# Weierstraß–Institut
# für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

## The degrees of ill-posedness in stochastic and deterministic noise models

Michael Nussbaum[1], Sergei Pereverzev[2]

submitted: 9th July 1999

[1] Weierstraß–Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39, D–10117 Berlin, Germany
E-Mail: nussbaum@wias-berlin.de

[2] Weierstrass Institute for Applied Analysis and Stochastic and
Kiev Institute of Mathematics
E-Mail: pereverz@wias-berlin.de

Preprint No. 509
Berlin 1999

---

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

*Key words and phrases.* Ill-posed problems, inverse estimation, operator equations, Gaussian noise, optimal difficulty, regularized inverse 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 degrees of ill-posedness for inverse estimation in Hilbert scales in the presence of deterministic and random noise are compared. For Gaussian random noise with different „smoothness“ the optimal order of the rate of convergence for above mentioned estimation is indicated.

**1.** Suppose we observe data $y_\delta^\xi$ of the form

$$y_\delta^\xi = Ax + \delta\xi, \tag{1}$$

where $x$ is an element of some subset $\mathfrak{M}$ of Hilbert space $X$, $A$ is a compact linear operator from $X$ to $X$, $\xi$ is a noise and $\delta$ is a small positive number used for measuring noise level. We are interested in recoverring an element $x$, and we wish to do this in such a way as to minimize the error occurring at the worst $x \in \mathfrak{M}$. A problem of such a kind arises when an unknown signal $x$ is to be recovered from an imperfect measurement $y_\delta^\xi$ of a given transform of the signal $Ax$. Or, in more mathematical language, on operator equations (1) that require a stable solution.

When $\xi$ is assumed to be a zero-mean Gaussian noise, this is a problem of minimax statistical estimation. Inverse problems (1) in such a ststistical setting can be found throughout Wahba's work (1977), in Nychka and Cox (1989), Johnstone and Silverman (1991) and, more recently, in Donoho (1995), Mair and Ruymgaart (1996), Lukas (1998). For direct density and regression estimation, where $A$ is the identity operator $I$, the reader is referred to Nussbaum (1985), Speckman (1985), Donoho at al. (1997).

When $\xi$ is assumed to be an element of $X$ chosen, not at random, but by an antagonistic opponent, subject to the constrain $\|\xi\|_X \le 1$, this is a problem of deterministic regularization or of optimal recovery. A few selected references from the huge literature on this topic are Tikhonov and Arsenin (1977), Ivanov et al. (1978), Lavrentiev et al. (1980), Morozov (1984), Vainikko and Veretennikov (1986), Traub, Wasilkowski and Wozniakowski (1988), Louis (1989). Suppose now that $A$ and $\mathfrak{M}$ are fixed, but we approach the problem two different ways: one time assuming the noise $\xi$ is random Gaussian, and the other time assuming the noise $\xi$ is deterministic, i.e. $\xi \in X$, $\|\xi\|_X \le 1$. In some cases both ways of stating the problem have been solved approximately, but with different rates of convergence, because the influence of the nature of the noise $\xi$ on the rate of convergence can be reasonably assumed.

This influence has been pointed out by Nussbaum (1994) in the special case that the operator

$$(2) \qquad Ax(t) = \int_0^t \frac{(t-\tau)^{a-1}}{a!} x(\tau) d\tau$$

is $a$-fold integration, $X = L_2(0,1)$ and $\mathfrak{M}$ is a Sobolev space $W_2^s(0,1)$. Namely, for deterministic noise the order of optimal error bound

$$(3) \qquad e_\delta^{\det}(A, \mathfrak{M}) := \inf_{\hat{x}} \sup_{x \in \mathfrak{M}} \sup_{\|\xi\| \le 1} \|x - \hat{x}(\xi)\|$$

for recovery $x \in \mathfrak{M} = W_2^s(0,1)$ from noisy data (1),(2) is $\delta^{\frac{s}{a+s}}$, where the inf in (3) is taken over all estimators $\hat{x}$ based on observations (1). On the other hand, it might be interesting to note that precisely in the example mentioned above, but for Gaussian white noise $\xi(t)$, which is the derivative of Weiner process $W(t)$, the optimal (minimax) rate of convergence for recovery $x \in \mathfrak{M} = W_2^s(0,1)$ is

$$(4) \qquad e_\delta^{\mathrm{stoc}}(A, \mathfrak{M}, \xi) := \inf_{\hat{x}} \sup_{x \in \mathfrak{M}} (\mathbf{E}\|x - \hat{x}(\xi)\|^2)^{1/2} \asymp \delta^{\frac{s}{a+s+1/2}},$$

where $\mathbf{E}$ means the expected value and "$\asymp$" means equivalent. Nussbaum (1994) offers an explanation of this phenomenon using the notion of the degree of ill-posedness of equations involving operator $A$. The degree of ill-posedness is a term coined by Wahba (1977) to quantify interplay between "nastiness" of operator $A$ and "dimensionality" of regularizing set $\mathfrak{M}$. For example, the problem (1),(2), with deterministic noise $\xi$ and $x \in W_2^s(0,1)$ is ill-posed of degree $(a, s)$ in the terminology of Wahba (1977). On the other hand, as it has been pointed out by Nussbaum (1994), for problem (1),(2) with Gaussian white noise in fact we observe

$$Y_\delta(t) = \int_0^t Ax(\tau) d\tau + \delta W(t), \quad t \in [0,1],$$

where $W(t)$ is Wiener process, and if $A$ is $a$-fold integration then th operator $x \to \int_0^t (Ax)$ has degree of ill-posedness $a+1$. The Wiener's noise $W(t)$ is smooth of degree 1/2, and an element $D^{1/2}W(t)$ is bounded only in $L_2$-norm, where $D^{1/2}$ is the derivative of order 1/2. Applying formally $D^{1/2}$ to our model with $Y_\delta$ we get an "effective ill-posed problem" of degree $(a+1-1/2, s)$ for $x \in W_2^s(0,1)$. The essence of explanation by Nussbaum (1994) is this: we can consider the problem (1),(2) with Gaussian white noise $\xi(t)$ and $x \in W_2^s(0,1)$ as problem (1) with operator $D^{1/2}\int_0^t (Ax)$ and "deterministic" noise $\delta D^{1/2}W(t)$. The optimal rate of convergence for such deterministic problem with degree of ill-posedness $(a+1/2, s)$ is $O(\delta^{\frac{s}{a+s+1/2}})$. This is just the order indicated in (4).

Note that the Sobolev scale of subspaces $W_2^s(0,1)$, $s \in \mathbb{R}$, is a particular case of Hilbert scale $\{X_s\}$ generated by some self-adjoint unbounded operator $L$. On the other hand, operator (2), in its turn, is a specific example of an operator acting along Hilbert scale. It is our purpose in this paper to show that above mentioned effect pointed out by Nussbaum (1994) for operator (2) and $X_s = W_2^s(0,1)$ takes place in general situation too. Moreover, we indicate a suitable "scale" of Gaussian noises $\xi^\beta$ with "interpolation" property such that $\xi^\beta$ is the Gaussian white noise for $\beta = 0$ and

$$e_\delta^{\mathrm{stoc}}(A, X_s, \xi^\beta) \asymp \delta^{\frac{s}{s+a-\beta+1/2}}, \quad \beta \in [0, a]. \tag{5}$$

In addition we establish that the employment of deterministic regularizaton spectral cut-off scheme allows to reach the optimal rate of convergence within the framework of deterministic as well as stochastic noise model. Moreover, we show that this scheme realizes the lower bound for the order of the difficulty of estimation with optimal precision.

**2.** A Hilbert scale $\{X_\mu, \ \mu \in \mathbb{R}\}$, it will be recalled, is a family of Hilbert spaces $X_\mu$ with inner product $(x,y)_\mu := (L^\mu x, L^\mu y)$, where $L$ is an unbounded self-adjoint strictly positive operator in a dense domain of $X$ and $(\cdot,\cdot)$ is the inner product in $X$. In more exact terms $X_\mu$ is defined as the completion of the intersection of domains of the operators $L^\nu$, $\nu \geq 0$, accomplished with the norm $\|\cdot\|_\mu$, such that $\|x\|_\mu := (x,x)_\mu^{1/2}$, $\|\cdot\|_0 = \|\cdot\|_X$.

To obtain a result for regularization of ill-posed problems (1) in Hilbert scale, it is often assumed that there exist constants $a, d_1, d_2 > 0$, such that

$$d_1 \|x\|_{-a} \leq \|Ax\|_0 \leq d_2 \|x\|_{-a} \tag{6}$$

holds for all $x \in X_0$. Moreover, the exact solution $x_0$ of the equation (1) for $\delta = 0$, which is assumed to exist, is in some fixed ball $X_s^R$ of $X_s$. i.e.

$$\|x_0\|_s \leq R. \tag{7}$$

Let us illustrate the assumptions (6), (7). Denote by $W_2^\mu$ the Sobolev space of 1-periodic functions (distributions) on the real line with the norm

$$\|x\|_\mu = \Big( |\hat{x}(0)|^2 + \sum_{\substack{m=-\infty,\\ m\neq 0}}^{\infty} |m|^{2\mu} |\hat{x}(m)|^2 \Big)^{1/2}$$

where $\hat{x}(m)$, $m = 0, \pm 1, \pm 2, \ldots$, are the Fourier coefficients of

$$x(t) = \sum_{m=-\infty}^{\infty} \hat{x}(m) e^{im2\pi t}.$$

It is easy to see that the family of Sobolev spaces $\{W_2^\mu\}$ is a Hilbert scale generated by the operator

$$Lx(t) = \hat{x}(0) + \sum_{\substack{m=-\infty,\\ m\neq 0}}^{\infty} |m| \hat{x}(m) e^{im2\pi t}.$$

Consider now Symm's integral equation

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

arising from solving the Dirichlet boundary value problem for the Laplace equation in some region with curve $\Gamma$ as the boundary. Assume that $\Gamma$ has a $C^{\infty}$-smooth 1-periodic parametrization $\gamma: [0,1] \to \Gamma$ with $|\gamma'(t)| \neq 0$ for $t \in [0,1]$. Then Symm's equation can be rewritten as

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

where $x(t) = z(\gamma(t))|\gamma'(t)|$, $y(t) = g(\gamma(t))$, and the operator $A$ defined above meets the condition (6) for $a = 1$ and $X_\mu = W_2^\mu$ (see, for example, Bruckner et al. (1996)). Note that some inverse boudary problems for partial differantial equations were considered recently from statistical viewpoint by Golubev and Khasiminskii (1997). It should be noted in addition that if injective operator $A$ does not meet the condition (6) for some standard Hilbert scale like scale of Sobolev spaces, one can construct a scale adapted to concrete operator $A$. Namely, any compact injective operator $A$ meets the condition (6) for $a = 1/2$ and Hilbert scale generated by the operator $L = (A^*A)^{-1}$ (see, for example, Natterer (1984)).

Regularization of ill-posed problems in Hilbert scales has been introduced by Natterer (1984). From his results it follows that within the framework of deterministic noise model under the conditions (6), (7)

$$e_\delta^{\det}(A, X_s^R) \asymp \delta^{\frac{s}{a+s}}. \tag{8}$$

Statistical inverse estimation in Hilbert scales has been studied by Mair and Ruymgaart (1996). But in their asymptotic considerations B.A. Mair and F.H. Ruymgaart didn't interfere in the case of noisy observations (1).

Let $\{\varphi_k\}_{k=1}^{\infty}$ be some orthonormal basis of Hilbert space $X$. If $\xi$ is a zero-mean $X$-valued Gaussian random noise defined on some probability space $(\Omega, \Sigma, \mathbb{P})$ then for any $k = 1, 2, \dots$, $(\xi, \varphi_k) = \lambda_k \xi_k$, $\lambda_k \in \mathbb{R}$, $\xi_k$ is i.i.d. $N(0,1)$ and $\xi$ can be represented in the form of series

$$\xi = \sum_{k=1}^{\infty} (\xi, \varphi_k) \varphi_k, \tag{9}$$

which converges in $X$ almost surely. For Gaussian white noise $\xi$ the inner products $(\xi, \varphi_k)$ are defined as i.i.d. $N(0,1)$ too but (9) is divergent series. If, as it usually is, we restrict ourselves to the case of random noise with finite second weak order then there is a very simple mathematical construction which allows to consider various types of Gaussian random noise from one standroint. Namely, as in Vakhania et al. (1987), we can associate with any random element $\xi$ having finite second weak order a linear bounded operator $\Lambda_\xi$ acting from $X$ to the space $L_2(\Omega) = L_2(\Omega, \Sigma, \mathbb{P})$

of functions $F: \Omega \to \mathbb{R}$, that are $\mathbb{P}$-square-summable on $\Omega$. Note that the inner product in $L_2(\Omega)$ is defined in term of expected value, i.e.

$$(F_1, F_2)_{L_2(\Omega)} = \mathbf{E}(F_1(\omega)F_2(\omega)) = \int_\Omega F_1(\omega)F_2(\omega)d\mathbb{P}(\omega).$$

The image of any element $\varphi \in X$ under the action of above mentioned operator $\Lambda_\xi$ is considered as inner product $(\varphi, \xi)$ in $X$, i.e. $(\varphi, \xi) := \Lambda_\xi \varphi \in L_2(\Omega)$. It suffices to define the operator $\Lambda_\xi$ only on elements of the basis $\{\varphi_k\}$. Let $\Lambda_\xi \varphi_k = \lambda_k \xi_k$, where $\lambda_k \in \mathbb{R}$, $\xi_k \in L_2(\Omega)$. If $\xi$ is zero-mean Gaussian random element then $\xi_k$ are i.i.d. $N(0,1)$ and $\Lambda_\xi$ has the following decomposition

$$\Lambda_\xi = \sum_{k=1}^{\infty} \lambda_k \xi_k (\varphi_k, \cdot). \tag{10}$$

For Gaussian white noise $\xi$ the operator $\Lambda_\xi$ has the form (10) too, but $\lambda_k = 1$, $k = 1, 2, \dots$.

If $\xi$ is non-random element of $X$ then formally operator $\Lambda_\xi$ is defined as $\Lambda_\xi = (\xi, \cdot)\varkappa_\Omega$, where $\varkappa_\Omega$ is characteristic function of $\Omega$, i.e. $\varkappa_\Omega(\omega) \equiv 1$, $\omega \in \Omega$. Thus, for deterministic as well as for stochastic noise $\xi$ the equivalent form of (1) is $\Lambda_{y_\delta^\xi} = \Lambda_{Ax} + \delta \Lambda_\xi$, but for deterministic noise $\xi$ $\Lambda_\xi$ is very poor 1-dimensional operator. If we know this operator and $\delta$ then we have the exact free term of initial equation $Ax_0 = y_0$, $y_0 = y_\delta^\xi - \delta\xi$, and inverse problem (1) can be solved with arbitrary however small error. Therefore, to avoid this trivial situation we assume that within the framework of deterministic setting the noise element $\xi$ is chosen by an antagonistic opponent and take the second sup in (3) over all possible choice of $\xi$. This is the main difference among deterministic noise model and stochastic one. Because in stochastic case we usually know the operator $\Lambda_\xi$, for example, for Gaussian white noise $\Lambda_\xi$ has the form (10) with $\lambda_k = 1$, $k = 1, 2, \dots$, but nevertheless we can guarantee only some fixed level of precision for recovery of unknown solution $x_0$. Now we estimate this level.

**3.** In this section to obtain a result for general Hilbert scales we assume the following conditions, which are usually fulfilled in special cases. First of all, following Ruymgaart (1993) and Mair and Ruymgaart (1996) we assume that in (1) $A$ is an injective operator and eigenvectors of operator $L$ generating a Hilbert scale $\{X_\mu\}$ coincide with eigenvectors of $A^*A$. This means that operators $L^{-1}$ and $A$ can be represented in the form.

$$L^{-1} = \sum_{k=1}^{\infty} l_k \psi_k (\psi_k, \cdot), \quad A = \sum_{k=1}^{\infty} a_k \varphi_k (\psi_k, \cdot), \tag{11}$$

where $\{\varphi_k\}$, $\{\psi_k\}$ are some orthonormal basises of $X$. Moreover, we assume that a Hilbert scale $\{X_\mu\}$ has the same embedding properties as a Sobolev scale $\{W_2^\mu(0,1)\}$.

What this means is

$$l_k \asymp k^{-1}, \quad k = 1, 2, \ldots \tag{12}$$

From (6), (11), (12) it follows, in particular, that

$$a_k \asymp k^{-a}, \quad k = 1, 2, \ldots \tag{13}$$

As to the noise $\xi$, we will assume that $\xi = \xi^\beta$, where $\xi^\beta$ is such that corresponding operator $\Lambda_{\xi^\beta}$ has singula-value decomposition (10) with $\lambda_k \asymp k^{-\beta}, \quad k = 1, 2, \ldots$ . Note that for $\beta = 0$ $\xi^\beta$ is a Gaussian white noise.

LEMMA 1. *Suppose we are given*

$$v_k = \theta_k + \delta \sigma_k \xi_k, \quad k = 1, 2, \ldots, \tag{14}$$

*where $\xi_k$ are i.i.d. $N(0,1)$, $\sigma_k \asymp k^b, b \geq 0$, and $\theta = (\theta_1, \theta_2, \ldots, \theta_k, \ldots)$ is unknown, but it is known that $\theta$ lies in*

$$B_R^s := \{\theta : \sum_{k=1}^{\infty} \gamma_k^2 \theta_k^2 \leq R^2, \; \gamma_k \asymp k^s\}.$$

*Then*

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

*where the inf is taken over all estimators $\hat\theta(v)$ based on observations (14).*

This lemma is proved, in fact, in Korostelev and Tsybakov (1993), ch.9, and in Belitser, Levit (1995). Moreover, it should be noted that Donoho (1995) obtained the analogue of Lemma 1 for more general situation when one uses $\|\cdot\|_{l_p}$ instead of $\|\cdot\|_{l_2}$ and $B_R^s$ is a Besov-body.

THEOREM 1. *Let the assumptions (6), (11), (12) be fulfilled. Then for $\beta \in [0, a]$*

$$e_\delta^{\text{stoc}}(A, X_s^R, \xi^\beta) \asymp \delta^{\frac{s}{s+a-\beta+1/2}}.$$

PROOF. Using (10), (11), (13) we can represent the observations (1) in the equivalent form (14), where $v_k = a_k^{-1}(y_\delta^\xi, \varphi_k)$, $\theta_k = (\psi_k, x_0)$, $\sigma_k = \lambda_k a_k^{-1} \asymp k^{a-\beta}$, $k = 1, 2, \ldots$, and any estimator $\theta$ based on (14) with such $v_k, \theta_k, \sigma_k$ gives the estimator $\hat x = \sum_k \theta_k \psi_k$ for $x_0$, and the converse. Applying Lemma 1 with $b = a - \beta$, $\gamma_k = l_k^{-s} \asymp k^s$ we obtain the assertion of the theorem:

$$\inf_{\hat x} \sup_{x \in X_s^R} (\mathbf{E}\|x - \hat x\|^2)^{1/2} = \inf_{\hat\theta} \sup_{\theta \in B_R^s} (\mathbf{E}\|\theta - \hat\theta\|_{l_2}^2)^{1/2} \asymp \delta^{\frac{s}{s+a-\beta+1/2}}.$$

From Theorem 1 it follows that the optimal rate of convergence obtained within the framework of stochastic setting for Gaussian noise $\xi^{\frac{1}{2}}$ coincides with optimal deterministic rate (8) in the sense of order. But Gaussian noise $\xi^{\frac{1}{2}}$ is not $X$-valued random element because in view of (10)

$$\mathbf{E}\|\xi^{\beta}\|^2 = \mathbf{E}(\sum_{k=1}^{\infty}(\xi^{\beta}, \varphi_k)^2) = \sum_{k=1}^{\infty} \lambda_k^2 \asymp \sum_{k=1}^{\infty} k^{-2\beta} < \infty$$

only for $\beta > 1/2$. It means that Gaussian structure of random noise allows to obtain the rate of convergence that can be reached within the framework of deterministic setting only for more "smooth" noise $\xi \in X$. On the other hand, for $X$-valued Gaussian random noise $\xi^{\frac{1}{2}+\varepsilon}$, where $\varepsilon$ is a small positive number, the optimal rate of convergence is even better than in the case when noise is assumed to be an element of $X$ chosen, not at random, but by antagonistic opponent. Thus, Gaussian $X$-valued noise $\xi^{\frac{1}{2}+\varepsilon}$ does not exhaust the potentialities of such antagonistic choise.

**4.** One of the estimators making possible to reach the rate of convergence indicated in Theorem 1 was constructed, in fact, by Pinsker (1980). Moreover, in some important cases Pinsker's estimator is optimal even up to the constant in the sense of quantity $e_{\delta}^{\text{stoc}}$ (see, for example, Nussbaum (1996)). To construct this estimator we must know the singular-value decomposition of $A$ and the values of $R, s, l_k, \lambda_k$. Then the realization of Pinsker's estimator reduces to solving some nonlinear equation depending on these constants. In this section we show that if one is interested only in optimal order of the rate of convergence then there is a more simple estimator realizing this order. This estimator, in addition, is order optimal within the framework of deterministic setting too. We mean the so-called spectral cut-off scheme when one takes as estimator for $x_0$ an element

$$x_{\delta}^{\alpha}(\xi) = \sum_{k:\, a_k \geq \alpha} a_k^{-1} \psi_k (\varphi_k, y_{\delta}^{\xi}). \tag{15}$$

This scheme is well-known. Statistical justification for it has been given recently by Mair and Ruymgaart (1996). But we would like to note once again that the case of noisy observations (1) is not considered in above mentioned paper.

THEOREM 2. *Under the conditions of Theorem 1 for* $\alpha = \delta^{\frac{2a}{2s+2a-2\beta+1}}$

$$\sup_{x \in X_s^R} (\mathbf{E}\|x - x_{\delta}^{\alpha}(\xi^{\beta})\|^2)^{1/2} \asymp e_{\delta}^{\text{stoc}}(A, X_s^R, \xi^{\beta}) \asymp \delta^{\frac{s}{s+a-\beta+1/2}}.$$

PROOF. In view of Theorem 1 it suffices to show that for any $x \in X_s^R$ and $\alpha = \delta^{\frac{2a}{2s+2a-2\beta+1}}$

$$\mathbf{E}\|x - x_{\delta}^{\alpha}(\xi^{\beta})\|^2 \leq c\delta^{\frac{2s}{s+a-\beta+1/2}},$$

where the constant $c$ does not depend on $\delta$ and $x$. From (1), (10), (11), (15) it follows that

$$\begin{aligned} \mathbf{E}\|x - x_\delta^\alpha(\xi^\beta)\|^2 &= \mathbf{E}\| \textstyle\sum_{k:a_k<\alpha} a_k^{-1}\psi_k(\varphi_k, Ax) \\ &\quad - \delta \textstyle\sum_{k:a_k\geq\alpha} a_k^{-1}\lambda_k\psi_k\xi_k\|^2 \qquad (16) \\ &= \textstyle\sum_{k:a_k<\alpha} a_k^{-2}(\varphi_k, Ax)^2 + \delta^2 \sum_{k:a_k\geq\alpha} a_k^{-2}\lambda_k^2. \end{aligned}$$

Note that for $\xi = \xi^\beta \quad \lambda_k \asymp k^{-\beta}$. Then from (13) we have

$$\delta^2 \sum_{k:a_k\geq\alpha} a_k^{-2}\lambda_k^2 \asymp \delta^2 \sum_{k\leq\alpha^{-1/a}} k^{2(a-\beta)} \asymp \delta^2\alpha^{-\frac{2a-2\beta+1}{a}}. \qquad (17)$$

On the other hand, if $x \in X_s^R$ then $x = L^{-s}v$, $\|v\| \leq R$, and from (11) it follows that

$$(\varphi_k, Ax)^2 = (\varphi_k, AL^{-s}v)^2 = a_k^2(\psi_k, L^{-s}v) = a_k^2 l_k^{-2s}(\psi_k, v)^2.$$

Then using (12), (13) we have

$$\begin{aligned} \textstyle\sum_{k:a_k<\alpha} a_k^{-2}(\varphi_k, Ax) &= \textstyle\sum_{k:a_k<\alpha} l_k^{-2s}(\psi_k, v)^2 \\ &\asymp \textstyle\sum_{k\geq\alpha^{-1/a}} k^{-2s}(\psi_k, v)^2 \qquad (18) \\ &\leq \alpha^{\frac{2s}{a}}\|v\|^2 \leq \alpha^{\frac{2s}{a}} R^2. \end{aligned}$$

Combining (16)–(18) with $\alpha^{\frac{2a}{2s+2a-2\beta+1}}$ we obtain the error estimate

$$\mathbf{E}\|x - x_\delta^\alpha(\xi^\beta)\|^2 \leq c\big(\alpha^{\frac{2s}{a}} + \delta^2\alpha^{-\frac{2a-2\beta+1}{a}}\big) \leq c\delta^{\frac{2s}{s+a-\beta+1/2}}$$

as claimed.

Note that within the framework of deterministic noise model under the conditions of Theorem 1 for $\alpha = \delta^{\frac{a}{s+a}}$

$$\begin{aligned} \|x - x_\delta^\alpha(\xi)\| &= \|\sum_{k:a_k<\alpha} a_k^{-1}\psi_k(Ax, \varphi_k)\| + \delta\|\sum_{k:a_k\geq\alpha} a_k^{-1}\psi_k(\varphi_k, \xi)\| \\ &\leq c\big(\alpha^{\frac{s}{a}}R + \delta(\sum_{k\leq\alpha^{-1/a}} k^{-2a}(\varphi_k, \xi)^2)^{1/2}\big) \asymp \alpha^{s/a} + \alpha^{-1}\delta \asymp \delta^{\frac{s}{a+s}}. \end{aligned}$$

In view of (8) we can see that in the case of deterministic noise the spectral cut-off scheme (15) is order optimal too.

**5.** From Theorem 2 and (13) it follows that to construct the estimator (15) making possible to reach the optimal rate of convergence indicated in Theorem 1 it suffices to use only finite amount of descretized observations of the form

$$y_{\delta,k}^\xi = (\varphi_k, y_\delta^\xi) = (\varphi_k, Ax) + \delta(\varphi_k, \xi), \quad k = 1, 2, \ldots, n, \qquad (19)$$

and the collection of elements $\varphi = \{\varphi_i\}_{i=1}^n$ plaies the role of the so-called design of the statistical experiment consisting in obtaining the values (19). Let us denote by $\Phi_n$ the set of all designs $f = \{f_i\}_{i=1}^n,\ f_i \in X$, determined by collections of no more than $n$ elements. As in Donoho et al. (1990) the number $n$ can be treated as the difficulty of the estimator with design $f \in \Phi_n$. It is natural to ask what is the minimal difficulty of estimation with optimal precision. If we concentrate on the case of linear estimators of $x$ from discretized observations $f(Ax+\delta\xi) = \{(f_i, y_\delta^\xi),\ i = 1, 2, \dots, n\}$ and it is a priori known that $x$ belongs to some set $\mathfrak{M} \subset X$ then the answer on above mentioned question is connected with behaviour of the quantity

$$\Delta_{n,\delta}(A, \mathfrak{M}, \xi) := \inf_{f\in\Phi_n} \inf_{S\in\mathcal{L}_n(X)} \sup_{x\in\mathfrak{M}} \{\mathbf{E}\|x - S\circ f(Ax+\delta\xi)\|^2\}^{1/2},$$

where $\mathcal{L}_n(X)$ is the set of all linear mapping from $\mathbb{R}^n$ to $X$. It should be noted that the quantity $\Delta_{n,\delta}(A, \mathfrak{M}, \xi)$ was considered earlier by Donoho et al. (1990) in the specific case when $A = I$ and $S \circ f$ is the orthogonal projector on $\mathrm{span}\{f_1, f_2, \dots, f_n\}$.

In this section we show that the estimator (15) and the design $\varphi$ are order optimal in the sense of difficulty.

THEOREM 3. *Let the assumptions of Theorem 1 be fulfilled. Then*

$$\Delta_{n,\delta}(A, X_s^R, \xi^\beta) \geq c\{n^{-s} + \delta^{\frac{s}{s+a-\beta+1/2}}\},$$

*where the constant $c$ does not depend on $\delta$ and $n$.*

PROOF. Let $\{e_k\}_{k=1}^n$ be the canonical basis of $\mathbb{R}^n$. Then for any $f \in \Phi_n,\ S \in \mathcal{L}_n(X),\ g \in X$

$$S\circ f(g) = S\Big(\sum_{k=1}^n e_k(f_k, g)\Big) = \sum_{k=1}^n Se_k(f_k, g) = \sum_{k=1}^n q_k(f_k, g),$$

$$q_k \in X, \quad k = 1, 2, \dots, n.$$

Moreover, by definition

$$\mathbf{E}(g, S\circ f(\xi^\beta)) = \sum_{k=1}^n (g, q_k)\mathbf{E}(f_k, \xi^\beta) = 0$$

and

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

Then using (12) and a simple fact that $\mathrm{rank}\{S \circ f(A\cdot)\} \le n$ we have

$$\begin{aligned}
\Delta_{n,\delta}(A, X_s^R, \xi^\beta) &\ge \inf_{f\in\Phi_n} \inf_{S\in\mathcal{L}_n(X)} \sup_{x\in X_s^R} \|x - S\circ f(Ax)\| \\
&\ge R \inf_{f\in\Phi_n} \inf_{S\in\mathcal{L}_n(X)} \sup_{g:\|g\|\le 1} \|L^{-s}g - S\circ f(AL^{-s}g)\| \\
&\ge R \inf_{\substack{B,\\ \mathrm{rank}\, B\le n}} \|L^{-s} - B\|_{X\to X} = R l_{n+1}^s \asymp n^{-s}.
\end{aligned} \tag{20}$$

On the other hand, it is easy to see that

$$\Delta_{n,\delta}(A, X_s^R, \xi^\beta) \ge e_\delta^{\mathrm{stoc}}(A, X_s^R, \xi^\beta), \tag{21}$$

and the assertion of the theorem follows now from (20), (21) and Theorem 1.

From the Theorem 3 it follows that under the conditions of Theorem 1 the lower bound for the difficulty of estimation of $x \in X_s^R$ from noisy observations $y_\delta^{\xi^\beta} = Ax + \delta\xi^\beta$ with optimal precision is $n \asymp \delta^{-\frac{1}{s+a-\beta+1/2}}$. On the other hand, from Theorem 2 it follows that the estimator (15) allows to reach the optimal level of precision using discretized observations (19) with $n$ such that $a_n \ge \delta^{\frac{a}{s+a-\beta+1/2}}$. Due to (13) we have that $n \le c\delta^{-\frac{1}{s+a-\beta+1/2}}$. It means that the estimator (15) is order optimal in the sense of difficulty.

## References


[1] BELITSER, E. and LEVIT, B. (1995). On minimax filtering on ellipsoids. *Math. Meth. Statist.* **4** 259-273.

[2] BRUCKNER, G., PRÖSSDORF, S. and VAINIKKO, G. (1996). Error bounds of discretization methods for boundary integral equations with noisy data. *Applic. Analysis* **63** 25-37.

[3] DONOHO, D.L., LIU, R.C. and MACGIBBON, B. (1990). Minimax risc over hyperrectangles and implications. *Ann. Statist.* **18** 1416-1437.

[4] DONOHO, D.L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1997). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam* 183-218. Springer, New York.

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

[6] GOLUBEV, G. and KHASIMINSKII, R. (1997). Statistical approach to inverse boundary value problems for partial differential equations. Preprint N 370, WIAS-Berlin.

[7] IVANOV, V. K., VASIN, V. V. and TANANA, V. P. (1978). *Theory of linear ill-posed problems and its applications.* Nauka, Moscow.

[8] JOHNSTONE, I.M. and SILVERMAN, B.W. (1991). Discretization effects in statistical inverse problems. *J. Complexity* **7** 1-34.

[9] KOROSTELEV, A.P. and TSYBAKOV, A.B. (1993). Minimax Theory of Image Reconstruction. *Lecture Notes in Statistics,* **v. 82**. Springer, New York.

[10] LAVRENTIEV, M. M., ROMANOV, V. G. and SHISHTATSKII, S. P. (1980). *Ill-posed problems of mathematical physics and analysis.* Nauka, Moscow.

[11] LOUIS, A. K. (1989). *Inverse und schleecht gestellte Probleme.* B. G. Teubner, Stuttgart.

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

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

[14] MOROZOV, V. A. (1984). *Methods for solving incorrectly posed problems.* Springer, New York.

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

[16] NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in $L_2$. *Ann. Statist.* **13** 984-997.

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

[18] NUSSBAUM, M. (1996). The Pinsker bound: a review. Preprint N 281, WIAS-Berlin.

[19] 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.

[20] PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** 120-133.

[21] RUYMGAART, F. H. (1993). A unified approach to inversion problems in statistics. *Math. Methods Statist.* **2** 130-146.

[22] SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Math. Statist.* **13** 970-984.

[23] TIKHONOV, A. N. and ARSENIN, V. Y. (1977). *Solutions of ill-posed problems.* Wiley, New-York.

[24] TRAUB, J. F., WASILKOWSKI, G. W. and WOZNIAKOWSKI, H. (1988). *Information-Based Complexity.* Academic Press, New-York.

[25] VAINIKKO, G. M. and VERETENNIKOV, A. Y. (1986). *Iteration procedures in ill-posed problems.* Nauka, Moscow.

[26] VAKHANIA, N. N., TARIELADZE, V. I. and CHOBANJAN, S. A. (1987). *Probability distributions on Banach spaces.* D. Reidel, Dordrecht.

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