (20) and (21).

References

[1] Abramovich, F., Silverman, B.W. (1998), Wavelet decomposition approaches to statistical inverse problems, *Biometrika* **85**, 115-129.

[2] Bakushinskii, A.B. (1967), A general method of constructing regularizing algorithms for a linear ill-posed equation in Hilbert space, *Zh. Vychisl. Mat. Mat. Fiz.* **7**, 672-677.

[3] Bakushinskii, A.B. (1969), On the construction of regularizing algorithms under random noise, *Soviet Math. Dokl.* **189**, 231-233.

[4] Barron, A., Birgé, L., Massart, P. (1999), Risk bounds for model selection via penalization, *Probab. Theory Related Fields* **113**, 301-413.

[5] Bickel, P.J., Ritov, Y. (1995), Estimating linear functionals of a PET image, *IEEE Trans. Med. Imaging* **14**, 81-87.

[6] Brown L., Low, M.G. (1996), A constrained risk inequality with application to nonparametric estimation, *Ann. Statist.* **24**, 2524-2535.

[7] Cavalier, L., Tsybakov, A.B. (1998), Sharp adaptation for inverse problems with random noise, *Manuscript.*

[8] Pao-Liu Chow, Ibragimov, I.A., Khasminskii, R.Z. (1999), Statistical approach to some ill–posed problems for linear partial differential equations, *Probab. Theory Related Fields* **113**, 421-441.

[9] Donoho, D.L. (1994), Statistical estimation and optimal recovery, *Ann. Statist.* **22**, 238-270.

[10] Donoho, D.L. (1995), Nonlinear solution of linear inverse problems by Wavelet-Vaguelette Decomposition, *Applied and Computational Harmonic Analysis* **2**, 101-126.

[11] Donoho, D.L., Johnstone I.M. (1994), Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425-455.

[12] Donoho, D.L., Low, M.G. (1992), Renormalization exponents and optimal pointwise rates of convergence, *Ann. Statist.* **20**, 944-970.

[13] Efromovich, S., Low, M.G. (1994), Adaptive estimates of linear functionals, *Probab. Theory Related Fields* **98**, 261-275.

[14] Engl, H.W., Neubauer, A. (1988), A parameter choice strategy for (iterated) Tikhonov regularization of ill-posed problems leading to superconvergence with optimal rates, *Applic. Anal.* **27**, 5-18.

[15] Engl, H.W., Hanke, M., Neubauer, A. (1996), Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht, Boston, London.

[16] Goldenshluger, A. (1998), On pointwise adaptive nonparametric deconvolution, *Bernoulli,* to appear.

[17] Golubev, G., Khasiminskii, R. (1997), Statistical approach to inverse boundary value problems for partial differential equations, Preprint No 370, WIAS - Berlin.

[18] Groetsch, C.W. (1977), Generalized Inverses of Linear Operators, Marcel Dekker, New York.

[19] Hegland, M. (1995), Variable Hilbert scale and their interpolation inequalities with applications to Tikhonov regularization, *Applic. Analysis* **59**,207-223.

[20] Ibragimov, I.A., Has'minskii, R.Z. (1984), On nonparametric estimation of the value of a linear functional in Gaussian white noise, *Theory Probab. Applic.* **29**, 18-32.

[21] Johnstone, I.M. (1999), Wavelet shrinkage for correlated data and inverse problems: adaptivity results, *Statistica Sinica* **9**, 51-83.

[22] Johnstone, I.M., Silverman, B.W. (1991), Discretization effects in statistical inverse problems, *J. Complexity* **7**, 1-34.

[23] Krein, S.G., Petunin, Ju.I., Semenov, E.M. (1982), Interpolation of Linear Operators, American Mathematical Society, Providence, Rhode Island.

[24] Lepskii, O. (1990), A problem of adaptive estimation in Gaussian white noise, *Theory Probab. Appl.* **36**, 454-466.

[25] Lepskii, O. (1991), Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates, *Theory Probab. Appl.* **36**, 682-697.

[26] Lepskii, O. (1992), Asymptotically minimax adaptive estimation II: Schemes without optimal adaptation. Adaptive estimators, *Theory Probab. Appl.* **37**, 433-448.

[27] Li, K.C. (1982), Minimaxity of the method of regularization of stochastic processes, *Ann. Statist.* **10**, 937-942.

[28] Lukas, M.A. (1998), Comparisons of parameter choice methods for regularization with discrete noisy data, *Inverse Problems* **14**, 161-184.

[29] Mair, B.A., and Ruymgaart, F.H. (1996), Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.* **56**, 1424-1444.

[30] Mathe, P., Pereverzev, S.V. (1999), Optimal discretization and degrees of ill-posedness for inverse estimation in Hilbert scales in the presence of random noise, Preprint No 469, WIAS - Berlin.

[31] Natterer, F. (1984), Error bounds for Tikhonov regularization in Hilbert scales, *Applic. Analysis* **19**, 29-37.

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

[33] Nussbaum, M. (1994), Degrees of ill-posedness in Stochastic and Deterministic Noise Models, Algorithms and Complexity for Continuous Problems, Dagstuhl-Seminar-Report 101.

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

[35] O'Sullivan, F. (1986), A statistical perspective on ill-posed problems, *Statist. Science.* **1**, 502-527.

[36] Plato, R., Vainikko, G. (1990), On the regularization of projection methods for solving ill-posed problems, *Numer. Math.* **57**, 63-79.

[37] Speckman, P. (1979), Minimax estimates of linear functionals in a Hilbert space, *Unpublished manuscript.*

[38] Stander, J., Silverman, B.W. (1995), Minimax estimation of linear functionals, particularly in nonparametric regression and positron emission tomography, *Comput. Statist.* **10**, 259-283.

[39] Sudakov, V.N., Khalfin, L.A. (1964), Statistical approach to ill-posed problems in mathematical physics, *Soviet Math. Dokl.* **157**, 1094-1096.

[40] Tautenhahn, U. (1996), Error estimates for regularization methods in Hilbert scales, *SIAM J. Numer. Anal.* **33**, 2120-2130.

[41] Tikhonov, A. N., Arsenin, V. Y. (1977), Solution of ill-posed problems, Wiley, New-York.

[42] Traub, J. F., Wasilkowski, G. W., Wozniakowski, H. (1988), Information-Based Complexity, Wiley, New York.

[43] Tsybakov, A. (1998), Pointwise and sup–norm sharp adaptive estimation of functions on the Sobolev classes, *Ann. Statist.* **26**, 2420-2469.

[44] Vainikko, G. M., Veretennikov, A. Y. (1986), Iteration procedures in ill-posed problems, Nauka, Moscow (in Russian).

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