Probab. Theory Relat. Fields 123, 323–354 (2002)
Digital Object Identifier (DOI) 10.1007/s004400100169

Laurent Cavalier · Alexandre Tsybakov

Sharp adaptation for inverse problems with random noise

Received: 19 January 2000 / Revised version: 30 April 2001 /
Published online: 14 June 2002 –

Abstract. We consider a heteroscedastic sequence space setup with polynomially increasing variances of observations that allows to treat a number of inverse problems, in particular multivariate ones. We propose an adaptive estimator that attains simultaneously exact asymptotic minimax constants on every ellipsoid of functions within a wide scale (that includes ellipoids with polynomially and exponentially decreasing axes) and, at the same time, satisfies asymptotically exact oracle inequalities within any class of linear estimates having monotone non-increasing weights. The construction of the estimator is based on a properly penalized blockwise Stein's rule, with weakly geometically increasing blocks. As an application, we construct sharp adaptive estimators in the problems of deconvolution and tomography.

1. Introduction

Let X be a separable Hilbert space with inner product $(\cdot, \cdot)$ and the norm $\|\cdot\|$. Consider the operator equation $g = Af$ where A is a known linear operator from $D \subseteq X$ into Range$(A) \subseteq X$. Inverse problem with random noise consists in statistical estimation of f from noisy observations of g. Symbolically, the statistical model can be written in the form

$$Y = Af + \varepsilon\xi, \tag{1.1}$$

where ξ is a random X-valued noise, $0 < \varepsilon < 1$ is a small parameter (the noise level) and Y is the observation. Often $D = X = L_2(T)$ where T is an interval in $\mathbb{R}^k$, $f : T \to \mathbb{R}$ and A is the integral operator defined by

$$Af(t) = \int_T K(t,x)f(x)dx, \tag{1.2}$$

L. Cavalier: CMI, Université de Provence, 39, rue F. Joliot-Curie, F-13453 Marseille, France. e-mail: `cavalier@cmi.univ-mrs.fr`

A.B. Tsybakov: Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, 4, place. Jussieu, BP 188, F-75252 Paris, France. e-mail: `tsybakov@ccr.jussieu.fr`

Mathematics Subject Classifications (2000): 62G05, 62G20

Key words or phrases: Adaptive curve estimation – Statistical inverse problems – Exact minimax constants – Oracle inequalities – Tomography – Deconvolution

where $K(t,x)$ is a given kernel. If $K(t,x) = K(t-x)$, we get a deconvolution problem. If A is a compact operator the problem is ill-posed since the inverse of A is not bounded.

In this paper ξ is gaussian, and the writing (1.1) is understood in the sense that for any $u \in X$, the random variable

$$Y(u) = (Af, u) + \varepsilon\xi(u) \tag{1.3}$$

is observable, where $\xi(u)$ is a gaussian random variable on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$, with mean 0 and variance $\|u\|^2$. We also assume that $\mathbf{E}\{\xi(u)\xi(v)\} = (u, v)$, for any $u, v \in X$, where $\mathbf{E}$ is the expectation w.r.t. $\mathbf{P}$.

The study of inverse problems with random noise was initiated in 1960-ies [Sudakov and Khalfin (1964), Bakushinskii (1969)] and has been in the focus of recent statistical literature. Several methods of statistical estimation were proposed: the Tikhonov-Phillips type regularization techniques, recursive estimation procedures in Hilbert space, projection (or Galerkin) methods [see Wahba (1977, 1990), Vapnik (1982), O'Sullivan (1986), Vainikko and Veretennikov (1986), Johnstone and Silverman (1990), Korostelev and Tsybakov (1993), Donoho (1995), Mair and Ruymgaart (1996), Efromovich and Kolchinskii (1998), Johnstone (1999), Mathé and Pereverzev (1999) and the references cited therein].

Here we deal with weighted projection methods. A natural way of projection for ill-posed problems is associated with the *singular value decomposition* (SVD) of A. Denote A^* the adjoint of A and assume that A^*A is a compact operator on X with eigenvalues $\{b_k^2\}$, $b_k > 0, k = 1, 2, \dots$, and with orthonormal system of eigenfunctions $\{\phi_k\}$. Clearly, $\|A\phi_k\| = b_k$. Set

$$\psi_k = \frac{A\phi_k}{\|A\phi_k\|} = b_k^{-1}A\phi_k.$$

The system $\{\psi_k\}$ is orthonormal. Furthermore,

$$A\phi_k = b_k\psi_k, \quad A^*\psi_k = b_k\phi_k. \tag{1.4}$$

We may also write, for any f in D,

$$Af = \sum_k b_k^{-1}(Af, \psi_k)A\phi_k = \sum_k b_k(f, \phi_k)\psi_k, \tag{1.5}$$

$$f = \sum_k b_k^{-1}(Af, \psi_k)\phi_k + u, \tag{1.6}$$

where $u \in \ker A$ and the series converge in $\|\cdot\|$. The relations (1.4) – (1.5) yield the SVD of A.

Typically (1.6) holds with $u = 0$, due to boundary or periodicity conditions. This is the case in the examples considered below, where we may write

$$f = \sum_k b_k^{-1}(Af, \psi_k)\phi_k. \tag{1.7}$$

Then the projection (or Galerkin) estimator $\hat{f}$ for f has the form of a truncated series (1.7) where the unknown coefficient (Af, ψ_k) is replaced by the observed value $y_k = Y(\psi_k)$:

$$\hat{f} = \sum_{k=1}^{n} b_k^{-1} y_k \phi_k$$

and n is the number to be chosen. One may consider a more general class of weighted projection estimates

$$\hat{f} = \sum_{k=1}^{\infty} \lambda_k b_k^{-1} y_k \phi_k \tag{1.8}$$

where λ_k are some weights, $0 \leq \lambda_k \leq 1$. In particular, a typical version of the Tikhonov-Phillips method corresponds to the weights

$$\lambda_k = \frac{1}{1 + Ck^{\tau}} \tag{1.9}$$

where $C > 0$ and $\tau > 0$. Optimizing over λ_k, in general, should produce estimators with better quality than the simple projection or Tikhonov-Phillips techniques. The quality of estimation is evaluated in terms of the mean squared risk w.r.t. the norm in X. To define the risk in a convenient form, we need some notation.

Using (1.3) and (1.4) we may write

$$y_k = b_k \theta_k + \varepsilon \xi_k, \qquad k = 1, 2, \ldots, \tag{1.10}$$

where $\xi_k = \xi(\psi_k)$ are i.i.d. standard normal random variables and $\theta_k = (f, \phi_k)$. We call (1.10) the *sequence space model* corresponding to (1.1).

Thus, we have a correspondence between (1.1) and (1.10) if the bases $\{\phi_j\}$, $\{\psi_j\}$ arise from the SVD of A. Note, however, that the model (1.10) is not confined to this case and it appears in many other situations. For example, some well-posed inverse problems with noise can be reduced to (1.10) with $b_k \to \infty$ (rather than $b_k \to 0$ characteristic for the ill-posed problems). Furthermore, the problems with direct observations and dependent noise can be reduced to the same model, see Johnstone (1999).

The mean squared risk of the linear estimator (1.8) is

$$\mathcal{R}(\hat{f}, f) = \mathbf{E}_f \|\hat{f} - f\|^2 = \mathbf{E}_\theta \left(\sum_k (\hat{\theta}_k - \theta_k)^2 \right) = \mathbf{E}_\theta \|\hat{\theta} - \theta\|^2,$$

where $\hat{\theta} = \{\hat{\theta}_k\}_{k=1}^{\infty}$, $\hat{\theta}_k = b_k^{-1} \lambda_k y_k$, $\theta = \{\theta_k\}_{k=1}^{\infty}$, and the notation $\|\cdot\|$ means the ℓ_2-norm when applied to θ-vectors in the sequence space. Here and later $\mathbf{E}_f$ and $\mathbf{E}_\theta$ denote the expectations w.r.t. Y and $y = (y_1, y_2, \ldots)$ respectively for models (1.1) and (1.10). Analyzing the risk $\mathcal{R}(\hat{f}, f)$ of the estimator (1.8) for the model (1.1) is equivalent, under our assumptions, to analyzing the risk $\mathbf{E}_\theta \|\hat{\theta} - \theta\|^2$ for the sequence space model (1.10).

The aim of this paper is twofold. First, given a class Λ of weight sequences $\{\lambda_k\}$, we propose adaptive estimators of f that mimic asymptotically the best linear oracle in Λ. The oracle inequalities are proved for an arbitrary subclass Λ of the class of all monotone non-increasing weights Λ_{mon} or piecewise constant weights Λ^*. Second, we consider the adaptive estimation of f in a minimax setting. We assume that f belongs to one of the functional classes corresponding to ellipsoids Θ in the space of coefficients $\{\theta_k\}$:

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

where $a = \{a_k\}$ is a non-negative sequence that tends to infinity, and $Q > 0$. Such classes arise naturally in various inverse problems, they include as special cases the (weighted) Sobolev classes and classes of analytic functions. We assume that the statistician does not know the parameters (a, Q) of the "true" ellipsoid, and only a general information on the possible values (a, Q) is available. This defines a scale of ellipsoids. We show that the same method of estimation guarantees sharp minimax adaptation, i.e. it achieves the exact asymptotics of minimax risk, whatever is the true ellipsoid in a given scale. The minimax results are obtained as a direct consequence of the oracle inequalities.

Minimax estimation for statistical inverse problem (1.1) (or for its sequence space analogue (1.10)) was discussed in a number of papers. Optimal rates of convergence in this problem are obtained for the L_2-risk [Johnstone and Silverman (1990), Korostelev and Tsybakov (1989, 1991, 1993), Koo (1993), Donoho (1995), Mair and Ruymgaart (1996)] and for the pointwise risk [Donoho and Low (1992), Korostelev and Tsybakov (1991, 1993), Chow, Ibragimov and Khasminskii (1999)]. Exact asymptotics of the minimax L_2-risks are known in the deconvolution problem with somewhat different setup [Ermakov (1989)], in the inverse Cauchy or Dirichlet problems for partial differential equations [Golubev and Khasminskii (1999a, b)] and in tomography, for minimax L_2-risks among linear estimators [Johnstone and Silverman (1990)]. Exact asymptotics for pointwise risks on the classes of analytic functions in tomography are due to Cavalier (1998a, b).

Adaptive minimax estimation in (1.1) has been studied quite recently. Adaptive rates of convergence under pointwise risk are analyzed by Goldenshluger (1998) (deconvolution problem) and Cavalier (1998a) (tomography). Johnstone (1999) studies adaptation in ℓ_2 by wavelet-vaguelette decomposition on the Besov scale of classes and proposes an estimator that mimics the optimal soft thresholding rule. Efromovich and Kolchinskii (1998) deal with adaptive rates for the L_2-risk when the operator A is not known and is estimated from an additional learning sample. A result on minimax adaptation in (1.1) with exact asymptotical constant among all estimators is due to Efromovich (1997) who considers the deconvolution problem with logarithmic convergence rates and supersmooth kernels (which corresponds to exponentially decreasing b_k in (1.10) and polynomially increasing a_k). Tsybakov (2000) considers the problem where both a_k and b_k are exponential and shows that the L_2-adaptive rates in this case are logarithmically worse than the optimal rates.

Here we consider the general sequence space setup (1.10) with polynomially decreasing b_k that allows to treat as special cases a number of inverse problems, in particular multivariate ones (deconvolution, tomography, inverse Cauchy problems for partial differential equations etc.). We propose an estimator that attains simultaneously exact asymptotic constants on every ellipsoid $\Theta(a, Q)$ within a wide scale (including both polynomial and exponential $\{a_k\}$) and, at the same time, satisfies asymptotically exact oracle inequalities for every $\theta \in \ell_2$ and with any class of estimates having monotone weights $\{\lambda_k\}$.

Our approach is designed for models that allow the sequence space representation (1.10). For ill-posed inverse problems it supposes the exact knowledge of the eigenfunctions ϕ_k, ψ_k in the SVD (1.4) – (1.5) and of the singular values b_k. This is the case in many problems of mathematical physics. If the SVD is not available, one can use a projection onto general Galerkin bases ϕ_k, ψ_k, which leads to the model of linear regression with growing number of parameters. This happens, for example, if the bases ϕ_k, ψ_k are imposed by the structure of a particular experiment and cannot be chosen by the statistician [see Mathé and Pereverzev (1999), Goldenshluger and Pereverzev (1999) for further discussion and results on the rates of convergence].

2. Linear minimax estimates in sequence space

Consider the sequence space model (1.10) where y_k are the observations, ξ_k are independent standard gaussian random variables, $0 < \varepsilon < 1$, $b = (b_1, b_2, \dots)$ is a known sequence, $b_k > 0,\ k = 1, 2, \dots,$ and $\theta = (\theta_1, \theta_2, \dots) \in \ell_2$ is an unknown parameter of interest.

Introduce the class of linear estimators :

$$\hat{\theta} = \hat{\theta}(h) = (\hat{\theta}_1, \hat{\theta}_2, \dots), \quad \hat{\theta}_k = h_k y_k, \quad k = 1, 2, \dots$$

where $h = (h_1, h_2, \dots)$ is an arbitrary sequence. Since $\hat{\theta}$ is uniquely determined by h we will sometimes use the name “estimator” for the sequence h itself.

The mean squared risk of the linear estimator $\hat{\theta}(h)$ has the form

$$R_\varepsilon(h, \theta) = \mathbf{E}_\theta \|\hat{\theta}(h) - \theta\|^2 = \sum_{k=1}^{\infty} ((1 - b_k h_k)^2 \theta_k^2 + \varepsilon^2 h_k^2) \tag{2.1}$$

if h is such that the right hand side is finite. For fixed $\theta \in \ell_2$ the minimum of $R_\varepsilon(h, \theta)$ is attained on the *linear oracle* $h^L = (h_1^L, h_2^L, \dots)$ where

$$h_k^L = \frac{b_k \theta_k^2}{\varepsilon^2 + b_k^2 \theta_k^2} = b_k^{-1} \frac{\theta_k^2}{\varepsilon^2 b_k^{-2} + \theta_k^2}.$$

The oracle cannot be realized from the data since it depends on the unknown θ. We write also $h^L = h^L(\theta)$.

The *linear minimax risk* $r_\varepsilon^L(\Theta)$ on the ellipsoid $\Theta = \Theta(a, Q)$ is defined by

$$r_\varepsilon^L(\Theta) = \inf_h \sup_{\theta \in \Theta} R_\varepsilon(h, \theta),$$

and the *minimax risk* $r_\varepsilon(\Theta)$ is defined by

$$r_\varepsilon(\Theta) = \inf_{\hat{t}} \sup_{\theta\in\Theta} \mathbf{E}_\theta \parallel \hat{t} - \theta \parallel^2,$$

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

The estimator $\hat{\theta}(h^*)$ is called *linear minimax estimator* on Θ if it satisfies

$$\sup_{\theta\in\Theta} R_\varepsilon(h^*, \theta) = \inf_h \sup_{\theta\in\Theta} R_\varepsilon(h, \theta).$$

Pinsker (1980) shows that linear minimax estimators on ellipsoids Θ are asymptotically minimax among all estimators. To define linear minimax estimators, introduce some notation. Let w_ε be a solution of the equation

$$\varepsilon^2 \sum_{k=1}^{\infty} b_k^{-2} a_k (1 - w_\varepsilon a_k)_+ = c_\varepsilon Q, \tag{2.2}$$

where $x_+ = \max(0, x)$. If the sequence $a_k \to \infty$ is monotone non-decreasing, the solution w_ε is unique and defined by

$$w_\varepsilon = \frac{\sum_{k=1}^n b_k^{-2} a_k}{Q\varepsilon^{-2} + \sum_{k=1}^n b_k^{-2} a_k^2}, \tag{2.3}$$

where $n = n_\varepsilon(\Theta)$ is finite integer:

$$n = \max\{k : \ a_k \le w_\varepsilon^{-1}\} = \max\{l : \ \varepsilon^2 \sum_{k=1}^{l} b_k^{-2} a_k (a_l - a_k) \le Q\}. \tag{2.4}$$

The following theorem is due to Pinsker (1980).

Theorem 1. *Let $\{a_k\}$ be a sequence of non-negative numbers, $a_k \to \infty$, and let $b_k > 0$, $k = 1, 2, \ldots$. Then the linear minimax estimator $h^* = \{h_k^*\}$ on $\Theta(a, Q)$ is given by*

$$h_k^* = b_k^{-1}(1 - w_\varepsilon a_k)_+, \tag{2.5}$$

and the linear minimax risk is

$$r_\varepsilon^L(\Theta) = \varepsilon^2 \sum_{k=1}^{\infty} b_k^{-2} (1 - w_\varepsilon a_k)_+. \tag{2.6}$$

Furthermore, if

$$\frac{\max_{k:a_k<d} b_k^{-2}}{\sum_{k:a_k<d} b_k^{-2}} = o(1), \quad d \to \infty, \tag{2.7}$$

then

$$r_\varepsilon(\Theta) = r_\varepsilon^L(\Theta)(1 + o(1)), \tag{2.8}$$

as $\varepsilon \to 0$.

Thus, under the condition (2.7), the linear minimax estimator given by (2.5) is asymptotically minimax among all estimators. Also the weights of the linear minimax estimator satisfy $h_k^* = 0$ for all $k > n_\varepsilon(\Theta)$.

3. The method of adaptation and oracle inequalities

An ideal goal of adaptation in the sequence space model (1.10) would be to find a data-driven estimator $\tilde{\theta}$ of θ that

(i) mimics asymptotically the risk of the linear oracle $h^L(\theta)$ for almost all $\theta \in \ell_2$ ("almost" means here that some θ should be obviously excluded, for example, $\theta = 0$),

and

(ii) attains asymptotically the minimax risk over any ellipsoid Θ.

We attain this goal only partly : we construct $\tilde{\theta}$ satisfying (ii) for a large scale of ellipsoids and satisfying slightly restricted versions of (i) where the linear oracle $h^L(\theta)$ is replaced by the linear monotone oracle or linear blockwise oracle with rather general blocks.

Consider the class of monotone sequences

$$\Lambda_{mon} = \{\lambda = \{\lambda_k\} \in \ell_2 : 1 \geq \lambda_1 \geq \ldots \geq \lambda_k \ldots \geq 0\}, \tag{3.1}$$

and the class of weights

$$\mathcal{H}_{mon} = \{h = \{h_k\} : h_k = b_k^{-1}\lambda_k, \{\lambda_k\} \in \Lambda_{mon}\}. \tag{3.2}$$

The linear monotone oracle $h^{mon} = h^{mon}(\theta)$ is defined as a solution of

$$R_\varepsilon(h^{mon}, \theta) = \inf_{h \in \mathcal{H}_{mon}} R_\varepsilon(h, \theta).$$

If the coefficients θ_k are monotone non-increasing, we have $h^L(\theta) = h^{mon}(\theta)$.

The class $\mathcal{H}_{mon}$ contains most of interesting examples of weight sequences $\{h_k\}$. The projection weights $h_k = b_k^{-1} I\{k \leq n\}$, where n is an integer, and the Tikhonov-Phillips weights (1.9) belong to $\mathcal{H}_{mon}$. Next, typically b_k are monotone non-increasing and a_k in the definition of the ellipsoid are monotone non-decreasing. Then the Pinsker weights (2.5) belong to $\mathcal{H}_{mon}$. It can be shown that some minimax solutions on other bodies in ℓ_2 than ellipsoids (e.g. hyperrectangles) are also in $\mathcal{H}_{mon}$.

We look for an adaptive estimator $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots)$ of the form

$$\tilde{\theta}_k = \tilde{h}_k \; y_k, \tag{3.3}$$

where $\tilde{h}_k = \tilde{h}_k(y)$ are some data-driven weights.

A well-known idea of choosing $\tilde{h}_k$ is based on the unbiased estimation of the risk (Mallows (1973), Akaike (1973) and Stein (1981)). In fact, the ℓ_2-error of the linear estimator $\hat{\theta}$ is

$$\| \hat{\theta}(h) - \theta \|^2 = \sum_k \theta_k^2 + \sum_k h_k^2 y_k^2 - 2 \sum_k h_k y_k \theta_k.$$

Thus, for any fixed h, the function

$$\mathcal{J}(h) = \sum_k (h_k^2 y_k^2 - 2h_k b_k^{-1}(y_k^2 - \varepsilon^2))$$

satisfies

$$\mathbf{E}_\theta(\mathcal{J}(h)) = \mathbf{E}_\theta\|\hat{\theta}(h) - \theta\|^2 - \sum_k \theta_k^2. \tag{3.4}$$

In other words, the function $\mathcal{J}(h)$ is (up to the summand $\sum_k \theta_k^2$ independent of h) an unbiased estimator of the risk $R_\varepsilon(h, \theta)$.

Now, given a class $\mathcal{H}$ of sequences $\{h_k\}$ (not necessarily $\mathcal{H} = \mathcal{H}_{mon}$), we may define the sequence of adaptive weight coefficients $\tilde{h}(\mathcal{H})$ as follows

$$\tilde{h}(\mathcal{H}) = \operatorname*{argmin}_{h \in \mathcal{H}} \mathcal{J}(h). \tag{3.5}$$

It is clear that $\tilde{h}(\mathcal{H})$ depends on the data $y_k,\ k = 1, 2, \ldots$, and not on the unknown parameter $\theta = \{\theta_k\}$. Denote $h^{\mathcal{H}} = h^{\mathcal{H}}(\theta)$ the oracle for the class $\mathcal{H}$:

$$R_\varepsilon(h^{\mathcal{H}}, \theta) = \inf_{h \in \mathcal{H}} R_\varepsilon(h, \theta).$$

For the problems of "direct" estimation (where $b_k \equiv 1,\ k = 1, 2, \ldots$) it is known that, under a proper choice of $\mathcal{H}$, the adaptive estimator $\tilde{\theta}$ defined by (3.3) and (3.5) achieves the required behavior : it is asymptotically minimax on ellipsoids Θ and has the asymptotic risk at least as small as that of the linear monotone oracle (Golubev (1987, 1990, 1992), Golubev and Nussbaum (1992), Oudshoorn (1997)). Other methods of adaptive weighting in the "direct" case that achieve the same properties are suggested by Efroimovich and Pinsker (1984) and Nemirovski (2000). In particular, Nemirovski (2000) uses a randomized method. Results on linear monotone oracles in the "direct" case for somewhat different setup can be found in Beran and Dümbgen (1998).

We show that, for polynomial (or quasipolynomial) $\{b_k\}$ the estimator $\tilde{\theta}$ defined by (3.3) and (3.5) can be modified to have the same adaptivity properties as in the direct case. Namely, we consider as $\mathcal{H}$ the class of coefficients with piecewise constant $\lambda_k = b_k h_k$ over suitably chosen blocks, and we apply a properly penalized Stein's rule in every block.

Define

$$\mathcal{H}^* = \{h = \{h_k\} : h_k = b_k^{-1}\lambda_k, \{\lambda_k\} \in \Lambda^*\}, \tag{3.6}$$

where Λ^* is the set of piecewise constant sequences,

$$\begin{aligned} \Lambda^* = \{\lambda \in \ell_2 : 0 \le \lambda_k \le 1, \lambda_k = \lambda_{\kappa_j}, \forall k \in [\kappa_j, \kappa_{j+1} - 1], \\ j = 0, \ldots, J-1, \lambda_k = 0, k > N\}, \end{aligned} \tag{3.7}$$

and $J, N, \kappa_j,\ j = 0, \ldots, J$, are integers such that $\kappa_0 = 1, \kappa_J = N+1, \kappa_j > \kappa_{j-1}$. Denote $I_j = \{k \in [\kappa_{j-1}, \kappa_j - 1]\}$ and $T_j = \kappa_j - \kappa_{j-1}$ for $j = 1, \ldots, J$.

Note that the solution $\tilde{h}^*$ of the minimization problem

$$\mathcal{J}(\tilde{h}^*) = \min_{h \in \mathcal{H}^*} \mathcal{J}(h)$$

is given by $\tilde{h}^* = (\tilde{h}_1^*, \tilde{h}_2^*, \ldots)$, where

$$\tilde{h}_k^* = \begin{cases} b_k^{-1}\left(1 - \frac{\sigma_j^2}{\|\bar{y}\|_{(j)}^2}\right)_+, & k \in I_j,\ j = 1, \ldots, J, \\ 0, & k > N, \end{cases}, \tag{3.8}$$

with $x_+ = \max(0, x)$,

$$\sigma_j^2 = \varepsilon^2 \sum_{k \in I_j} b_k^{-2}, \qquad \|\bar{y}\|_{(j)}^2 = \sum_{k \in I_j} \bar{y}_k^2, \tag{3.9}$$

$$\bar{y}_k = b_k^{-1} y_k = \theta_k + \varepsilon\, b_k^{-1} \xi_k, \qquad \bar{y} = \{\bar{y}_k\}, \tag{3.10}$$

and

$$\Delta_j = \frac{\max_{k \in I_j} b_k^{-2}}{\sum_{k \in I_j} b_k^{-2}}. \tag{3.11}$$

The weights (3.8) define a blockwise Stein's rule. The blockwise Stein's estimator is $\tilde{\theta}^* = (\tilde{\theta}_1^*, \tilde{\theta}_2^*, \ldots)$ where $\tilde{\theta}_k^* = \tilde{h}_k^* y_k$.

We now modify the weights $\tilde{h}^*$ and define $\tilde{h} = (\tilde{h}_1, \tilde{h}_2, \ldots)$ by

$$\tilde{h}_k = \begin{cases} b_k^{-1}\left(1 - \frac{\sigma_j^2(1+\varphi_j)}{\|\bar{y}\|_{(j)}^2}\right)_+, & k \in I_j,\ j = 1, \ldots, J, \\ 0, & k > N, \end{cases}$$

where $\varphi_j > 0$ is some penalty term.

Finally, the adaptive estimate that we propose has the form $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \ldots)$ where

$$\tilde{\theta}_k = \begin{cases} \bar{y}_k\left(1 - \frac{\sigma_j^2(1+\varphi_j)}{\|\bar{y}\|_{(j)}^2}\right)_+, & k \in I_j,\ j = 1, \ldots, J, \\ 0, & k > N. \end{cases} \tag{3.12}$$

This estimator can be interpreted as a penalized blockwise Stein's rule. The penalizing factor $(1+\varphi_j)$ forces the estimator to contain fewer nonzero coefficients $\tilde{\theta}_k$ than for the usual blockwise Stein's rule (3.8): our estimator is more "sparse". However, we consider the case where the values φ_j are small and $\max_{1 \le j \le J} \varphi_j \to 0$, as $\varepsilon \to 0$. Therefore, the difference from Stein's rule is not very strong. The choice of the penalty φ_j in the examples considered below is $\varphi_j = \Delta_j^\gamma$, where $0 < \gamma < 1/2$.

The assumption $\gamma < 1/2$ is important, as shows an inspection of the proof: $\gamma = 1/2$ already will not suffice to get the same order of remainder terms in oracle inequalities. Intuitively, this effect is easy to explain. If b_k decreases as a power of k we have: standard deviation(Z_j)/expectation(Z_j) $\sim \Delta_j^{1/2}$ where Z_j is the stochastic error term corresponding to jth block. Hence, to control the variability of stochastic terms, one needs a penalty φ_j that is slightly larger than $\Delta_j^{1/2}$. The choice $\varphi_j = \Delta_j^\gamma$, where $0 < \gamma < 1/2$, is sufficient. Other choices are possible that give similar first order asymptotics but with somewhat different remainder terms, for example, the penalty $\varphi_j = C\sqrt{\Delta_j \log \frac{1}{\Delta_j}}$, with $C > 0$ large enough (cf. Birgé and Massart (2001) for the case $b_k \equiv 1$).

Assume that the sequence $\{b_k\}$, the blocks $I_1, I_2, \dots I_J$ and the penalties φ_j satisfy the following conditions.

(A1) *There exists a constant $c_1 > 0$ independent of ε, such that*

$$\sum_{j=1}^{J} \left(\max_{k \in I_j} b_k^{-2} \right) \exp \left(- \frac{\varphi_j^2}{16 \Delta_j (1 + 2\sqrt{\varphi_j})^2} \right) \le c_1.$$

(A2) *For all $j = 1, \dots J$, we have*

$$\Delta_j \le \frac{1 - \varphi_j}{4}.$$

Assumption (A1) is satisfied if b_k are polynomially decreasing and the blocks I_j are growing sufficiently fast as j grows. On the other hand, it does not hold if b_k are exponentially decreasing, since in this case $\Delta_j \not\to 0$ as $j \to \infty$, φ_j are bounded (in view of (A2)), and thus the sum in (A1) is not bounded as $J = J(\varepsilon) \to \infty$.

A natural simplification of these assumptions would consist to suppose that the b_k are of the same order within a block I_j: $\max_{I_j} b_k^{-2} / \min_{I_j} b_k^{-2} \le C$ for a constant C. Then instead of Δ_j we can substitute in the above formulas $1/T_j$ where T_j is the size of the jth block. A motivation of the more general formulation involving Δ_j lies in multivariate applications. We will see in particular that in the context of tomography (Section 5.2) we have $\max_{I_j} b_k^{-2} / \min_{I_j} b_k^{-2} \to \infty$ and nevertheless the assumptions (A1) and (A2) hold.

The next proposition yields a first oracle inequality. It states that the adaptive estimator (3.12) is at least as good as the blockwise constant oracle $h^{\mathcal{H}^*}(\theta)$ (up to small residual terms), for any $\theta \in \ell_2$.

Proposition 1. *Let $\mathcal{H}^*$ be the class of all piecewise constant rules (3.6) and let $\tilde{\theta}$ be the estimator defined in (3.12). Assume (A1) and (A2). Then for any $\theta \in \ell_2$ and any $0 < \varepsilon < 1$ we have*

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1 + \varphi(\varepsilon)) \inf_{h \in \mathcal{H}^*} R_\varepsilon(h, \theta) + 8 c_1 \varepsilon^2,$$

where $\varphi(\varepsilon) = \max_{1 \le j \le J} (2\varphi_j + 16\Delta_j / \varphi_j)$.

Remark that we choose $\varphi_j \to 0$ as $j \to \infty$ and $J \to \infty$. Then, to satisfy (A1), one needs that $\Delta_j = o(\varphi_j^2)$. Thus, (A2) holds automatically for j large enough, and $\varphi(\varepsilon) \le C \max_{1 \le j \le J} \varphi_j$ for some $C > 0$.

Proposition 1 has the following asymptotic corollary.

Corollary 1. *Let $b_k \to 0$ as $k \to \infty$, and $\varphi(\varepsilon) \to 0$ as $\varepsilon \to 0$. Let the sequence $\theta \in \ell_2$ have the infinite number of non-zero coefficients θ_k. Then, under the assumptions of Proposition 1,*

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1 + o(1)) \inf_{h \in \mathcal{H}^*} R_\varepsilon(h, \theta),$$

as $\varepsilon \to 0$.

Proof of Proposition 1 and Corollary 1 is given in Section 6.

Remark 1. Proposition 1 and Corollary 1 deal with general blocks I_j. For example, these results apply to the wavelet setup considered by Johnstone (1999) where the blocks are the resolution levels of size $T_j = 2^j$, the coefficients b_k are constant within each block and decrease with j as a power of T_j: $b_k \equiv T_j^{-\beta}$, $\forall k \in I_j$, with some $\beta > 0$. In this case assumptions (A1) and (A2) are valid with $\varphi_j = \Delta_j^{\gamma}$, and we get that the estimator (3.12) is asymptotically at least as good as the levelwise constant wavelet oracle. Moreover, Proposition 1 allows to cover a more realistic situation where $C_1 T_j^{-\beta} \le b_k \le C_2 T_j^{-\beta}$ for some $C_2 > C_1 > 0$.

The next step of our argument is to show that an oracle inequality similar to that of Proposition 1 holds for $\tilde{\theta}$, but with $\mathcal{H}_{mon}$ in place of $\mathcal{H}^*$. This is obtained as a consequence of Proposition 1 and of Lemma 1 stated below.

The following additional assumption is needed.

(A3) *There exists* $0 < \eta_\varepsilon < 1/2$ *such that*

$$\max_{1\le j\le J-1} \frac{\sigma_{j+1}^2}{\sigma_j^2} \le 1 + \eta_\varepsilon .$$

In the examples that we consider below $\eta_\varepsilon \to 0$ as $\varepsilon \to 0$.

Lemma 1. *Let* $r > 0$, $N \ge \max\{m : \sum_{k=1}^m b_k^{-2} \le r^2\varepsilon^{-2}\eta_\varepsilon^{-2}\}$, *and let (A3) hold. If* $\|\theta\| \le r$ *then, for any* $0 < \varepsilon < 1$,

$$\inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta) \le (1-\eta_\varepsilon)^{-2} \inf_{h\in\mathcal{H}_{mon}} R_\varepsilon(h,\theta) + \sigma_1^2. \tag{3.13}$$

Furthermore, if $h \in \mathcal{H}_{mon}$ *and* $\theta \in \ell_2$ *are such that* $R_\varepsilon(h,\theta) \le r^2$, *then there exists* $\bar{h} \in \mathcal{H}^*$ *such that, for any* $0 < \varepsilon < 1$,

$$R_\varepsilon(\bar{h},\theta) \le (1-\eta_\varepsilon)^{-2} R_\varepsilon(h,\theta) + \sigma_1^2. \tag{3.14}$$

Proof of Lemma 1 is given in Section 6. It is inspired by the argument in Nemirovskii (2000), Section 6.3.3.

We need that the term σ_1^2 were small enough w.r.t. the main term $\inf_{h\in\mathcal{H}_{mon}} R_\varepsilon(h,\theta)$. If this main term is $O(\varepsilon^s)$, $\varepsilon \to 0$, for some $0 < s < 2$ and $b_k^{-1} = O(k^\beta)$, $k \to \infty$, for some $\beta > 0$, the asymptotic negligibility of σ_1^2 is easily obtained by choosing T_1 of logarithmic order.

Proposition 1 and Lemma 1 entail the following oracle inequalities.

Proposition 2. *Let* $\tilde{\theta}$ *be the estimator defined in (3.12). Assume (A1)–(A3), and let* $r > 0$, $N \ge \max\{m : \sum_{k=1}^m b_k^{-2} \le r^2\varepsilon^{-2}\eta_\varepsilon^{-2}\}$. *Then :*

(i) For any θ *such that* $\|\theta\| \le r$ *and any* $0 < \varepsilon < 1$ *we have*

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1+\Gamma_\varepsilon) \inf_{h\in\mathcal{H}_{mon}} R_\varepsilon(h,\theta) + c_2(\varepsilon^2 + \sigma_1^2), \tag{3.15}$$

where $\Gamma_\varepsilon = (2\eta_\varepsilon + \varphi(\varepsilon))/(1-2\eta_\varepsilon)$ *and* $c_2 > 0$ *does not depend on* θ, ε.

(ii) For $h \in \mathcal{H}_{mon}$ and $\theta \in \ell_2$ such that $R_\varepsilon(h,\theta) \le r^2$ and any $0 < \varepsilon < 1$ we have

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1 + \Gamma_\varepsilon) R_\varepsilon(h,\theta) + c_2(\varepsilon^2 + \sigma_1^2). \tag{3.16}$$

We introduce now a construction of *weakly geometrically increasing blocks I_j* where the conditions (A1) – (A3) are satisfied, provided b_k decreases as a power of k. In the next section we will show that with this construction the estimator (3.12) is sharp minimax adaptive on a large scale of classes.

Let ν_ε be an integer valued function of ε such that $\nu_\varepsilon \ge 2$ and $\nu_\varepsilon \to \infty$ as $\varepsilon \to 0$. A typical choice would be $\nu_\varepsilon \sim \log(1/\varepsilon)$ or $\nu_\varepsilon \sim \log\log(1/\varepsilon)$. Let

$$\rho_\varepsilon = \frac{1}{\log \nu_\varepsilon}.$$

Clearly, $\rho_\varepsilon \to 0$ as $\varepsilon \to 0$. Define the sequence $\{\kappa_j\}$ by

$$\kappa_j = \begin{cases} 1 & j = 0, \\ \nu_\varepsilon & j = 1, \\ \kappa_{j-1} + \lfloor \nu_\varepsilon \rho_\varepsilon (1+\rho_\varepsilon)^{j-1} \rfloor & j = 2, 3, \dots, \end{cases} \tag{3.17}$$

where $\lfloor x \rfloor$ is the maximal integer that is strictly less than x. Let N^* be any integer satisfying

$$N^* \ge \max\{m : \sum_{k=1}^{m} b_k^{-2} \le \varepsilon^{-2} \rho_\varepsilon^{-3}\}. \tag{3.18}$$

Then, for ε small enough, $N^* \ge \max\{m : \sum_{k=1}^m b_k^{-2} \le r^2 \varepsilon^{-2} \rho_\varepsilon^{-2}\}$, $\forall r > 0$. The following assumptions will be used: **(B1)** *The blocks are $I_j = [\kappa_{j-1}, \kappa_j - 1]$, $j = 1, \dots, J$, such that the values κ_j satisfy (3.17), and $J = \min\{j : \kappa_j > N^*\}$ where N^* satisfies (3.18).*

Clearly, $N = k_J - 1 \ge N^*$ if (B1) holds.

(B2) *The penalty is $\varphi_j = \Delta_j^\gamma$, where $0 < \gamma < 1/2$.*

We also assume that the values b_k decrease as a power of k :

(B3) *There exist $\beta \ge 0, b_{max} > 0, b_{min} > 0$ such that $b_{min} k^{-\beta} \le b_k \le b_{max} k^{-\beta}$, $k = 1, 2, \dots$*

The next result follows from Proposition 1.

Corollary 2. *Let $\mathcal{H}^*$ be defined in (3.6) and $\tilde{\theta}$ be the estimator defined in (3.12). Assume (B1) – (B3). Then for any $\theta \in \ell_2$ and any $0 < \varepsilon < \varepsilon_1$ we have*

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le \left(1 + c_3(\rho_\varepsilon \nu_\varepsilon)^{-\gamma}\right) \inf_{h \in \mathcal{H}^*} R_\varepsilon(h,\theta) + c_4 \varepsilon^2,$$

where the constants $0 < \varepsilon_1 < 1$, $c_3 > 0$ and $c_4 > 0$ do not depend on θ, ε.

Proof of Corollary 2 consists in checking the conditions (A1) and (A2) and it is given in the Appendix.

Note that (B1) – (B3) do not imply (A3), and to get oracle inequalities over $\mathcal{H}_{mon}$, as in Proposition 2, we need a stronger condition on b_k :

(B4) *The coefficients b_k are positive and there exist $\beta \geq 0, b_* > 0$ such that*

$$b_k = b_* k^{-\beta}(1 + o(1)), \ k \to \infty.$$

Corollary 3. *Let $\tilde{\theta}$ be the estimator defined in (3.12). Assume (B1),(B2) and (B4), and let $r > 0$ be fixed. Then :*

(i) For any θ such that $\|\theta\| \leq r$ and any $0 < \varepsilon < \varepsilon_2$ we have

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \leq (1 + \tilde{\eta}_\varepsilon) \inf_{h \in \mathcal{H}_{mon}} R_\varepsilon(h, \theta) + c_5 \varepsilon^2 \nu_\varepsilon^{2\beta+1},$$

where the constants $0 < \varepsilon_2 < 1$ and $c_5 > 0$ do not depend on θ, ε, and $\tilde{\eta}_\varepsilon = o(1), \ \varepsilon \to 0$, $\tilde{\eta}_\varepsilon$ does not depend on θ. (ii) For $h \in \mathcal{H}_{mon}$ and $\theta \in \ell_2$ such that $R_\varepsilon(h, \theta) \leq r^2$ and any $0 < \varepsilon < \varepsilon_2$ we have

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \leq (1 + \tilde{\eta}_\varepsilon) R_\varepsilon(h, \theta) + c_5 \varepsilon^2 \nu_\varepsilon^{2\beta+1}.$$

Proof of this corollary is given in the Appendix.

Remark 3. Since $\rho_\varepsilon \nu_\varepsilon \to \infty$ and $\tilde{\eta}_\varepsilon \to 0$, the oracle inequalities of Corollaries 2, 3 are asymptotically exact. Note also that Proposition 2 and Corollary 3 yield asymptotically exact oracle inequalities on smaller classes than $\mathcal{H}_{mon}$. In particular, our estimator $\tilde{\theta}$ is asymptotically at least as good as the optimal projection estimator i.e. estimator in the class $\mathcal{H}_{proj} = \{h : \ h_k = b_k^{-1} I(k \leq n), \ n = 1, 2, \ldots\}$ and the optimal Tikhonov-Phillips type estimator in the class $\mathcal{H}_{TP} = \{h : \ h_k = \frac{b_k^{-1}}{1+Ck^\tau}, \ \tau > 0, C > 0\}$.

Remark 4. Beran and Dümbgen (1998) show that the estimator $\tilde{h}(\mathcal{H}_{mon})$ for the case where $b_k \equiv 1$ can be computed numerically and has a piecewise-constant structure on random blocks. We believe that this is also true for general b_k and it is possible to prove a result similar to Corollary 3 for this estimator (in the "direct" case where $b_k \equiv 1$ such a result is implicit in Golubev (1990)). Although, since the values of $\tilde{h}(\mathcal{H}_{mon})$ on the blocks are monotone decreasing, this estimator will not mimic the blockwise constant oracle, unless θ has a special form, and thus it will not be useful, for example, in the wavelet context.

Remark 5. After this paper has been submitted, the paper of Cai (1999) was published that considers the estimator (3.12) for the "direct" case where $b_k \equiv 1$. The approach of Cai (1999) is different from ours: he suggests a fixed penalty φ_j for all j and logarithmically small blocks (rather than weakly geometrically increasing blocks as we do), and does not consider the oracle inequalities for monotone oracles. He studies rates of convergence rather than exact asymptotics of the risks.

4. Minimax sharp adaptation

In this section we apply the results of Section 3 to show that the estimator (3.12) with weakly geometrically increasing blocks I_j is sharp adaptive in a minimax sense on the classes of ellipsoids.

Theorem 2. *Let $\Theta = \Theta(a, Q)$ be an ellipsoid with monotone non-decreasing $a = \{a_k\}$, $a_k \to \infty$ and $Q > 0$. Let the blocks I_j satisfy (B1), the penalties satisfy (B2), and the coefficients b_k satisfy (B4). Assume also that ν_ε is chosen so that*

$$\frac{\varepsilon^2 \nu_\varepsilon^{2\beta+1}}{r_\varepsilon(\Theta)} = o(1), \quad \varepsilon \to 0. \tag{4.1}$$

Then the estimator $\tilde{\theta} = \{\tilde{\theta}_k\}$ defined in (3.12) is asymptotically minimax on Θ among all estimators, i.e.

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 = r_\varepsilon(\Theta)(1 + o(1)), \tag{4.2}$$

as $\varepsilon \to 0$.

Proof. This is a simple consequence of Corollary 3 and Theorem 1. In fact, note that under the assumptions of Theorem 2 the minimax sequence of weights h^* defined in (2.5) satisfies $h^* \in \mathcal{H}_{mon}$. Next, since a_k is monotone non-decreasing, $a_k \to \infty$, and b_k satisfies (B3), we have

$$\lim_{\varepsilon \to 0} r_\varepsilon(\Theta) = 0, \tag{4.3}$$

by Theorem 2 of Pinsker (1980). Hence,

$$\sup_{\theta \in \Theta} R_\varepsilon(h^*, \theta) = r_\varepsilon^L(\Theta) = r_\varepsilon(\Theta)(1 + o(1)) = o(1),$$

as $\varepsilon \to 0$ where we used (2.8) and (4.3). Thus, the assumptions of Corollary 3 (ii) are satisfied for $h = h^*$, $\theta \in \Theta$ and $r = 1$ if ε is small enough, and we may write

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1 + o(1)) \sup_{\theta \in \Theta} R_\varepsilon(h^*, \theta) + c_6 \varepsilon^2 \nu_\varepsilon^{2\beta+1}. \tag{4.4}$$

This, together with (4.1), yields

$$\sup_{\theta \in \Theta} \mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le r_\varepsilon^L(\Theta)(1 + o(1)),$$

which is equivalent to (4.2), in view of (2.8) and of the definition of $r_\varepsilon(\Theta)$. □

Note that Theorem 2 states a sharp adaptivity property of $\tilde{\theta}$: this estimator is sharp asymptotically minimax on every ellipsoid $\Theta = \Theta(a, Q)$ satisfying (4.1), while no prior knowledge about a and Q is required to define $\tilde{\theta}$. Moreover, no "ellipsoidal" structure appears in the definition of $\tilde{\theta}$. In fact, minimax results similar to Theorem 2 can be formulated for other classes than ellipsoids (for example, for hyperrectangles), provided the minimax solution h^* belongs to $\mathcal{H}_{mon}$.

Remark also that the condition (4.1) is quite weak. It suffices to choose ν_ε smaller than some iterated logarithm of $1/\varepsilon$, in order to satisfy these conditions for most of usual examples of ellipsoids Θ.

Corollary 4. *Let $\Theta = \Theta(a, Q)$ be any ellipsoid with monotone non-decreasing $a = \{a_k\}$ such that $k^{\alpha_0} \le a_k \le \exp(\alpha k^r)$, $\forall k$, for some $\alpha_0 > 0, \alpha > 0, r \ge 1$, $Q > 0$. Assume (B1), (B2) and (B4) with $\nu_\varepsilon = \max(\lfloor \log\log 1/\varepsilon \rfloor, 2)$. Then the estimator $\tilde\theta$ defined in (3.12) satisfies (4.2).*

Proof. The fastest convergence rate of the minimax risk $r_\varepsilon(\Theta)$ to 0 is attained for the ellipsoid Θ with $a_k = \exp(\alpha k^r)$, and this rate equals $\varepsilon^2 \left(\log \frac{1}{\varepsilon}\right)^{\frac{2\beta+1}{r}}$. Therefore, for any ellipsoid Θ satisfying the assumptions of Corollary 4, we have

$$\varepsilon^2 \left(\log \frac{1}{\varepsilon}\right)^{\frac{2\beta+1}{r}} / r_\varepsilon(\Theta) = O(1), \ \varepsilon \to 0.$$

This, together with the definition of ν_ε, yields (4.1). □

One can also get uniform results over certain scales of ellipsoids. We give now such uniform results for the Sobolev scale (polynomially increasing a_k) and the scale of classes of analytic functions (exponentially increasing a_k).

Theorem 3. *(Sobolev scale of classes). Let $\Theta_\alpha(Q) = \Theta(a, Q)$ where $a = \{k^\alpha\}$, $\alpha > 0$, $Q > 0$. Assume (B1), (B2) and (B4), with $\nu_\varepsilon = \max(\lfloor \log\log 1/\varepsilon \rfloor, 2)$. Then the estimator $\tilde\theta$ defined in (3.12) satisfies*

$$\lim_{\varepsilon \to 0} \sup_{\alpha \in [\alpha_1, \alpha_2], Q \in [Q_1, Q_2]} \sup_{\theta \in \Theta_\alpha(Q)} \frac{\mathbf{E}_\theta \|\tilde\theta - \theta\|^2}{r_\varepsilon^*(\alpha, Q)} = 1, \tag{4.5}$$

for any $0 < \alpha_1 < \alpha_2 < \infty$, $0 < Q_1 < Q_2 < \infty$, where

$$r_\varepsilon^*(\alpha, Q) = C(\alpha, Q) \varepsilon^{\frac{4\alpha}{2\alpha+2\beta+1}},$$

$$C(\alpha, Q) = (2\beta+1)^{-1} (Q(2\alpha+2\beta+1))^{\frac{2\beta+1}{2\alpha+2\beta+1}} \left(\frac{\alpha b_*^{-2}}{\alpha+2\beta+1}\right)^{\frac{2\alpha}{2\alpha+2\beta+1}}.$$

Proof. The fact that $r_\varepsilon^L(\Theta(Q)) = r_\varepsilon^*(\alpha, Q)(1 + o(1))$ is shown in Belitser and Levit (1995). It is easy to prove in the same way that somewhat stronger relation holds :

$$\sup_{(\alpha, Q) \in \mathcal{U}} \left| \frac{r_\varepsilon^L(\Theta_\alpha(Q))}{r_\varepsilon^*(\alpha, Q)} - 1 \right| = o(1), \quad \varepsilon \to 0, \tag{4.6}$$

where $\mathcal{U} = [\alpha_1, \alpha_2] \times [Q_1, Q_2]$. Moreover,

$$\sup_{(\alpha, Q) \in \mathcal{U}} \frac{\varepsilon^2 \nu_\varepsilon^{2\beta+1}}{r_\varepsilon^*(\alpha, Q)} = o(1), \quad \varepsilon \to 0. \tag{4.7}$$

Now, we apply the same argument as in the proof of Theorem 2, taking in (4.4) the weights $h^* = h^*(\alpha, Q)$ that are computed via (2.5) for $\Theta_\alpha(Q)$. Using (4.6), (4.7) and the fact that $o(1)$ and c_6 in (4.4) do not depend on α and Q, we find

$$\sup_{(\alpha, Q) \in \mathcal{U}} \sup_{\theta \in \Theta_\alpha(Q)} \frac{\mathbf{E}_\theta \|\tilde\theta - \theta\|^2}{r_\varepsilon^*(\alpha, Q)} \le (1+o(1)) \sup_{(\alpha, Q) \in \mathcal{U}} \frac{r_\varepsilon^L(\Theta_\alpha(Q))}{r_\varepsilon^*(\alpha, Q)} + o(1) = 1 + o(1),$$

as $\varepsilon \to 0$. On the other hand, using (2.8) for $\alpha = \alpha_1$, $Q = Q_1$ and (4.6), we obtain

$$\sup_{(\alpha,Q)\in\mathcal{U}} \sup_{\theta\in\Theta_\alpha(Q)} \frac{\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2}{r_\varepsilon^*(\alpha,Q)} \geq \sup_{\theta\in\Theta_{\alpha_1}(Q_1)} \frac{\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2}{r_\varepsilon^*(\alpha_1,Q_1)}$$
$$\geq \frac{r_\varepsilon(\Theta_{\alpha_1}(Q_1))}{r_\varepsilon^*(\alpha_1,Q_1)} = 1+o(1), \quad \varepsilon \to 0. \qquad \square$$

We now consider the classes of analytical and supersmooth functions that correspond to ellipsoids $\Theta(a, Q)$ with $a_k = \exp(\alpha k^r)$, $\alpha > 0, r \geq 1$.

Theorem 4. *Let $\Theta_{\alpha,r}(Q) = \Theta(\{\exp(\alpha k^r)\}, Q)$, where $\alpha > 0, r \geq 1, Q > 0$. Assume (B1), (B2) and (B4) with $\nu_\varepsilon = \max(\lfloor \log\log 1/\varepsilon \rfloor, 2)$. Then the estimator $\tilde{\theta}$ defined in (3.12) satisfies*

$$\lim_{\varepsilon\to 0} \sup_{(\alpha,r,Q)\in W} \sup_{\theta\in\Theta_{\alpha,r}(Q)} \frac{\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2}{r_\varepsilon^*(\alpha,r,Q)} = 1, \tag{4.8}$$

where

$$r_\varepsilon^*(\alpha,r,Q) = \frac{b_*^{-2}}{2\beta+1}\varepsilon^2\left(\frac{1}{\alpha}\log\frac{1}{\varepsilon}\right)^{\frac{2\beta+1}{r}}$$

and the supremum in (4.8) is taken over (α, r, Q) in $W = [\alpha_1, \alpha_2] \times [r_1, r_2] \times [Q_1, Q_2]$, where $0 < \alpha_1 < \alpha_2 < \infty$, $1 \leq r_1 < r_2 < \infty$, $0 < Q_1 < Q_2 < \infty$.

Proof. Direct calculations using (2.6) yield

$$\sup_{(\alpha,r,Q)\in W} \left| \frac{r_\varepsilon^L(\Theta_{\alpha,r}(Q))}{r_\varepsilon^*(\alpha,r,Q)} - 1 \right| = o(1), \quad \varepsilon \to 0.$$

The rest of the proof follows the same lines as in Theorem 3, and we omit it. $\square$

5. Examples

5.1. Deconvolution and estimation of derivatives

The above results allow to construct sharp adaptive estimators for deconvolution problem in Gausssian white noise. Optimal rates of convergence for this problem were obtained in different settings by Ermakov (1989), Donoho and Low (1992), Koo (1993), Korostelev and Tsybakov (1993), Donoho (1995), Johnstone (1999). Sharp asymptotics of the minimax risk in the Pinsker type framework is calculated by Ermakov (1989) and Efromovich (1997). Goldenshluger (1998) and Johnstone (1999) propose rate adaptive estimation methods for deconvolution. Efromovich (1997) shows that for the case of convolution kernels with exponentially decreasing Fourier transforms the usual projection estimate is sharp adaptive. Here we construct sharp adaptive estimators for deconvolution with polynomially decreasing Fourier coefficients of a kernel.

Consider the model

$$dY(t) = g * f(t)dt \ + \ \varepsilon \, dW(t), \ \ t \in [0, 1], \tag{5.1}$$

where we observe the process $\{Y(t), t \in [0, 1]\}$, g is a known 1-periodic filter in $L_2([0, 1])$ (convolution kernel), f is a 1-periodic signal in $L_2([0, 1])$, $0 < \varepsilon < 1$ is the level of the noise and $W(t)$ is the standard Wiener process. Let $\{\phi_k(t)\}$ be the usual trigonometric basis on $[0, 1]$:

$$\phi_1(t) \equiv 1, \quad \phi_{2k}(t) = \sqrt{2}\cos(2\pi kt), \quad \phi_{2k+1}(t) = \sqrt{2}\sin(2\pi kt), \quad k = 1, 2, \ldots. \tag{5.2}$$

The model (5.1) is equivalent to the sequence space model

$$y_k = b_k \, \theta_k \ + \ \varepsilon \, \xi_k, \ \ k = 1, 2, \ldots, \tag{5.3}$$

where $\xi_k = \int_0^1 \phi_k(t)dW(t)$ are standard normal random variables and $\theta_k = \int_0^1 f(t)\phi_k(t)dt$, $b_k = \int_0^1 g(t)\phi_k(t)dt$.

Assume that the filter g has the Fourier coefficients b_k that satisfy the assumption (B3).

Introduce the Sobolev class of functions

$$\mathcal{W}(\alpha, Q) = \{f = \sum_{k=1}^{\infty} \theta_k \phi_k : \theta \in \Theta_\alpha^*(Q)\}$$

where $\Theta_\alpha^*(Q) = \Theta(a, Q)$ with the sequence $a = \{a_k\}$ such that

$$a_k = \begin{cases} (k-1)^\alpha & \text{for } k \text{ odd}, \\ k^\alpha & \text{for } k \text{ even}, \end{cases} \qquad k = 1, 2, \ldots,$$

where $\alpha > 0$, $Q > 0$. If α is an integer, $W(\alpha, Q) = \{f : \int_0^1 (f^{(\alpha)}(t))^2 dt \le \pi^{2\alpha} Q\}$ where $f^{(\alpha)}$ denotes the weak derivative of f of order α.

Consider also the classes of infinitely many times differentiable functions

$$\mathcal{A}(\alpha, r, Q) = \{f = \sum_{k=1}^{\infty} \theta_k \phi_k : \theta \in \Theta_{\alpha,r}(Q)\}$$

where $\alpha > 0$, $r \ge 1$, $Q > 0$. The case $r = 1$ corresponds to usual classes of analytical functions.

The following result is a straightforward modification of Theorems 3, 4.

Theorem 5. *Let the Fourier coefficients b_k of the filter g satisfy (B4), the blocks I_j satisfy (B1)and the penalty satisfies (B2), with* $\nu_\varepsilon = \max(\lfloor \log\log 1/\varepsilon \rfloor, 2)$. *Then the estimator $\tilde{f} = \sum_{k=1}^\infty \tilde{\theta}_k \phi_k$, where $\tilde{\theta}$ is defined in (3.12) and $\{\phi_k\}$ is the trigonometric basis (5.2), satisfies*

$$\lim_{\varepsilon \to 0} \sup_{\alpha \in [\alpha_1, \alpha_2], Q \in [Q_1, Q_2]} \sup_{f \in \mathcal{W}(\alpha, Q)} \frac{\mathbf{E}_\theta \|\tilde{f} - f\|^2}{r_\varepsilon^*(\alpha, Q)} = 1,$$

and

$$\lim_{\varepsilon\to 0} \sup_{(\alpha,r,Q)\in W} \sup_{\theta\in\mathcal{A}(\alpha,r,Q)} \frac{\mathbf{E}_\theta\|\tilde f - f\|^2}{r_\varepsilon^*(\alpha, r, Q)} = 1,$$

for any $0 < \alpha_1 < \alpha_2 < \infty$, $0 < Q_1 < Q_2 < \infty$, *where* $r_\varepsilon^*(\alpha, Q)$, $r_\varepsilon^*(\alpha, r, Q)$, W *are as in Theorems 3 and 4.*

Thus, $\tilde f$ is a sharp adaptive estimator in minimax sense simultaneously on Sobolev classes, classes of analytical functions and classes of supersmooth functions.

A special case of Theorem 5 corresponds to the adaptive estimation of derivatives. Assume that g in (5.1) is such that the derivative $(g * f)^{(\beta)} = f$, where $\beta < \alpha$ is an integer. In view of the periodicity assumptions, this implies that the Fourier coefficients of $g * f$ have the form $b_k\theta_k$ with $b_1 = 0$, and

$$b_k = \begin{cases} (-1)^{\beta/2}\pi^{-\beta}(k-1)^{-\beta} & \text{for } k \text{ odd,} \\ (-1)^{\beta/2}\pi^{-\beta}k^{-\beta} & \text{for } k \text{ even,} \end{cases} \qquad k = 2, 3, \ldots,$$

if β is even (if β is odd, similar expression is obtained after some reordering of the Fourier coefficients). Thus, (B4) is satisfied, and Theorem 5 applies in this particular case. Note that for this case and for the scale of Sobolev classes a different method of minimax adaptive estimation is suggested by Efromovich (1998).

5.2. Tomography

The problem of tomography is to reconstruct a 2-dimensional function f from observations of its integrals over lines. This problem appears in different fields, for example, in radiology. For references see Deans (1983) and Natterer (1986). Statistical aspects of the tomography problem have been studied by Johnstone and Silverman (1990), Korostelev and Tsybakov (1989,1991,1993), Donoho and Low (1992), Cavalier (1998a, b) among others. The main models analyzed in a statistical context are positron emission tomography (a density estimation type model) and X-ray tomography (a regression type model). Here we consider the X-ray tomography problem that can be formulated as the problem of estimating f in (1.1) from the noisy data Y where A is the Radon transform operator.

Let $H = \{x \in \mathbb{R} : \|x\| \le 1\}$ be the unit disk in $\mathbb{R}^2$, and let μ denote the Lebesgue measure in $\mathbb{R}^2$. Consider the integrals of a function $f : H \to \mathbb{R}$ over all the lines that intersect H. We parametrize the lines by the length $u \in [0, 1]$ of the perpendicular from the origin to the line and by the orientation $\varphi \in [0, 2\pi)$ of this perpendicular.

Suppose that the function $f(x_1, x_2)$ belongs to $L_1(H, \mu) \cap L_2(H, \mu)$. Define the Radon transform Rf of the function f by

$$Rf(u,\varphi) = \frac{\pi}{2(1-u^2)^{\frac{1}{2}}} \int_{-\sqrt{1-u^2}}^{\sqrt{1-u^2}} f(u\cos\varphi - t\sin\varphi, u\sin\varphi + t\cos\varphi)dt,$$
$$(u,\varphi) \in S, \tag{5.4}$$

where $S = \{(u, \varphi) : 0 \le u \le 1,\ 0 \le \varphi < 2\pi\}$. With this definition, the Radon transform $Rf(u, \varphi)$ is π times the average of f over the line segment (parametrized by (u, φ)) that intersects H. It is natural to consider Rf as an element of $L_2(S, \mu_0)$ where μ_0 is the measure defined by $d\mu_0(u, \varphi) = 2\pi^{-1}(1-u^2)^{\frac{1}{2}} du\, d\varphi$.

The SVD of the Radon transform is well known (see Deans (1983) or Johnstone and Silverman (1990) for further references). To introduce it, define the set of double indices $\Gamma = \{\nu = (j, k) : j \geq 0, k \geq 0\}$. An orthonormal complex-valued basis for $L_2(H, \mu)$ is given by

$$\tilde{\phi}_\nu(r, \theta) = \pi^{-\frac{1}{2}}(j+k+1)^{\frac{1}{2}} Z_{j+k}^{|j-k|}(r) e^{i(j-k)\theta}, \ \nu = (j, k) \in \Gamma, \ (r, \theta) \in H, \quad (5.5)$$

where Z_a^b denotes the Zernike polynomial of degree a and order b. The corresponding orthonormal functions in $L_2(S, \mu_0)$ are

$$\tilde{\psi}_\nu(u, \varphi) = \pi^{-\frac{1}{2}} U_{j+k}(u) e^{i(j-k)\varphi}, \quad \nu = (j, k) \in \Gamma, \ (u, \varphi) \in S, \tag{5.6}$$

where $U_m(\cos\theta) = \sin((m+1)\theta)/\sin\theta$ are the Chebyshev polynomials of the second kind. We have $R\tilde{\phi}_\nu = b_\nu \tilde{\psi}_\nu$, with the singular values

$$b_\nu = \pi^{-1}(j+k+1)^{-\frac{1}{2}}, \quad \nu = (j, k) \in \Gamma. \tag{5.7}$$

Since we work with real functions, we identify the complex bases (5.5) and (5.6) with the equivalent real orthonormal bases $\{\phi_\nu\}$, $\{\psi_\nu\}$ in a standard way,

$$\phi_\nu = \begin{cases} \sqrt{2}\mathrm{Re}(\tilde{\phi}_\nu) & \text{if } j > k, \\ \tilde{\phi}_\nu & \text{if } j = k, \\ \sqrt{2}\mathrm{Im}(\tilde{\phi}_\nu) & \text{if } j < k. \end{cases} \tag{5.8}$$

Consider the statistical model (1.3) where $A = R$ is the Radon transform operator. Using the SVD (5.5) – (5.7), and arguing as in Section 1, we reduce (1.3) to the sequence space model

$$y_\nu = b_\nu \theta_\nu + \varepsilon\, \xi_\nu, \ \nu = (j, k), j \geq 0, k \geq 0, \tag{5.9}$$

where $\theta_\nu = (f, \phi_\nu)$, and ξ_ν are i.i.d. standard gausssian random variables.

Following Johnstone and Silverman (1990), consider the class of functions with polynomially decreasing coefficients θ_ν, i.e. the set

$$\mathcal{F}(\alpha, Q) = \left\{ f = \sum_{\nu \in \Gamma} \theta_\nu \phi_\nu : \sum_{\nu \neq 0} (j+1)^{2\alpha}(k+1)^{2\alpha}\theta_\nu^2 \leq Q \right\}.$$

Johnstone and Silverman (1990) show that $\mathcal{F}(\alpha, Q)$ can be identified with the set of functions f which have 2α weak derivatives (provided 2α is an integer) that are square integrable on H with respect to the modified dominating measure $d\mu_{2\alpha+1}(x) = (1 - \|x\|^2)^{2\alpha} d\mu(x)$. This is weaker than the square-integrability with respect to μ assumed for the usual Sobolev spaces.

Define $a_\nu = (j+1)^\alpha(k+1)^\alpha$, for $\nu = (j, k) \in \Gamma$, $j + k \neq 0$, and $a_0 = 0$. To obtain a non-decreasing sequence $\{a_\nu\}$ we order the set of indices Γ in the direction of increasing $|\nu| = (j+1)(k+1)$ (the multiple ν with the same value $|\nu|$ are ordered in the direction of increasing first coordinate j).

Define the blocks $I_\ell = \{\nu \in \Gamma : \kappa_{\ell-1} \le |\nu| \le \kappa_\ell - 1\}$, which satisfy (B1). Let $n = |G|$ where $G \in \Gamma$ is given by $G = \max\{\eta \in \Gamma : \sum_{(\eta)} b_\nu^{-2} a_\nu (a_\eta - a_\nu) \le Q\varepsilon^{-2}\}$, with $(\eta) = \{\nu \in \Gamma : 0 \le \nu \le \eta\}$ and the inequality $\nu \le \eta$ for $\nu, \eta \in \Gamma$ means the ordering as defined above.

The following lemma is an easy extension of Lemma 4.3 in Johnstone and Silverman (1990).

Lemma 2. *For $r \ge 0$, as $|\eta| \to \infty$, we have*

$$\sum_{(\eta)} (j+1)^r (k+1)^r = (r+1)^{-1} |\eta|^{r+1} \log|\eta| + O(|\eta|^{r+1}), \tag{5.10}$$

$$\sum_{(\eta)} (j+k+1)(j+1)^r (k+1)^r = \frac{\pi^2}{3} (r+2)^{-1} |\eta|^{r+2} + O(|\eta|^{r+1} \log|\eta|), \tag{5.11}$$

and

$$\sum_{(\eta)} (j+k+1)^2 (j+1)^r (k+1)^r = c(r+3)^{-1} |\eta|^{r+3} + O(|\eta|^{r+2} \log|\eta|), \tag{5.12}$$

where c is a positive constant.

Using Lemma 2 we obtain

$$\sum_{(G)} b_\nu^{-2} a_\nu (a_G - a_\nu) = \frac{\pi^4 \alpha}{3(\alpha+2)(2\alpha+2)} |G|^{2\alpha+2} + O(|G|^{2\alpha+1} \log|G|).$$

Therefore

$$n = \left(\frac{3(\alpha+2)(2\alpha+2)Q}{\pi^4 \alpha} \varepsilon^{-2} \right)^{\frac{1}{2\alpha+2}} (1 + o(1)),$$

and, according to (2.4),

$$w_\varepsilon = n^{-\alpha}(1 + o(1)).$$

Now, we compute the linear minimax risk $r_\varepsilon^L(\Theta)$. From (2.6) and Lemma 2 we get

$$r_\varepsilon^L(\Theta) = \varepsilon^2 \sum_{(G)} b_\nu^{-2} (1 - w_\varepsilon a_\nu) = r_\varepsilon^T(\alpha, Q)(1 + o(1))$$

where

$$r_\varepsilon^T(\alpha, Q) = \frac{1}{2} \left(\frac{\pi^4 \alpha}{3(\alpha+2)} \right)^{\frac{2\alpha}{2\alpha+2}} ((2\alpha+2)Q)^{\frac{2}{2\alpha+2}} \; \varepsilon^{\frac{4\alpha}{2\alpha+2}}.$$

These expressions coincide with those in Johnstone and Silverman (1990), up to constant factors. However, the model that we consider here is different: it is a regression type model, while Johnstone and Silverman (1990) study the positron emission tomography (a density type model).

Using Lemma 2 it is easy to check (2.7) and to conclude (by Theorem 1) that the optimal linear estimator attains the minimax risk among all the estimators.

Note that we cannot use the results of Section 4 to prove that $\tilde{f} = \sum_\nu \tilde{\theta}_\nu \phi_\nu$ (with $\tilde{\theta}$ as in (3.12)) is a minimax adaptive estimator on the scale of classes $\mathcal{F}(\alpha, Q)$. In fact, the conditions (B3) or (B4) are not satisfied in this two-dimensional structure. However, the conditions (A1) – (A3) hold and we can apply directly the oracle inequalities of Proposition 2. The details of this agrument are omitted: we use Lemma 2 and act as in the proof of Corollaries 2 and 3 to check (A1) – (A3). As a result, the following analog of Theorem 3 is obtained.

Theorem 6. *Let $\tilde{f} = \sum_{\nu\in\Gamma} \tilde{\theta}_\nu \phi_\nu$ be the estimator of the function f, where $\tilde{\theta}$ is defined in (3.12), $\{\phi_k\}$ is the basis (5.8) and b_ν is as in (5.7). Let the blocks I_j satisfy (B1) with $\nu_\varepsilon = \max(\lfloor \log\log 1/\varepsilon \rfloor, 2)$. Then*

$$\lim_{\varepsilon\to 0} \sup_{\alpha\in[\alpha_1,\alpha_2], Q\in[Q_1,Q_2]} \sup_{f\in\mathcal{F}(\alpha,Q)} \frac{\mathbf{E}_f \|\tilde{f} - f\|^2}{r_\varepsilon^T(\alpha, Q)} = 1,$$

for any $0<\alpha_1<\alpha_2<\infty$, $0<Q_1<Q_2<\infty$.

6. Proofs

Proof of Proposition 1. We have

$$\inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta) \geq \sum_{j=1}^J \min_{t_j} \left(\sum_{k\in I_j} ((1-b_k t_j)^2 \theta_k^2 + \varepsilon^2 t_j^2) \right) + \sum_{k>N} \theta_k^2$$

$$= \sum_{j=1}^J \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} + \sum_{k>N} \theta_k^2, \tag{6.1}$$

where $\|\theta\|_{(j)}^2 = \sum_{k\in I_j} \theta_k^2$. Also,

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 = \sum_{j=1}^J \mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 + \sum_{k>N} \theta_k^2. \tag{6.2}$$

To prove Proposition 1 we bound the risks

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 = \mathbf{E}_\theta \sum_{k\in I_j} (\tilde{\theta}_k - \theta_k)^2$$

by the respective summands in the last but one sum in (6.1), modulo small remainders terms.

Fix $j \in \{1, \ldots, J\}$. The risk $\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2$ will be upper bounded differently for the two following cases :

$$\mathbf{1}^\circ. \quad \|\theta\|_{(j)}^2 < \varphi_j \sigma_j^2/2,$$

$$\mathbf{2}^\circ. \quad \|\theta\|_{(j)}^2 \geq \varphi_j \sigma_j^2/2.$$

Bound on the risk under 1°. If 1° holds, write using Lemma 6,

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 \leq \|\theta\|_{(j)}^2 + \mathbf{E}_\theta \left(W(\|\bar{y}\|_{(j)}^2) I(\bar{A}_j) \right), \tag{6.3}$$

where

$$W(x) = -x + 2\sigma_j^2 + \frac{\sigma_j^4(4\Delta_j(1+\varphi_j) - (1-\varphi_j^2))}{x},$$

and

$$A_j = \{\|\bar{y}\|_{(j)}^2 < \sigma_j^2(1+\varphi_j)\}.$$

It is easy to see that the derivative $W'(x) < 0$ for all $x > \sigma_j^2(1+\varphi_j)$ and also $W(\sigma_j^2(1+\varphi_j)) = 4\Delta_j\sigma_j^2$. Thus,

$$\mathbf{E}_\theta \left(W(\|\bar{y}\|_{(j)}^2) I(\bar{A}_j) \right) \leq 4\Delta_j \sigma_j^2 P(\bar{A}_j). \tag{6.4}$$

On the other hand, under 1°,

$$1 - \varphi_j/2 \leq \frac{\sigma_j^2}{\sigma_j^2 + \|\theta\|_{(j)}^2}. \tag{6.5}$$

Substituting (6.4), (6.5) and the result of Lemma 4 (see the Appendix) into (6.3) we obtain

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 \leq \frac{1}{1-\varphi_j/2} \frac{\|\theta\|_{(j)}^2 \sigma_j^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} + 8\Delta_j \sigma_j^2 \exp\left(-\frac{\varphi_j^2}{16\Delta_j(1+2\sqrt{\varphi_j})^2} \right), \tag{6.6}$$

Bound on the risk under 2°. From Lemma 5 in the Appendix we get

$$\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 \leq \mathbf{E}_\theta \|\bar{\theta} - \theta\|_{(j)}^2 = \sum_{k \in I_j} \mathbf{E}_\theta (\bar{\theta}_k - \theta_k)^2,$$

where

$$\bar{\theta}_k = \bar{y}_k \left(1 - \frac{\sigma_j^2(1+\varphi_j)}{\|\bar{y}\|_{(j)}^2} \right).$$

Next, using (A.6) in the Appendix we find

$$\begin{aligned}
\mathbf{E}_\theta (\bar{\theta}_k - \theta_k)^2 &= \varepsilon^2 b_k^{-2} + 2\mathbf{E}_\theta \left((\theta_k - \bar{y}_k) \frac{\bar{y}_k \sigma_j^2 (1+\varphi_j)}{\|\bar{y}\|_{(j)}^2} \right) + \mathbf{E}_\theta \left(\frac{\bar{y}_k^2 \sigma_j^4 (1+\varphi_j)^2}{\|\bar{y}\|_{(j)}^4} \right) \\
&= \varepsilon^2 b_k^{-2} + 2\varepsilon^2 b_k^{-2} \sigma_j^2 (1+\varphi_j) \mathbf{E}_\theta \left(\frac{2\bar{y}_k^2 - \|\bar{y}\|_{(j)}^2}{\|\bar{y}\|_{(j)}^4} \right) + \mathbf{E}_\theta \left(\frac{\bar{y}_k^2 \sigma_j^4 (1+\varphi_j)^2}{\|\bar{y}\|_{(j)}^4} \right).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 \\
&\quad \le \sigma_j^2 - (1-\varphi_j^2)\sigma_j^4 \mathbf{E}_\theta \left(\frac{1}{\|\bar{y}\|_{(j)}^2} \right) + 4\varepsilon^2 \sigma_j^2 (1+\varphi_j) \mathbf{E}_\theta \left(\frac{\sum_{k\in I_j} \bar{y}_k^2 b_k^{-2}}{\|\bar{y}\|_{(j)}^4} \right) \\
&\quad \le \sigma_j^2 - (1-\varphi_j^2)\sigma_j^4 \mathbf{E}_\theta \left(\frac{1}{\|\bar{y}\|_{(j)}^2} \right) + 4\varepsilon^2 \sigma_j^2 (1+\varphi_j) \max_{k\in I_j} b_k^{-2} \mathbf{E}_\theta \left(\frac{1}{\|\bar{y}\|_{(j)}^2} \right) \\
&\quad = \sigma_j^2 - (1-\varphi_j^2 - 4\Delta_j(1+\varphi_j))\sigma_j^4 \mathbf{E}_\theta \left(\frac{1}{\|\bar{y}\|_{(j)}^2} \right),
\end{aligned}
$$

where the second term is negative in view of (A2). By Jensen's inequality

$$
\mathbf{E}_\theta \left(\frac{1}{\|\bar{y}\|_{(j)}^2} \right) \ge \frac{1}{\mathbf{E}_\theta \left(\|\bar{y}\|_{(j)}^2 \right)} = \frac{1}{\sigma_j^2 + \|\theta\|_{(j)}^2}.
$$

Thus, under 2° we have

$$
\begin{aligned}
\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 &\le \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} \left(\frac{\sigma_j^2 + \|\theta\|_{(j)}^2}{\|\theta\|_{(j)}^2} - \frac{\sigma_j^2}{\|\theta\|_{(j)}^2} + \frac{(\varphi_j^2 + 8\Delta_j)\sigma_j^2}{\|\theta\|_{(j)}^2} \right) \\
&\le \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} \left(1 + \frac{2(\varphi_j^2 + 8\Delta_j)}{\varphi_j} \right). \qquad (6.7)
\end{aligned}
$$

Final bound on the risk. We have $0 < \varphi_j < 1$, in view of (A2). Thus,

$$
\frac{1}{1-\varphi_j/2} \le 1 + \varphi_j \le 1 + \frac{2(\varphi_j^2 + 8\Delta_j)}{\varphi_j}.
$$

Using this remark and (6.6), (6.7), we obtain for any $\theta \in \ell_2$,

$$
\begin{aligned}
\mathbf{E}_\theta \|\tilde{\theta} - \theta\|_{(j)}^2 \le (1 + 2\varphi_j + 16\Delta_j/\varphi_j) \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} \\
+ 8\Delta_j \sigma_j^2 \exp\left(- \frac{\varphi_j^2}{16\Delta_j (1 + 2\sqrt{\varphi_j})^2} \right).
\end{aligned}
$$

Summing up over j and using (6.2), we find

$$
\mathbf{E}_\theta \|\tilde{\theta} - \theta\|^2 \le (1 + \varphi(\varepsilon)) \sum_{j=1}^{J} \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} + \sum_{k>N} \theta_k^2 + 8c_1\varepsilon^2.
$$

In view of (6.1), this proves Proposition 1. □

Proof of Corollary 1. Denote

$$k(\varepsilon) = \max\{k : \theta_k^2 b_k^2 \geq \varepsilon^2\}.$$

Note that $k(\varepsilon) \to \infty$, as $\varepsilon \to 0$. In fact, assume that this is not true. Then, there exists a sequence $\varepsilon_i \to 0$ (as $i \to \infty$) and an integer $M < \infty$ (independent of i) such that $\sup_i k(\varepsilon_i) \leq M$. We get $\theta_k^2 b_k^2 \leq \varepsilon_i^2,\ \forall i,\ \forall k > M$, and thus $\theta_k = 0,\ \forall k > M$, since $b_k > 0$. This contradicts the assumption that there exists the infinite number of non-zero coefficients θ_k. Next,

$$\inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta) \geq \inf_h R_\varepsilon(h,\theta) = \sum_k \frac{\varepsilon^2 b_k^{-2}\theta_k^2}{\varepsilon^2 b_k^{-2} + \theta_k^2} \geq \varepsilon^2 \frac{b_{k(\varepsilon)}^{-2}}{\varepsilon^2 b_{k(\varepsilon)}^{-2}\theta_{k(\varepsilon)}^{-2} + 1} \geq \frac{\varepsilon^2 b_{k(\varepsilon)}^{-2}}{2}.$$

Since $b_k \to 0$ as $k \to \infty$ and $\varphi(\varepsilon) \to 0$ as $\varepsilon \to 0$, we get

$$\begin{aligned}(1+\varphi(\varepsilon)) \inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta) + 8c_1\varepsilon^2 &\leq (1+\varphi(\varepsilon) + 16c_1 b_{k(\varepsilon)}^2) \inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta)\\ &= (1+o(1)) \inf_{h\in\mathcal{H}^*} R_\varepsilon(h,\theta).\end{aligned}$$

This, together with Proposition 1, proves the Corollary. □

Proof of Lemma 1. Let $h \in \mathcal{H}_{mon}$ be given. This means that $\lambda \in \Lambda_{mon}$ is given where $\lambda_k = b_k h_k$. Define

$$\bar{\lambda}_k = \begin{cases} 1 & k \in I_1, \\ \lambda_{\kappa_{j-1}} & k \in I_j,\ j = 2, \ldots, J, \\ 0 & k > N, \end{cases}$$

and $\bar{h}_k = b_k^{-1}\bar{\lambda}_k$. It suffices to show (3.14) since for the proof of (3.13) we can consider only the case where $\inf_{h\in\mathcal{H}_{mon}} R_\varepsilon(h,\theta) \leq r^2$. Indeed, the sequence $h^0 = \{h_k^0\}$ such that $h_k^0 = b_k^{-1} I\{k \in I_1\}$ satisfies $R_\varepsilon(h^0,\theta) \leq \sigma_1^2 + \|\theta\|^2 \leq \sigma_1^2 + r^2$, and $h^0 \in \mathcal{H}^*$. Hence, if $\inf_{h\in\mathcal{H}_{mon}} R_\varepsilon(h,\theta) > r^2$, (3.13) is straightforward.

Thus, assume that $h \in \mathcal{H}_{mon}$, $\theta \in \ell_2$ are such that $R_\varepsilon(h,\theta) \leq r^2$. Let us prove that

$$(1-\bar{\lambda}_k)^2 \leq (1-\eta_\varepsilon)^{-2}(1-\lambda_k)^2. \tag{6.8}$$

We first show that $\lambda_k < \eta_\varepsilon$ for $k > N$. In fact, let $M = \max\{k : \lambda_k \geq \eta_\varepsilon\}$. Then,

$$r^2 \geq R_\varepsilon(h,\theta) \geq \varepsilon^2 \sum_{k=1}^M b_k^{-2}\lambda_k^2 \geq \varepsilon^2\eta_\varepsilon^2 \sum_{k=1}^M b_k^{-2}$$

which implies $M \leq \max\{m : \sum_{k=1}^m b_k^{-2} \leq r^2\varepsilon^{-2}\eta_\varepsilon^{-2}\} \leq N$. Thus, $\lambda_k < \eta_\varepsilon$ for $k > N$. For $k > N$ we have $\bar{\lambda}_k = 0$ and thus (6.8) holds since $(1-\bar{\lambda}_k)^2 = 1 \leq (1-\eta_\varepsilon)^{-2}(1-\lambda_k)^2$.

Next, we have $0 \leq \lambda_k \leq \bar{\lambda}_k \leq 1$ for all $k \leq N$. Hence (6.8) holds for $k \leq N$.

Using (6.8) we get

$$\sum_{k=1}^\infty ((1-\bar{\lambda}_k)^2\theta_k^2 + \varepsilon^2\bar{h}_k^2) \leq (1-\eta_\varepsilon)^{-2} \sum_{k=1}^\infty (1-\lambda_k)^2\theta_k^2 + \varepsilon^2 \sum_{k=1}^\infty \bar{h}_k^2. \tag{6.9}$$

It remains to show that

$$\varepsilon^2 \sum_{k=1}^{\infty} \bar{h}_k^2 \le (1-\eta_\varepsilon)^{-2} \varepsilon^2 \sum_{k=1}^{\infty} h_k^2 + \sigma_1^2. \tag{6.10}$$

We have

$$\sum_{k=1}^{N} h_k^2 \le \sum_{k=1}^{\infty} h_k^2, \quad \sum_{k=1}^{\infty} \bar{h}_k^2 = \sum_{k=1}^{N} \bar{h}_k^2.$$

Remark that

$$\varepsilon^2 \sum_{k=1}^{N} \bar{h}_k^2 = \varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \bar{\lambda}_k^2 = \sigma_1^2 + \sum_{j=1}^{J-1} \bar{\lambda}_{\kappa_j}^2 \sigma_{j+1}^2.$$

Under assumption (A3) this gives

$$\varepsilon^2 \sum_{k=1}^{N} \bar{h}_k^2 \le (1+\eta_\varepsilon) \sum_{j=1}^{J-1} \bar{\lambda}_{\kappa_j}^2 \sigma_j^2 + \sigma_1^2 \le (1-\eta_\varepsilon)^{-2} \sum_{j=1}^{J-1} \bar{\lambda}_{\kappa_j}^2 \sigma_j^2 + \sigma_1^2. \tag{6.11}$$

Now, by monotonicity, $\lambda_k^2 \ge \lambda_{\kappa_j}^2 = \bar{\lambda}_{\kappa_j}^2$, $k \in I_j$. Hence,

$$\varepsilon^2 \sum_{k=1}^{N} h_k^2 \ge \sum_{j=1}^{J-1} \bar{\lambda}_{\kappa_j}^2 \sigma_j^2. \tag{6.12}$$

Using (6.11) and (6.12) we obtain (6.10). Finally, note that (6.9) and (6.10) entail (3.14). □

Appendix

Lemma 3. *For any $t > 0$:*

$$p_1(t) = P\left(\sum_{k \in I_j} \theta_k b_k^{-1} \xi_k \ge t \right) \le \exp\left(-\frac{t^2}{2 \sum_{k \in I_j} \theta_k^2 b_k^{-2}} \right), \tag{A.1}$$

$$p_2(t) = P\left(\sum_{k \in I_j} b_k^{-2} (\xi_k^2 - 1) \ge t \right) \le \exp\left(-\frac{t^2}{4(\sum_{k \in I_j} b_k^{-4} + t \max_{k \in I_j} b_k^{-2})} \right). \tag{A.2}$$

Proof. Inequality (A.1) is straightforward since $\sum_{k \in I_j} \theta_k b_k^{-1} \xi_k$ is a gaussian random variable with mean zero and variance $\sum_{k \in I_j} \theta_k^2 b_k^{-2}$. Inequality (A.2) is proved by a standard argument using exponential Markov inequality (see e.g. Cavalier, Golubev, Picard and Tsybakov (2000)). □.

Lemma 4. *Under the assumptions of Proposition 1 and if* 1° *holds, we have*

$$P(\bar{A}_j) \le 2\exp\left(-\frac{\varphi_j^2}{16\Delta_j(1+2\sqrt{\varphi_j})^2}\right), \; j = 1, \dots, J.$$

Proof. From the definition of $\bar{y}$ in (3.10) we have

$$\begin{aligned} P(\bar{A}_j) &= P\left(\|\bar{y}\|^2_{(j)} \ge \sigma_j^2(1+\varphi_j)\right) \\ &= P\left(2\varepsilon\sum_{k\in I_j}\theta_k b_k^{-1}\xi_k + \varepsilon^2\sum_{k\in I_j} b_k^{-2}(\xi_k^2-1) \ge \sigma_j^2\varphi_j - \|\theta\|^2_{(j)}\right). \end{aligned}$$

Using 1°, for any $0 < \delta < 1$ we get

$$\begin{aligned} P(\bar{A}_j) &\le P\left(2\varepsilon\sum_{k\in I_j}\theta_k b_k^{-1}\xi_k + \varepsilon^2\sum_{k\in I_j} b_k^{-2}(\xi_k^2-1) \ge \frac{\sigma_j^2\varphi_j}{2}\right) \\ &\le p_1\left(\frac{\delta\sigma_j^2\varphi_j}{4\varepsilon}\right) + p_2\left(\frac{(1-\delta)\sigma_j^2\varphi_j}{2\varepsilon^2}\right). \end{aligned} \tag{A.3}$$

Applying (A.1) of Lemma 3 and 1° we find

$$\begin{aligned} p_1\left(\frac{\delta\sigma_j^2\varphi_j}{4\varepsilon}\right) &\le \exp\left(-\frac{(\delta\sigma_j^2\varphi_j)^2}{32\varepsilon^2\sum_{k\in I_j}\theta_k^2 b_k^{-2}}\right) \\ &\le \exp\left(-\frac{\delta^2\|\theta\|^2_{(j)}\sigma_j^2\varphi_j}{16\varepsilon^2\max_{k\in I_j} b_k^{-2}\|\theta\|^2_{(j)}}\right) \\ &= \exp\left(-\frac{\delta^2\varphi_j}{16\Delta_j}\right), \end{aligned} \tag{A.4}$$

whenever $\|\theta\|_{(j)} \neq 0$. If $\|\theta\|_{(j)} = 0$, (A.5) is obvious.

Next, (A.2) of Lemma 3 yields

$$\begin{aligned} & p_2\left(\frac{(1-\delta)\sigma_j^2\varphi_j}{2\varepsilon^2}\right) \\ &\le \exp\left(-\frac{(1-\delta)^2\varphi_j^2\Big(\sum_{k\in I_j} b_k^{-2}\Big)^2}{16\Big(\sum_{k\in I_j} b_k^{-4} + (1-\delta)\varphi_j\sum_{k\in I_j} b_k^{-2}\max_{k\in I_j} b_k^{-2}/2\Big)}\right) \\ &\le \exp\left(-\frac{(1-\delta)^2\varphi_j^2}{16\Delta_j(1+\varphi_j/2)}\right). \end{aligned} \tag{A.5}$$

Choose

$$\delta = \frac{\sqrt{\varphi_j}}{1+2\sqrt{\varphi_j}}.$$

Then

$$\frac{(1-\delta)^2}{1+\varphi_j/2} \geq \frac{1}{(1+2\sqrt{\varphi_j})^2}.$$

Using this inequality and (A.3) – (A.5) we obtain the lemma. □

Lemma 5. *Let*

$$\bar{\theta}_k = \bar{y}_k\left(1-\frac{c}{\|\bar{y}\|^2_{(j)}}\right), \qquad \tilde{\theta}_k = \bar{y}_k\left(1-\frac{c}{\|\bar{y}\|^2_{(j)}}\right)_+, \qquad c>0.$$

Then

$$\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2_{(j)} \leq \mathbf{E}_\theta\|\bar{\theta}-\theta\|^2_{(j)}.$$

Proof of this lemma follows the same lines as the proof of Theorem 6.2 in Lehmann (1983), where the case $b_k \equiv 1$ is considered.

Lemma 6. *We have*

$$\mathbf{E}_\theta\left((\theta_k-\bar{y}_k)\frac{\bar{y}_k}{\|\bar{y}\|^2_{(j)}}\right) = \varepsilon^2 b_k^{-2}\mathbf{E}_\theta\left(\frac{2\bar{y}_k^2-\|\bar{y}\|^2_{(j)}}{\|\bar{y}\|^4_{(j)}}\right), \tag{A.6}$$

and, for any $j=1,\ldots,J$,

$$\begin{aligned}&\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2_{(j)}\\ &\leq \|\theta\|^2_{(j)}+\mathbf{E}_\theta\left[\left(-\|\bar{y}\|^2_{(j)}+2\sigma_j^2+\frac{\sigma_j^4(4\Delta_j(1+\varphi_j)-(1-\varphi_j^2))}{\|\bar{y}\|^2_{(j)}}\right)I(\bar{A}_j)\right].\end{aligned} \tag{A.7}$$

Proof. Equation (A.6) follows from integration by parts (cf. Stein (1981)). Next, using Stein's (1981) unbiased risk estimator for $\mathbf{E}_\theta(\tilde{\theta}_k-\theta_k)^2$ and summing up over $k \in I_j$ we find after some algebra

$$\begin{aligned}\mathbf{E}_\theta\|\tilde{\theta}-\theta\|^2_{(j)} &= \sigma_j^2+\mathbf{E}_\theta\left[(\|\bar{y}\|^2_{(j)}-2\sigma_j^2)I(A_j)\right]\\ &\quad+\mathbf{E}_\theta\left[\left(\frac{4\varepsilon^2\sigma_j^2(1+\varphi_j)\sum_{k\in I_j}b_k^{-2}\bar{y}_k^2}{\|\bar{y}\|^4_{(j)}}-\frac{\sigma_j^4(1-\varphi_j^2)}{\|\bar{y}\|^2_{(j)}}\right)I(\bar{A}_j)\right].\end{aligned}$$

Since

$$\varepsilon^2\sum_{k\in I_j}b_k^{-2}\bar{y}_k^2 \leq \varepsilon^2\max_{k\in I_j}b_k^{-2}\|\bar{y}\|^2_{(j)} = \sigma_j^2\Delta_j\|\bar{y}\|^2_{(j)},$$

and $\mathbf{E}_\theta\|\bar{y}\|^2_{(j)} = \|\theta\|^2_{(j)}+\sigma_j^2$, we obtain (A.7). □

Proof of Corollary 2. Suppose w.l.o.g. that ε is small enough, so that $0<\rho_\varepsilon<1$. Note that

$$\lfloor x\rfloor \geq (1-\rho_\varepsilon)x, \quad \forall x \geq \rho_\varepsilon^{-1}. \tag{A.8}$$

We first study the asymptotics of Δ_j when (B1)–(B3) hold. Clearly, as $\varepsilon \to 0$,

$$\Delta_1 = \frac{\max_{1\le k\le \nu_\varepsilon} b_k^{-2}}{\sum_{k=1}^{\nu_\varepsilon} b_k^{-2}} \le \frac{b_{min}^{-2}}{b_{max}^{-2}} \frac{\nu_\varepsilon^{2\beta}}{\sum_{k=1}^{\nu_\varepsilon} k^{2\beta}} = O\left(\frac{1}{\nu_\varepsilon}\right), \tag{A.9}$$

Next, for $j \ge 2$, assuming that $\kappa_j - 1 > \kappa_{j-1}$ (which is true for ε small enough),

$$\sum_{k\in I_j} b_k^{-2} \ge b_{max}^{-2} \sum_{k=\kappa_{j-1}}^{\kappa_j-1} k^{2\beta} \ge \kappa_{j-1}^{2\beta} b_{max}^{-2}(\kappa_j - \kappa_{j-1}). \tag{A.10}$$

Thus, for $j \ge 2$,

$$\Delta_j = \frac{\max_{k\in I_j} b_k^{-2}}{\sum_{k\in I_j} b_k^{-2}} \le \frac{b_{min}^{-2}}{b_{max}^{-2}} \left(\frac{\kappa_j}{\kappa_{j-1}}\right)^{2\beta} \frac{1}{\kappa_j - \kappa_{j-1}}. \tag{A.11}$$

Using (A.8) we find

$$\kappa_j - \kappa_{j-1} = \lfloor \nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^{j-1}\rfloor \ge (1-\rho_\varepsilon)\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^{j-1}, \tag{A.12}$$

provided ε is small enough, so that $\nu_\varepsilon\rho_\varepsilon \ge \rho_\varepsilon^{-1}$.

We now show that the ratio κ_j/κ_{j-1} in (A.11) is bounded. By definition, for $j \ge 2$

$$\kappa_j = \nu_\varepsilon + \sum_{s=2}^{j} \lfloor \nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^{s-1}\rfloor \le \nu_\varepsilon\left(1 + \rho_\varepsilon\sum_{s=1}^{j-1}(1+\rho_\varepsilon)^s\right) = \nu_\varepsilon(1+\rho_\varepsilon)^j, \tag{A.13}$$

and, using (A.8) (we suppose that ε is small enough, so that $\nu_\varepsilon\rho_\varepsilon \ge \rho_\varepsilon^{-1}$),

$$\kappa_{j-1} = \nu_\varepsilon + \sum_{s=2}^{j-1} \lfloor \nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^{s-1}\rfloor \ge \nu_\varepsilon(1+\rho_\varepsilon)^j\left(\frac{1-\rho_\varepsilon}{1+\rho_\varepsilon}\right).$$

Thus,

$$\left(\frac{\kappa_j}{\kappa_{j-1}}\right)^{2\beta} \le \left(\frac{1+\rho_\varepsilon}{1-\rho_\varepsilon}\right)^{2\beta} = 1 + o(1), \quad \varepsilon \to 0. \tag{A.14}$$

This, together with (A.11)–(A.12) yields $\max_{2\le j\le J}\Delta_j = O\left(\frac{1}{\rho_\varepsilon\nu_\varepsilon}\right)$, $\varepsilon \to 0$. Taking into account (A.9) and (B2), we find that (A2) is satisfied for ε small enough, and

$$\varphi(\varepsilon) = \max_{1\le j\le J}(2\Delta_j^\gamma + 16\Delta_j^{1-\gamma}) = O((\rho_\varepsilon\nu_\varepsilon)^{-\gamma}),\ \varepsilon \to 0. \tag{A.15}$$

We now check that (A1) holds for ε small enough, whenever (B1)–(B3) hold.

Fix an arbitrary $C > 0$. For $j = 1$, using (A.9), we have $\max_{k\in I_1} b_k^{-2} \exp(-C\Delta_1^{2\gamma-1}) \le b_{min}^{-2}\nu_\varepsilon^{2\beta}\exp(-C\Delta_1^{2\gamma-1}) \to 0,\ \varepsilon \to 0$, for any $C > 0, 0 < \gamma < 1/2$. Now, if $j \ge 2$, using (A.11), (A.12), and (A.14) we find, for ε small enough,

$$\Delta_j^{2\gamma-1} \ge (\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^j/d_1)^{1-2\gamma},$$

and, by (B3) and (A.13),

$$\max_{k\in I_j} b_k^{-2} \le b_{min}^{-2}(\nu_\varepsilon(1+\rho_\varepsilon)^j)^{2\beta} = \rho_\varepsilon^{-2\beta} b_{min}^{-2}(\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^j)^{2\beta},$$

where $d_1 > 0$ is a constant. Thus, there exist constants $d_2 > 0$ and $\varepsilon_0 \in (0,1)$ such that for all $0 < \varepsilon < \varepsilon_0$ we have $\max_{k\in I_j} b_k^{-2}\exp(-C\Delta_j^{2\gamma-1}) \le \rho_\varepsilon^{-2\beta}$ $\exp(-d_2\Delta_j^{2\gamma-1}),\ j \ge 2$, and

$$\sum_{j=2}^{J}\left(\max_{k\in I_j} b_k^{-2}\right)\exp(-C\Delta_j^{2\gamma-1})$$
$$\le \rho_\varepsilon^{-2\beta}\sum_{j=1}^{\infty}\exp\left(-d_2(\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^j/d_1)^{1-2\gamma}\right) = o(1),$$

as $\varepsilon \to 0$, since $(1+\rho_\varepsilon)^j \ge j\rho_\varepsilon$ and in view of the definition of $\rho_\varepsilon, \nu_\varepsilon$. We see that (A1) is satisfied, and even more: the sum in (A1) is $o(1)$, as $\varepsilon \to 0$.

Thus, we have shown that the assumptions of Proposition 1 are satisfied under the assumptions of Corollary 2. Hence, the oracle inequality of Proposition 1 holds with $\varphi(\varepsilon)$ satisfying (A.15). This yields the inequality of Corollary 2. □

Proof of Corollary 3. Using Proposition 2, it suffices to check the conditions (A1) – (A3) and to show that

$$\eta_\varepsilon = o(1), \quad \sigma_1^2 = O(\varepsilon^2\nu_\varepsilon^{2\beta+1}),\ \varepsilon \to 0. \tag{A.16}$$

But (A1) and (A2) have already been checked in the previous proof, and the second relation in (A.16) is straightforward. Thus, it remains to show that (A3) holds with $\eta_\varepsilon = o(1)$.

First,

$$\frac{\sigma_2^2}{\sigma_1^2} = O\left(\frac{\kappa_2^{2\beta}(\kappa_2-\kappa_1)}{\kappa_1^{2\beta+1}}\right) = O(\rho_\varepsilon) = o(1), \quad \varepsilon \to 0. \tag{A.17}$$

Now, in view of (B4) there exists $\alpha(\varepsilon) > 0,\ \alpha(\varepsilon) \to 0$, independent of k, such that $|b_k^{-2} - b_*^{-2}k^{2\beta}| \le b_*^{-2}k^{2\beta}\alpha(\varepsilon)$ for all $k \ge \nu_\varepsilon$. We assume that ε is small enough, so that $\alpha(\varepsilon) < 1$. If $k \in I_j,\ j \ge 2$, we have $k \ge \nu_\varepsilon$, and thus for $j \ge 2$,

$$\frac{\sigma_{j+1}^2}{\sigma_j^2} \le \left(\frac{1+\alpha(\varepsilon)}{1-\alpha(\varepsilon)}\right)\frac{\sum_{k\in I_{j+1}}k^{2\beta}}{\sum_{k\in I_j}k^{2\beta}} \le \frac{\kappa_{j+1}^{2\beta}(\kappa_{j+1}-\kappa_j)}{\kappa_{j-1}^{2\beta}(\kappa_j-\kappa_{j-1})}(1+o(1)), \quad \varepsilon \to 0. \tag{A.18}$$

Using (A.12) and (A.14) we obtain

$$\left(\frac{\kappa_{j+1}}{\kappa_{j-1}}\right)^{2\beta}\frac{\kappa_{j+1}-\kappa_j}{\kappa_j-\kappa_{j-1}} \le \left(\frac{1+\rho_\varepsilon}{1-\rho_\varepsilon}\right)^{4\beta}$$
$$\times\frac{\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^j}{(1-\rho_\varepsilon)\nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)^{j-1}} = 1+o(1), \quad \varepsilon \to 0.$$

This and (A.17) – (A.18) entail (A3) with $\eta_\varepsilon = o(1)$. □

References

1. Akaike, H.: *Information theory and an extension of the maximum likelihood principle*, Proc. 2nd Intern. Symp. Inf. Theory, Petrov P.N. and Csaki F. eds. Budapest, 267–281 (1973)
2. Bakushinskii, A.B.: On the construction of regularizing algorithms under random noise, *Soviet Math. Doklady.*, **189**, 231–233 (1969)
3. Belitser, E.N., Levit, B.Ya.: On minimax filtering on ellipsoids, *Math. Meth. Statist.*, **4**, 259–273 (1995)
4. Beran, R., Dümbgen, L.: Modulation estimators and confidence sets, *Annals of Statistics.*, **26**, 1826–1856 (1998)
5. Birgé, L., Massart, P.: Gaussian model selection, *JEMS.*, to appear (2001)
6. Cai, T.: Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *Annals of Statistics.*, **27**, 2607–2625 (1999)
7. Cavalier, L.: Estimation non paramétrique pour deux problèmes liés à l'image: ensembles de niveau et tomographie. PhD Thesis, Université Paris VI. (1998a)
8. Cavalier, L.: Asymptotically efficient estimation in a problem related to tomography, *Math. Meth. Statist.*, **7**, 445–456 (1998b)
9. Cavalier, L., Golubev, Yu., Picard, D., Tsybakov, A.B.: Oracle inequalities for inverse problems. Preprint. Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6 - Paris 7. http://www.proba.jussieu.fr/mathdoc/preprints/index.html# 2000
10. Chow, P.-L., Ibragimov, I.A., Khasminskii, R.Z.: Statistical approach to some ill-posed problems for linear partial differential equations, *Probab. Theory and Related Fields.*, **113**, 421–441 (1999)
11. Deans, S.R.: *The Radon Transform and some of its Applications.*, Wiley, N.Y. (1983)
12. Donoho, D.L.: Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition, *Journal of Applied and Computational Harmonic Analysis.*, **2**, 101–126 (1995)
13. Donoho, D., Low, M.: Renormalization exponents and optimal pointwise rates of convergence, *Ann. Statist.*, **20**, 944–970 (1992)
14. Efromovich, S.: 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 (1997)
15. Efromovich, S.: Simultaneous sharp adaptive estimation of functions and their derivatives, *Ann. Statist.*, **26**, 273–278 (1998)
16. Efromovich, S., Kolchinskii, V.: Projection estimation for inverse problems with unknown operator. Manuscript (1998)
17. Efroimovich, S.Yu., Pinsker, M.S.: Learning algorithm for nonparametric filtering, *Automation and Remote Control.*, **11**, 1434–1440 (1984)
18. Ermakov, M.S.: Minimax estimation of the solution of an ill-posed convolution type problem, *Problems of Information Transmission.*, **25**, 191–200 (1989)
19. Goldenshluger, A.: On pointwise adaptive nonparametric deconvolution, *Bernoulli.*, **5**, 907–926 (1998)
20. Goldenshluger, A., Pereverzev, S.V.: Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations, *Probability Theory and Related Fields.*, **118**, 169–186 (1999)
21. Golubev, G.K.: Adaptive asymptotically minimax estimates of smooth signals, *Problems of Information Transmission.*, **23**, 57–67 (1987)
22. Golubev, G.K.: Quasi-linear estimates of signals in L_2, *Problems of Information Transmission*, **26**, 15–20 (1990)

23. Golubev, G.K.: Nonparametric estimation of smooth probability densities in L_2, *Problems of Information Transmission.*, **28**, 44–54, (1992)
24. Golubev, G.K., Nussbaum, M.: Adaptive spline estimates in a nonparametric regression model, *Theory of Probability and its Applications.*, **37**, 521–529 (1992)
25. Golubev, G.K., Khasminskii, R.Z.: Statistical approach to some inverse boundary problems for partial differential equations, *Problems of Information Transmission.*, **35**, 51–66 (1999a)
26. Golubev, G.K., Khasminskii, R.Z.: Statistical approach to Cauchy problem for Laplace equation, In: *Festschrift for W.van Zwet.*, IMS Lecture Notes-Monograph Series, **36**, (1999b)
27. Johnstone, I.M.: Wavelet shrinkage for correlated data and inverse problems: adaptivity results, *Statistica Sinica.*, **9**, 51–83 (1999)
28. Johnstone, I.M., Silverman, B.W.: Speed of estimation in positron emission tomography and related inverse problems, *Annals of Statistics.*, **18**, 251–280 (1990)
29. Koo J.Y.: Optimal rates of convergence for nonparametric statistical inverse problems, *Annals of Statistics.*, **21**, 590–599 (1993)
30. Korostelev, A.P., Tsybakov, A.B.: The minimax accuracy of image reconstruction in the problem of tomography. In: Fifth Vilnius Conf. on Probability and Statist. Abstracts, vol.1, 270–271 (1989)
31. Korostelev, A.P., Tsybakov, A.B.: Optimal rates of convergence of estimators in a probabilistic setup of tomography problem, *Problems of Information Transmission.*, **27**, 73–81 (1991)
32. Korostelev, A.P., Tsybakov, A.B.: *Minimax Theory of Image Reconstruction.*, Lecture Notes in Statistics, v.82, Springer, New York. (1993)
33. Lehmann, E.L. *Theory of Point Estimation.*, Wiley, N.Y. e.a. (1983)
34. Mair, B., Ruymgaart, F.H.: Statistical estimation in Hilbert scale, *SIAM J. Appl. Math.*, **56**, 1424–1444 (1996)
35. Mathé, P., Pereverzev, S.V.: Optimal discretization and degrees of ill-posedness for inverse estimation in Hilbert scales in the presence of random noise. Preprint 469, WIAS - Berlin. (1999)
36. Mallows, C.L.: Some comments on C_p, *Technometrics.*, **15**, 661–675 (1973)
37. Natterer, F.: *The Mathematics of Computerized Tomography.*, Teubner, Stuttgart. (1986)
38. Nemirovski, A.: *Topics in Non-Parametric Statistics.*, In: Lectures on Probability Theory and Statistics. Ecole d'Eté de Probabilités de St. Flour XXVIII - 1998. Lecture Notes in Mathematics, v.1738. Springer, N.Y. e.a. (2000)
39. O'Sullivan, F.: A statistical perspective on ill-posed problems, *Statist. Science.*, **1**, 502–527 (1986)
40. Oudshoorn, K.: Asymptotically minimax estimation of a function with jumps, *Bernoulli.*, **4**, 15–33 (1997)
41. Pinsker, M.S.: Optimal filtering of square integrable signals in Gaussian white noise, *Problems of Information Transmission.*, **16**, 120–133 (1980)
42. Stein, C.M.: Estimation of the mean of a multivariate normal distribution, *Ann. Statist.*, **9**, 1135–1151 (1981)
43. Sudakov, V.N., Khalfin, L.A.: Statistical approach to ill-posed problems in mathematical physics, *Soviet Math. Doklady.*, **157**, 1094–1096 (1964)
44. Tsybakov, A.B.: On the best rate of adaptive estimation in some inverse problems, *C.R. Acad. Sci. Paris, sér. 1.*, **330**, 835–840 (2000)
45. Vainikko, G., Veretennikov, A.Yu.: *Iteration Procedures in Ill-Posed Problems.*, Nauka, Moscow (in Russian) (1986)
46. Vapnik, V.N.: *Estimation of Dependencies Based on Empirical Data.*, Springer, N.Y. (1982)

47. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy, *SIAM J. Numer. Anal.*, **14**, 651–667 (1977)
48. Wahba, G.: *Spline Models for Observational Data*., SIAM, Philadelphia. (1990)