ISSN 0032-9460, Problems of Information Transmission, 2008, Vol. 44, No. 4, pp. 321–332. © Pleiades Publishing, Inc., 2008.
Original Russian Text © E.N. Belitser, F.N. Enikeeva, 2008, published in Problemy Peredachi Informatsii, 2008, Vol. 43, No. 4, pp. 39–51.

METHODS OF SIGNAL PROCESSING

Adaptive Filtering of a Random Signal in Gaussian White Noise

E. N. Belitser[a] and F. N. Enikeeva[b]

[a] *Mathematical Institute, Utrecht University, The Netherlands*
E.N.Belitser@uu.nl
[b] *Kharkevich Institute for Information Transmission Problems, RAS, Moscow*
enikeeva@iitp.ru

Received March 31, 2008

Abstract—We consider the problem of estimating an infinite-dimensional vector θ observed in Gaussian white noise. Under the condition that components of the vector have a Gaussian prior distribution that depends on an unknown parameter β, we construct an adaptive estimator with respect to β. The proposed method of estimation is based on the empirical Bayes approach.

DOI: 10.1134/S0032946008040054

1. INTRODUCTION

We observe the vector $X = (X_1, X_2, \ldots)$ with components given by

$$X_i = \theta_i + n^{-1/2}\xi_i, \quad i = 1, 2, \ldots, \tag{1}$$

where ξ_i are independent $\mathcal{N}(0,1)$ random variables and $n \to \infty$. Our goal is to estimate the unknown infinite-dimensional vector $\theta = (\theta_1, \theta_2, \ldots)$, $\theta \in \ell_2$.

This model is a discrete version of the well-known white noise model. Indeed, consider observations of an unknown signal $f \in \mathcal{L}_2(0,1)$ in the white noise model,

$$dx(t) = f(t)\,dt + \varepsilon\,dw(t), \quad t \in [0,1]. \tag{2}$$

These observations can be regarded as the sequence of the Fourier coefficients of f with respect to some orthonormal basis $\{\varphi_k\}_{k\in\mathbb{N}}$ in $\mathcal{L}_2(0,1)$. Thus, we obtain a discrete model

$$\widetilde{X}_i = \widetilde{\theta}_i + \varepsilon\widetilde{\xi}_i, \quad i \in \mathbb{N},$$

where $w(t)$ is the standard Wiener process, $\varepsilon > 0$ is a small parameter, $\widetilde{X}_i = \int_0^1 \varphi_i(t)\,dx(t)$, $\widetilde{\theta}_i = \int_0^1 \varphi_i(t)f(t)\,dt$, and $\widetilde{\xi}_i = \int_0^1 \varphi_i(t)\,dw(t)$ are i.i.d. $\mathcal{N}(0,1)$ random variables. Thus, if we set $\varepsilon = n^{-1/2}$, then the problem of estimating an unknown signal from observations (1) is equivalent to the problem of estimating a signal in the Gaussian white noise model (2) (see [1,2]). That is why, sometimes we call the vector θ in model (1) a signal. The Gaussian white noise model (2) for $\varepsilon = n^{-1/2}$ is a good approximation for many problems of estimation theory, for example, for the problem of density estimation [3] and the problem of estimating the function of nonparametric regression [4] from n observations. In [3,4] the equivalence of the corresponding models is proved.

There are two different approaches to signal estimation. The first approach assumes that the signal is unknown and constant, and the second one assumes that the signal is a random process. Depending on the approach, we need different prior information on the unknown signal. For the first approach, it is usually assumed that the vector θ belongs to some given compact symmetric subset $\Theta \subset \ell_2$. In the classical case, this subset is an ellipsoid that belongs to ℓ_2:

$$\Theta = \Theta_\beta(Q) = \Big\{\theta : \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2 \le Q\Big\}.$$

The parameter $\beta > 0$ has the meaning of the degree of smoothness of the signal, since the relation $\theta \in \Theta_\beta(Q)$ imposes a condition on the smoothness of the signal f in the space of functions (it is related to the Sobolev space of smoothness β). The quality of the estimate $\widehat{\theta} = \widehat{\theta}(X)$ for a fixed θ is measured by the mean-square risk function $R(\widehat{\theta}, \theta) = \mathbf{E}_\theta \|\widehat{\theta} - \theta\|^2$, where $\|\cdot\|$ is the norm in the space ℓ_2. The quality of the estimator $\widehat{\theta} = \widehat{\theta}(X)$ for the class Θ is characterized by the value of the maximum mean-square risk $R(\widehat{\theta}, \Theta) = \sup\limits_{\theta\in\Theta} R(\widehat{\theta}, \theta)$. Then the optimal estimator in terms of the minimax risk $r(\Theta) = \inf\limits_{\widehat{\theta}} \sup\limits_{\theta\in\Theta} R(\widehat{\theta}, \theta)$ should be constructed. In [5], under the prior information $\theta \in \Theta_\beta(Q)$, the minimax estimator is constructed and the exact asymptotic of the minimax risk is obtained: $\lim\limits_{n\to\infty} n^{2\beta/(2\beta+1)} r(\Theta_\beta(Q)) = \gamma(\beta, Q)$, where $\gamma(\beta, Q)$ is the so-called Pinsker constant.

In the second approach, the signal is random, and the classical Bayesian approach is mostly often applied. This approach requires the knowledge of the prior distribution of the signal $\theta \sim \pi$. This problem is called the filtering problem for a signal (with realizations in ℓ_2), and the quality of estimation is characterized by the Bayesian risk $R_\pi(\widehat{\theta}) = \mathbf{E}_\pi R(\widehat{\theta}, \theta)$, where the mean $\mathbf{E}_\pi$ is with respect to the prior measure π. It is well known that the optimal (Bayesian) estimate in terms of the mean-square risk is the posterior mean $\widehat{\theta} = \mathbf{E}(\theta \,|\, X)$.

Note that these two approaches are closely related. On one hand, the minimax estimate is Bayesian with respect to the so-called less informative prior. On the other hand, Bayesian estimate can be considered (with respect to the prior distribution π) as an optimal estimate of the nonrandom signal θ if the quality of estimation is measured by the Bayesian risk $R_\pi(\widehat{\theta})$. The author of [5] combined these two approaches and showed, among many other results, that, asymptotically, we can choose the distribution $\pi = \pi(n, \beta, Q)$ as the less informative (asymptotically) prior, where θ_i are independent random variables with zero means and specially chosen variances $\tau_i^2 = \tau_i^2(n, \beta, Q)$. The knowledge of the parameter β is essential here, since the convergence rate of the risk of the corresponding Bayesian estimator (which is the same as the minimax convergence rate) depends on β: $n^{-2\beta/(2\beta+1)}$, whereas the parameter Q is only involved in the expression for the optimal constant of the risk $\gamma(\beta, Q)$. If we restrict ourselves with the convergence rate of the minimax risk over the Sobolev ellipsoid $\Theta_\beta(Q)$, then, as we show in the present paper, the prior distribution $\pi = \pi_\beta$ defined by relations (3) given below leads to a Bayesian estimator $\widehat{\theta}(\beta)$, which is also minimax with respect to the convergence rate. One of the main features of this approach is its independence of n, which is traditional for the Bayesian approach, though this is not a necessary requirement on the prior distribution in the framework of the minimax approach (in the proof of lower bounds). Note also that in the case of the trigonometric basis $\{\varphi_k\}_{k\in\mathbb{N}}$, the corresponding signal f in the equivalent Gaussian white noise model is a stationary Gaussian process, and its Bayesian estimate is the well-known Wiener filter in the classical problem of filtering of the stationary Gaussian process with Gaussian white noise. Thus, the parameter β that appears in the definition of the prior π_β can be regarded as smoothness of the signal θ, and the distribution π_β itself describes the worst situation from the minimax point of view since it is the least advantageous in terms of the convergence rate of the minimax risk.

If the parameter β is unknown, then the problem of adaptive estimation of the signal θ arises. The minimax setup of this problem was studied in [6–11]. The authors of those papers proposed various methods to construct adaptive estimators of a nonrandom signal θ. Here we consider a Bayesian version of the problem of adaptive signal estimation. We assume that the signal θ is random and is distributed according to the prior distribution (3) with an unknown smoothness parameter $\beta > 0$. We also assume that realizations of the signal θ belong to ℓ_2. Our goal is to estimate the signal θ adaptively with respect to the parameter β, using the Bayesian ℓ_2-risk $R_{\pi_\beta}(\widehat{\theta})$ as a measure of estimation quality. It is clear that the Bayesian estimator $\widehat{\theta}(\beta)$ cannot be used, since the smoothness parameter β is unknown. Below, we use the Bayesian estimator $\widehat{\theta}(\beta)$ (5) as a benchmark, which is usually called an oracle. The main goal is to construct an optimal adaptive estimator of the signal, that is, a measurable function of observations independent of β such that its risk is asymptotically not worse than the risk of the oracle. Note that our results can also be interpreted for the problem of estimating a nonrandom signal, namely, as if we were estimating a nonrandom signal θ in the best way in terms of the Bayesian risk with a prior distribution π_β, where the parameter β is unknown.

We use the empirical Bayes approach proposed in [12] to construct our estimator. We construct the estimate of the parameter β based on observations X and use it instead of β in formula (5). Note that in the minimax setup it is hard to interpret the pronlem of estimating the smoothness of a signal, since it is unclear what the smoothness of the signal θ is, for example, if it is known that $\theta \in \Theta_\beta(Q)$ (it is possible that at the same time $\theta \in \Theta_{\beta'}(Q')$ for $\beta' > \beta$ and for another Q'). From the Bayesian point of view, the problem of estimating the smoothness parameter β does make sense; in this case β is an unknown parameter of the prior distribution π_β.

The problem of estimating the smoothness parameter β is an auxiliary problem, but it is interesting on its own. This is a specific problem of parametric estimation with infinitely many non-identically distributed observations and with a nontraditional asymptotic behavior (in our case, the parameter n, $n \to \infty$, is not the number of observations but some parameter involved in the variance of observations). We use a version of the maximum likelihood method, where the likelihood is the distribution of the observation X, which can be found by integrating the mutual distribution (X, θ) with respect to the unobserved component θ.

The empirical Bayes approach to the problem of adaptive estimation of a functional in the model of Gaussian white noise with prior distribution $\tau_i^2(\beta) = e^{-\beta i}$ of the parameter θ was considered in [13]. This prior distribution corresponds to the class of analytic functions f, where the parameter β accounts for the "degree of analyticity" of a function. In [14] there was considered the Bayesian approach to the adaptive estimation in the minimax setup for Sobolev classes; i.e., the signal was assumed to be nonrandom and to belong to a Sobolev class of an unknown smoothness.

2. MAIN RESULTS

2.1. Statement of the Problem and Notation

Let us recall the statement of the problem. We have to estimate an unknown parameter $\theta = (\theta_1, \theta_2, \dots)$ from its observations in Gaussian white noise

$$X_i = \theta_i + n^{-1/2}\xi_i, \quad i = 1, 2, \dots,$$

where $\xi_i \sim \mathcal{N}(0,1)$, $n \to \infty$. We assume that θ is a random variable such that its realizations θ belong to ℓ_2. The distribution $\pi = \pi_\beta$ of the vector θ is such that $\beta > 0$, the components θ_i are independent, and

$$\theta_i \sim \mathcal{N}(0, \tau_i^2(\beta)), \qquad \tau_i^2(\beta) = i^{-(2\beta+1)}, \quad i \in \mathbb{N}. \tag{3}$$

Let $\widehat{\theta}$ be an estimator of the parameter θ, $\mathbf{E}_\theta$ be the conditional mean of X given θ, $\mathbf{E}_\pi$ be the conditional mean with respect to the prior distribution $\pi = \pi_\beta$, and $\mathbf{E}$ be the mean with respect to the joint distribution of (X, θ). Then we can define the Bayesian mean-square risk of the estimator $\widehat{\theta}$ as

$$R_\pi(\widehat{\theta}) = \mathbf{E}_\pi\, R(\widehat{\theta}, \theta) = \mathbf{E}_\pi\, \mathbf{E}_\theta\, \|\widehat{\theta} - \theta\|^2 = \mathbf{E}\, \|\widehat{\theta} - \theta\|^2, \tag{4}$$

and the maximal risk of the estimator $\widehat{\theta}$ over the Sobolev ellipsoid is given by

$$R(\widehat{\theta}, \Theta_\beta(Q)) = \sup_{\theta \in \Theta_\beta(Q)} R(\widehat{\theta}, \theta).$$

It is not difficult to obtain a Bayesian estimator $\widehat{\theta}(\beta)$ with respect to the prior distribution π_β. From properties of the normal distribution, it follows that the posterior distribution $\pi_\beta(\theta \,|\, X)$ is given by

$$\theta_i \,|\, X \sim \mathcal{N}\Big(\frac{\tau_i^2(\beta) X_i}{\tau_i^2(\beta) + n^{-1}}, \frac{\tau_i^2(\beta) n^{-1}}{\tau_i^2(\beta) + n^{-1}}\Big)$$

and X_i are (conditionally) independent. Thus, the Bayesian estimator $\widehat{\theta}(\beta)$ is given by

$$\widehat{\theta}_i(\beta) = \mathbf{E}(\theta_i \,|\, X) = \mathbf{E}(\theta_i \,|\, X_i) = \frac{\tau_i^2(\beta)}{\tau_i^2(\beta) + n^{-1}} X_i, \quad i \in \mathbb{N}, \tag{5}$$

where $\widehat{\theta}_i(\beta)$ are components of the vector $\widehat{\theta}(\beta)$. Note that $\widehat{\theta}(\beta)$ is not an adaptive estimator, since the parameter β is used in its construction.

The following two results prove that the prior distribution π_β in the Bayesian setup represents the property $\theta \in \Theta_\beta(Q)$ in the minimax setup at least in the sense of convergence rate. Thus, the parameter β of the prior distribution can be considered as a smoothness parameter in the nonparametric class $\Theta_\beta(Q)$. The first result shows that the Bayes estimator $\widehat{\theta}(\beta)$ is a minimax estimator with respect to the convergence rate under the condition that θ belongs to the Sobolev ellipsoid $\Theta_\beta(Q)$.

For $r \geq 0$, $0 < q < \infty$, and $pq > r + 1$, define the function

$$B(p, q, r) = \int_0^\infty \frac{u^r}{(1 + u^p)^q}\, du = p^{-1} \mathrm{Beta}\Big(q - \frac{r+1}{p}, \frac{r+1}{p}\Big), \tag{6}$$

where $\mathrm{Beta}(\alpha, \beta) = \int_0^1 u^{\alpha - 1}(1 - u)^{\beta - 1} du$ is the beta function.

Proposition 1. *We have the inequality*

$$\limsup_{n \to \infty} n^{2\beta/(2\beta+1)} R(\widehat{\theta}(\beta), \Theta_\beta(Q)) \leq \big(QC(\beta) + B(2\beta + 1, 2, 0)\big),$$

where $C(\beta) = \dfrac{(1 + \beta^{-1})^{2(\beta+1)/(2\beta+1)}}{(2 + \beta^{-1})^2}$ *and the function* B *is defined by* (6).

Proof. We have

$$\begin{aligned}\mathbf{E}_\theta \,\|\theta - \widehat{\theta}\|^2 &= \mathbf{E}_\theta \sum_{i=1}^\infty \Big(\frac{\tau_i^2(\beta) X_i}{\tau_i^2(\beta) + n^{-1}} - \theta_i\Big)^2 \\ &= \sum_{i=1}^\infty \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta) + n^{-1})^2} + \sum_{i=1}^\infty \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta) + n^{-1})^2}.\end{aligned}$$

Thus, as $n \to \infty$,

$$\begin{aligned}\sum_{i=1}^\infty \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta) + n^{-1})^2} &= \sum_{i=1}^\infty \frac{i^{2(2\beta+1)}\theta_i^2}{(n + i^{2\beta+1})^2} \leq Q \max_{i \in \mathbb{N}} \frac{i^{2\beta+2}}{\big(n + i^{2\beta+1}\big)^2} \\ &= QC(\beta) n^{-2\beta/(2\beta+1)}(1 + o(1)), \\ \sum_{i=1}^\infty \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta) + n^{-1})^2} &= \sum_{i=1}^\infty \frac{n}{\big(n + i^{2\beta+1}\big)^2} \\ &= n^{-2\beta/(2\beta+1)} B(2\beta + 1, 2, 0)(1 + o(1)). \quad \triangle\end{aligned}$$

In the next lemma the asymptotic Bayesian mean-square risk is obtained for the Bayesian estimator (5). The convergence rate of this estimator is equal to the minimax convergence rate.

Proposition 2. *Let the estimate* $\widehat{\theta}$ *be defined by* (5). *Then*

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)} R_\pi(\widehat{\theta}) = \frac{\pi}{(2\beta+1)^2\sin(\pi/(2\beta+1))}.$$

Proof. From properties of beta functions (and gamma functions), it follows that for $r \ge 0$ and $p > r+1$ we have

$$B(p,1,r) = p^{-1}\mathrm{Beta}\Big(1-\frac{r+1}{p},\frac{r+1}{p}\Big) = \frac{\pi}{p^2\sin(\pi(r+1)/p)}. \tag{7}$$

From (7) it follows that

$$R_\pi(\widehat{\theta}) = \sum_{i=1}^{\infty}\frac{\tau_i^2(\beta)n^{-1}}{\tau_i^2(\beta)+n^{-1}} = \sum_{i=1}^{\infty}\frac{1}{n+i^{2\beta+1}} = B(2\beta+1,1,0)n^{-2\beta/(2\beta+1)}(1+o(1))$$
$$= \frac{\pi n^{-2\beta/(2\beta+1)}}{(2\beta+1)^2\sin(\pi/(2\beta+1))}(1+o(1))$$

as $n\to\infty$. △

Thus, the Bayesian mean-square risk for the Bayesian estimator (5) for a known β has asymptotically the same convergence rate as that for the minimax risk over the ellipsoid $\Theta_\beta(Q)$. Moreover, it can be shown that if the parameter θ belongs to a Sobolev subset $\Theta_\beta = \Big\{\theta : \sum\limits_{i=1}^{\infty} i^{2\beta}\theta_i^2 < \infty\Big\}$ of the space ℓ_2, then the Bayes estimator $\widehat{\theta}(\beta)$ (5) belongs to the same subset with probability tending to 1 as $n\to\infty$. Thus, the parameter β involved in the definition of the prior distribution π_β can be considered as the smoothness of the signal θ, since the prior distribution π_β describes the worst situation from the viewpoint of minimax approach, being the least advantageous in terms of the convergence rate of the minimax risk.

Remark 1. We can construct a family of prior distributions $\{\pi_\delta(\beta),\ \delta > 1-2\beta\}$, each of them leading to a Bayesian estimator $\theta \sim \pi_\delta(\beta)$ with the minimax convergence rate such that the components θ_i are independent and

$$\theta_i \sim \mathcal{N}(0,\tau_i^2(\beta)), \qquad \tau_i^2(\beta) = \tau_i^2(\beta,\delta,n) = n^{\frac{\delta-1}{2\beta+1}} i^{-(2\beta+\delta)}, \quad i\in\mathbb{N}. \tag{8}$$

In the present paper we consider only one prior distribution from this family, namely, that defined by $\delta = 1$. This distribution is independent of n, as is conventional in the Bayesian tradition. However, for formal constructions of estimators, the condition of independence of n of the prior distribution is not necessarily required.

2.2. Empirical Bayes Approach

Consider now the situation where the smoothness parameter β is unknown. Denote by $\beta_0 > 0$ the real value of the unknown parameter β. Recall that the marginal distribution of the vector X is as follows: the random variables X_i are independent and $X_i \sim N\big(0,\tau_i^2(\beta_0)+n^{-1}\big)$, $i\in\mathbb{N}$.

Let $L_n(\beta) = L_n(\beta, X)$ be the marginal likelihood of the sample $X = (X_i)_{i\in\mathbb{N}}$:

$$L_n(\beta) = \prod_{i=1}^{\infty}\frac{1}{\sqrt{2\pi\big(\tau_i^2(\beta)+n^{-1}\big)}}\exp\Big\{-\frac{X_i^2}{2\big(\tau_i^2(\beta)+n^{-1}\big)}\Big\}.$$

Note that maximizing the function $L_n(\beta)$ is equivalent to minimizing the function $Z_n(\beta) = -2\log L_n(\beta)$. To avoid difficulties in finding the minimum of the function $Z_n(\beta)$ for $\{Z_n(\beta) = \pm\infty\}$, it is convenient to introduce a new variable $\bar{Z}_n(\beta) = Z_n(\beta, \bar{\beta}) = -2\log \dfrac{L_n(\beta)}{L_n(\bar{\beta})}$ for some reference value $\bar{\beta} > 0$. This variable is almost surely finite. For any set $S_n \subseteq (0, +\infty)$, define the marginal likelihood estimator of the parameter β over the set S_n,

$$\widehat{\beta} = \widehat{\beta}(S_n) = \widehat{\beta}(S_n, X, n) = \arg\min_{\beta \in S_n} \bar{Z}_n(\beta). \tag{9}$$

This means that $Z_n(\widehat{\beta}(S_n)) \le Z_n(\beta')$ for any $\beta' \in S_n$. Thus, $Z_n(\widehat{\beta}(S_n), \beta') \le 0$ for any $\beta' \in S_n$. Denote, for brevity,

$$a_i = a_i(\beta, \beta') = \frac{1}{\tau_i^2(\beta) + n^{-1}} - \frac{1}{\tau_i^2(\beta') + n^{-1}}, \tag{10}$$

$$b_i = b_i(\beta, \beta') = \frac{\tau_i^2(\beta) + n^{-1}}{\tau_i^2(\beta') + n^{-1}}. \tag{11}$$

Then $Z_n(\beta, \beta') = \sum\limits_{i=1}^{\infty} a_i(\beta, \beta') X_i^2 + \sum\limits_{i=1}^{\infty} \log b_i(\beta, \beta')$, and for any $\beta' \in S_n$ we have

$$\sum_{i=1}^{\infty} a_i(\widehat{\beta}(S), \beta') X_i^2 \le \sum_{i=1}^{\infty} \log \left[b_i(\widehat{\beta}(S), \beta') \right]^{-1}. \tag{12}$$

Throughout what follows, we assume that the following condition on the set S_n holds.

Condition S. We have $S_n = \{\varkappa_n + k\varepsilon_n,\ k = 0, 1, \ldots, M_n - 1\}$. The sequences $\varkappa_n, \varepsilon_n > 0$ and $M_n \in \mathbb{N}$ are such that $\varkappa_n \to 0$, $\varepsilon_n \to 0$, and $M_n\varepsilon_n \to \infty$ as $n \to \infty$.

Thus, the cardinality of the set S_n is $|S_n| = M_n < \infty$.

For an unknown parameter θ, we construct an empirical Bayes estimator $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots)$ as follows:

$$\widehat{\theta}_i = \widehat{\theta}_i(\widehat{\beta}) = \frac{\tau_i^2(\widehat{\beta}) X_i}{\tau_i^2(\widehat{\beta}) + n^{-1}}, \quad i \in \mathbb{N}. \tag{13}$$

2.3. Main Result

Theorem. *Let the estimator $\widehat{\theta}$ be defined by relation* (13)*. Assume that Condition S holds and the parameters ε_n, $\varkappa_n$, and M_n are such that $\varepsilon_n = o(1/\log n)$ and*

$$\varkappa_n^{-1} M_n^{1/2} \exp\left\{ -\frac{\varepsilon_n^2 (\log n)^2 n^{1/(2\beta_0+2)}}{128(2\beta_0+1)^2} \right\} = o(n^{-2\beta_0/(2\beta_0+1)})$$

as $n \to \infty$. Then

$$R(\widehat{\theta}, \theta) = R(\widehat{\theta}(\beta_0), \theta)(1 + o(1)) = \frac{n^{-2\beta_0/(2\beta_0+1)}\pi}{(2\beta_0+1)^2 \sin(\pi/(2\beta_0+1))}(1 + o(1))$$

as $n \to \infty$.

Remark 2. The sequences ε_n, $\varkappa_n$, and M_n can be chosen in different ways to satisfy the conditions of the theorem. For example, $\varepsilon_n = 1/(\log n)^2$, $\varkappa_n = 1/\log n$, and $M_n = \lfloor (\log n)^3 \rfloor$ will do. In fact, it does not make sense to choose the sequence $\varkappa_n$ converging to zero faster than $1/\log n$, since already for $\beta_0 = \varkappa_n = 1/\log n$ the risk does not converge to zero. It either makes no sense to choose the sequence M_n converging to infinity faster than a sequence with $\log n = o(M_n\varepsilon_n)$, since for $\beta_0 = \log n$ the parametric convergence rate of the risk is n^{-1}.

Proof. It is obvious that

$$\begin{aligned} R(\widehat{\theta}, \theta) &= \mathbf{E}\,\|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 \\ &= \mathbf{E}\left[\|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 I\{|\widehat{\beta} - \beta_0| \ge 2\varepsilon_n\}\right] + \mathbf{E}\left[\|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 I\{|\widehat{\beta} - \beta_0| < 2\varepsilon_n\}\right] = T_1 + T_2. \end{aligned}$$

Note that

$$T_2 = \mathbf{E}\left[\|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 I\{|\widehat{\beta} - \beta_0| < 2\varepsilon_n\}\right] \le \max_{\beta\colon |\beta - \beta_0| < 2\varepsilon_n} \mathbf{E}\,\|\widehat{\theta}(\beta) - \theta\|^2.$$

Recall that $\mathbf{E}(X_i - \theta_i)^2 = n^{-1}$ and $\mathbf{E}\,\theta_i^2 = i^{-(2\beta_0+1)}$. Then

$$\begin{aligned} \mathbf{E}\,\|\widehat{\theta}(\beta) - \theta\|^2 &= \sum_{i=1}^{\infty} \frac{\tau_i^4(\beta)\,\mathbf{E}(X_i - \theta_i)^2}{(\tau_i^2(\beta) + n^{-1})^2} + \sum_{i=1}^{\infty} \frac{n^{-2}\,\mathbf{E}\,\theta_i^2}{(\tau_i^2(\beta) + n^{-1})^2} \\ &= \sum_{i=1}^{\infty} \frac{n}{(i^{2\beta+1} + n)^2} + \sum_{i=1}^{\infty} \frac{i^{4\beta - 2\beta_0 + 1}}{(i^{2\beta+1} + n)^2} \\ &= \sum_{i=1}^{\infty} \frac{1}{i^{2\beta+1} + n} + \sum_{i=1}^{\infty} \frac{i^{4\beta-2\beta_0+1} - i^{2\beta+1}}{(i^{2\beta+1} + n)^2} \end{aligned}$$

as $n \to \infty$. For $\varepsilon_n = o(1/\log n)$ we have

$$\max_{\beta\colon |\beta - \beta_0| < 2\varepsilon_n} \sum_{i=1}^{\infty} \frac{1}{i^{2\beta+1} + n} = \mathbf{E}\,\|\widehat{\theta}(\beta_0) - \theta\|^2 (1 + o(1)).$$

Next, the following estimate holds uniformly over $|\beta - \beta_0| < 2\varepsilon_n$ for $\varepsilon_n = o(1/\log n)$:

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{|i^{4\beta-2\beta_0+1} - i^{2\beta+1}|}{(i^{2\beta+1} + n)^2} &\le \sum_{i=1}^{n} \frac{|i^{2\beta-2\beta_0} - 1|}{i^{2\beta+1} + n} + \sum_{i=n+1}^{\infty} \frac{1}{i^{2\beta_0+1}} \\ &\le o(1) \sum_{i=1}^{n} \frac{1}{i^{2\beta_0+1} + n} + \sum_{i=n+1}^{\infty} \frac{1}{i^{2\beta_0+1}} = o\big(n^{-2\beta_0/(2\beta_0+1)}\big) \end{aligned}$$

as $n \to \infty$. Thus,

$$T_2 \le \max_{\beta\colon |\beta - \beta_0| < 2\varepsilon_n} \mathbf{E}\,\|\widehat{\theta}(\beta) - \theta\|^2 \le \mathbf{E}\,\|\widehat{\theta}(\beta_0) - \theta\|^2 (1 + o(1))$$

as $n \to \infty$. Now we have to show that

$$T_1 = o\big(n^{-2\beta_0/(2\beta_0+1)}\big)$$

as $n \to \infty$. By the Cauchy–Bunyakovskii–Schwartz inequality,

$$T_1 = \mathbf{E}\left[\|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 I\{|\widehat{\beta} - \beta_0| \ge 2\varepsilon_n\}\right] \le \Big(\mathbf{E}\,\|\widehat{\theta}(\widehat{\beta}) - \theta\|^4\Big)^{1/2} \Big(\mathbf{P}\,\{|\widehat{\beta} - \beta_0| \ge 2\varepsilon_n\}\Big)^{1/2}.$$

Recall the following fact. Let $Z_1, Z_2, \ldots$ be independent, $Z_i \sim \mathcal{N}(0, \sigma_i^2)$, $\sum\limits_{i=1}^{\infty} \sigma_i^2 < \infty$; then we have $\mathbf{E}\Big(\sum\limits_{i=1}^{\infty} Z_i^2\Big)^2 \le 3\Big(\sum\limits_{i=1}^{\infty} \sigma_i^2\Big)^2$. Applying twice the elementary inequality $(a + b)^2 \le 2a^2 + 2b^2$ and the above property, we obtain

$$\begin{aligned}
\mathbf{E}\,\|\widehat{\theta}(\widehat{\beta})-\theta\|^4 &= \mathbf{E}\left[\sum_{i=1}^{\infty}\left(\frac{\tau_i^2(\widehat{\beta})(X_i-\theta_i)}{\tau_i^2(\widehat{\beta})+n^{-1}}-\frac{n^{-1}\theta_i}{\tau_i^2(\widehat{\beta})+n^{-1}}\right)^2\right]^2 \\
&\le 2\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{\tau_i^4(\widehat{\beta})\xi_i^2 n^{-1}}{(\tau_i^2(\widehat{\beta})+n^{-1})^2}+\frac{n^{-2}\theta_i^2}{(\tau_i^2(\widehat{\beta})+n^{-1})^2}\right]^2 \\
&\le 4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{\tau_i^4(\widehat{\beta})\xi_i^2 n^{-1}}{(\tau_i^2(\widehat{\beta})+n^{-1})^2}\right]^2+4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{n^{-2}\theta_i^2}{(\tau_i^2(\widehat{\beta})+n^{-1})^2}\right]^2 \\
&= 4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{n\xi_i^2}{(i^{2\widehat{\beta}+1}+n)^2}\right]^2+4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{i^{4\widehat{\beta}+2}\theta_i^2}{(i^{2\widehat{\beta}+1}+n)^2}\right]^2 \\
&\le 4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\frac{n\xi_i^2}{(i^{2\varkappa_n+1}+n)^2}\right]^2+4\,\mathbf{E}\left[\sum_{i=1}^{\infty}\theta_i^2\right]^2 \\
&\le 12\left[\sum_{i=1}^{\infty}\frac{n}{(i^{2\varkappa_n+1}+n)^2}\right]^2+12\left[\sum_{i=1}^{\infty}i^{-(2\beta_0+1)}\right]^2 \\
&\le 12\left[\sum_{i=1}^{\infty}i^{-(2\varkappa_n+1)}\right]^2+12\left[\sum_{i=1}^{\infty}i^{-(2\beta_0+1)}\right]^2 \\
&\le 12(1+(2\varkappa_n)^{-1})^2+12(1+(2\beta_0)^{-1})^2\le 4\varkappa_n^{-2}
\end{aligned}$$

for sufficiently large n.

Since $\varepsilon_n = o(1/\log n)$, by Lemma 2 for sufficiently large n we have

$$\mathbf{P}\{\widehat{\beta}=\beta\}\le\exp\left\{-\frac{\varepsilon_n^2(\log n)^2 n^{1/(2\beta_0+2)}}{64(2\beta_0+1)^2}\right\}$$

uniformly over all β such that $|\beta-\beta_0|\ge 2\varepsilon_n$. Therefore,

$$\mathbf{P}\left\{|\widehat{\beta}-\beta_0|\ge 2\varepsilon_n\right\}=\sum_{\beta:\,|\beta-\beta_0|\ge 2\varepsilon_n}\mathbf{P}\left\{\widehat{\beta}=\beta\right\}\le M_n\exp\left\{-\frac{\varepsilon_n^2(\log n)^2 n^{1/(2\beta_0+2)}}{64(2\beta_0+1)^2}\right\}$$

for sufficiently large n. Taking into account the latter relation and the conditions of the theorem, we obtain

$$T_1\le 2\varkappa_n^{-1}M_n^{1/2}\exp\left\{-\frac{\varepsilon_n^2(\log n)^2 n^{1/(2\beta_0+2)}}{128(2\beta_0+1)^2}\right\}=o(n^{-2\beta_0/(2\beta_0+1)}),$$

and the theorem follows. △

3. AUXILIARY RESULTS

Recall that $X_i\sim\mathcal{N}(0,n^{-1}+\tau_i^2(\beta_0))$, where β_0 is a real value of the unknown parameter β.

Lemma 1. *For all $\beta,\beta'\in S_n$ such that either $\beta'<\beta$ or $\beta<\beta'<\beta_0$ and for any $0<\lambda\le 1/2$, we have*

$$\mathbf{P}\{\widehat{\beta}=\beta\}\le\prod_{i=1}^{\infty}\left(\frac{\tau_i^2(\beta')+n^{-1}}{\tau_i^2(\beta)+n^{-1}}\right)^{\lambda}\left(1+2\lambda\frac{(\tau_i^2(\beta')-\tau_i^2(\beta))(\tau_i(\beta_0)+n^{-1})}{(\tau_i^2(\beta')+n^{-1})(\tau_i^2(\beta)+n^{-1})}\right)^{-1/2}. \qquad (14)$$

Proof. Denote $a_i = a_i(\beta, \beta')$ and $b_i = b_i(\beta, \beta')$. The proof follows from the Markov inequality. Indeed, since $\beta' \in S_n$, from the definition of $\widehat{\beta}$ we have

$$\begin{aligned}\mathbf{P}\{\widehat{\beta} = \beta\} &= \mathbf{P}\left\{Z_n(\beta, \beta'') \le 0,\ \forall\, \beta'' \in S_n\right\} \le \mathbf{P}\left\{Z_n(\beta, \beta') \le 0\right\} \\ &= \mathbf{P}\left\{-\sum_{i=1}^{\infty} a_i X_i^2 \ge \sum_{i=1}^{\infty} \log b_i\right\} \\ &\le \mathbf{E} \exp\left\{-\lambda \sum_{i=1}^{\infty} a_i X_i^2\right\} \exp\left\{\sum_{i=1}^{\infty} \log(b_i^{-\lambda})\right\}.\end{aligned}$$

Recall that $X_i \sim \mathcal{N}(0, n^{-1} + \tau_i^2(\beta_0))$. There is the following simple property of a Gaussian random variable $\eta \sim \mathcal{N}(\mu, \sigma^2)$:

$$\mathbf{E} \exp\{\varkappa \eta^2\} = (1 - 2\varkappa\sigma^2)^{-1/2} \exp\left\{\frac{\varkappa \mu^2}{1 - 2\varkappa\sigma^2}\right\} \quad \text{for} \quad \varkappa < \frac{1}{2\sigma^2}.$$

Taking into account this formula and the fact that $-\lambda a_i < 1/(2(n^{-1} + \tau_i^2(\beta_0)))$ for all i by the conditions of the lemma, we obtain

$$\begin{aligned}\mathbf{E} \exp\left\{-\lambda \sum_{i=1}^{\infty} a_i X_i^2\right\} &= \prod_{i=1}^{\infty} \left(1 + 2\lambda a_i(\beta, \beta')(\tau_i^2(\beta_0) + n^{-1})\right)^{-1/2} \\ &= \prod_{i=1}^{\infty} \left(1 + 2\lambda \frac{(\tau_i^2(\beta') - \tau_i^2(\beta))(\tau_i(\beta_0) + n^{-1})}{(\tau_i^2(\beta') + n^{-1})(\tau_i^2(\beta) + n^{-1})}\right)^{-1/2},\end{aligned}$$

which proves the lemma. △

Lemma 2. *Let $\beta \in S_n$, $|\beta - \beta_0| \ge 2\varepsilon_n$, where $\varepsilon_n = o(1/\log n)$. Then there exists a positive $M = M(\beta_0)$ such that for all $n \ge M$ we have*

$$\mathbf{P}\{\widehat{\beta} = \beta\} \le \exp\left\{-\frac{1}{64} n^{1/(2\beta_0 + \varepsilon_n + 1)} \left(\frac{\varepsilon_n \log n}{2\beta_0 + 1}\right)^2\right\}.$$

Proof. For $\lambda = 1/2$, relation (14) is given by $\mathbf{P}\{\widehat{\beta} = \beta\} \le \prod\limits_{i=1}^{\infty} (1 + z_i)^{1/2}$, where

$$\begin{aligned}z_i \equiv z_{i,n}(\beta, \beta_0, \beta') &= \frac{(\tau_i^2(\beta_0) - \tau_i^2(\beta'))(\tau_i^2(\beta) - \tau_i^2(\beta'))}{\tau_i^2(\beta)\tau_i^2(\beta') + \tau_i^2(\beta')\tau_i^2(\beta_0) - \tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta') + n^{-2}} \\ &= \frac{n^2(i^{2\beta'} - i^{2\beta})(i^{2\beta'} - i^{2\beta_0})}{n^2 i^{2\beta_0 + 2\beta'} + n^2 i^{2\beta' + 2\beta} - n^2 i^{4\beta'} + 2n i^{2\beta_0 + 2\beta' + 2\beta + 1} + i^{2\beta_0 + 4\beta' + 2\beta + 2}}.\end{aligned}$$

Let $\beta' \in S_n$ belong to the interval between β_0 and β (independently of which of these two variables is larger). Since

$$z_i = -1 + \frac{2n^{-1}\tau_i^2(\beta') + n^{-2} + \tau_i^4(\beta')}{\tau_i^2(\beta)\tau_i^2(\beta') + \tau_i^2(\beta')\tau_i^2(\beta_0) - \tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta') + n^{-2}},$$

we have $-1 \le z_i \le 0$ for all $i \in \mathbb{N}$. Thus, for any $\beta' \in S_n$ in the interval between β_0 and β (such that $\beta_0 \le \beta' \le \beta$ or $\beta \le \beta' \le \beta_0$), for all $N_1 \in \mathbb{N}$ and $N_2 \in \mathbb{N} \cup \{+\infty\}$ with $N_1 \le N_2$, we have

$$\mathbf{P}\{\widehat{\beta} = \beta\} \le \prod_{i=1}^{\infty} (1 + z_i)^{1/2} \le \exp\left\{\frac{1}{2} \sum_{i=1}^{\infty} z_i\right\} \le \exp\left\{\frac{1}{2} \sum_{i=N_1}^{N_2} z_i\right\}. \qquad (15)$$

Rewrite z_i as $z_i \equiv \dfrac{x_i}{y_i}$, where

$$x_i = n^2\big(i^{2\beta'} - i^{2\beta}\big)\big(i^{2\beta'} - i^{2\beta_0}\big),$$
$$y_i = n^2 i^{2\beta_0+2\beta'} + n^2 i^{2\beta'+2\beta} - n^2 i^{4\beta'} + 2n i^{2\beta_0+2\beta'+2\beta+1} + i^{2\beta_0+4\beta'+2\beta+2}.$$

First consider the case $\beta - \beta_0 \ge 2\varepsilon_n$ and, correspondingly, $\beta_0 < \beta' < \beta$. For all $i \ge N_1$ we have

$$\begin{aligned} x_i &= -n^2 i^{2\beta}\big(1 - i^{2\beta'-2\beta}\big) i^{2\beta'}\big(1 - i^{2\beta_0-2\beta'}\big) \\ &\le -n^2 i^{2(\beta+\beta')}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big). \end{aligned}$$

Similarly we can estimate y_i:

$$\begin{aligned} y_i &= n^2 i^{2\beta'}\big(i^{2\beta_0} - i^{2\beta'}\big) + i^{2\beta+2\beta'}\big(n^2 + 2n i^{2\beta_0+1} + i^{2\beta'+2\beta_0+2}\big) \\ &\le i^{2\beta+2\beta'}\big(n^2 + 2n i^{2\beta_0+1} + i^{2\beta'+2\beta_0+2}\big). \end{aligned}$$

If $i \le n^{\frac{1}{\beta'+\beta_0+1}}$, then $y_i \le 4n^2 i^{2(\beta+\beta')}$. Thus, using these estimates, we get

$$\begin{aligned} z_i = \frac{x_i}{y_i} &\le -\frac{n^2 i^{2(\beta+\beta')}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big)}{4n^2 i^{2(\beta+\beta')}} \\ &= -\frac{1}{4}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big), \end{aligned}$$

where $N_1 \le i \le N_2 \le n^{\frac{1}{\beta'+\beta_0+1}}$. Therefore, for all $N_1 \le i \le N_2 \le n^{1/(\beta_0+\beta'+1)}$, we have

$$\begin{aligned} \sum_{i=N_1}^{N_2} z_i &\le -\sum_{N_1}^{N_2} \frac{1}{4}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big) \\ &= -\frac{N_2 - N_1 + 1}{4}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big). \end{aligned}$$

Thus, using (15) and the latter inequality, for all $N_1 \le N_2 \le n^{1/(\beta_0+\beta'+1)}$ we obtain

$$\mathbf{P}\{\widehat{\beta} = \beta\} \le \exp\left\{\frac{1}{2}\sum_{i=N_1}^{N_2} z_i\right\} \le \exp\left\{-\frac{(N_2 - N_1 + 1)}{8}\Big(1 - N_1^{2(\beta'-\beta)}\Big)\Big(1 - N_1^{2(\beta_0-\beta')}\Big)\right\}. \qquad (16)$$

If $\beta \in S_n$ and $\beta - \beta_0 \ge 2\varepsilon_n$, then there exists $\beta' \in S_n$ between β_0 and β such that $\beta_0 + \varepsilon_n \le \beta' \le \min\{\beta_0 + 2\varepsilon_n, \beta - \varepsilon_n\}$ for sufficiently large n. Indeed, $\beta_0 \ge \varkappa_n + 2\varepsilon_n$ for sufficiently large n as $\varkappa_n \to 0$ and $\varepsilon_n \to 0$; thus, $\beta' = \min\{\beta \in S_n :\ \beta \ge \beta_0 + \varepsilon_n\}$ satisfies these inequalities. Note that for this choice of β' we have $\beta - \beta' \ge \varepsilon_n$ and $\beta' - \beta_0 \ge \varepsilon_n$.

Choose $N_2 = \lfloor n^{1/(\beta_0+\beta'+1)} \rfloor$ and $N_1 = \lfloor N_2/2 \rfloor + 1$. Since $\varepsilon_n \to 0$ and $\beta' \le \beta_0 + 2\varepsilon_n$, there exists $M_1 = M_1(\beta_0)$ such that for all $n \ge M_1$ we have

$$\log 2 - \frac{\log n}{\beta' + \beta_0 + 1} < -\frac{\log n}{2(2\beta_0 + 1)}.$$

Thus, for any $n \ge M_1$ we have

$$\begin{aligned} 1 - N_1^{-2(\beta-\beta')} &\ge 1 - (N_2/2)^{-2(\beta-\beta')} \\ &\ge 1 - \exp\{-2(\beta-\beta')[\log N_2 - \log 2]\} \\ &= 1 - \exp\left\{2(\beta-\beta')[\log 2 - \frac{\log n}{\beta' + \beta_0 + 1}]\right\} \\ &\ge 1 - \exp\left\{-2(\beta-\beta')\frac{\log n}{2(2\beta_0+1)}\right\} \\ &\ge 1 - \exp\left\{-\varepsilon_n \frac{\log n}{2\beta_0+1}\right\} \end{aligned}$$

(recall that $\beta - \beta' \ge \varepsilon_n$). Applying the inequality $1 - e^{-x} \ge x/2$ for $0 \le x \le 3/2$ and the fact that for $\varepsilon_n = o(1/\log n)$ there exists $M_2 = M_2(\beta_0) > 0$ such that $\varepsilon_n \frac{\log n}{2\beta_0+1} < 3/2$ for all $n \ge M_2$, we obtain that for any $n \ge \max(M_1, M_2)$ we have

$$1 - N_1^{-2(\beta-\beta')} \ge \frac{\varepsilon_n \log n}{2(2\beta_0+1)}.$$

Similarly, $1 - N_1^{-2(\beta'-\beta_0)} \ge \frac{\varepsilon_n \log n}{2(2\beta_0+1)}$ for all $n \ge \max(M_1, M_2)$. Moreover, we have

$$N_2 - N_1 + 1 \ge \frac{1}{2}N_2 \ge \frac{1}{2}n^{\frac{1}{\beta'+\beta_0+1}} \ge \frac{1}{2}n^{\frac{1}{2\beta_0+\varepsilon_n+1}}.$$

Substituting the obtained estimates into (16) proves the lemma for $\beta - \beta_0 \ge 2\varepsilon_n$.

Now consider the case $\beta - \beta_0 \le -2\varepsilon_n$. We can apply the same reasoning slightly changed. Choose $\beta' = \max\{\beta \in S_n : \beta \le \beta_0 - \varepsilon_n\}$, $N_2 = \lfloor n^{1/(\beta+\beta'+1)} \rfloor$, and $N_1 = \lfloor N_2/2 \rfloor + 1$. Then $n^2 i^{2\beta'+2\beta_0}$ is the main term in x_i and y_i. We have

$$x_i = n^2\big(i^{2\beta'} - i^{2\beta}\big)\big(i^{2\beta'} - i^{2\beta_0}\big) \le -n^2 i^{2\beta_0+2\beta'}\Big(1 - N_1^{2(\beta'-\beta_0)}\Big)\Big(1 - N_1^{2(\beta-\beta')}\Big)$$

for all $i \ge N_1$, and

$$y_i \le 4n^2 i^{2\beta_0+2\beta'}$$

for all $i \le N_2$. Then $z_i = \frac{x_i}{y_i} \le -\frac{1}{4}\Big(1 - N_1^{2(\beta'-\beta_0)}\Big)\Big(1 - N_1^{2(\beta-\beta')}\Big)$. Estimating the factors on the right-hand side of this relation in a similar way and substituting the estimates into (16), we obtain the statement of the lemma for $\beta - \beta_0 \le -2\varepsilon_n$. △

Remark 3. We can formally calculate the Fisher information for $X_i \sim \mathcal{N}(0, \tau_i^2(\beta) + n^{-1})$, which is

$$I(\beta) = 2\sum_{i=1}^{\infty}\left(\frac{n \log i}{n + i^{2\beta+1}}\right)^2.$$

It is easy to see that it is of the order of $O\Big((\log n)^2 n^{\frac{1}{2\beta+1}}\Big)$. Thus, apparently, the choice of a sequence ε_n such that $\varepsilon_n^2 = o(n^{-1/(2\beta+1)} \log^{-2} n)$ gives the optimal rate of convergence of the estimate $\widehat{\beta}$ to β_0. However, this plays no role in the proof of the main result.

REFERENCES

1. Ibragimov, I.A. and Khas'minskii, R.Z., *Asimptoticheskaya teoriya otsenivaniya*, Moscow: Nauka, 1979. Translated under the title *Statistical Estimation. Asymptotic Theory*, New York: Springer, 1981.

2. Johnstone, I., *Function Estimation and Gaussian Sequence Models*, draft of a monograph, 1999. Available at `http://www-stat.stanford.edu/~imj/baseb.pdf`.

3. Nussbaum, M., Asymptotic Equivalence of Density Estimation and Gaussian White Noise, *Ann. Statist.*, 1996, vol. 24, no. 6, pp. 2399–2430.

4. Brown, L.D. and Low, M.G., Asymptotic Equivalence of Nonparametric Regression and White Noise, *Ann. Statist.*, 1996, vol. 24, no. 6, pp. 2384–2398.

5. Pinsker, M.S., Optimal Filtering of Square-Integrable Signals in Gaussian Noise, *Probl. Peredachi Inf.*, 1980, vol. 16, no. 2, pp. 52–68 [*Probl. Inf. Trans.* (Engl. Transl.), 1980, vol. 16, no. 2, pp. 120–133].

6. Efroimovich, S.Yu. and Pinsker, M.S., A Self-training Algorithm for Nonparametric Filtering, *Avtomat. i Telemekh.*, 1984, no. 11, pp. 58–65.

7. Lepskii, O.V., A Problem of Adaptive Estimation in Gaussian White Noise, *Teor. Veroyatnost. i Primenen.*, 1990, vol. 35, no. 3, pp. 459–470 [*Theory Probab. Appl.* (Engl. Transl.), 1990, vol. 35, no. 3, pp. 454–466].

8. Lepskii, O.V., Asymptotically Minimax Adaptive Estimation. I: Upper Bounds. Optimally Adaptive Estimates, *Teor. Veroyatnost. i Primenen.*, 1991, vol. 36, no. 4, pp. 645–659 [*Theory Probab. Appl.* (Engl. Transl.), 1991, vol. 36, no. 4, pp. 682–697].

9. Lepskii, O.V., Asymptotically Minimax Adaptive Estimation. II: Schemes without Optimal Adaptation. Adaptive Estimates, *Teor. Veroyatnost. i Primenen.*, 1992, vol. 37, no. 3, pp. 468–481 [*Theory Probab. Appl.* (Engl. Transl.), 1992, vol. 37, no. 3, pp. 433–448].

10. Golubev, G.K. and Nussbaum, M., Adaptive Spline Estimates in a Nonparametric Regression Model, *Teor. Veroyatnost. i Primenen.*, 1992, vol. 37, no. 3, pp. 554–561 [*Theory Probab. Appl.* (Engl. Transl.), 1992, vol. 37, no. 3, pp. 521–529].

11. Cavalier, L. and Tsybakov, A.B., Penalized Blockwise Stein's Method, Monotone Oracles and Sharp Adaptive Estimation, *Math. Methods Statist.*, 2001, vol. 10, no. 3, pp. 247–282.

12. Robbins, H., An Empirical Bayes Approach to Statistics, in *Proc. 3rd Berkeley Symp. on Mathematical Statistics and Probability, 1954–1955*, vol. 1, Berkeley: Univ. of California Press, 1956, pp. 157–163.

13. Belitser, E. and Levit, B., On the Empirical Bayes Approach to Adaptive Filtering, *Math. Methods Statist.*, 2003, vol. 12, no. 2, pp. 131–154.

14. Belitser, E. and Ghosal, S., Adaptive Bayesian Inference on the Mean of an Infinite-Dimensional Normal Distribution, *Ann. Statist.*, 2003, vol. 31, no. 2, pp. 536–559.