

Adaptive estimation for inverse problems: a logarithmic effect in L_2

A.B. TSYBAKOV

Université Paris VI

Abstract

This note presents a method of adaptive estimation for a class of inverse problems with random noise. We show that the proposed method attains the best rate of adaptive estimation in L_2 , but this rate is slower than the "non-adaptive" minimax rate by a logarithmic factor.

Mathematics Subject Classifications: 62G05, 62G20

Key Words: Adaptive curve estimation, Statistical inverse problems, Heavily ill-posed problems, Optimal rate of adaptation

1 Introduction

Assume that one observes a sequence of random variables $y = (y_1, y_2, \dots)$ where

$$y_k = b_k \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \dots, \quad (1)$$

$b_k > 0$ are known numbers, ξ_k are i.i.d. standard gaussian random variables, $0 < \varepsilon < 1$ is the noise level and $\theta = (\theta_1, \theta_2, \dots) \in l_2$ is an unknown sequence to be estimated.

The model (1) is used as a sequence space framework for statistical linear inverse problems where one wants to estimate the solution f of the operator equation $g = Af$ from noisy observations of g (see e.g. Korostelev and Tsybakov (1993), Donoho (1995), Mair and Ruymgaart (1996)). Here A is a linear operator defined on a Hilbert space X with scalar product (\cdot, \cdot) and $f \in X$. Then (1) is related to the singular value decomposition: b_k^2 are interpreted as the eigenvalues of the operator A^*A , while $\theta_k = (f, \varphi_k)$, where A^* is the adjoint of A and $\{\varphi_k\}$ are the orthonormalized eigenfunctions of A^*A .

We assume that there exist $b_{\min} > 0, b_{\max} > 0, \beta > 0, q \in \mathbb{R}$, such that

$$b_{\min} k^q e^{-\beta k} \leq b_k \leq b_{\max} k^q e^{-\beta k}, \quad \forall k. \quad (2)$$

Condition (2) on the eigenvalues holds for several well-known problems of mathematical physics that can be called *heavily ill-posed*. We mention here the Cauchy problem for the Laplace equation in a rectangle, the inverse Dirichlet problem for the Laplace equation in the disk, the problem of satellite geodesy (see Golubev and Khasminskii (1999a,b), Pereverzev and Schock (1999) and the references therein).

The model (1) can be written in the equivalent form

$$\bar{y}_k = \theta_k + \varepsilon b_k^{-1} \xi_k, \quad k = 1, 2, \dots, \quad (3)$$

where $\bar{y}_k = y_k/b_k$, and can be also interpreted as the model with direct observations and correlated data (cf. Johnstone (1999)).

Assume that the sequence θ belongs to the following ellipsoid in l_2 :

$$\Theta(\alpha, r, Q) = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq Q \right\},$$

where $a_k = a_k(\alpha) = k^r e^{\alpha k}$ with $\alpha > 0, Q > 0$ and $r \in \mathbb{R}$.

Pinsker (1980) obtains linear minimax estimators of θ on $\Theta(\alpha, r, Q)$ in the model (3) for general b_k , in particular, for those satisfying (2). Golubev and Khasminskii (1999b) study exact asymptotic behavior on $\Theta(\alpha, r, Q)$ of the minimax risk among all estimators for $q = r = 1$. They show that linear minimax estimates as in Pinsker (1980) are not sharp minimax in this model. Ermakov (1989), Efromovich (1997), Efromovich and Kolchinskii (1998) study minimax estimation in statistical inverse problems for some related setups.

Here we consider the problem of adaptive estimation of θ . We assume that the statistician does not know the values α, r, Q . It is only known that there exist some (unknown) positive constants $\alpha_{\min}, \alpha_{\max}, r_{\max}, Q_{\max}$ such that $\alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}$ and $0 < Q \leq Q_{\max}$.

We construct an estimator (called *adaptive*) such that it attains the fastest possible rate of convergence uniformly on the scale of classes $\{\Theta(\alpha, r, Q), \alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}, 0 < Q \leq Q_{\max}\}$. The most interesting fact is that the optimal rate of adaptation for this problem is slower (by a logarithmic factor) than the minimax "non-adaptive" rate. Such a logarithmic phenomenon in adaptation was known before only for the problems of estimation at a fixed point. It was first found by Lepskii (1990), and studied in the context of statistical inverse problems by Goldenshluger (1998), Goldenshluger and Pereverzev (1999). On the other hand, it was believed that such effects do not occur in L_2 , where not only the rates but also exact constants are often the same in adaptive and non-adaptive settings (cf. Efromovich and Pinsker (1985), Cavalier and Tsybakov (2002)). The results below may be viewed as a counter-example showing that the logarithmic deterioration of adaptive rate may happen

in L_2 . This effect is based on the exponentiality of both a_k and b_k . The key assumption is that the exponents of a_k and b_k contain the same powers of k . If this is not the case, or at least one of a_k or b_k has a power behavior rather than exponential one, the logarithmic factor in the adaptive rate disappears and the optimal rate of adaptation coincides with the usual minimax rate.

2 Results

For an arbitrary estimator $\hat{\theta}$ define the maximal risk

$$R_\varepsilon(\hat{\theta}, \alpha, r, Q) = \sup_{\theta \in \Theta(\alpha, r, Q)} \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2$$

where \mathbf{E}_θ is the expectation w.r.t. the distribution of y satisfying (1), and $\|\cdot\|$ is the l_2 -norm.

Let $\bar{y}_k = y_k/b_k$. For any integer N define the projection estimator

$$\hat{\theta}(N) = (\bar{y}_1, \dots, \bar{y}_N, 0, 0, \dots).$$

Set

$$\psi_\alpha(\varepsilon) = \left(\log \frac{1}{\varepsilon} \right)^{-\frac{\beta r + \alpha q}{\beta + \alpha}} \varepsilon^{\alpha/(\beta + \alpha)}$$

Theorem 1 *Let $\alpha > 0$, $r \in \mathbb{R}$, $Q > 0$, and let $N = N(\varepsilon) = \left\lceil \frac{1}{\beta + \alpha} \log \frac{1}{\varepsilon} + \frac{q-r}{\beta + \alpha} \log \log \frac{1}{\varepsilon} \right\rceil$ where $\lceil \cdot \rceil$ denotes the integer part. Then*

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon(\hat{\theta}(N(\varepsilon)), \alpha, r, Q) \psi_\alpha^{-2}(\varepsilon) \leq C < \infty, \quad (4)$$

and

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\hat{\theta}} R_\varepsilon(\hat{\theta}, \alpha, r, Q) \psi_\alpha^{-2}(\varepsilon) \geq c > 0 \quad (5)$$

where $\inf_{\hat{\theta}}$ denotes the infimum over all estimators.

This theorem entails that $\psi_\alpha(\varepsilon)$ is the minimax rate of convergence for estimation of θ in the model (1) on the class $\Theta(\alpha, r, Q)$.

For the upper bound (4) note that, for ε small enough, $a_{N(\varepsilon)}^2 \leq a_k^2$, $\forall k > N(\varepsilon)$, and thus

$$\mathbf{E}_\theta \|\hat{\theta}(N) - \theta\|^2 = \varepsilon^2 \sum_{k=1}^N b_k^{-2} + \sum_{k=N+1}^{\infty} \theta_k^2 \leq K \varepsilon^2 N^{-2q} e^{2\beta N} + Q a_N^{-2} = O \left(\left(\log \frac{1}{\varepsilon} \right)^{-\frac{2(\beta r + \alpha q)}{\beta + \alpha}} \varepsilon^{2\alpha/(\beta + \alpha)} \right),$$

as $\varepsilon \rightarrow 0$, where $K > 0$ is a constant that does not depend on ε and θ . The lower bound (5) is proved by reduction to the problem of testing 2 simple hypotheses. We omit the proof,

since it is not far from the argument in Golubev and Khasminskii (1999b) where the case $q = r = 1$ is treated.

Note that the parameter Q was not used in the construction of the rate optimal estimator $\hat{\theta}(N(\varepsilon))$. However, the knowledge of the parameters α and r was crucial. If these parameters are not available consider an adaptive estimator of θ . Denote

$$N_{\max} = N_{\max}(\varepsilon) = \left\lceil \frac{1}{\beta} \log \frac{1}{\varepsilon} \right\rceil. \quad (6)$$

It is clear that $N_{\max}(\varepsilon) > N(\varepsilon)$ for ε small enough.

Define the estimator:

$$\tilde{\theta} = \hat{\theta}(\hat{N}),$$

where \hat{N} is given by a thresholding of Lepski type:

$$\hat{N} = \min\{1 \leq N \leq N_{\max} : \|\hat{\theta}(N) - \hat{\theta}(N')\| \leq 10\eta_N(\varepsilon), \forall N' \geq N, 1 \leq N' \leq N_{\max}\},$$

with $\eta_N(\varepsilon) = \left(\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^N b_k^{-4}} \right)^{1/2}$. Denote

$$\tilde{\psi}_\alpha(\varepsilon) = \left(\log \frac{1}{\varepsilon} \right)^{-\frac{\beta r + \alpha q}{\beta + \alpha}} \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}} \right)^{\alpha/(\beta + \alpha)}$$

Theorem 2 *Let $W = \{(\alpha, r, Q) : \alpha_{\min} \leq \alpha \leq \alpha_{\max}, |r| \leq r_{\max}, 0 < Q \leq Q_{\max}\}$. Then*

$$\limsup_{\varepsilon \rightarrow 0} \sup_{(\alpha, r, Q) \in W} R_\varepsilon(\tilde{\theta}, \alpha, r, Q) \tilde{\psi}_\alpha^{-2}(\varepsilon) \leq C_{\max} < \infty.$$

Thus we have the upper bound $\tilde{\psi}_\alpha(\varepsilon)$ on the rate of convergence of the adaptive estimator $\tilde{\theta}$ on $\Theta(\alpha, r, Q)$ which is worse than the minimax rate $\psi_\alpha(\varepsilon)$ by a logarithmic factor. The next theorem shows that $\tilde{\psi}_\alpha(\varepsilon)$ is, in fact, the optimal rate of adaptation: it cannot be improved even in the simplest case where α takes only two possible values.

Theorem 3 *For any $r > 0, Q > 0$ and any $\alpha_2 > \alpha_1 > 0$ we have*

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\hat{\theta}} \max_{\alpha \in \{\alpha_1, \alpha_2\}} R_\varepsilon(\hat{\theta}, \alpha, r, Q) \tilde{\psi}_\alpha^{-2}(\varepsilon) \geq c_{\min} > 0.$$

3 Proofs of the theorems

Note that

$$\sum_{k=1}^N k^{-4q} e^{4\beta k} = \frac{e^{4\beta}}{e^{4\beta} - 1} N^{-4q} e^{4\beta N} (1 + o(1)), \quad N \rightarrow \infty.$$

Thus, there exist $C_{b1} > 0, C_{b2} > 0$ depending only on $b_{\max}, b_{\min}, q, \beta$ such that

$$C_{b1} N^{-4q} e^{4\beta N} \leq \sum_{k=1}^N b_k^{-4} \leq C_{b2} N^{-4q} e^{4\beta N}, \quad \forall N. \quad (7)$$

Similarly, there exists $C_{b3} > 0$ depending only on $b_{\max}, b_{\min}, q, \beta$ such that

$$\sum_{k=1}^N b_k^{-2} \leq C_{b3} N^{-2q} e^{2\beta N}, \quad \forall N. \quad (8)$$

Proof of Theorem 2. Consider the random event

$$\mathcal{A} = \left\{ \max_{N=1, \dots, N_{\max}} \varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2 / \eta_N^2(\varepsilon) \leq 17 \right\}$$

It is shown in the Appendix that

$$\mathbf{P}(\overline{\mathcal{A}}) \leq N_{\max} e^{1/4} \varepsilon^4 \quad (9)$$

for ε small enough. Fix $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ and denote

$$N_* = \min\{1 \leq N \leq N_{\max} : N^{-2r} e^{-2\alpha N} \leq C_0 \eta_N^2(\varepsilon)\}$$

where $C_0 > 0$ is a constant that will be chosen later. Using (7) it is easy to show that, for ε small enough,

$$C_1 \log(1/\varepsilon) \leq N_* \leq C_2 \log(1/\varepsilon) \quad (10)$$

where $C_1 > 0, C_2 > 0$ depend only on $b_{\max}, b_{\min}, q, \beta, r_{\max}, \alpha_{\min}, \alpha_{\max}, Q_{\max}$.

Assume that $\theta \in \Theta(\alpha, r, Q)$ and that \mathcal{A} holds. Then, for any $N \geq N_*$, we get

$$\begin{aligned} \|\hat{\theta}(N) - \theta\|^2 &\leq \varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2 + Q \max_{k \geq N} a_k^{-2}(\alpha) \\ &\leq 17 \eta_N^2(\varepsilon) + Q_{\max} N^{-2r} e^{-2\alpha N} \end{aligned} \quad (11)$$

where we assumed that ε is small enough to have monotone increasing a_k for $k \geq N_*$. Since, in addition, $\eta_N(\varepsilon)$ is monotone increasing in N , and using the definition of N_* , we find

$$N^{-2r} e^{-2\alpha N} \leq C_0 \eta_N^2(\varepsilon), \quad \forall N \geq N_*.$$

This and (11) yield, for ε small enough,

$$\|\hat{\theta}(N) - \theta\|^2 \leq (17 + Q_{\max} C_0) \eta_N^2(\varepsilon), \quad \forall N \geq N_*, \quad (12)$$

provided $\theta \in \Theta(\alpha, r, Q)$ and \mathcal{A} holds. Choose $C_0 = 8/Q_{\max}$. Then

$$\|\hat{\theta}(N) - \theta\|^2 \leq 25\eta_N^2(\varepsilon), \quad \forall N \geq N_*. \quad (13)$$

Consequently,

$$\|\hat{\theta}(N) - \hat{\theta}(N_*)\| \leq \|\hat{\theta}(N) - \theta\| + \|\hat{\theta}(N_*) - \theta\| \leq 10\eta_N(\varepsilon), \quad \forall N \geq N_*,$$

and thus, by definition of \hat{N} , we have $N_* \geq \hat{N}$. This entails, again by definition of \hat{N} , that $\|\hat{\theta}(\hat{N}) - \hat{\theta}(N_*)\| \leq 10\eta_*(\varepsilon)$ where $\eta_*(\varepsilon) = \eta_{N_*}(\varepsilon)$. Now, applying (13) with $N = N_*$ and the last inequality we conclude that

$$\|\tilde{\theta} - \theta\| = \|\hat{\theta}(\hat{N}) - \theta\| \leq \|\hat{\theta}(\hat{N}) - \hat{\theta}(N_*)\| + \|\hat{\theta}(N_*) - \theta\| \leq 15\eta_*(\varepsilon),$$

provided $\theta \in \Theta(\alpha, r, Q)$ and \mathcal{A} holds. Using this, (9), (6) and Lemmas 1, 2 from the Appendix we get, for ε small enough,

$$\begin{aligned} \sup_{\theta \in \Theta(\alpha, r, Q)} \mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 &\leq (15\eta_*(\varepsilon))^2 + \sup_{\theta \in \Theta(\alpha, r, Q)} \left(\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^4 \right)^{1/2} \mathbf{P}^{1/2}(\overline{\mathcal{A}}) \\ &\leq 225\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_*} b_k^{-4} + \sqrt{N_{\max}} e^{1/8} \varepsilon^2} \sup_{(\alpha, r, Q) \in W} \sup_{\theta \in \Theta(\alpha, r, Q)} \left(\sum_{N=1}^{N_{\max}} \mathbf{E}_\theta \|\hat{\theta}(N) - \theta\|^4 \right)^{1/2} \\ &\leq C_3(\tilde{\psi}_\alpha^2(\varepsilon) + \varepsilon^2 \log^{C_4}(1/\varepsilon)) \end{aligned}$$

where the constants $C_3 > 0, C_4 > 0$ do not depend on α, r, Q . Since

$$\sup_{(\alpha, r, Q) \in W} \frac{\varepsilon^2 \log^{C_4}(1/\varepsilon)}{\tilde{\psi}_\alpha^2(\varepsilon)} = o(1), \quad \varepsilon \rightarrow 0,$$

Theorem 2 is proved.

Proof of Theorem 3. Denote $\theta^{(0)} = (0, 0, \dots)$ the zero sequence and $\theta^{(1)} = (0, \dots, 0, v, 0, \dots)$ the sequence that has all elements 0 except for the m th element which equals $v = v_0 \tilde{\psi}_{\alpha_1}(\varepsilon)$. Here

$$m = \left\lceil \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left(q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \right\rceil$$

and v_0 is a positive number. We choose v_0 small enough to guarantee that $\theta^{(1)} \in \Theta(\alpha_1, r, Q)$. In fact, $\theta^{(1)} \in \Theta(\alpha_1, r, Q)$ if

$$v^2 m^{2r} e^{2\alpha m} \leq Q.$$

Let us check this inequality. We have, for ε small enough

$$m \leq \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left(q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \leq K_1 \log \frac{1}{\varepsilon} \quad (14)$$

where $K_1 = (3/2 + |q| + |r|)/\beta$. Hence

$$\begin{aligned} v^2 m^{2r} e^{2\alpha m} &\leq v_0^2 \tilde{\psi}_{\alpha_1}^2(\varepsilon) \left(K_1 \log \frac{1}{\varepsilon} \right)^{2r} \exp \left(\frac{2\alpha_1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{2\alpha_1}{\beta + \alpha_1} \left(q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} \right) \\ &= v_0^2 K_1^{2r} \leq Q \end{aligned}$$

for v_0 small enough. This shows that $\theta^{(1)} \in \Theta(\alpha_1, r, Q)$ for v_0 small enough. Denote P_0, P_1 the probability measures generated by observations (1) with $\theta = \theta^{(0)}$ and $\theta = \theta^{(1)}$ respectively and let $E_j = \mathbf{E}_{\theta^{(j)}}$, $j = 0, 1$. Then, for any estimator $\hat{\theta}$,

$$\begin{aligned} \max_{\alpha \in \{\alpha_1, \alpha_2\}} \sup_{\theta \in \Theta(\alpha, r, Q)} \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2 \tilde{\psi}_\alpha^{-2}(\varepsilon) &\geq \max \left\{ E_0 \|\hat{\theta}\|^2 \tilde{\psi}_{\alpha_2}^{-2}(\varepsilon), E_1 \|\hat{\theta} - \theta^{(1)}\|^2 \tilde{\psi}_{\alpha_1}^{-2}(\varepsilon) \right\} \\ &= v_0^2 \max \left\{ E_0(\lambda^2 d^2(\hat{\theta}, \theta^{(0)})), E_1(d^2(\hat{\theta}, \theta^{(1)})) \right\} \end{aligned} \quad (15)$$

where $d(\theta, \theta') = \|\theta - \theta'\| v_0^{-1} \tilde{\psi}_{\alpha_1}^{-1}(\varepsilon)$ and $\lambda = \varepsilon^{-\gamma} \left(\sqrt{\log \frac{1}{\varepsilon}} \right)^{\gamma_1}$, with $\gamma = \frac{\beta(\alpha_2 - \alpha_1)}{(\beta + \alpha_1)(\beta + \alpha_2)} > 0$, and $\gamma_1 = \frac{\beta(\alpha_1 - \alpha_2)}{(\beta + \alpha_1)(\beta + \alpha_2)} \left(r - q + \frac{1}{2} \right)$. Apply now Theorem 6(i) of Tsybakov (1998), with $M = 1$, $w(t) = t^2$, $\delta = 1/4$. Note that $d(\theta^{(0)}, \theta^{(1)}) = 1$, which yields (A.1) of Tsybakov (1998). Next, check (A.2) of Tsybakov (1998). Set $\tau = \varepsilon^\gamma$. Then, denoting ζ a standard normal random variable and assuming that v_0 is small enough, we get

$$P_1 \left(\frac{dP_0}{dP_1} \geq \tau \right) = \mathbf{P} \left(\exp \left(\zeta \varepsilon^{-1} b_m v - \frac{\varepsilon^{-2} b_m^2 v^2}{2} \right) \geq \tau \right) \geq \mathbf{P} \left(\zeta \geq -\sqrt{\frac{\gamma}{2} \log \frac{1}{\varepsilon}} \right) \rightarrow 1,$$

as $\varepsilon \rightarrow 0$. To prove the last inequality note first that

$$m \geq \frac{1}{\beta + \alpha_1} \log \frac{1}{\varepsilon} + \frac{1}{\beta + \alpha_1} \left(q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} - 1.$$

This and (14) yield

$$\begin{aligned} \varepsilon^{-1} b_m v &\leq b_{\max} v_0 \varepsilon^{-1} m^q e^{-\beta m} \tilde{\psi}_{\alpha_1}(\varepsilon) \\ &\leq b_{\max} v_0 \varepsilon^{-1} \left(K_1 \log \frac{1}{\varepsilon} \right)^q \exp \left(-\frac{\beta}{\beta + \alpha_1} \log \frac{1}{\varepsilon} - \frac{\beta}{\beta + \alpha_1} \left(q - r - \frac{1}{2} \right) \log \log \frac{1}{\varepsilon} + \beta \right) \tilde{\psi}_{\alpha_1}(\varepsilon) \\ &= K_2 v_0 \sqrt{\log \frac{1}{\varepsilon}} \end{aligned}$$

where $K_2 = b_{\max} e^\beta K_1^q$. Hence, for v_0 small enough, so that $K_2^2 v_0^2 < \gamma/2$, we obtain

$$\begin{aligned} \mathbf{P} \left(\exp \left(\zeta \varepsilon^{-1} b_m v - \frac{\varepsilon^{-2} b_m^2 v^2}{2} \right) \geq \tau \right) &= \mathbf{P} \left(\zeta \varepsilon^{-1} b_m v - \frac{\varepsilon^{-2} b_m^2 v^2}{2} \geq -\gamma \log \frac{1}{\varepsilon} \right) \\ &\geq \mathbf{P} \left(\zeta \varepsilon^{-1} b_m v \geq -\frac{\gamma}{2} \log \frac{1}{\varepsilon} \right) \geq \mathbf{P} \left(\zeta \geq -\frac{\gamma}{2K_2 v_0} \sqrt{\log \frac{1}{\varepsilon}} \right) \\ &\geq \mathbf{P} \left(\zeta \geq -\sqrt{\frac{\gamma}{2}} \log \frac{1}{\varepsilon} \right). \end{aligned}$$

Hence, for ε small enough, $P_1 \left(\frac{dP_0}{dP_1} \geq \tau \right) \geq 1/2$, which yields (A.2) of Tsybakov (1998). Thus, we can apply (A.3) of Tsybakov (1998), and we get, for ε small enough,

$$\max \left\{ E_0(\lambda^2 d^2(\hat{\theta}, \theta^{(0)})), E_1(d^2(\hat{\theta}, \theta^{(1)})) \right\} \geq \tau \lambda^2 / (32 + 8\tau \lambda^2) \rightarrow 1/8,$$

as $\varepsilon \rightarrow 0$, since $\tau \lambda^2 \rightarrow \infty$. This and (15) prove Theorem 3.

APPENDIX

Proof of (9).

We have

$$\mathbf{P}(\overline{\mathcal{A}}) \leq N_{\max} \max_{N=1, \dots, N_{\max}} \mathbf{P} \left\{ \varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2 > 17\eta_N^2(\varepsilon) \right\} \quad (16)$$

Note that for all N and all $t > 0$

$$\begin{aligned} \mathbf{P} \left\{ \sum_{k=1}^N b_k^{-2} (\xi_k^2 - 1) > t \right\} &\leq \exp(-ht) \prod_{k=1}^N \mathbf{E} \exp(hb_k^{-2} (\xi_k^2 - 1)) \\ &= \exp(-ht) \prod_{k=1}^N g(hb_k^{-2}) \end{aligned}$$

where $g(x) = e^{-x}(1 - 2x)^{-1/2}$ and it is assumed that $h \max_{1 \leq k \leq N} b_k^{-2} < 1/4$. Note that $g(x) \leq \exp(4x^2)$, $\forall 0 < x < 1/4$, and thus

$$\mathbf{P} \left\{ \sum_{k=1}^N b_k^{-2} (\xi_k^2 - 1) > t \right\} \leq \exp(-ht + 4h^2 \sum_{k=1}^N b_k^{-4}).$$

Let $h = \left(4\sqrt{\sum_{k=1}^N b_k^{-4}}\right)^{-1}$. Then $h \max_{1 \leq k \leq N} b_k^{-2} < 1/4$ and we have

$$\mathbf{P}\left\{\sum_{k=1}^N b_k^{-2}(\xi_k^2 - 1) > t\right\} \leq \exp\left(-\frac{t}{4\sqrt{\sum_{k=1}^N b_k^{-4}}} + \frac{1}{4}\right). \quad (17)$$

Now, (7) entails that for ε small enough

$$\log(1/\varepsilon) \sqrt{\sum_{k=1}^N b_k^{-4}} \geq \sum_{k=1}^N b_k^{-2}, \quad \forall N.$$

Using this and (17) we find

$$\begin{aligned} & \mathbf{P}\left\{\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2 > 17\eta_N^2(\varepsilon)\right\} \\ &= \mathbf{P}\left\{\sum_{k=1}^N b_k^{-2}(\xi_k^2 - 1) > 17\log(1/\varepsilon) \sqrt{\sum_{k=1}^N b_k^{-4}} - \sum_{k=1}^N b_k^{-2}\right\} \\ &\leq \mathbf{P}\left\{\sum_{k=1}^N b_k^{-2}(\xi_k^2 - 1) > 16\log(1/\varepsilon) \sqrt{\sum_{k=1}^N b_k^{-4}}\right\} \\ &\leq e^{1/4} \exp(-4\log(1/\varepsilon)) = e^{1/4} \varepsilon^4. \end{aligned}$$

This proves (9).

Lemma 1 *There exist constants $C_5 > 0, C_6 > 0$ such that*

$$\sup_{(\alpha, r, Q) \in W} \sup_{\theta \in \Theta(\alpha, r, Q)} \left(\sum_{N=1}^{N_{\max}} \mathbf{E}_{\theta} \|\hat{\theta}(N) - \theta\|^4 \right)^{1/2} \leq C_5 \log^{C_6}(1/\varepsilon).$$

Proof.

$$\begin{aligned} \|\hat{\theta}(N) - \theta\|^4 &= \left(\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2 + \sum_{k=N+1}^{\infty} \theta_k^2\right)^2 \\ &\leq 2\left(\left(\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2\right)^2 + \left(\sum_{k=N+1}^{\infty} \theta_k^2\right)^2\right) \\ &\leq 2\left(\left(\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2\right)^2 + Q_{\max}^2 \max_{k>N} k^{-2r} e^{-2\alpha k}\right). \end{aligned}$$

Let ε be small enough to have monotonicity of $k^{-2r}e^{-2\alpha k}$ for $k \geq N_{\max}$, whatever are $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, $|r| \leq r_{\max}$. Then

$$\max_{k > N} k^{-2r} e^{-2\alpha k} \leq N_{\max}^{-2r} e^{-2\alpha N_{\max}} + \max_{k < N_{\max}} k^{-2r} e^{-2\alpha k} \leq N_{\max}^{-2r} + \max_{k < N_{\max}} k^{-2r} \leq 2N_{\max}^{2r_{\max}}.$$

Using this we get

$$\|\hat{\theta}(N) - \theta\|^4 \leq 2((\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2)^2 + 2Q_{\max}^2 N_{\max}^{2r_{\max}}). \quad (18)$$

It is easy to see that

$$\begin{aligned} \mathbf{E}\{(\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2)^2\} &= \varepsilon^4 \sum_{k,j=1}^N b_k^{-2} b_j^{-2} \mathbf{E}(\xi_k^2 \xi_j^2) \\ &\leq 3\varepsilon^4 \sum_{k,j=1}^N b_k^{-2} b_j^{-2} = 3\varepsilon^4 \left(\sum_{k=1}^N b_k^{-2}\right)^2 \leq 3C_{b3}^2 \varepsilon^4 N^{-4q} e^{4\beta N} \end{aligned}$$

where we used (8) in the last inequality. This and (6) yield

$$\mathbf{E}\{(\varepsilon^2 \sum_{k=1}^N b_k^{-2} \xi_k^2)^2\} \leq 3C_{b3}^2 N^{4|q|} \leq 3C_{b3}^2 N_{\max}^{4|q|}. \quad (19)$$

This, together with (18) and (6) proves the lemma.

Lemma 2 *There exists a constant $C_7 > 0$ such that for any $(\alpha, r, Q) \in W$ we have*

$$\varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_*} b_k^{-4}} \leq C_7 \tilde{\psi}_\alpha^2(\varepsilon).$$

Proof. By definition of N_* ,

$$(N_* - 1)^{-2r} e^{-2\alpha(N_* - 1)} > C_0 \varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_* - 1} b_k^{-4}}.$$

This and (7) entail

$$(N_* - 1)^{-2(r-q)} e^{-2(\alpha+\beta)(N_* - 1)} > C_0 \sqrt{C_{b1}} \varepsilon^2 \log \frac{1}{\varepsilon}. \quad (20)$$

Using (7) and then (20) we get

$$\begin{aligned} \sqrt{\sum_{k=1}^{N_*} b_k^{-4}} &\leq \sqrt{C_{b2} N_*^{-2q}} e^{2\beta N_*} = \sqrt{C_{b2}} e^{2\beta} N_*^{-2q} e^{2\beta(N_*-1)} \\ &\leq \sqrt{C_{b2}} e^{2\beta} (C_0 \sqrt{C_{b1}})^{-\beta/(\alpha+\beta)} N_*^{-2q} (N_* - 1)^{-2(r-q)\beta/(\alpha+\beta)} \left(\varepsilon^2 \log \frac{1}{\varepsilon} \right)^{-\beta/(\alpha+\beta)}. \end{aligned}$$

Combining this and (10) we finally get

$$\begin{aligned} \varepsilon^2 \log \frac{1}{\varepsilon} \sqrt{\sum_{k=1}^{N_*} b_k^{-4}} &\leq C_8 (\varepsilon^2 \log \frac{1}{\varepsilon})^{\alpha/(\alpha+\beta)} N_*^{-\frac{2(\beta r + \alpha q)}{\beta + \alpha}} \\ &\leq C_8 \max\{C_1^{-\frac{2|\beta r + \alpha q|}{\beta + \alpha}} C_2^{\frac{2|\beta r + \alpha q|}{\beta + \alpha}}\} \tilde{\psi}_\alpha^2(\varepsilon) \leq C_7 \tilde{\psi}_\alpha^2(\varepsilon) \end{aligned}$$

where $C_7 > 0, C_8 > 0$ depend only on $b_{\max}, b_{\min}, q, \beta, r_{\max}, \alpha_{\min}, \alpha_{\max}, Q_{\max}$. This proves the lemma.

Acknowledgement. This research was supported by Centre for Mathematics and its Applications of Australian National University.

References

- [1] Cavalier, L. and Tsybakov A.B. (2002) Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, **123**, 323-354.
- [2] Donoho, D. L. (1995) Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition. *Journal of Applied and Computational Harmonic Analysis*, **2**, 101-126.
- [3] Efroimovich S.Yu. and Pinsker M.S. (1984) Learning algorithm for nonparametric filtering. *Automation and Remote Control*, **11**, 1434-1440.
- [4] Efroimovich S. (1997). Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. *IEEE Trans. Inform. Theory*, **43**, 1184-1191.
- [5] Efroimovich S. and Kolchinskii V. (1998) Projection estimation for inverse problems with unknown operator. Preprint.
- [6] Ermakov M.S. (1989) Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission*, **25**, 191-200.
- [7] Goldenshluger A. (1998) On pointwise adaptive nonparametric deconvolution. *Bernoulli*, **5**, 907-926.
- [8] Goldenshluger A. and Pereverzev S.V. (1999) Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. Preprint 506, WIAS, Berlin.

- [9] Golubev G.K. and Khasminskii R.Z. (1999a) Statistical approach to some inverse boundary problems for partial differential equations. *Problems of Information Transmission*, **35**, 51-66.
- [10] Golubev G.K. and Khasminskii R.Z. (1999b) Statistical approach to Cauchy problem for Laplace equation. *Festschrift for W.van Zwet*.
- [11] Johnstone I.M. (1999) Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statistica Sinica* **9**, 51-83.
- [12] Korostelev, A. P. and Tsybakov, A. B.(1993) *Minimax Theory of Image Reconstruction*. Lecture Notes in Statistics, v.82, Springer, New York.
- [13] Lepskii O.V. (1990) On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.*, **35**, 454-466.
- [14] Mair B. and Ruymgaart F.H. (1996) Statistical estimation in Hilbert scale. *SIAM J. Appl. Math.*, **56**, 1424-1444.
- [15] Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems of Information Transmission*, **16**, 120-133.
- [16] Pereverzev S.V. and Schock E. (1999) Error estimates for band-limited spherical regularization wavelets in an inverse problem of satellite geodesy. *Inverse Problems*, **15**, 881-890.
- [17] Tsybakov A.B. (1998) Pointwise and sup-norm adaptive estimation of functions on the Sobolev classes. *Annals of Statistics*, **26**, 2420-2469.

ALEXANDRE TSYBAKOV
 LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES
 UNIVERSITÉ PARIS 6
 4, PL.JUSSIEU, BP 188, 75252 PARIS, FRANCE
 EMAIL: TSYBAKOV@CCR.JUSSIEU.FR