Asymptotically Minimax Nonparametric Regression in $\mathbf{L}_2$

Eduard N. Belitser Boris Y. Levit

University of Utrecht

Summary. Additive nonparametric regression with equidistant observation design is considered. The Pinsker-type minimax results are derived and the linear asymptotically minimax estimators are exhibited, based on approximation of the initial nonparametric model by a linear models of dimension which is increasing with the number of observations. The proof of optimality of these linear estimators within the class of all possible estimators is based on the rather elementary but very useful van Trees inequality.

AMS 1991 *subject classification:* 62G05, 62G20.

Key words: ellipsoid, asymptotic minimax risk, Bayes risk, asymptotically minimax estimator, second-order rate minimax estimator.

1 Introduction

In the nonparametric regression context, the notion of asymptotic optimality usually associates with the "optimal rate of convergence". Minimax rates of convergence have been extensively studied (Ibragimov and Hasminskii (1980), (1982); Stone (1980), (1982) and many others).

Different estimators turn out to be optimal in the sense of the best rate of convergence. We mention only some of them: kernel estimators (Ibragimov and Hasminskii (1980), Korostelev (1993)), projection estimators (Ibragimov and Hasminskii (1981)), spline estimators (Speckman (1985), Nussbaum (1985)), wavelets (Donoho and Johnstone (1992)). From the practical point of view stochastic approximation estimators considered in Belitser and Korostelev (1992) are also of interest.

However, comparing estimators on the basis of their rates of convergence does not make it possible to distinguish among estimators optimal in that sence. Also from a more practical point of view, such approach does not give a standard recipe for choosing parameters of the estimator involved: the bandwidth for the kernel method, the number of terms for the orthogonal series method, etc. Thus two estimators, optimal in the sense of the rate of convergence, can compare in actual applications quite poorly.

The minimax approach becomes more accurate if the constants involved in the lower and upper bounds are found, especially when these constants happen to coincide. The problem of finding the exact constants tends to be of an increasing interest. First result of this kind in nonparametric estimation problem was obtained by Pinsker (1980) for a white noise model. The essence of Pinsker's method consists in showing that minimax linear estimators are asymptotically minimax in the class of all estimators.

In the nonparametric regression context such approach was studied by, among others, Speckman (1985), Nussbaum (1985), Golubev and Nussbaum (1990), Efroimovich (1994). In the paper of Speckman (1985) the minimax linear estimator is a spline. The first result about asymptotics of the minimax risk within the class of all estimators in a regression context is due to Nussbaum (1985), where normality of the errors was assumed, nonparametric class is Sobolev class and a smoothing spline proved to be asymptotically minimax among all estimators. Exact lower bounds for the minimax risk were obtained in the paper of Golubev and Nussbaum (1990) for nonequidistant designs of observations without assumption of normality of the errors. In a recent paper Efroimovich (1994) studied exact asymptotic behaviour of the minimax risk for random design nonparametric regression models also without assumption of normality.

We establish exact asymptotics of the minimax risk in case of Gaussian errors, by methods related to but different from those of the papers mentioned above. Our treatment of the lower bound is based on the elementary but rather powerful van Trees inequality (van Trees (1968)). For further references and applications of the van Trees inequality see Borovkov (1984), Gill and Levit (1992). Another approach for obtaining the lower bound based on asymptotic equivalence of the original model and the white noise model has been also actively pursued recently, see Brown and Low (1992), Nussbaum (1995).

Our approach with respect to the upper bound is based on equivalence of the initial nonparametric model to a sequence of linear models of increasing dimensions. Namely, with the class of regression functions $f(x)$ under consideration, our problem of estimating $f(x)$ is equivalent to that of estimating an infinite-dimensional parameter $(\theta_i,\ i = 1, 2, \ldots)$ based on observations:

$$Z_i = \theta_i + \tilde{\theta}_i + n^{-1/2}\xi_i\,, \quad i = 1, 2, \ldots, n\,.$$

Here ξ_l's are Gaussian random variables, $\mathbf{E}\xi_i = 0$, $\mathbf{E}[\xi_l \xi_k] = \sigma^2\delta_{lk}$ (δ_{kl} denotes the Kroneker symbol), $\tilde{\theta}_i$'s are "nuisance" parameters , which are negligibly small (see Remark 7) provided $f(x)$ belongs to correspondent classes of smooth functions.

2 The model and main results

Below we study the problem of estimating a nonparametric regression function $f(x)$, $x \in [0, 1]$ on the basis of the observations

$$Y_i = f(t_{in}) + \epsilon_i\,, \quad i = 1, 2, \ldots, n\,, \tag{2.1}$$

where ϵ_i's are iid Gaussian random variables with zero mean and variance σ^2.

The design is assumed to be equidistant: $t_{in} = i/n, \quad i = 1, 2, \ldots, n$. For simplicity some variables and dependence subscript n will frequently be dropped from notation.

Let $\mathbf{L}_2 = \mathbf{L}_2[0,1]$ be the Hilbert space of square-integrable functions on $[0,1]$ and $\{\phi_k(x),\ k = 1, 2, \ldots\}$ be its orthonormal trigonometrical basis, i.e.

$$\phi_j(x) = \begin{cases} 1, & j = 1 \\ \sqrt{2}\sin(2\pi k x), & j = 2k \\ \sqrt{2}\cos(2\pi k x), & j = 2k + 1\,. \end{cases}$$

We assume that $f(x) \in \mathbf{L}_2[0,1]$. Hence it can be represented as follows:

$$f(x) = \sum_{k=1}^{\infty} \theta_k \phi_k(x)\,, \qquad \text{where} \qquad \theta_k = \int_0^1 f(x)\phi_k(x)dx\,.$$

Here convergence is meant in $\mathbf{L}_2$-sence.

Let $(a_k,\ k = 1, 2, \ldots)$ be a positive numerical sequence converging to infinity. Now we define the nonparametric class which is ellipsoid in Hilbert space l_2 (cf. Efroimovich and Pinsker (1982)):

$$\Theta = \Theta(Q) = \left\{ f(\cdot) \in \mathbf{L}_2 :\ \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq Q \right\}.$$

We are interested in the asymptotic behaviour of the minimax risk

$$r_n = r_n(\Theta) = \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2\,.$$

Here *inf* is taken over all estimators and *sup* is taken over all regression curves from class Θ. We call an estimator $\hat{f}_n$ *asymptotically minimax* if

$$R_n(\hat{f}_n) \stackrel{\text{def}}{=} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 = r_n(\Theta)(1 + o(1)) \quad \text{as} \quad n \to \infty\,.$$

All asymptotic equations below refer to, unless otherwise specified, $n \to \infty$.

To formulate the results about the minimax risk, we introduce some notations. Since $\lim_{k\to\infty} a_k = \infty$, the equation

$$\sum_{k=1}^{\infty} \sigma^2 a_k (1 - c_n a_k)_+ = c_n Q n \tag{2.2}$$

has the unique solution $c_n = c_n(\sigma^2, \Theta) > 0$. Here x_+ denotes nonnegative part of x. Denote

$$\mathcal{I} = \mathcal{I}_n(\sigma^2, \Theta) = \{k :\ 0 \leq c_n a_k \leq 1\}\,, \qquad N = N_n(\sigma^2, \Theta) = \operatorname{card} \mathcal{I}\,, \tag{2.3}$$

$$d_n = d_n(\sigma^2, \Theta) = n^{-1} \sum_{k=1}^{\infty} \sigma^2 (1 - c_n a_k)_+. \tag{2.4}$$

Note that in fact $\mathcal{I}$ is the set of indexes k's for which correspondent terms in sum (2.2) are nonzero.

Remark 1. Suppose that sequence $(a_k,\ k=1,2,\ldots)$ is nondecreasing. Then from (2.2) it is easy to see that $\mathcal{I}=\{1,2,\ldots,N\}$ and

$$c_n=\frac{\sum_{k=1}^N a_k}{Qn\sigma^{-2}+\sum_{k=1}^N a_k^2}\,,$$

where N is the number of nonzero terms in sum (2.2). One can verify that

$$N=\max\{k:\ a_k\le c_n^{-1}\}=\max\left\{l:\ \sum_{k=1}^{l}(a_ka_l-a_k^2)\le Q\sigma^{-2}n\right\}.$$

Remark 2. All results below remain valid under the weaker condition $a_k\ge 0$, $k=1,2,\ldots$, instead of strict positivity of a_k's since only finitely many zero a_k's are possible.

Denote next

$$\psi_n(\gamma)=\psi_n(\gamma,\sigma^2,\Theta)=\exp\left\{\frac{-\gamma n^2c_n^2}{\sigma^4\sum_{k=1}^\infty a_k^2(1-c_na_k)_+^2}\right\}.$$

We define also two conditions:

$$\mathcal{F}_1=\mathcal{F}_1(\sigma^2):\quad \{\text{ for any } \gamma>0\quad \psi_n(\gamma)=o(c_n)\},$$

$$\mathcal{F}_2=\mathcal{F}_2(\sigma^2):\quad \left\{c_n\sum_{k\in\mathcal{I}}a_k=o(N)\right\}.$$

Here $c_n=c_n(\sigma^2,\Theta)$, $\mathcal{I}=\mathcal{I}_n(\sigma^2,\Theta)$ and $N=N_n(\sigma^2,\Theta)$ are defined by (2.2)–(2.3).

Next theorem gives the asymptotic lower bound for the minimax risk.

Theorem 1 *If condition $\mathcal{F}_1$ or $\mathcal{F}_2$ is fulfilled, then*

$$r_n(\Theta)\ge d_n(\sigma^2,\Theta)(1+o(1))\,,$$

where d_n is defined by (2.2), (2.4).

Remark 3. Consider the topology generated by the following norm:

$$\|g\|=\left(\sum_{k=1}^\infty a_k^2g_k^2\right)^{1/2},$$

where g_k's are Fourier coefficients of $g(\cdot)\in\mathbf{L}_2[0,1]$. If we substitute any ball S, $S\subseteq\Theta$ of radius Q in definition of r_n instead of Θ, then Theorem 1 and Corollary 7 still hold. The proof is in essence the same. Note that the lower bounds do not depend on center of ball S.

Now we construct the estimator which is going to be efficient for ellipsoids satisfying certain regularity condition. Define

$$\hat{f}_n^M(x) = \sum_{k=1}^{n} \lambda_k \hat{\theta}_k \phi_k(x) \quad \text{with} \quad \hat{\theta}_k = n^{-1} \sum_{i=1}^{n} \phi_k(i/n) Y_i \,, \tag{2.5}$$

$$\lambda_k = (1 - c_n a_k)_+ \,, \tag{2.6}$$

where $c_n = c_n(\sigma^2, \Theta)$ is given by (2.2). We see that the estimator $\hat{f}_n^M(x)$ is a generalized kernel estimator

$$\hat{f}_n^M(x) = \sum_{i=1}^{n} K_n(x, i/n) Y_i \,,$$

where its kernel is defined as follows:

$$K_n(x, i/n) = n^{-1} \sum_{k=1}^{n} (1 - c_n a_k)_+ \, \phi_k(i/n) \phi_k(x) \,. \tag{2.7}$$

We introduce conditions under either of which we derive the upper bound for the minimax risk:

$$\mathcal{F}_3 : \quad \left\{ \max_{1 \le k \le n} \sum_{l=1}^{\infty} a_{k+ln}^{-2} = o(n^{-1}) \right\} ,$$

$$\mathcal{F}_4 = \mathcal{F}_4(\sigma^2) : \quad \left\{ \sum_{k=n}^{\infty} a_k^{-2} = o(d_n) \right\} ,$$

where $d_n = d_n(\sigma^2, \Theta)$ is defined by (2.2), (2.4).

Theorem 2 *If condition $\mathcal{F}_3$ or $\mathcal{F}_4$ is fulfilled, then*

$$\sup_{\Theta} \mathbf{E}_f \| \hat{f}_n^M - f \|^2 \le d_n(\sigma^2, \Theta)(1 + o(1)) \,,$$

where estimator $\hat{f}_n^M$ is defined by (2.5)–(2.6) *and d_n is defined by* (2.2), (2.4).

Remark 4. If ϵ_k's are iid zero mean random variables with variance σ^2 (not necessarily Gaussian), then Theorem 2 remains unchanged. If ϵ_k's are iid zero mean random variables with distribution density $p_\epsilon(x)$ and finite Fisher information

$$I_\epsilon = \int \left(\frac{\partial \log p_\epsilon(x)}{\partial x} \right)^2 p_\epsilon(x) dx \,,$$

then Theorem 1 and Corollary 7 still hold with I_ϵ^{-1} in place of σ^2. Thus, only for Gaussian errors the lower and upper bounds coincide asymptotically. The lower bound is apparently asymptotically exact and the minimax estimator is likely to be no longer linear (cf. Efroimovich (1994) for a related model).

We immediately conclude from Theorem 2 and Theorem 2 the following result.

Corollary 1 *Let either of conditions $\mathcal{F}_1$, $\mathcal{F}_2$ and either of conditions $\mathcal{F}_3$, $\mathcal{F}_4$ be fulfilled. Then*

$$r_n(\Theta) = d_n(\sigma^2, \Theta)(1 + o(1))$$

and estimator $\hat{f}_n^M$ is asymptotically minimax.

Consider the problem of the robust estimation of unknown regression function $f(x) \in \Theta$.

Corollary 2 *Let either of conditions $\mathcal{F}_1$, $\mathcal{F}_2$ and either of conditions $\mathcal{F}_3$, $\mathcal{F}_4$ be fulfilled. Then*

$$\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 = d_n(\sigma^2, \Theta)(1 + o(1)),$$

where Π is set of all distributions of noises with zero mean and variance σ^2.

Surely, on the one hand

$$\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 \geq \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 \geq d_n(\sigma^2, \Theta)(1 + o(1)),$$

where p_ϵ in the right-hand side is taken to be Gaussian. On the other hand, according to Remark 4, we have

$$\inf_{\hat{f}_n} \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n - f\|^2 \leq \sup_{\Theta} \sup_{p_\epsilon \in \Pi} \mathbf{E}_{f,p_\epsilon} \|\hat{f}_n^M - f\|^2 \leq d_n(\sigma^2, \Theta)(1 + o(1)),$$

where $\hat{f}_n^M$ is the estimator defined by (2.5)–(2.6).

3 Examples

If an ellipsoid Θ is such that for some positive constant $C = C(\Theta)$ and positive decreasing to zero sequence ψ_n the asymptotics

$$r_n(\Theta) = C(\Theta)\psi_n^2(1 + o(1))$$

holds, then, clearly, ψ_n is the rate of convergence, asymptotically minimax estimator is asymptotically minimax within a constant and constant $C(\Theta)$ is optimal. We describe below examples where this is the case.

Example 1. Let, for a given α, $\alpha > 1/2$,

$$\Theta = \left\{ f(\cdot) \in \mathbf{L}_2 : \sum_{k=1}^{\infty} k^{2\alpha}(\theta_{2k}^2 + \theta_{2k+1}^2) \leq Q \right\}. \tag{3.1}$$

We have to impose the condition $\alpha > 1/2$ in order to provide $\Theta \in \mathcal{F}_1 \cap \mathcal{F}_3$.

Corollary 3 *Let the ellipsoid Θ be defined by* (3.1). *Then*

$$r_n = Q^{\frac{1}{2\alpha+1}} \sigma^{\frac{4\alpha}{2\alpha+1}} (2\alpha/(\alpha+1))^{\frac{2\alpha}{2\alpha+1}} (2\alpha+1)^{\frac{1}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}} (1 + o(1))$$

and estimator $\hat{f}_n^M$ is asymptotically minimax.

Proof. Condition $\Theta \in \mathcal{F}_3$ can be verified straightforward:

$$\max_{1\le k\le n}\sum_{k=1}^{n} a_{k+ln}^{-2} \le 2^{2\alpha}\sum_{l=1}^{\infty}(ln)^{-2\alpha} \le Cn^{-2\alpha} = o(n^{-1})\,.$$

We calculate now the asymptotic value of d_n. In this case it is easy to prove that $c_n N^\alpha \to 1$ as $n\to\infty$. Therefore, $N = c_n^{-1/\alpha}(1+o(1))$. The equation (2.2) to define c_n is as follows:

$$2\sum_{k=1}^{N}(k^\alpha - c_n k^{2\alpha}) = Q\sigma^{-2}nc_n\,.$$

So, making use of asymptotic equality

$$\sum_{k=1}^{N} k^\alpha = \frac{N^{\alpha+1}}{\alpha+1}(1+o(1)) \quad \text{as} \quad N\to\infty\,,\;\; \alpha>-1\,, \tag{3.2}$$

we obtain the asymptotic relations:

$$c_n = \left(\frac{2\alpha\sigma^2}{(\alpha+1)(2\alpha+1)Qn}\right)^{\frac{\alpha}{2\alpha+1}}(1+o(1))\,,$$

$$N = \Big((2\alpha+1)(\alpha+1)Qn/(2\alpha\sigma^2)\Big)^{\frac{1}{2\alpha+1}}(1+o(1)). \tag{3.3}$$

Using this, (2.4) and again (3.2), we find that

$$d_n = n^{-\frac{2\alpha}{2\alpha+1}}Q^{\frac{1}{2\alpha+1}}(2\alpha+1)^{\frac{1}{2\alpha+1}}(2\alpha\sigma^2/(\alpha+1))^{\frac{2\alpha}{2\alpha+1}}(1+o(1))\,.$$

Next, one makes sure easily that $\Theta\in\mathcal{F}_1$:

$$\psi_n(\gamma) = O\left(\exp\left\{-\gamma_1 n^{1/(2\alpha+1)}\right\}\right) = o(c_n)$$

for some $\gamma_1>0$. Finally, applying Corollary 1, we get the statement of Corollary 3. □

Remark 5. Let, for a natural α and $f(\cdot)\in\mathbf{L}_2([0,1])$, $D^\alpha f$ denote the derivative of order α in distributional sense, and let

$$\widetilde{W}_2^\alpha(Q) = \left\{f(\cdot)\in\mathbf{L}_2 :\; \|D^\alpha f\|^2\le Q,\; D^l f(0) = D^l f(1),\; l=0,1,\ldots,\alpha-1\right\}$$

be the αth order periodic Sobolev space on the unit interval. Then the following asymptotic equation holds (cf. Nussbaum (1985) and Golubev and Nussbaum (1990)):

$$r_n(\widetilde{W}_2^\alpha(Q)) = r_n(\Theta(Q/(2\pi)^{2\alpha}))(1+o(1)) = \gamma(\alpha,Q)\sigma^{\frac{4\alpha}{2\alpha+1}}n^{-\frac{2\alpha}{2\alpha+1}}(1+o(1))\,,$$

where $\Theta(Q)$ is defined by (3.1) and $\gamma(\alpha,Q) = (Q(2\alpha+1))^{\frac{1}{2\alpha+1}}(\alpha/\pi(\alpha+1))^{\frac{2\alpha}{2\alpha+1}}$ is Pinsker's constant. Indeed, the upper bound follows from the relation:

$$\widetilde{W}_2^\alpha(Q)\subseteq\Theta(Q/(2\pi)^{2\alpha})\,.$$

The proof of the lower bound carries through literally since

$$\sum_{k=1}^{N} \theta_k \phi_k(x) \in \widetilde{W}_2^\alpha(Q) \qquad \text{if} \quad \theta \in \Theta(Q/(2\pi)^{2\alpha}) \,.$$

Thus, nonparametric class (3.1) can be viewed as an extension of class $\widetilde{W}_2^\alpha$ for nonperiodic functions and nonnatural α.

For the ellipsoid defined by (3.1), it is not difficult to get the expression for the kernel (2.7):

$$K_n(x, i/n) = n^{-1} \left(1 + 2 \sum_{k=1}^{N} (1 - c_n k^\alpha) \cos(2\pi k(x - i/n)) \right) .$$

Note that this is a discrete version of the kernel described in Golubev (1987) (p. 49).

Consider an estimator $\tilde{f}_n^M$ defined by (2.5)–(2.6) with $c_n = N_n^{-\alpha}$, where N_n is an arbitrary sequence satisfying (3.3). This estimator is asymptotically minimax over class (3.1). Indeed, following the proof of Theorem 2, one obtains

$$R_n(\tilde{f}_n^M) = \sup_\Theta \left\{ \sum_{k=1}^{n} \left(n^{-1}\sigma^2 \lambda_k^2 + (1 - \lambda_k)^2 \theta_k^2 \right) \right\} + o(n^{-1}) \,.$$

Further, the same reasoning as in (4.3) leads to

$$\sup_\Theta \left\{ \sum_{k=1}^{n} \left(n^{-1}\sigma^2 \lambda_k^2 + (1 - \lambda_k)^2 \theta_k^2 \right) \right\} \le Q c_n^2 + n^{-1} \sum_{k=1}^{\infty} \sigma^2 (1 - c_n a_k)_+^2 \,.$$

Combining last relations and using (3.2), one computes

$$R_n(\tilde{f}_n^M) \le Q N_n^{-2\alpha} + n^{-1} 2 \sum_{k=1}^{N_n} \sigma^2 (1 - k^\alpha N_n^{-\alpha})^2 + O(n^{-1}) = d_n(1 + o(1)) \,.$$

For the ellipsoid defined by (3.1), with $\alpha = 1$, one can get by routine calculations the expression for the kernel corresponding to the estimator $\tilde{f}_n^M$:

$$K_n(x, i/n) = \frac{\sin^2(N_n \pi (x - i/n))}{n N_n \sin^2(\pi(x - i/n))}$$

which is well known Feier kernel. For $\alpha = 2$, the kernel of the estimator $\tilde{f}_n^M$ is as follows:

$$K_n(x, i/n) = \frac{\sin(2N_n\pi(x - i/n)) \cos(\pi(x - i/n))}{2nN_n^2 \sin^3(\pi(x - i/n))} - \frac{\cos(2N_n\pi(x - i/n))}{nN_n \sin^2(\pi(x - i/n))} \,.$$

Example 2. Let

$$\Theta = \left\{ f(\cdot) \in \mathbf{L}_2 : \sum_{k=1}^{\infty} e^{2\beta k} (\theta_{2k}^2 + \theta_{2k+1}^2) \le Q \right\} . \tag{3.4}$$

In this case it has been possible to describe the minimax risk up to the rate of second-order term.

Corollary 4 *Let the ellipsoid* Θ *be defined by* (3.4). *Then*

$$r_n = \sigma^2\beta^{-1}n^{-1}\log n + O(n^{-1}) \tag{3.5}$$

and estimator $\hat{f}_n^M$ *is asymptotically minimax (also second-order rate minimax).*

Proof. From (2.3) one can see that

$$e^{\beta} \leq c_n e^{N\beta} \leq 1\,. \tag{3.6}$$

The equation (2.2) in this case is written as follows:

$$2\sum_{k=1}^{N}(e^{\beta k} - c_n e^{2\beta k}) = Q\sigma^{-2}nc_n\,.$$

Using this and (3.6), one obtains the asymptotics

$$N = (2\beta)^{-1}\log n + O(1)\,.$$

According to Corollary 7 and the proof of Theorem 2, we have

$$d_n(\sigma^2, \Theta(Q/\pi^2)) \leq r_n \leq d_n(\sigma^2, \Theta(Q)) + \delta_n\,, \tag{3.7}$$

where

$$|\delta_n| \leq Qa_n^{-2} + 2Q\sum_{k=n+1}^{\infty} a_k^{-2} = O\left(e^{-\beta n}\right)\,.$$

Further, using asymptotics for N and again (3.6), we have

$$d_n = \frac{\sigma^2}{n} + \frac{2\sigma^2}{n}\sum_{k=1}^{N}(1 - c_n e^{\beta k}) = \frac{\sigma^2}{\beta}\frac{\log n}{n} + O(n^{-1})\,.$$

From this and (3.7) we finally obtain

$$r_n = \sigma^2\beta^{-1}n^{-1}\log n + O(n^{-1})\,.$$

Corollary 4 is proved. □

The optimal constant does not depend on the "size" Q of ellipsoid Θ. That is why stronger result is available. Namely,

Corollary 5 *Let the ellipsoid* Θ *be defined by* (3.4). *Then for any vicinity* $V \subseteq \Theta$

$$\inf_{\hat{f}_n}\sup_{V} \mathbf{E}_f\|\hat{f}_n - f\|^2 = \sigma^2\beta^{-1}n^{-1}\log n + O(n^{-1})\,,$$

where the meant topology is generated by the norm defined in Remark 3.

Indeed, let S be such a ball that $S \subseteq V$. Then, on the one hand, according to Remark 3, we have

$$\inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2 \geq \inf_{\hat{f}_n} \sup_{S} \mathbf{E}_f \|\hat{f}_n - f\|^2 = \sigma^2 \beta^{-1} n^{-1} \log n + O(n^{-1})$$

and, on the other hand,

$$\inf_{\hat{f}_n} \sup_{V} \mathbf{E}_f \|\hat{f}_n - f\|^2 \leq \inf_{\hat{f}_n} \sup_{\Theta} \mathbf{E}_f \|\hat{f}_n - f\|^2 = \sigma^2 \beta^{-1} n^{-1} \log n + O(n^{-1}) .$$

Remark 6. Let the ellipsoid Θ be defined by (3.4). Consider the projection estimator $\hat{f}_n^P$ defined by (2.5) with

$$\lambda_k = \begin{cases} 1, & k \leq N_n \\ 0, & k > N_n , \end{cases}$$

where N_n is any positive sequence satisfying the inequality:

$$|N_n - \beta^{-1} \log n| \leq (1-\mu)\beta^{-1} \log\log n \quad \text{for some } \mu > 0 .$$

The estimator $\hat{f}_n^P$, while being simpler than the estimator $\hat{f}_n^M$ above, is still asymptotically minimax, i.e.

$$R_n(\hat{f}_n^P) = \sigma^2 \beta^{-1} n^{-1} \log n \, (1 + o(1)) .$$

If $N_n = \beta^{-1} \log n$, then the estimator $\hat{f}_n^P$ is asymptotically second-order minimax:

$$R_n(\hat{f}_n^P) = \sigma^2 \beta^{-1} n^{-1} \log n + n^{-1} Q (1 + o(1)) .$$

On the other hand, consider the estimator $\hat{f}_n^{VP}$ corresponding to the de la Vallee Poussin kernel (cf. Ibragimov and Hasminskii (1982)) which is estimator (2.5) with

$$\lambda_k = \begin{cases} 1, & k \leq N_n/2, \\ \frac{N_n - k}{N_n/2}, & 1 + N_n/2 \leq k \ \leq N_n, \\ 0, & k > N_n \ . \end{cases}$$

One can choose the sequence N_n optimally as $N_n = (2\beta)^{-1} \log n$. It is well known (see Ibragimov and Hasminskii (1982)) that such estimator allows to obtain the optimal rates, with properly chosen N_n, for all nonparametric classes considered above. However, this estimator is not asymptotically minimax as one can see by comparing (3.5) to the maximal risk of the estimator $\hat{f}_n^{VP}$:

$$R_n(\hat{f}_n^{VP}) = \frac{4}{3} \sigma^2 \beta^{-1} n^{-1} \log n \, (1 + o(1)) .$$

Corollary 6 *Let the ellipsoid Θ be defined by (3.4). Then the estimator $\hat{f}_n^P$ defined in Remark 6 is locally asymptotically minimax and adaptive with respect to σ^2 and vicinity.*

Surely, we take for example $N_n = \beta^{-1} \log n$ and while constructing estimator $\hat{f}_n^P$ we need not to know σ^2 and vicinity. Now the statement of this Corollary follows from previous Corollary and the expression for the maximal risk $R_n(\hat{f}_n^P)$.

The kernel corresponding to the estimator $\hat{f}_n^P$ has the following form:

$$K_n(x, i/n) = \frac{\sin((2N_n + 1)\pi(x - i/n))}{n \sin(\pi(x - i/n))} .$$

4 Proofs

Denote

$$\phi_k = (\phi_k(1/n), \ldots, \phi_k(n/n))^T .$$

Proposition 1

$$\phi_{k+mn} = \begin{cases} \phi_k, & m = 2l\,, \qquad l = 1, 2, \ldots \\ (-1)^{n-k+1}\phi_{n-k}, & m = 2l - 1\,, \quad l = 1, 2, \ldots\,; \end{cases}$$
$$\phi_k^T \phi_l = n\, \delta_{kl}\,, \quad 1 \le k \le n\,, \qquad 1 \le l \le n\,,$$

where $\phi_0 = 0$ and δ_{kl} denotes the Kroneker symbol.

We skip the elementary proof of this Proposition.

It turns out that the asymptotic behaviour of the minimax risk is closely related to the solution of some deterministic minimax problem (see Remark 7 below). We describe this problem. Let $x, y \in \mathbf{R}^\infty$. Denote

$$R_n(x, \theta) = R_n(x, \theta, \sigma^2) = \sum_{k=1}^{\infty} \left(\sigma^2 x_k^2/n + (1 - x_k)^2 \theta_k^2\right) \tag{4.1}$$

and we would like to find $\inf_x \sup_{\theta\in\Theta} R_n(x, \theta)$. Now we provide the key result about this minimax problem.

Lemma 1 *Let $c_n = c_n(\sigma^2, \Theta)$, $d_n = d_n(\sigma^2, \Theta)$ and $R_n(x, \theta)$ be defined by* (2.2), (2.4), (4.1) *respectively. Then the following properties hold:*

(i)

$$\inf_x \sup_{\theta\in\Theta} R_n(x, \theta) = \sup_{\theta\in\Theta} \inf_x R_n(x, \theta)$$

with the saddle point (x^o, θ^o):

$$x_k^o = (1 - c_n a_k)_+\,, \quad (\theta_k^o)^2 = \frac{\sigma^2(1 - c_n a_k)_+}{c_n a_k n}\,; \tag{4.2}$$

(ii)

$$\inf_x \sup_{\theta\in\Theta} R_n(x,\theta) = d_n = \sup_{\theta\in\Theta} n^{-1} \sum_{k=1}^{\infty} \frac{\sigma^2\theta_k^2}{\theta_k^2+\sigma^2 n^{-1}}\,.$$

Proof. Using the equation (2.2) in the form $n^{-1}\sum_{k=1}^{\infty}\sigma^2 c_n a_k(1-c_n a_k)_+ = c_n^2 Q$, we obtain

$$\begin{aligned}
\inf_x \sup_{\theta\in\Theta} R_n(x,\theta) &\le \sup_{\theta\in\Theta} R_n(x^o,\theta) \le Q \sup_{k\ge 1}(1-x_k^o)^2/a_k^2 + \sum_{k=1}^{\infty} n^{-1}\sigma^2(x_k^o)^2 \\
&\le Qc_n^2 + n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+^2 \\
&= n^{-1}\sum_{k=1}^{\infty}\sigma^2((c_n a_k(1-c_n a_k)_+ + (1-c_n a_k)_+^2)) \\
&= n^{-1}\sum_{k=1}^{\infty}\sigma^2(1-c_n a_k)_+ = d_n\,. \qquad (4.3)
\end{aligned}$$

Now note that equation (2.2) can be also rewritten as

$$\sum_{k=1}^{\infty} a_k^2(\theta_k^o)^2 = Q\,, \qquad (4.4)$$

i.e. $\theta^o \in \Theta$. Taking into account (4.3) and (4.4), we easily obtain

$$\begin{aligned}
d_n &\ge \sup_{\theta\in\Theta} R_n(x^o,\theta) \ge \inf_x \sup_{\theta\in\Theta} R_n(x,\theta) \\
&\ge \sup_{\theta\in\Theta}\inf_x R_n(x,\theta) = \sup_{\theta\in\Theta} n^{-1}\sum_{k=1}^{\infty}\frac{\sigma^2\theta_k^2}{\theta_k^2+\sigma^2 n^{-1}} \\
&\ge \inf_x R_n(x,\theta^o) = n^{-1}\sum_{k=1}^{\infty}\frac{\sigma^2(\theta_k^o)^2}{(\theta_k^o)^2+\sigma^2 n^{-1}} = d_n\,.
\end{aligned}$$

The both properties of Lemma 1 follow from the last relations. Lemma 1 is proved. □

Proof of Theorem 1. We assume that $\hat f$ is an estimator with the realizations in $\mathbf{L}_2$ because otherwise the statement of Theorem 1 becomes trivial. Then we have by Parseval identity:

$$r_n = \inf_{\hat f}\sup_{\Theta}\mathbf{E}_f\|\hat f - f\|^2 = \inf_{\hat\theta}\sup_{\Theta}\mathbf{E}_f\sum_{k=1}^{\infty}(\hat\theta_k-\theta_k)^2\,. \qquad (4.5)$$

First we consider a somewhat simpler case: condition $\mathcal{F}_2$ is fulfilled. Let m_k, $k = 1,2,\ldots$, be a set of positive numbers such that $\sum_{k=1}^{\infty} a_k^2 m_k^2 \le Q$, i.e. $m=(m_1,m_2,\ldots)\in\Theta$. Introduce

$$\nu_k(x) = (1/m_k)\nu_0(x/m_k), \quad k=1,2,\ldots\,,$$

where $\nu_0(x)$ is a probability density on the interval $[-1,1]$ with a finite Fisher information

$$I_0 = \int_{-1}^{1}(\nu_0'(x))^2\nu_0^{-1}(x)dx$$

such that $\nu(-1) = \nu(1) = 0$ and $\nu_0(x)$ is continuously differentiable for $|x| < 1$. The functions $\nu_k(x)$'s are probability densities with supports $[-m_k, m_k]$ respectively. It is easy to calculate the Fisher information of the distribution defined by the density $\nu_k(x)$:

$$I(\nu_k) = I_0 m_k^{-2} .$$

It should be indicated that the minimum of $\int_{-1}^{1}(q'(t))^2 q^{-1}(t)dt$ over all differentiable densities $q(t)$ with support $[-1,1]$ is attained by function $q(t) = \cos^2(\pi t/2)$ (see, for example, Borovkov (1984)). Therefore, one can always choose any $I_0 \geq \pi^2$.

We select a prior measure $d\mu$ on $\mathbf{R}^\infty$ such that θ_k, $k = 1, 2, \ldots$, are distributed independently with densities $\nu_k(x)$, $k = 1, 2, \ldots$, respectively.

Since assumption $m \in \Theta$ provides that $supp\,\mu \subseteq \Theta$, we estimate the minimax risk, using (4.5), from below as follows:

$$r_n \geq \inf_{\hat\theta} \int_\Theta \sum_{k=1}^{\infty} \mathbf{E}_f(\hat\theta_k - \theta_k)^2 d\mu(\theta) = \inf_{\hat\theta} \sum_{k=1}^{\infty} \mathbf{E}(\hat\theta_k - \theta_k)^2 . \tag{4.6}$$

We will write $\mathbf{E}$ for the expectation with respect to the joint distribution of $Y_1, \ldots, Y_n$ and $\theta_1, \theta_2, \ldots$. To estimate $\mathbf{E}(\hat\theta_k - \theta_k)^2$, we apply the van Trees inequality (for details see Borovkov (1984), Gill and Levit (1992), van Trees (1968)):

$$\mathbf{E}(\hat\theta_k - \theta_k)^2 \geq \frac{1}{\mathbf{E}I(\theta_k) + I(\nu_k)} , \tag{4.7}$$

where $I(\theta_k)$ is the Fisher information about θ_k contained in observations $Y_1, Y_2, \ldots, Y_n$. It is easy to calculate

$$\mathbf{E}I(\theta_k) = \int \mathbf{E}_f \left[\sum_{i=1}^{n} \frac{\partial \log p_\epsilon(Y_i - f(i/n))}{\partial \theta_k}\right]^2 d\nu_k(\theta_k) = \sigma^{-2} \sum_{i=1}^{n} \phi_k^2(i/n) = \sigma^{-2} n ,$$

which follows from Proposition 1. Recalling that $I(\nu_k) = I_0 m_k^{-2}$, we obtain

$$\mathbf{E}(\hat\theta_k - \theta_k)^2 \geq \frac{1}{I_0 m_k^{-2} + \sigma^{-2} n} = \frac{\sigma^2 n^{-1} m_k^2 I_0^{-1}}{m_k^2 I_0^{-1} + \sigma^2 n^{-1}} .$$

We make use of the last inequality and (4.6):

$$r_n \geq \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{m_k^2 I_0^{-1}}{m_k^2 I_0^{-1} + \sigma^2 n^{-1}} . \tag{4.8}$$

The inequality (4.8) holds for any $m \in \Theta$. At this point we make use of the vector θ^o defined by (4.2). Relation (4.4) provides that $\theta^o \in \Theta$. Substituting θ^o in (4.8) results in

$$r_n \geq \frac{\sigma^2}{n} \sum_{k=1}^{\infty} \frac{(\theta_k^o)^2 I_0^{-1}}{(\theta_k^o)^2 I_0^{-1} + \sigma^2 n^{-1}} = \frac{N\sigma^2}{n} - \frac{\sigma^2}{n} \sum_{k \in \mathcal{I}} \frac{c_n a_k}{(1 - I_0^{-1}) c_n a_k + I_0^{-1}} .$$

Further, combining this with (2.3), (2.4) and condition $\Theta \in \mathcal{F}_2$, we finally get

$$r_n \geq \frac{N\sigma^2}{n} - \frac{c_n I_0 \sigma^2}{n} \sum_{k \in \mathcal{I}} a_k = \frac{N\sigma^2}{n}(1 + o(1)) = d_n(\sigma^2, \Theta)(1 + o(1)),$$

which proves the first part of Theorem 1.

Suppose now that condition $\mathcal{F}_1$ is fulfilled. For arbitrary $0 < \delta < 1$ we can find $R_\delta > 0$ and an absolutely continuous probability density $\nu_\delta(x)$ such that $\nu_\delta(x)$ is positive inside an interval $(-R_\delta, R_\delta)$, equals to zero outside this interval, has finite Fisher information $I(\nu_\delta)$ and satisfies the following properties:

$$\mathbf{E}X^2 = 1 - \delta/2 \quad \text{and} \quad I(\nu_\delta) = \mathbf{E}[(\log \nu_k(X_k))']^2 \leq 1 + \delta\,,$$

where X is a random variable with probability density $\nu_\delta(x)$. Note that under the imposed conditions on density $\nu_\delta(x)$ the relation between $\mathbf{E}X^2$ and $I(\nu_\delta)$ is not arbitrary since the inequality $\mathbf{E}X^2 \geq 1/I(\nu_\delta)$ should hold. Introduce for arbitrary $m_k > 0$, $k = 1, 2, \ldots$,

$$\nu_k(x) = (1/m_k)\nu_0(x/m_k), \quad k = 1, 2, \ldots\,.$$

These are the probability densities with supports $(-R_\delta m_k, m_k R_\delta)$ respectively and if $X_k = m_k X$ then X_k is a random variable with density $\nu_k(x)$. We have

$$\mathbf{E}X_k^2 = m_k^2(1 - \delta/2), \quad I(\nu_k) = I(\nu_\delta)/m_k^2 \leq (1 + \delta)/m_k^2\,. \tag{4.9}$$

Now we select a prior measure $d\mu(\theta)$ on $\mathbf{R}^\infty$ such that θ_k, $k = 1, 2, \ldots$, are distributed independently with the densities $\nu_k(x)$, $k = 1, 2, \ldots$, respectively. In view of (4.5), we evaluate the minimax risk

$$\begin{aligned} r_n &\geq \inf_{\hat{\theta}} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta(\hat{\theta}_k - \theta_k)^2 d\mu(\theta) = \inf_{\hat{\theta} \in supp\, \mu} \int_\Theta \sum_{k=1}^\infty \mathbf{E}_\theta(\hat{\theta}_k - \theta_k)^2 d\mu(\theta) \\ &\geq \inf_{\hat{\theta}} \sum_{k=1}^\infty \mathbf{E}(\hat{\theta}_k - \theta_k)^2 - \sup_{\hat{\theta} \in supp\, \mu} \int_{\Theta^C} \sum_{k=1}^\infty \mathbf{E}_\theta(\hat{\theta}_k - \theta_k)^2 d\mu(\theta) \\ &\geq \inf_{\hat{\theta}} \sum_{k=1}^\infty \mathbf{E}(\hat{\theta}_k - \theta_k)^2 - 4R_\delta^2 \mu(\Theta^C) \sum_{k=1}^\infty m_k^2\,. \end{aligned} \tag{4.10}$$

Due to the assumptions on probability density $\nu_k(x)$, we can apply the van Trees inequality (4.7) to the Bayes risk $\mathbf{E}(\hat{\theta}_k - \theta_k)^2$ (see Borovkov (1984), Gill and Levit (1992), van Trees (1968)). Thus, by (4.7) and (4.9), we obtain

$$\inf_{\hat{\theta}} \sum_{k=1}^\infty \mathbf{E}(\hat{\theta}_k - \theta_k)^2 \geq \frac{\sigma^2}{n(1+\delta)} \sum_{k=1}^\infty \frac{m_k^2}{m_k^2(1+\delta)^{-1} + \sigma^2 n^{-1}}\,. \tag{4.11}$$

Assuming that $m = (m_1, m_2, \ldots) \in \Theta$ (i.e. $\sum_{k=1}^\infty a_k^2 m_k^2 \leq Q$) and recalling (4.9), we have

$$|a_k^2(\theta_k^2 - \mathbf{E}\theta_k^2)| \leq a_k^2 m_k^2 |R_\delta - 1 + \delta/2|\,,$$

$$Q-\sum_{k=1}^{\infty}a_k^2\mathbf{E}\theta_k^2 = Q-(1-\delta/2)\sum_{k=1}^{\infty}a_k^2 m_k^2 \geq Q\delta/2\,.$$

Using these relations and Hoeffding inequality (see Pollard (1984)), we evaluate $\mu(\Theta^C)$:

$$\mu(\Theta^C)=\mu\left\{\sum_{k=1}^{\infty}a_k^2(\theta_k^2-\mathbf{E}\theta_k^2)>Q-\sum_{k=1}^{\infty}a_k^2\mathbf{E}\theta_k^2\right\}\leq \exp\left\{\frac{-\gamma}{\sum_{k=1}^{\infty}a_k^4 m_k^4}\right\}, \tag{4.12}$$

where $\gamma=(Q\delta)^2/(8(R_\delta^2-1+\delta/2)^2)$.

We choose $m_k^2=(\theta_k^o)^2$, $k=1,2,\ldots$, where θ_k^o's are defined by (4.2). Note that $\theta^o\in\Theta$ (see (4.4)) and θ_k^o's are those for which, according to Lemma 1,

$$\sup_{\theta\in\Theta}\frac{\sigma^2}{n}\sum_{k=1}^{\infty}\frac{\theta_k^2}{\theta_k^2+\sigma^2 n^{-1}}$$

is attained. Recalling the definition of $\psi_n(\gamma)$, we see that right-hand side of the inequality (4.12) becomes $\psi_n(\gamma)$. Therefore, combining (4.10), (4.11) and (4.12) gives

$$r_n\geq\frac{\sigma^2}{n(1+\delta)}\sum_{k=1}^{\infty}\frac{(\theta_k^o)^2}{(\theta_k^o)^2(1+\delta)^{-1}+\sigma^2n^{-1}}-4R_\delta^2\psi_n(\gamma)\sum_{k=1}^{\infty}(\theta_k^o)^2\,. \tag{4.13}$$

According to (2.4), (4.2) and condition $\Theta\in\mathcal{F}_1$,

$$\psi_n(\gamma)\sum_{k=1}^{\infty}(\theta_k^o)^2=\psi_n(\gamma)n^{-1}\sum_{k=1}^{\infty}\frac{\sigma^2(1-c_na_k)_+}{c_na_k}=o(d_n)\,.$$

Combining (4.13) with the last relation and Lemma 1, we get

$$r_n\geq\frac{\sigma^2}{n(1+\delta)}\sum_{k=1}^{\infty}\frac{(\theta_k^o)^2}{(\theta_k^o)^2+\sigma^2n^{-1}}+R_\delta^2 o(d_n)=(1+\delta)^{-1}d_n+R_\delta^2 o(d_n)\,.$$

Since this inequality holds for any $\delta\in(0,1)$, we conclude that Theorem 1 is proved. □

Corollary 7 *For any ellipsoid $\Theta(Q)$ the following inequality is true:*

$$r_n\geq d_n(\sigma^2,\Theta(Q/\pi^2))\,.$$

Proof. Because we do not use any condition of Theorem 1 up to (4.8), we invoke now the inequality (4.8) with $I_0=\pi^2$. Recall that (4.8) holds for any $m\in\Theta(Q)$ and therefore

$$\begin{aligned} r_n &\geq \sup_{m\in\Theta(Q)}\frac{\sigma^2}{n}\sum_{k=1}^{\infty}\frac{m_k^2/\pi^2}{m_k^2/\pi^2+\sigma^2n^{-1}} \\ &= \sup_{m\in\Theta(Q/\pi^2)}\frac{\sigma^2}{n}\sum_{k=1}^{\infty}\frac{m_k^2}{m_k^2+\sigma^2n^{-1}}\,. \end{aligned}$$

Finally, by property (ii) of Lemma 1, we obtain the statement of Corollary. $\square$

Proof of Theorem 2. Since, by (2.5) and second property of Proposition 1,

$$\mathbf{E}_f(\lambda_k\hat{\theta}_k-\theta_k)^2=\frac{\sigma^2\lambda_k^2}{n}+(1-\lambda_k)^2\theta_k^2+2(\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\,,$$

$$\tilde{\theta}_k=\mathbf{E}_f\hat{\theta}_k-\theta_k=n^{-1}\sum_{m=1}^{n}\phi_k(m/n)f(m/n)-\theta_k\,,$$

we evaluate

$$\begin{aligned}\sup_{\Theta}\mathbf{E}_f\|\hat{f}_n^M-f\|^2&=\sup_{\Theta}\left\{\sum_{k=1}^{n}\mathbf{E}_f(\lambda_k\hat{\theta}_k-\theta_k)^2+\sum_{k=n+1}^{\infty}\theta_k^2\right\}\\ &\le\ \sup_{\Theta}\left\{\sum_{k=1}^{\infty}\left(n^{-1}\sigma^2\lambda_k^2+(1-\lambda_k)^2\theta_k^2\right)\right\}\\ &+\ \ 2\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\right)\right\}+\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\}.\end{aligned}$$

According to Lemma 1, the first term of the last inequality is exactly $d_n(\sigma^2,\Theta)$. Therefore, it is sufficient to show

$$\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\right)\right\}=o(d_n)\,,\tag{4.14}$$

$$\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\}=o(d_n)\,.$$

The last relation follows immediately from the definition of set $\mathcal{F}_3$ (or set $\mathcal{F}_4$):

$$\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2\right\}\le\sup_{\Theta}\left\{\sum_{k=n+1}^{\infty}\theta_k^2a_k^2\right\}\max_{k>n}a_k^{-2}=o(d_n)\,.$$

Suppose we have the following relation:

$$\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2=o(d_n)\,.\tag{4.15}$$

Then, taking into account that $\lambda_k=(1-c_na_k)_+$ and $Qc_n^2\le d_n$ (which follows from (4.3)), we prove (4.14) by Cauchy-Shwarz inequality:

$$\begin{aligned}&\sup_{\Theta}\left\{\sum_{k=1}^{n}\left((\lambda_k-1)\lambda_k\theta_k\tilde{\theta}_k+\lambda_k^2\tilde{\theta}_k^2\right)\right\}\\ &\le\ \sup_{\Theta}\left\{\left(\sum_{k=1}^{\infty}a_k^2\theta_k^2\right)^{1/2}c_n\left(\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2\right)^{1/2}\right\}+\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2\\ &\le\ d_n^{1/2}\sup_{\Theta}\left(\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2\right)^{1/2}+\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2=o(d_n)\,.\end{aligned}$$

It remains to show (4.15). For any k such that $1 \le k \le n-1$, by Proposition 1 and Cauchy-Shwarz inequality, we have

$$
\begin{aligned}
\tilde{\theta}_k &= n^{-1}\sum_{m=1}^{n}\phi_k(m/n)f(m/n)-\theta_k = n^{-1}\sum_{l=1}^{\infty}\theta_l\sum_{m=1}^{n}\phi_k(m/n)\phi_l(m/n)-\theta_k \\
&= \sum_{l=n+1}^{\infty}\theta_l\, n^{-1}\sum_{m=1}^{n}\phi_k(m/n)\phi_l(m/n) = \sum_{l=1}^{\infty}(\theta_{k+2ln}+(-1)^{k+1}\theta_{2ln-k})
\end{aligned}
$$

and

$$
\tilde{\theta}_n = \sum_{l=1}^{\infty}\theta_{(2l+1)n}\,.
$$

Therefore,

$$
\begin{aligned}
\tilde{\theta}_k &\le \left(\sup_{\Theta}\sum_{l=1}^{\infty}\theta^2_{k+2ln}a^2_{k+2ln}+\theta^2_{2ln-k}a^2_{2ln-k}\right)^{1/2}\left(\sum_{l=1}^{\infty}a^{-2}_{k+2ln}+a^{-2}_{2ln-k}\right)^{1/2} \\
&\le Q^{1/2}\left(\sum_{l=1}^{\infty}a^{-2}_{k+2ln}+a^{-2}_{2ln-k}\right)^{1/2}
\end{aligned}
$$

and hence, by definition of set $\mathcal{F}_3$ (or $\mathcal{F}_4$), we arrive to (4.15):

$$
\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2 \le Q\left(2\max_{1\le k\le n}\sum_{l=1}^{\infty}a^{-2}_{k+ln}\right)\sum_{k=1}^{n}\lambda_k = o(n^{-1})\sum_{k=1}^{n}\lambda_k = o(d_n)
$$

or

$$
\sup_{\Theta}\sum_{k=1}^{n}\lambda_k^2\tilde{\theta}_k^2 \le \sup_{\Theta}\sum_{k=1}^{n}\tilde{\theta}_k^2 \le 2Q\sum_{k=n+1}^{\infty}a_k^{-2} = o(d_n)\,.
$$

This completes the proof of Theorem 2. □

Remark 7. Denote $Y=(Y_1,\ldots,Y_n)^T$ and $\epsilon=(\epsilon_1,\ldots,\epsilon_n)^T$. We rewrite the model (2.1) as follows:

$$
Y = \sum_{k=1}^{\infty}\theta_k\phi_k+\epsilon\,.
$$

Now we multiply this equality by ϕ_l^T/n, $l=1,2,\ldots,n$. Then, using Proposition 1, we get:

$$
Z_l = \theta_l+\tilde{\theta}_l+n^{-1/2}\xi_l\,,\quad l=1,2,\ldots,n\,,
$$

where

$$
Z_l=\phi_l^T Y/n \quad \text{and} \quad \xi_l=n^{-1/2}\phi_l^T\epsilon\,,
$$

i.e. ξ_l's are Gaussian random variables with zero mean and covariances $\mathbf{E}[\xi_l\xi_k]=\sigma^2\delta_{lk}$. The regularity conditions ($\mathcal{F}_1-\mathcal{F}_4$) imply that, as the proofs of Theorems 1 and 2 show, the original model and the model

$$
Z_l' = \theta_l+n^{-1/2}\xi_l\,,\quad l=1,2,\ldots,n\,,
$$

are asymptotically equivalent in the sense that the best linear estimators and the minimax risks for both models coinside asymptotically. Note that (4.1) is nothing else but the risk of the linear estimator $\hat{\theta}_k'=\lambda_k Z_k'$, $k=1,2,\ldots,n$.

References

[1] Belitser, E.N. and Korostelev, A.P. (1992). Pseudovalues and minimax filtering algorithms for the nonparametric median. *Advances in Soviet Math.*, **12**, 115-124.

[2] Borovkov, A.A.(1984). *Mathematical Statistics. Parameter Estimation.* (in Russian) Nauka, Moscow.

[3] Brown, L.D. and Low, M.G. (1992). Asymptotic equivalence of nonparametric regression and white noise. Technical report, Cornell University (to appear in *Ann. Statist.*).

[4] Donoho, D.L. and Johnstone, I.M.(1992). Minimax estimation via wavelet shrinkage. Technical report 402, Stanford University.

[5] Efroimovich, S.Y.(1994). On nonparametric regression for iid observations in general setting. Mss.

[6] Efroimovich, S.Y. and Pinsker, M.S.(1982). Estimation of square-integrable density of a random variable. *Problems Inform. Transmission*, **18**, No. 3, 175-189.

[7] Gill, R.D. and Levit, B.Y.(1995). Applications of the van Trees inequality: A Bayesian Cramer-Rao bound. *Bernoulli*, bf 1, 59-79.

[8] Golubev, G.K. (1987). Adaptive asymptotically minimax estimators of smooth signals. *Problems Inform. Transmission*, **23**, No. 1, 47-55.

[9] Golubev, G.K. and Nussbaum, M.(1990). A risk bound in Sobolev class regression. *Ann. Statist.*, **18**, 758-778.

[10] Ibragimov, I.A. and Hasminskii, R.Z.(1980). Asymptotic efficiency bounds for nonparametric estimation of a regression function in $\mathbf{L}_p$ (in Russian). *Zap. Nauchn. Sem. LOMI*, **97**, 88-101.

[11] Ibragimov, I.A. and Hasminskii, R.Z.(1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New-York.

[12] Ibragimov, I.A. and Hasminskii, R.Z.(1982). Bounds for the risk of nonparametric regression estimates. *Theory Probab. Appl.*, **27**, 84-99.

[13] Korostelev, A.P.(1993). Exact asymptotic minimax estimate of a nonparametric regression in the uniform norm. *Theory Probab. Appl.*, **38**, 875-882. to appear.

[14] Nussbaum, M.(1985). Spline smoothing in regression models and asymptotic efficiency in $\mathbf{L}_2$. *Ann. Statist.*, **13**, 984-992.

[15] Nussbaum, M.(1995). Asymptotic equivalence of density estimation and Gaussian white noise. Preprint 174, Weierstrass Institute, Berlin.

[16] Pinsker, M.S.(1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission.*, **16**, No. 2, 120-133.

[17] Pollard, D.(1984). *Convergence of Stochastic Processes.* Springer-Verlag, New-York.

[18] Speckman, P.(1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.*, **13**, 970-983.

[19] Stone, C.J.(1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, **8**, 1348-1360.

[20] Stone, C.J.(1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

[21] van Trees, H.L.(1968). *Detection, Estimation and Modulation Theory, Part 1.* Wiley, New York.

E. Belitser and B. Levit
Mathematical Institute
University of Utrecht
Budapestlaan 6
3584 CD Utrecht
The Netherlands