DOI 10.1007/s10958-020-04651-1
*Journal of Mathematical Sciences, Vol. 244, No. 5, February, 2020*

# MINIMAX NONPARAMETRIC ESTIMATION ON MAXISETS

**M. Ermakov***

UDC 519.2

*We study nonparametric estimation of a signal in Gaussian white noise on maxisets. We point out minimax estimators in the class of all linear estimators and strong asymptotically minimax estimators in the class of all estimators. We show that balls in Sobolev spaces are maxisets for the Pinsker estimators. Bibliography: 22 titles.*

## 1. Introduction

One of the most popular models of nonparametric estimation is nonparametric estimation of a signal in Gaussian white noise. This problem has been explored in numerous papers for a wide range of functional spaces and for completely different setups (see [4, 16, 9, 21] and references therein). Strong asymptotically minimax estimators are known for this setup only if a priori information is provided that a signal belongs to an ellipsoid in $\mathbb{L}_2$ [13, 18, 9, 21, 17], balls in $\mathbb{L}_\infty$ [1, 3, 12, 14], or to some bodies in Besov spaces defined in terms of wavelets [9]. The goal of this paper is to pay attention to the fact that strong asymptotically minimax estimators can also be obtained for other sets of functions. The definition of these sets coincides with the definition of a ball in the Besov space $B^\alpha_{2\infty}$ in terms of a trigonometric system of functions and some norm (see [19]). We denote these sets by $\mathbb{B}(\alpha, P_0)$ with $\alpha > 0$ and $P_0 > 0$.

The balls $B^\alpha_{2\infty}(P_0)$ have remarkable properties in nonparametric estimation. These sets carry a rather reasonable information on the signal smoothness:

*on these sets, the most known linear nonparametric estimators have given rates of convergence* [10, 11]*;*

*for linear statistical estimators, these sets are the largest sets with a given rate of convergence* [10, 11, 15, 19].

The appearing strong asymptotically minimax estimators are penalized maximum likelihood estimators for some quadratic penalty function [5, 22]. Thus, we conclude that likelihood estimation with a quadratic penalty function is optimal not only in the Bayes sense but in the minimax sense as well. These asymptotically minimax estimators are also trigonometric spline estimators [9, 21, 22]. Results of this paper can also be treated as a solution of the inverse problem: For the Bayes estimators and maximum penalized likelihood estimators, one needs to find the largest nonparametric sets such that these estimators are asymptotically minimax on these sets.

The nonasymptotic setup is also explored. In this setup, we show that our estimator is minimax on maxisets for the class of all linear estimators.

The results can easily be extended to the setup of minimax estimation of solutions of linear inverse ill-posed problem. For this setup, a minimax estimator can be treated as some version of the Tikhonov regularization algorithm [20].

We show that the order of rates of convergence of the Pinsker estimator on $\mathbb{B}(\alpha, P_0)$ is worse than the order of rates of convergence of widespread linear estimators. We prove that balls in Sobolev spaces are maxisets for the Pinsker estimators.

*Institute of Problems of Mechanical Engineering of RAS, St.Petersburg State University, St.Petersburg, Russia, e-mail: `erm2512@gmail.com`.

Translated from *Zapiski Nauchnykh Seminarov POMI*, Vol. 466, 2017, pp. 120–133. Original article submitted November 14, 2017.

The results are provided in terms of a sequence model. Let we observe a random sequence $y = \{y_j\}_{j=1}^{\infty}$,

$$y_j = x_j + \varepsilon\,\sigma_j\,\xi_j, \quad \varepsilon > 0, \quad 1 \le j < \infty,$$

where $\sigma_j > 0$ are known constants, $\xi_j, 1 \le j < \infty$, are independent Gaussian random variables, $\mathbf{E}\xi_j = 0$, and $\mathbf{E}\xi_j^2 = 1$.

The problem is to estimate the parameter $x = \{x_j\}_{j=1}^{\infty}$.

Denote $\sigma = \{\sigma_j\}_{j=1}^{\infty}$ and $\xi = \{\xi_j\}_{j=1}^{\infty}$.

For estimation with a fixed $\varepsilon > 0$, minimax estimators in the class of all linear estimators are obtained if a priori information is provided in the following form:

$$x \in \mathbb{B}(a, P_0) = \Big\{ x = \{x_i\}_{i=1}^{\infty} \,:\, \sup_k a_k^{-1} \sum_{j=k}^{\infty} x_j^2 \le P_0 \Big\}, \tag{1.1}$$

where $a = \{a_k\}_{k=1}^{\infty}$ and $a_k > 0$ is a decreasing sequence.

Asymptotically minimax estimators in the class of all estimators will be obtained if a priori information is provided that the signal belongs to the sets $\mathbb{B}(\alpha, P_0) = \mathbb{B}(\tilde{a}, P_0)$ with $\tilde{a} = \{k^{-2\alpha}\}$, $\alpha > 0$. The analysis of the proof shows that the results can be extended to other sequences $a_k$. However, this requires more accurate reasoning. For trigonometric orthogonal system of functions, $\sup_k a_k^{-1} \sum_{j=k}^{\infty} x_j^2$ can be considered as the square of some norm in the Besov space $\mathbb{B}^{\alpha}_{2\infty}$. For Besov bodies in $\mathbb{B}^{r}_{2\infty}$ generated by wavelets, asymptotically minimax estimators has been obtained by Johnstone [9]. For this setup, the proof is reduced to another extremal problem, and the solution of this problem is completely different.

There are numerous results on strong adaptive asymptotically minimax estimation [9, 21]. Note that the results on adaptive estimation in the Pinsker model [9, 21] are easily carried over to the setup of this paper with the sets $\mathbb{B}(\alpha, P_0)$.

Below we recall the definition of maxisets.

For an estimator $\hat{x}_\varepsilon$, for the loss function $\|\hat{x}_\varepsilon - x\|^2$, for rates of convergence $\varepsilon^\gamma, \gamma > 0$, and for a constant $C > 0$, the maxiset is

$$\mathbb{MS}(\hat{x}_\varepsilon, \gamma)(C) = \{\, x : \sup_\varepsilon \varepsilon^{-2\gamma} \mathbf{E}_x \|\hat{x}_\varepsilon - x\|^2 < C \,\}.$$

Here $\|x\|$ denotes the norm of a vector $x = \{x_j\}_{j=1}^{\infty}$ in a Hilbert space, $\|x\|^2 = \sum_{j=1}^{\infty} x_j^2$.

In what follows, we denote by $c$ and $C$ positive constants and write $a_\varepsilon \asymp b_\varepsilon$ if $c < a_\varepsilon / b_\varepsilon < C$ for all $\varepsilon > 0$.

## 2. Main results

We say that a linear estimator $\hat{x}_\varepsilon = \{\hat{x}_{\varepsilon j}\}_{j=1}^{\infty}$ is minimax in the class of linear estimators $\hat{x}_{\varepsilon\lambda} = \{\hat{x}_{\varepsilon j \lambda_j}\}_{j=1}^{\infty}$, $\hat{x}_{\varepsilon j \lambda_j} = \lambda_j\, y_j$, $\lambda_j \in \mathbb{R}^1$, $1 \le j < \infty$, $\lambda = \{\lambda_j\}_{j=1}^{\infty}$ if

$$R_{l\varepsilon} \doteq \sup_{x \in \mathbb{B}(a,P_0)} \mathbf{E}_x \,\|\hat{x}_\varepsilon - x\|^2 = \inf_\lambda \sup_{x \in \mathbb{B}(a,P_0)} \mathbf{E}_x \,\|\hat{x}_{\varepsilon\lambda} - x\|^2. \tag{2.1}$$

We say that an estimator $\hat{x}_\varepsilon$ is asymptotically minimax if

$$R_\varepsilon \doteq \sup_{x \in B(\alpha,P_0)} \mathbf{E}_x \,\|\hat{x}_\varepsilon - x\|^2 = \inf_{\tilde{x}_\varepsilon \in \Psi} \sup_{x \in B(\alpha,P_0)} \mathbf{E}_x \,\|\tilde{x}_\varepsilon - x\|^2 \,(1 + o(1)) \tag{2.2}$$

as $\varepsilon \to 0$. Here $\Psi$ is the set of all estimators.

The minimax estimator in the class of linear estimators is obtained under the following assumptions.

**A1**. There is a $c > 0$ such that $c < \sigma_j^2 < \infty$ for all $j$.

**A2**. For all $j > 1$,

$$\frac{\sigma_j^2\,(a_{j-1} - a_j)}{\sigma_{j-1}^2\,(a_j - a_{j+1})} > 1. \tag{2.3}$$

This implies that the sequence $\sigma_j^2(a_{j-1} - a_j)$ is strictly increasing.

**Theorem 2.1.** *Let assumptions* A1 *and* A2 *be satisfied. Then the linear estimator* $\hat{x}_\lambda$ *with*

$$\lambda_j = \frac{P_0\,(a_j - a_{j+1})}{P_0\,(a_j - a_{j+1}) + \varepsilon^2\,\sigma_j^2} \tag{2.4}$$

*is minimax on the set of all linear estimators.*

*The minimax risk equals*

$$R_{l\varepsilon} = \varepsilon^2 \sum_{j=1}^{\infty} \frac{P_0\,\sigma_j^2\,(a_j - a_{j+1})}{P_0\,(a_j - a_{j+1}) + \varepsilon^2\,\sigma_j^2}. \tag{2.5}$$

**Remark 2.1.** The estimator $\hat{x}_\lambda$ is the maximum penalized likelihood estimator [5, 22] with quadratic penalty function

$$P_0^{-1} \sum_{j=1}^{\infty} (a_j - a_{j+1})^{-1} \sigma_j^2 x_j^2$$

and a Bayes estimator with a priori measure corresponding to independent Gaussian random coordinates $x_j$ with $\mathbf{E}x_j = 0$ and $\mathbf{E}x_j^2 = P_0(a_j - a_{j+1})$, $1 \le j < \infty$.

In Theorem 2.2, we replace A2 by a simpler assumption.

**B1**. For all $j > j_0$,

$$\frac{\sigma_j^2\, j^{2\alpha+1}}{\sigma_{j-1}^2\,(j-1)^{2\alpha+1}} > 1. \tag{2.6}$$

This implies that the sequence $\sigma_j^2 j^{2\alpha+1}$ is strictly increasing.

**Theorem 2.2.** *Let assumptions* A1 *and* B1 *be satisfied. Then the linear estimator* $\hat{x}_\lambda$ *with*

$$\lambda_j = \frac{2\alpha P_0\, j^{-2\alpha-1}}{2\alpha P_0\, j^{-2\alpha-1} + \varepsilon^2\,\sigma_j^2} \tag{2.7}$$

*is asymptotically minimax on the set of all estimators.*

*The asymptotically minimax risk equals*

$$R_\varepsilon = \varepsilon^2 \sum_{j=1}^{\infty} \frac{2\alpha P_0\, j^{-2\alpha-1}\,\sigma_j^2}{2\alpha P_0\, j^{-2\alpha-1} + \varepsilon^2\,\sigma_j^2}(1 + o(1)). \tag{2.8}$$

**Remark 2.2.** The estimator $\hat{x}_\lambda$ is the maximum penalized likelihood estimator [5, 22] with quadratic penalty function

$$(2\alpha P_0)^{-1} \sum_{j=1}^{\infty} j^{1+2\alpha}\,\sigma_j^2 x_j^2$$

and a Bayes estimator with a priori measure corresponding to independent Gaussian random coordinates $x_j$ with $\mathbf{E}x_j = 0$ and $\mathbf{E}x_j^2 = 2\alpha P_0\, j^{-1-2\alpha}$, $1 \le j < \infty$.

Theorems 2.1 and 2.2 are easily extended to the setup of estimation of a solution of a linear ill-posed inverse problem with Gaussian random noise. Maxisets for this setup were studied by Loubes and Rivoirard [15].

Assume that we observe a random vector

$$y = R\,x + \varepsilon\,\xi \tag{2.9}$$

with a self-adjoint linear operator $R : \mathbb{H} \to \mathbb{H}$ in a separable Hilbert space $\mathbb{H}$. The remaining notation is the same as in the previous setup.

Assume that the linear operator $R$ admits a singular value decomposition (see [21, 9, 8, 15]) with eigenvalues $r_j$, $1 \le j < \infty$. Then we can consider this setup in the following form.

We observe a random vector

$$z_j = r_j\, x_j + \varepsilon\, \sigma_j\, \xi_j, \quad 1 \le j < \infty,$$

where $\xi_j$, $1 \le j < \infty$, are i.i.d. Gaussian r.v.'s with $\mathbf{E}\xi_j = 0$ and $\mathbf{E}\xi_j^2 = 1$. The problems of estimation of $x = \{x_j\}_{j=1}^{\infty}$ are the same. Dividing by $r_j$, we obtain the setup of signal estimation.

Below we provide two asymptotics of minimax risks for linear ill-posed inverse problems.

**Example 2.1.** *Let $\alpha > 0$ and $\gamma > 0$. Let $|r_j| = Cj^{-\gamma}(1 + o(1))$ and $\sigma_j = 1, 1 \le j < \infty$. Then*

$$R_\varepsilon = \varepsilon^{\frac{4\alpha}{1+2\alpha+2\gamma}} \frac{\pi}{2\alpha \sin\left(\frac{\pi(2\gamma+1)}{2\alpha}\right)} (2\alpha P_0)^{\frac{2\gamma+1}{2\gamma+2\alpha+1}}\, C^{-\frac{2\alpha}{2\gamma+2\alpha+1}} (1 + o(1)). \tag{2.10}$$

**Example 2.2.** *Let $\alpha > 0$, $\gamma > 0$, $B > 0$, and $\kappa \in R^1$. Let $|r_j| = Cj^{-\kappa} \exp\{-Bj^{\gamma}\}$ and $\sigma_j = 1, 1 \le j < \infty$. Then*

$$R_\varepsilon = P_0\, B^{2\alpha/\gamma}\, |\log \varepsilon\,|^{-2\alpha/\gamma}\, (1 + o(1)). \tag{2.11}$$

Note that these asymptotics coincide with the asymptotics of risks of the corresponding Bayes estimators.

Johnstone ([9, Chap. 3, Theorem 3.10]) compared strong asymptotics of minimax risks for trigonometric spline estimators and Pinsker estimators in the case where the unknown signal belongs to a ball in a Sobolev space. Trigonometric spline estimators are strong asymptotically minimax estimators on maxisets $\mathbb{B}(\alpha, P_0)$. Thus, we can consider this result as a comparison of risk asymptotics on Sobolev balls for strong asymptotically minimax estimators on maxisets $\mathbb{B}(\alpha, P_0)$ and for Pinsker estimators. Below we provide a similar comparison for strong asymptotically minimax estimators on maxisets and for Pinsker estimators in the case where the unknown signal belongs to a maxiset.

A Pinsker estimator $\tilde{\theta}_{\varepsilon\mu} = \{\tilde{\theta}_{\varepsilon j}\}_{j=1}^{\infty}$ is a linear estimator

$$\tilde{\theta}_{\varepsilon j} = \lambda_{\varepsilon j}\, y_j$$

with

$$\lambda_{\varepsilon j} = (1 - \mu\, b_j)_+,$$

where $b_j = j^{\beta}, \beta > 0$, and parameter $\mu$ is defined by the equation

$$\varepsilon^2 \sum_{j=1}^{\infty} b_j^2((\mu b_j)^{-1} - 1)_+ = P.$$

A Pinsker estimator is asymptotically minimax on ellipsoids

$$\mathbb{S}(\beta, P) = \left\{x\, :\, \sum_{j=1}^{\infty} b_j^2\, x_j^2 \le P,\, x = \{x_j\}_1^{\infty}\right\}$$

with $P > 0$.

Denote

$$R_\varepsilon(\alpha, \beta) = \inf_{\mu} \sup_{\theta \in \mathbb{B}(\alpha, P_0)} \mathbf{E}_\theta ||\tilde{\theta}_{\varepsilon\mu} - \theta||^2$$

and

$$C = \frac{2\alpha^2}{(1+\alpha)(1+2\alpha)}.$$

**Theorem 2.3.** *Let $0 < \alpha < \beta$. Then*

$$R_\varepsilon(\alpha,\beta) = C^{\frac{2\alpha}{1+2\alpha}}\, C_1^{\frac{1}{1+2\alpha}} \left((2\alpha)^{-\frac{2\alpha}{1+2\alpha}} + (2\alpha)^{\frac{1}{1+2\alpha}}\right) \varepsilon^{\frac{4\alpha}{1+2\alpha}}, \tag{2.12}$$

*where $C_1 = \frac{\beta}{\beta-\alpha} P_0$.*

*Let $\alpha > \beta > 0$. Then*

$$R_\varepsilon(\alpha,\beta) = C^{\frac{2\beta}{1+2\beta}}\, C_1^{\frac{1}{1+2\beta}} \left((2\beta)^{-\frac{2\beta}{1+2\beta}} + (2\beta)^{\frac{1}{1+2\beta}}\right) \varepsilon^{\frac{4\beta}{1+2\beta}}, \tag{2.13}$$

*where*

$$C_1 = \sum_{j=1}^{\infty} j^{2\beta}\Big(j^{-2\alpha} - (j+1)^{-2\alpha}\Big).$$

*If $\alpha = \beta$, then*

$$R_\varepsilon(\alpha,\beta) = \left((2\alpha^2)^{\frac{1}{1+2\alpha}} + 2^{-\frac{2\alpha}{1+2\alpha}}\alpha^{\frac{1-2\alpha}{1+2\alpha}}\right)(1+2\alpha)^{-\frac{1}{1+2\alpha}} P_0^{\frac{1}{1+2\alpha}} C^{\frac{2\alpha}{1+2\alpha}} \varepsilon^{\frac{4\alpha}{1+2\alpha}} |2\log\varepsilon|^{\frac{1}{1+2\alpha}}. \tag{2.14}$$

It is of principal interest to compare risks of the Pinsker estimator and of asymptotically minimax estimators on maxisets if $\alpha = \beta$. For this setup, we compare the risks of estimators on sets having almost the same smoothness. We show that in this case, risks of the Pinsker estimators have an additional logarithmic term in the asymptotic. Thus, Pinsker estimators do not belong to the class of linear estimators having maxisets $\mathbb{B}(\alpha, P_0)$. It turns out that balls in the Sobolev space $\mathbb{S}^\beta$ are maxisets for Pinsker estimators.

**Theorem 2.4.** *There exists a $C > 0$ such that, for all $\varepsilon > 0$,*

$$R_\varepsilon(\beta, x) = \varepsilon^{-\frac{4\beta}{1+2\beta}} \inf_\mu \mathbf{E}_x \|\tilde{x}_{\varepsilon\mu} - x\|^2 < C < \infty \tag{2.15}$$

*if and only if $x$ belongs to a ball in the Sobolev space*

$$\mathbb{S}^\beta = \Big\{x : \sum_{j=1}^{\infty} b_j^2 x_j^2 < P,\ x = \{x_j\}_{j=1}^{\infty}\Big\}, \quad P > 0.$$

In the theory of linear ill-posed inverse problems, one of the most usual assumptions is that the solution $x$ satisfies a source condition [2, 15],

$$x \in \{x : x = Bu,\ \|u\| \le 1,\ u \in \mathbb{H}\},$$

where $B$ is a self-adjoint compact linear operator. This implies that the solution $x$ belongs to an ellipsoid. Theorems 2.3 and 2.4 show that the optimal linear solution on such sets has worse rates of convergence on wider sets $\mathbb{B}(\alpha, P_0)$ than a wide class of linear estimators.

## 3. Proofs of theorems

*Proof of Theorem* 2.1. We begin with the proof of the lower bound. Denote $\theta_j^2 = P_0\,(a_j - a_{j+1})$ and $\theta = \{\theta_j\}_{j=1}^{\infty}$.

Then

$$\inf_\lambda \sup_{x\in\mathbb{B}(a,P_0)} \mathbf{E}_x \|\hat{x}_\lambda - x\|^2 \ge \inf_\lambda \mathbf{E}_\theta \|\hat{x}_{\varepsilon\lambda} - \theta\|^2 = \varepsilon^2 \sum_{j=1}^{\infty} \frac{\theta_j^2\sigma_j^2}{\theta_j^2 + \varepsilon^2\sigma_j^2}, \tag{3.1}$$

and the infimum is attained for

$$\lambda_j = \frac{\theta_j^2}{\theta_j^2 + \varepsilon^2 \sigma_j^2} = \frac{P_0\,(a_j - a_{j+1})}{P_0\,(a_j - a_{j+1}) + \varepsilon^2\,\sigma_j^2}.$$

Our proof of the upper bound is based on the following reasoning. Let $x = \{x_j\}_{j=1}^{\infty} \in \mathbb{B}(a, P_0)$. For all $k$ denote

$$u_k = a_k^{-1} \sum_{j=k}^{\infty} x_j^2.$$

Then $x_k^2 = a_k\,u_k - a_{k+1}\,u_{k+1}$.

For the sequence of $\lambda_j$ defined in Theorem 2.1 we have the relations

$$\begin{aligned}
\mathbf{E}_x \sum_{j=1}^{\infty} (\lambda_j\,y_j - x_j)^2 &= \varepsilon^2 \sum_{j=1}^{\infty} \lambda_j^2\,\sigma_j^2 + \sum_{j=1}^{\infty} (1-\lambda_j)^2\,x_j^2 \\
&= \varepsilon^2 \sum_{j=1}^{\infty} \lambda_j^2\,\sigma_j^2 + \sum_{j=1}^{\infty} (\theta_j^2\,\sigma_j^{-2}\,\varepsilon^{-2} + 1)^{-2}\,(a_j\,u_j - a_{j+1}\,u_{j+1}) \\
&= \varepsilon^2 \sum_{j=1}^{\infty} \lambda_j^2\,\sigma_j^2 + (\theta_1^2\,\sigma_1^{-2}\,\varepsilon^{-2} + 1)^{-2}\,u_1 \\
&\quad - \sum_{j=2}^{\infty} u_j\,a_j\,\Big((\theta_{j-1}^2\,\sigma_{j-1}^{-2}\,\varepsilon^{-2} + 1)^{-2} - (\theta_j^2\,\sigma_j^{-2}\,\varepsilon^{-2} + 1)^{-2}\Big).
\end{aligned} \tag{3.2}$$

By assumption A2, the last terms in the right hand-side of (3.2) are negative. Therefore, the supremum of the right hand-side of (3.2) is attained for $u_j = P_0$, $1 \le j < \infty$. This completes the proof of Theorem 2.1. □

*Proof of Theorem* 2.2. The upper bound follows from Theorem 2.1. Below we prove the lower bound. This proof has a lot of common features with the proof of lower bound in the Pinsker theorem [9, 18, 21].

Fix values $\delta_1$, $0 < \delta_1 < 1$, and $\delta$, $0 < \delta < P_0$. Define a family of natural numbers $k_\varepsilon, \varepsilon > 0$, such that $\varepsilon^{-2}\sigma_{k_\varepsilon}^2\,2rP_0k_\varepsilon^{-2r-1} = 1 + o(1)$ as $\varepsilon \to 0$. Define a sequence $\eta = \{\eta_j\}_{j=1}^{\infty}$ of Gaussian i.i.d.r.v.'s $\eta_j = \eta_{j\delta\delta_1}$ such that $\mathbf{E}[\eta_j] = 0$, $\mathbf{Var}[\eta_j] = (P_0 - \delta)(2r)^{-1}j^{-2r-1}$ if $\delta k_\varepsilon \le j \le \delta^{-1}k_\varepsilon$, and $\eta_j = 0$ if either $j < \delta_1 k_\varepsilon$ or $j > \delta_1^{-1}k_\varepsilon$.

Denote by $\mu$ the probability measure of the random vector $\eta$. Denote by $\tilde{x}$ the Bayes estimator with a priory measure $\mu$.

Define the conditional probability measure $\nu_\delta$ of the random vector $\eta$ given that $\eta \in \mathbb{B}(\alpha, P_0)$. Denote by $\bar{x}$ the Bayes estimator of $x$ with a priori measure $\nu_\delta$. Denote by $\theta$ the random variable having probability measure $\nu_\delta$.

For any estimator $\hat{x}$,

$$\begin{aligned}
\sup_{x \in \mathbb{B}(\alpha, P_0)} \mathbf{E}_x\,\|\hat{x} - x\|^2 \ge \mathbf{E}_{\nu_\delta}\,\mathbf{E}_\theta\,\|\hat{x} - \theta\|^2 \ge \mathbf{E}_\mu\,\mathbf{E}_\eta\,\|\tilde{x} - \eta\|^2 - \mathbf{E}_\mu\,\mathbf{E}_\eta\,(\|\bar{x} - \eta\|^2, \\
\eta \notin \mathbb{B}(\alpha, P_0))\,\mathbf{P}_\mu^{-1}\,(\eta \in \mathbb{B}(\alpha, P_0)).
\end{aligned} \tag{3.3}$$

We have the relation

$$\mathbf{E}_\mu\,\mathbf{E}_\eta\,\|\tilde{x} - \eta\|^2 = I(P_0 - \delta)(1 + o(1)), \tag{3.4}$$

where

$$I(P_0 - \delta) = \varepsilon^2 \sum_{j=l_1}^{l_2} \frac{\sigma_j^2}{1 + (2\alpha\,(P_0 - \delta_1))^{-1}\varepsilon^2\,\sigma_j^2\,j^{2\alpha+1}}$$

and $l_1 = [\delta_1 k_\varepsilon]$, $l_2 = [\delta_1^{-1} k_\varepsilon]$. Here $[a]$ denotes the integral part of a number $a \in \mathbb{R}^1$.

Since

$$\|\bar{x}\|^2 \le \sup_{x \in \mathbb{B}(\alpha, P_0)} \|x\|^2 \le P_0,$$

we have the estimates

$$\begin{aligned}\mathbf{E}_\mu \, \mathbf{E}_\eta \, (\|\bar{x} - \eta\|^2, \, \eta \notin \mathbb{B}(\alpha, P_0)) &\le 2 \, \mathbf{E}_\mu \, \mathbf{E}_\eta \, (\|\bar{x}\|^2 + \|\eta\|^2, \; \eta \notin \mathbb{B}(\alpha, P_0)) \\ &\le 2P_0 \, \mathbf{P}_\mu \, (\eta \notin \mathbb{B}(\alpha, P_0)) + \sum_{j=l_1}^{l_2} (\mathbf{E}_\mu \, \eta_j^4)^{1/2} \; \mathbf{P}_\mu^{1/2}(\eta \notin B(\alpha, P_0)).\end{aligned} \tag{3.5}$$

Since $\mathbf{E}_\mu[\eta_j^4] \le C j^{-2(\alpha-2}$,

$$\sum_{j=l_1}^{l_2} (\mathbf{E}_\mu \, \eta_j^4)^{1/2} \le C \delta_1^{-r} \, k_\varepsilon^{-2r}. \tag{3.6}$$

It remains to estimate

$$\mathbf{P}_\mu \, (\eta \notin \mathbb{B}(\alpha, P_0)) = \mathbf{P} \Big( \max_{l_1 \le i \le l_2} \; i^{2r} \sum_{j=i}^{l_2} \eta_j^2 - P_0 \, (1 - \delta_1/2) > P_0 \, \delta_1/2 \Big) \le \sum_{i=l_1}^{l_2} J_i, \tag{3.7}$$

where

$$J_i = \mathbf{P} \Big( i^{2\alpha} \sum_{j=i}^{l_2} \eta_j^2 - P_0 \, (1 - \delta/2) > P_0 \, \delta/2 \Big).$$

To estimate $J_i$, we apply the following proposition (see [6]). □

**Proposition 3.1.** *Let $\xi = \{\xi_i\}_{i=1}^l$ be a Gaussian random vector with i.i.d.r.v.'s $\xi_i$ such that $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi^2 = 1$. Let $A$ be an $(l \times l)$-matrix and let $\Sigma = A^T A$. Then*

$$\mathbf{P}(||A \, \xi||^2 > \mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2) \, t} + 2\|\Sigma\| \, t) \le \exp\{-t\}. \tag{3.8}$$

*Here* $\mathrm{tr}\,(\Sigma)$ *denotes the trace of the matrix* $\Sigma$.

Define a matrix $\Sigma = \{\sigma_{lj}\}_{l,j=i}^{l_2}$ with entries $\sigma_{jj} = 2\alpha \, (P_0 - \delta) \, j^{-2\alpha-1} \, i^{2\alpha}$ and $\sigma_{lj} = 0$ if $l \ne j$. Then

$$2\sqrt{\mathrm{tr}(\Sigma^2) t} + 2\|\Sigma\| t = \frac{P_0 - \delta}{\alpha(4\alpha + 1)} \sqrt{i^{-1} t}(1 + o(1)) + i^{-1} t \doteq V_i(t). \tag{3.9}$$

We put $t = k_\varepsilon^{1/2}$. Then $V_i(t) < C k_\varepsilon^{-1/2}$, $1 \le i \le l_2$, and it follows from (3.8) that

$$J_i < \exp\{-k_\varepsilon^{-1/2}\}; \tag{3.10}$$

therefore,

$$\sum_{j=l_1}^{l_2} J_i \le \delta_1^{-1} \, k_\varepsilon \, \exp\{-k_\varepsilon^{1/2}\}. \tag{3.11}$$

To complete the proof, it remains to estimate $R_\varepsilon - I(P_0 - \delta)$. By a straightforward estimation, it is easy to verify that

$$|I(P_0) - I(P_0 - \delta)| < C \delta I(P_0). \tag{3.12}$$

We have the relations

$$
\begin{aligned}
\varepsilon^2 \sum_{j=1}^{l_1} \frac{\sigma_j^2}{1+(2\alpha P_0)^{-1}\varepsilon^2\sigma_j^2 j^{2\alpha+1}} &\asymp \varepsilon^2 \sum_{j=1}^{l_1} \sigma_j^2 \\
< C\delta_1\varepsilon^2 \sum_{j=l_1}^{k_\varepsilon} \sigma_j^2 &\asymp C\delta_1\varepsilon^2 \sum_{j=l_1}^{k_\varepsilon} \frac{\sigma_j^2}{1+(2\alpha P_0)^{-1}\varepsilon^2\sigma_j^2 j^{2\alpha+1}}
\end{aligned}
\tag{3.13}
$$

and

$$
\begin{aligned}
\varepsilon^2 \sum_{j=l_2}^{\infty} \frac{\sigma_j^2}{1+(2\alpha P_0)^{-1}\varepsilon^2\sigma_j^2 j^{2\alpha+1}} &\asymp \varepsilon^2 \sum_{j=l_2}^{\infty} j^{-2\alpha-1} \\
\le \varepsilon^2\delta_1^{2\alpha} C \sum_{k_\varepsilon}^{l_2} j^{-2\alpha-1} &\asymp \varepsilon^2\delta_1^{2\alpha} C \sum_{k_\varepsilon}^{l_2} \frac{\sigma_j^2}{1+(2\alpha P_0)^{-1}\varepsilon^2\sigma_j^2 j^{2\alpha+1}}.
\end{aligned}
\tag{3.14}
$$

Now (3.12)–(3.14) imply that $R_\varepsilon - I(P_0-\delta) \to 0$ for some $\delta=\delta(\varepsilon)\to 0$ and $\delta_1=\delta_1(\varepsilon)\to 0$ as $\varepsilon \to 0$.

*Proof of Theorem* 2.3. The reasoning is based on the following lemma. □

**Lemma 3.1.** *The following relation holds*:

$$
\sup_{x\in\mathbb{B}(\alpha,P_0)} \mathbf{E}_x\|\tilde{x}_\varepsilon - x\|^2 = \mathbf{E}_{\theta_\varepsilon}||\tilde{x}_\varepsilon-\theta_\varepsilon||^2, \tag{3.15}
$$

*where* $\theta_\varepsilon=\{\theta_{\varepsilon k}\}_{k=1}^\infty$, $\theta_{\varepsilon k}^2 = P_0(a_k-a_{k+1})$.

*Proof of Lemma* 3.1. Denote $u_k = a_k^{-1}\sum_{j=k}^{\infty}\theta_j^2$. Then

$$
\theta_k^2 = a_k\,u_k - a_{k+1}\,u_{k+1}.
$$

Denote $l=[\mu^{-\frac{1}{\beta}}]$.

We represent

$$
\mathbf{E}_x\,\|\tilde{x}_\varepsilon - x\|^2 = \mu^2\sum_{j=1}^{l} b_j^2\,x_j^2 + \sum_{j=l+1}^{\infty} x_j^2 + \varepsilon^2\sum_{j=1}^{l}\lambda_j^2 \doteq J_1+J_2+J_3. \tag{3.16}
$$

Note that

$$
\begin{aligned}
J_1+J_2 &= \mu^2\sum_{j=1}^{l} b_j^2\,(a_j u_j - a_{j+1}\,u_{j+1}) + a_{l+1}\,u_{l+1} \\
&= \mu^2\,a_1\,b_1^2\,u_1^2 - \mu^2\,a_{l+1}\,b_l^2\,u_{l+1}^2 + \mu^2\sum_{j=2}^{l} a_j\,u_j\,(b_j^2-b_{j-1}^2) + a_{l+1}\,u_{l+1}.
\end{aligned}
\tag{3.17}
$$

The maximum of the right-hand side of (3.17) is attained for $u_j = P_0$, $1\le j<\infty$, with $x_j^2 = P_0(a_j-a_{j+1})$.

By straightforward calculations, we show that $J_3 = C\varepsilon^2 l$.

If $\beta>\alpha$, then

$$
J_1+J_2 = \frac{\beta}{\beta-\alpha} l^{-2\alpha}(1+o(1)).
$$

If $\alpha>\beta$, then

$$
J_1+J_2 = P_0 l^{-2\beta} C_1(1+o(1)).
$$

If $\alpha = \beta$, then

$$J_1 + J_2 = \alpha P_0 l^{-2\alpha} \log l.$$

Minimizing $J_1 + J_2 + J_3$ with respect to $l$, we prove Theorem 2.3. □

*Proof of Theorem* 2.4. It suffices to prove the necessary conditions.

We have the relations

$$\begin{aligned} \mathbf{E}_x \, \|\tilde{x}_{\varepsilon\mu} - x\|^2 &= \varepsilon^2 \sum_{j=1}^{l} (1 - l^{-\beta} j^{\beta}) + l^{-2\beta} \sum_{j=1}^{l} j^{2\beta} x_j^2 + \sum_{j=l}^{\infty} x_j^2 \\ &\geq C \, \varepsilon^2 \, l + l^{-2\beta} \sum_{j=1}^{l} j^{2\beta} x_j^2 \doteq J_\varepsilon(l, x). \end{aligned} \tag{3.18}$$

It is easy to see that if

$$\sum_{j=1}^{l} j^{2\beta} x_j^2 \to \infty \quad \text{as} \quad l \to \infty, \tag{3.19}$$

then

$$\lim_{\varepsilon \to 0} \varepsilon^{-\frac{4\beta}{1+2\beta}} \inf_{l} J_\varepsilon(l, x) = \infty. \tag{3.20}$$

□

Supported in part by the RFBR (project 17-01-00828).

Translated by M. Ermakov.

## REFERENCES

1. K. Bertin, “Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes,” *Bernoulli*, **10**, 873–888 (2004).
2. L. Cavalier, “Inverse problems in statistics. Inverse problems and high-dimensional estimation,” *Lect. Notes Statist. Proc.*, **203**, 3–96, Springer (2011).
3. D. L. Donoho, “Asymptotic minimax risk for sup-norm loss: solution via optimal recovery,” *Probab. Th. Rel. Fields*, **99**, 145–170 (1994).
4. D. L. Donoho and I. M. Johnstone, “Minimax estimation via wavelet shrinkage,” *Ann. Statist.*, **26**, 879–921 (1998).
5. P. B. Eggermont and V. N. LaRiccia, *Maximum Penalized Likelyhood Estimation.* II, Springer (2009).
6. D. Hsu, S. M. Kakade, and T. Zang, “A tail inequality for quadratic forms of subgaussian random vector,”*Electronic Commun. Probab.*, **17**, 1–6 (2012).
7. I. Ibragimov and R. Khasminskii, *Statistical Estimation: Asymptotic Theory*, Springer (1981).
8. Yu. I. Ingster, T. Sapatinas, and I. A. Suslina, “Minimax signal detection in ill-posed inverse problems,” *Ann. Statist.*, **40**, 1524–1549 (2012).
9. I. M. Johnstone, *Gaussian Estimation. Sequence and Wavelet Models.* Book Draft `http://statweb.stanford.edu/~imj/` (2015).
10. G. Kerkyacharian and D. Picard, “Density estimation by kernel and wavelets methods: optimality of Besov spaces,” *Statist. Probab. Lett.*, **18**, 327–336 (1993).
11. G. Kerkyacharian and D. Picard, “Minimax or maxisets?” *Bernoulli*, **8**, 219–253 (2002).
12. A. P. Korostelev, “An asymptotically minimax regression estimator in the uniform norm up to a constant,” *Teor. Veroyatn. Primen.*, **38**, 857–882 (1993).

13. J. A. Kuks and V. Olman, "A minimax linear estimator of regression coefficients," *Izv. Akad. Nauk Eston. SSR*, **20**, 480-482 (1971).
14. O. V. Lepski and A. B. Tsybakov, "Asymptotically exact nonparametric hypothesis testing in supnorm and at a fixed point," *Probab. Th. Rel. Fields*, **117**, 17–48 (2000).
15. J. M. Loubes and V. Rivoirard, "Review of rates of convergence and regularity conditions for inverse problems," *Intern. J. Tomogr. Statist.*, **11**, 61–82 (2009).
16. A. S. Nemirovskii, "Nonparametric estimation of smooth regression functions," *Soviet J. Computer and Systems Sciences*, **23**, 1–11 (1985).
17. M. Nussbaum, "Spline smoothing in regression models and asymptotic efficiency in $L_2$," *Ann. Statist.*, **13**, 984–997 (1985).
18. M. S. Pinsker, "Optimal filtering of square integrable signals in Gaussian white noise," *Problems of Information Transmission*, **16**, 120–133 (1980).
19. V. Rivoirard, "Maxisets for linear procedures," *Statist. Probab. Lett.*, **67**, 267–275 (2004).
20. A. Tikhonov, "Regularization of incorrectly posed problems," *Soviet Math. Dokl.*, **4**, 1624–1627 (1963).
21. A. Tsybakov, *Introduction to Nonparametric Estimation*, Springer Series in Statistics, **130**, Springer (2009).
22. G. Wahba, *Spline Models for Observational Data*, SIAM (1990).