Adaptive nonparametric estimation in heteroscedastic regression models.

Part 2: Asymptotic efficiency.

By Leonid Galtchouk and Sergey Pergamenshchikov *

Louis Pasteur University of Strasbourg and University of Rouen

Abstract

In the paper we study asymptotic properties of the adaptive procedure proposed in the paper Galtchouk, Pergamenshchikov, 2007, for nonparametric estimation of unknown regression. We prove that this procedure is asymptotically efficient for some quadratic risk, i.e. we show that the asymptotic quadratic risk for this procedure coincides with the Pinsker constant which gives a sharp lower bound for quadratic risk over all possible estimates. [1] [2]

*The second author is partially supported by the RFFI-Grant 04-01-00855.

[1] *AMS 2000 Subject Classification* : primary 62G08; secondary 62G05, 62G20

[2] *Key words*: adaptive estimation, asymptotic bounds, efficient estimation, heteroscedastic regression, nonparametric estimation, non-asymptotic estimation, oracle inequality, Pinsker's constant.

1 Introduction

The paper deals with the estimation problem in the heteroscedastic non-parametic regression model

$$y_j = S(x_j) + \sigma_j(S)\,\xi_j\,, \tag{1.1}$$

where the design points $x_j = j/n$, $S(\cdot)$ is an unknown function to be estimated, $(\xi_j)_{1\le j\le n}$ is a sequence of centered i.i.d. random variables with unit variance and $\mathbf{E}\xi_1^4 = \xi^* < \infty$, $(\sigma_j(S))_{1\le j\le n}$ are unknown scale functionals depending on unknown regression function S and the design points.

Typically, the notion of asymptotic optimality is associated with the optimal convergence rate of the minimax risk (see for example, Ibragimov, Hasminskii,1981; Stone,1982). An important question in optimality results is to study the exact asymptotic behaviour of the minimax risk. Such results have been obtained only in a limited number of investigations. As to the nonparametric estimation problem for heteroscedastic regression models we should mention the papers Efromovich, 2007, Efromovich, Pinsker, 1996, and Galtchouk, Pergamenshchikov, 2005, concerning the exact asymptotic behaviour of the $\mathcal{L}_2$-risk and paper by Brua, 2007, devoted to the efficient pointwise estimation for heteroscedastic regressions.

We remind that an example of heteroscedastic regression models is given by econometrics (see, for example, Goldfeld, Quandt, 1972, p. 83), where for consumer budget problems one uses some parametric version of model (1.1) with the scale coefficients defined as

$$\sigma_j^2(S) = c_0 + c_1 x_j + c_2 S^2(x_j)\,, \tag{1.2}$$

where c_0, c_1 and c_2 are some positive unknown constants.

The purpose of the article is to study asymptotic properties of the adaptive estimation procedure proposed in Galtchouk, Pergamenshchikov, 2007, for which a non-asymptotic oracle inequality was proved for quadratic risks. We will prove that this oracle inequality is asymptotically sharp, i.e. the asymptotic quadratic risk is minimal. It means the adaptive estimation procedure is efficient under some conditions on the scales $(\sigma_j(S))_{1\le j\le n}$ which are satisfied in the case (1.2). Note that in Efromovich, 2007, Efromovich, Pinsker, 1996, an efficient adaptive procedure is constructed for heteroscedastic regression when the scale coefficient is independent of S, i.e. $\sigma_j(S) = \sigma_j$. In Galtchouk, Pergamenshchikov, 2005, for the model (1.1) the asymptotic efficiency was proved under strong conditions on the scales which are not satisfied in the case (1.2). Moreover in the cited papers the efficiency was proved for the gaussian random variables $(\xi_j)_{1\le j\le n}$ that is very restrictive for applications of proposed methods to practical problems.

In the paper we modify the risk by introducing into a additional supremum with respect to a classe of unknown noise distributions like to Galtchouk, Pergamenshchikov, 2006. This modification allow us to eliminate from the risk dependence on the noise distribution. Moreover for this risk a efficient procedure is robust with respect to changing of noise distributions.

It is well known to prove the asymptotic efficiency one has to show that the asymptotic quadratic risk coincides with the lower bound which is equal to the Pinsker constant. In the paper two problems are resolved: in the first one an upper bound for the risk is obtained by making use of the non-asymptotic oracle inequality from Galtchouk, Pergamenshchikov, 2007, in the second one we prove that this upper bound coincides with the Pinsker constant. Let us remind that the adaptive procedure proposed in Galtchouk, Pergamenshchikov, 2007, is based on weighted mean-squares estimates, where the weights are corresponding modifications of the Pinsker weights for the homogene case (when $\sigma_1(S) = \ldots = \sigma_n(S) = 1$) relative to a certain smoothness of the function S and this procedure chooses an estimator best for the quadratic risk among these estimates. To obtain the Pinsker constant for the model (1.1) one has to prove a sharp asymptotic lower bound for the quadratic risk in the case when the noise variance depends on the unknown regression function. This lower bound is obtained by making use of an inequality of kind of the van Trees inequality (see, Gill, Levit, 1995). First we prove the inequality for a parametric regression with the noise variance depending on the unknown regression (see Section 6) and further we apply the inequality to the nonparametric regression by standard reducing to a parametric case.

The paper is organized as follows. In Section 2 we construct a adaptive estimation procedure. In Section 3 we formulate principal conditions. The main result is given in Section 4. The upper bound for the quadratic risk is given in Section 5. In Section 6 we find the lower bound for a parametric model. In Section 7 we study the parametric family. In Section 8 we obtain the lower bound for model (1.1). An appendix contains some technical results.

2 Adaptive procedure

In this section we describe the adaptive procedure proposed in [6]. We make use of the standard trigonometric basis $(\phi_j)_{j\ge1}$ in $\mathcal{L}_2[0,1]$, i.e.

$$\phi_1(x)=1\,,\quad \phi_j(x)=\sqrt{2}\,Tr_j(2\pi[j/2]x)\,,\ j\ge 2\,,\tag{2.1}$$

where the function $Tr_j(x)=\cos(x)$ for even j and $Tr_j(x)=\sin(x)$ for odd j; $[x]$ denotes the integer part of x. We remind that if n is odd then the functions $(\phi_j)_{1\le j\le n}$ are orthonormal with respect to the empirical inner product generated by the sieve $(x_j)_{1\le j\le n}$ in (1.1), i.e. for any $1\le i,j\le n$,

$$(\phi_i\,,\,\phi_j)_n=\frac{1}{n}\sum_{l=1}^{n}\phi_i(x_l)\phi_j(x_l)=\mathbf{Kr}_{ij}\,,$$

where $\mathbf{Kr}_{ij}$ is Kronecker's symbol. Thanks to this basis we pass to the discrete Fourier transformation of model (1.1), i.e.

$$\hat{\vartheta}_{j,n}=\vartheta_{j,n}+(1/\sqrt{n})\xi_{j,n}\,,\tag{2.2}$$

where $\hat{\theta}_{j,n}=(Y,\phi_j)_n$, $\theta_{j,n}=(S,\phi_j)_n$ and

$$\xi_{j,n}=\frac{1}{\sqrt{n}}\sum_{l=1}^{n}\sigma_l(S)\xi_l\phi_j(x_l)\,.$$

Here $Y=(y_1,\ldots,y_n)'$ and $S=(S(x_1),\ldots,S(x_n))'$. The prime denotes the transposition.

We estimate the function S by the weighted least squares estimator

$$\hat{S}_\lambda=\sum_{j=1}^{n}\lambda(j)\hat{\vartheta}_{j,n}\phi_j\,,\tag{2.3}$$

where the weight vector $\lambda = (\lambda(1), \ldots, \lambda(n))'$ belongs to some finite set Λ from $[0,1]^n$ with $n \geq 3$. Here we make use of the weight family Λ introduced in [6], i.e.

$$\Lambda = \{\lambda_\alpha \,, \, \alpha \in \mathcal{A}_\varepsilon\} \,, \quad \mathcal{A}_\varepsilon = \{1, \ldots, k_*\} \times \{t_1, \ldots, t_m\} \,, \tag{2.4}$$

where $k_* = [1/\sqrt{\varepsilon}]$, $t_i = i\varepsilon$, $m = [1/\varepsilon^2]$ and $\varepsilon = 1/\ln n$.

For any $\alpha = (\beta, t) \in \mathcal{A}_\varepsilon$ we define the weight vector $\lambda_\alpha = (\lambda_\alpha(1), \ldots, \lambda_\alpha(n))'$ as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + \left(1 - (j/\omega(\alpha))^\beta\right) \mathbf{1}_{\{j_0 < j \leq \omega(\alpha)\}} \,, \tag{2.5}$$

where $j_0 = j_0(\alpha) = [\omega(\alpha)/\ln n]$, $\omega(\alpha) = (A_\beta \, t)^{1/(2\beta+1)} n^{1/(2\beta+1)}$ and

$$A_\beta = (\beta+1)(2\beta+1)/(\pi^{2\beta}\beta) \,.$$

To find the optimal weights we choose the cost function equals to the penalized mean integrated squared error in which unknown parameters are replaced by some estimators. The cost function is as follows

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\vartheta}^2_{j,n} - 2 \sum_{j=1}^n \lambda(j) \, \tilde{\vartheta}_{j,n} + \rho \hat{P}_n(\lambda) \,, \tag{2.6}$$

where

$$\tilde{\vartheta}_{j,n} = \hat{\vartheta}^2_{j,n} - \frac{1}{n} \hat{\varsigma}_n \quad \text{with} \quad \hat{\varsigma}_n = \sum_{j=l_n+1}^n \hat{\vartheta}^2_{j,n} \tag{2.7}$$

and $l_n = [n^{1/3} + 1]$. The penalty term we define as

$$\hat{P}_n(\lambda) = \frac{|\lambda|^2 \hat{\varsigma}_n}{n} \,, \quad |\lambda|^2 = \sum_{j=1}^n \lambda^2(j) \quad \text{and} \quad \rho = \frac{1}{3 + \ln^\gamma n} \,.$$

for some $\gamma > 0$. Finally, we set

$$\hat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda) \quad \text{and} \quad \hat{S}_* = \hat{S}_{\hat{\lambda}} \,. \tag{2.8}$$

The goal of this paper is to study asymptotic ($n \to \infty$) properties of this estimation procedure.

3 Conditions

First we impose some conditions on unknown function S in model (1.1).

Let $\mathcal{C}^k_{per,1}(\mathbb{R})$ be the set of 1-periodic k times differentiable $\mathbb{R} \to \mathbb{R}$ functions. We assume that S belongs to the following set

$$W^k_r = \{f \in \mathcal{C}^k_{per,1}(\mathbb{R}) \,:\, \sum_{j=0}^{k} \|f^{(j)}\|^2 \le r\}\,, \tag{3.1}$$

where $\|\cdot\|$ denotes the norm in $\mathcal{L}_2[0,1]$, i.e.

$$\|f\|^2 = \int_0^1 f^2(t)\mathrm{d}t\,. \tag{3.2}$$

Moreover, we suppose that $r > 0$ and $k \ge 1$ are unknown parameters.

Note that, we can represent the set W^k_r as an ellipse in $\mathcal{L}_2[0,1]$, i.e.

$$W^k_r = \{f \in \mathcal{L}_2[0,1] \,:\, \sum_{j=1}^{\infty} a_j \vartheta_j^2 \le r\}\,, \tag{3.3}$$

where

$$\vartheta_j = (f, \phi_j) = \int_0^1 f(t)\phi_j(t)\mathrm{d}t \tag{3.4}$$

and

$$a_j = \sum_{l=0}^{k} \|\phi_j^{(l)}\|^2 = \sum_{i=0}^{k} (2\pi[j/2])^{2i}\,. \tag{3.5}$$

Here $(\phi_j)_{j\ge1}$ is the trigonometric basis defined in (2.1).

Now we decribe the conditions on the scale coefficients $(\sigma_j(S))_{j\ge1}$.

H$_1$) *$\sigma_j(S) = g(x_j, S)$ for some unknown function $g : [0,1] \times \mathbf{L}_1[0,1] \to \mathbb{R}_+$, which is square integrable with respect to x such that*

$$\lim_{n\to\infty} \sup_{S\in W_r^k} \left| n^{-1} \sum_{j=1}^{n} g^2(x_j, S) \; - \; \varsigma(S) \right| = 0\,, \tag{3.6}$$

where $\varsigma(S) := \int_0^1 g^2(x,S)\mathrm{d}x$. Moreover,

$$g_* = \inf_{0\le x\le 1} \inf_{S\in W_r^k} g^2(x,S) \; > 0 \tag{3.7}$$

and

$$\sup_{S\in W_r^k} \varsigma(S) < \infty\,. \tag{3.8}$$

H$_2$) *For any $x \in [0,1]$ the operator $g^2(x,\cdot) \; : \; \mathbf{C}[0,1] \to \mathbb{R}$ is differentiable in the Fréchet sense for any fixed function f_0 from $\mathbf{C}[0,1]$, i.e. for any f from some vicinity of f_0 in $\mathbf{C}[0,1]$*

$$g^2(x,f) = g^2(x,f_0) + \mathbf{L}_{x,f_0}(f - f_0) + \Upsilon(x, f_0, f)\,,$$

where the Fréchet derivative $\mathbf{L}_{x,f_0} \; : \; \mathbf{C}[0,1] \to \mathbb{R}$ is a bounded linear operator and the residual term $\Upsilon(x, f_0, f)$ for each $x \in [0,1]$ satisfies the following property

$$\lim_{|f-f_0|_*\to 0} \frac{|\Upsilon(x, f_0, f)|}{|f - f_0|_*} = 0\,,$$

where $|f|_* = \sup_{0\le t\le 1} |f(t)|$.

H$_3$) *There exists some positive constant C^* such that for any function S from $\mathbf{C}[0,1]$ the operator $\mathbf{L}_{x,S}$ defined in condition **H**$_2$) satisfies the following inequality for any function f from $\mathbf{C}[0,1]$*

$$|\mathbf{L}_{x,S}(f)| \le C^* \left(|S(x)f(x)| + |f|_1 + \|S\| \, \|f\| \right)\,, \tag{3.9}$$

where $|f|_1 = \int_0^1 |f(t)|\mathrm{d}t$.

$\mathbf{H}_4$) *The function* $g_0^2(\cdot) = g^2(\cdot, S_0)$ *corresponding to* $S_0 \equiv 0$ *is continuous on the interval* $[0,1]$. *Moreover,*

$$\lim_{\delta\to 0}\,\sup_{0\le x\le 1}\,\sup_{|S|_*\le\delta}\,|g^2(x,S)-g^2(x,S_0)| \,=\, 0\,.$$

Now we give some examples of functions satisfying conditions $\mathbf{H}_1$)-$\mathbf{H}_4$).

We fix some $c_0 > 0$. Let $G\,:\,[0,1]\times\mathbb{R}\to[c_0\,,+\infty)$ be a function such that

$$\lim_{\delta\to 0}\,\max_{|u-v|\le\delta}\,\sup_{y\in\mathbb{R}}|G(u,y)-G(v,y)| = 0\,. \tag{3.10}$$

and

$$G'_* = \sup_{0\le x\le 1}\,\sup_{y\in\mathbb{R}}|G_y(x,y)|/|y| < \infty\,. \tag{3.11}$$

Moreover, let $V\,:\,\mathbb{R}\to\mathbb{R}_+$ be a continuously differentiable function such that

$$v'_* = \sup_{y\in\mathbb{R}}\,|\dot V(y)|/(1+|y|) < \infty\,.$$

We set

$$g^2(x,S) = G(x,S(x)) + \int_0^1\,V(S(t))\mathrm{d}t\,. \tag{3.12}$$

In this case

$$\varsigma(S) = \int_0^1 G(t,S(t))\mathrm{d}t + \int_0^1\,V(S(t))\mathrm{d}t$$

and for any $S\in W_r^k$

$$\left|n^{-1}\sum_{j=1}^n g^2(x_j,S)-\varsigma(S)\right| \le \sum_{j=1}^n\int_{x_{j-1}}^{x_j}\,\big|G(x_j,S(x_j))-G(t,S(t))\big|\,\mathrm{d}t$$
$$\le \Delta_n + \frac{G'_*}{n}\int_0^1\,|S(t)|\,|\dot S(t)|\mathrm{d}t \le \Delta_n + \frac{G'_*}{n}\,r\,,$$

where

$$\Delta_n = \max_{|u-v|\le 1/n}\,\sup_{y\in\mathbb{R}}|G(u,y)-G(v,y)|\,.$$

Therefore by condition (3.10) we obtain $\mathbf{H}_1$).

Moreover, the Fréchet derivative in this case is given by

$$\mathbf{L}_{x,S}(f) = G_y(x, S(x))f(x) + \int_0^1 \dot{V}(S(t))f(t)\mathrm{d}t\,.$$

It is easy to see that this operator satisfies the inequality (3.9) with

$$C^* = G'_* + v'_*\,.$$

For example, we can take in (3.12)

$$G(x,y) = c_0 + c_1 x + c_2 y^2 \quad \text{and} \quad V(x) = c_3 x^2 \qquad (3.13)$$

with some coefficients $c_0 > 0$, $c_i \geq 0, i = 1,2,3$. Therefore, we obtain the function (1.2) if we put in (3.12)-(3.13) $c_3 = 0$, i.e. $V \equiv 0$.

4 Main results

Denote by $\mathcal{P}_*$ the family of unknown noise density. Remind that the noise random variables $(\xi_j)_{1\le j\le n}$ are centered with unit variance and $\mathbf{E}\xi_1^4 \leq \xi^*$, where $\xi^* \geq 3$. For any estimate $\hat{S}$ we define the following quadratic risk

$$\mathcal{R}_n(\hat{S}, S) = \sup_{p\in\mathcal{P}_*} \mathbf{E}_{S,p}\|\hat{S} - S\|_n^2\,, \qquad (4.1)$$

where $\mathbf{E}_{S,p}$ is the expectation with respect to the distribution $\mathbf{P}_{S,p}$ of the observations $(y_1, \ldots, y_n)$ with the fixed function S and the fixed density p of random variables $(\xi_j)_{1\le j\le n}$ in model (1.1), $\|S\|_n^2 = (S,S)_n$.

In Galtchouk, Pergamenshchikov, 2007, we shown the following non-asymptotic Oracle inequality for procedure (2.8).

Theorem 4.1. *Assume that in model* (1.1) *the function* S *belongs to* W_r^1. *Then, for any odd* $n \geq 3$, *any* $0 < \rho < 1/3$ *and* $r > 0$, *the estimate* $\hat{S}_*$ *satisfies the following oracle inequality*

$$\mathcal{R}_n(\hat{S}_*, S) \leq (1 + \kappa(\rho)) \min_{\lambda \in \Lambda} \mathcal{R}_n(\hat{S}_\lambda, S) + n^{-1} \mathcal{B}_n(\rho)\,, \tag{4.2}$$

where

$$\kappa(\rho) = (6\rho - 2\rho^2)/(1 - 3\rho)$$

and the function $\mathcal{B}_n(\rho)$ *is such that, for any* $\delta > 0$,

$$\lim_{n\to\infty} \mathcal{B}_n(\rho)/n^\delta = 0\,. \tag{4.3}$$

Now we formulate the main asymptotic results. To this end for any function $S \in W_r^k$ we set

$$\gamma_k(S) = \Gamma_k^* \, r^{1/(2k+1)} \, (\varsigma(S))^{2k/(2k+1)}\,, \tag{4.4}$$

where

$$\Gamma_k^* = (2k+1)^{1/(2k+1)} \, (k/(\pi\,(k+1)))^{2k/(2k+1)}\ .$$

It is well known (see, for example, Nussbaum, 1985) that for any function $S \in W_r^k$ the optimal convergence rate is $n^{2k/(2k+1)}$.

Theorem 4.2. *Assume that in model* (1.1) *the sequence* $(\sigma_j(S))$ *fulfils the condition* $\mathbf{H}_1$). *Then the estimator* $\hat{S}_*$ *from* (2.8) *satisfies the inequality*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}_n(\hat{S}_*, S)/\gamma_k(S) \leq 1\,. \tag{4.5}$$

The following result gives the sharp lower bound for risk (4.1) and show that $\gamma_k(S)$ is *the Pinsker* constant.

Theorem 4.3. *Assume that in model* (1.1) *the sequence* $(\sigma_j(S))$ *satisfies the conditions* $\mathbf{H}_2)$– $\mathbf{H}_4)$. *Then, for any estimate* $\hat{S}_n$*, the risk* $\mathcal{R}_n(\hat{S}_n, S)$ *admits the following asymptotic lower bound*

$$\liminf_{n\to\infty} n^{2k/(2k+1)} \inf_{\hat{S}_n} \sup_{S\in W_r^k} \mathcal{R}_n(\hat{S}_n, S)/\gamma_k(S) \geq 1\,. \tag{4.6}$$

Remark 4.1. *Note that in Galtchouk, Pergamenshchikov, 2005 an asymptotically efficient estimate was constructed and results similar to Theorems 4.2 and 4.3 were claimed for the model* (1.1). *In fact the upper bound is true there under some additional condition on the smoothness of the function* S*, i.e. on the parameter* k*. In the cited paper this additional condition is not formulated since erroneous inequality* $(A.6)$. *To avoid the use of this inequality we modify the estimating procedure by introducing the penalty term* $\rho\,\hat{P}_n(\lambda)$ *in the cost function* (2.6). *By this way we remove all additional conditions on the smoothness parameter* k.

5 Upper bound

In this section we prove Theorem 4.2. To this end we will make use of oracle inequality (4.2). We have to find an estimator from the family (2.3)-(2.4) for which we can show the upper bound (4.5). We start with the construction of such an estimator. First we put

$$\tilde{l}_n = \inf\{i \geq 1\,:\, i\varepsilon \geq \overline{r}(S)\} \wedge m \quad \text{and} \quad \overline{r}(S) = r/\varsigma(S)\,. \tag{5.1}$$

Then we choose an index from the set $\mathcal{A}_\varepsilon$ as

$$\tilde{\alpha} = (k, \tilde{t}_n)\,,$$

where k is the parameter of the set W_r^k and $\tilde{t}_n = \tilde{l}_n \varepsilon$. Finally, we set

$$\tilde{S} = \hat{S}_{\tilde{\lambda}} \quad \text{and} \quad \tilde{\lambda} = \lambda_{\tilde{\alpha}} \,. \tag{5.2}$$

Now we show the upper bound (4.5) for this estimator.

Theorem 5.1. *Assume that condition* $\mathbf{H}_1$) *hold. Then*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n(\tilde{S}, S)/\gamma_k(S) \, \le 1 \,. \tag{5.3}$$

Remark 5.1. *Note that the estimator* $\tilde{S}$ *belongs to estimate family* (2.3)-(2.4)*, but we can't use directly this estimator because the parameters* k*,* r *and* $\overline{r}(S)$ *are unknown. We can use this upper bound only through the oracle inequality* (4.2) *proved for procedure* (2.8)*.*

Proof. To prove the theorem we will adapt to the heteroscedastic case the corresponding proof from Nussbaum, 1985.

First, from (2.3) we obtain that, for any $p \in \mathcal{P}_*$,

$$\mathbf{E}_{S,p} \, \|\tilde{S} - S\|_n^2 = \sum_{j=1}^n (1 \, - \, \tilde{\lambda}_j)^2 \vartheta_{j,n}^2 + \frac{1}{n} \sum_{j=1}^n \tilde{\lambda}_j^2 \varsigma_{j,n} \,, \tag{5.4}$$

where

$$\varsigma_{j,n} = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \phi_j^2(x_l) \,.$$

Setting now $\tilde{\omega} = \omega(\tilde{\alpha})$, $\tilde{j}_0 = [\tilde{\omega}/\ln n]$, $\tilde{j}_1 = [\tilde{\omega} \ln n]$ and

$$\varsigma_n = \frac{1}{n} \sum_{l=1}^n \sigma_l^2(S) \,,$$

we rewrite (5.4) as follows

$$\mathbf{E}_{S,p} \, \|\tilde{S} - S\|_n^2 = \sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (1 \, - \, \tilde{\lambda}_j)^2 \vartheta_{j,n}^2 + \varsigma_n \, n^{-1} \sum_{j=1}^n \tilde{\lambda}_j^2 + \Delta_1(n) + \Delta_2(n) \tag{5.5}$$

with

$$\Delta_1(n) \,=\, \sum_{j=\tilde{j}_1}^{n} \vartheta^2_{j,n} \quad \text{and} \quad \Delta_2(n) \,=\, n^{-1} \sum_{j=1}^{n} \tilde{\lambda}^2_j \left(\varsigma_{j,n} - \varsigma_n \right) \,.$$

Note that we have decomposed the first term in the right-hand of (5.4) into the sum

$$\sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (1 \,-\, \tilde{\lambda}_j)^2 \vartheta^2_{j,n} \,+\, \Delta_1(n)\,.$$

This decomposition allows us to show that $\Delta_1(n)$ is negligible and further to approximate the first term by a similar term in which the coefficients $\vartheta_{j,n}$ will be replaced by the Fourier coefficients ϑ_j of the function S.

Taking into account the definition of $\omega(\alpha)$ in (2.5) we can bound $\tilde{\omega}$ as

$$\tilde{\omega} \ge (A_k)^{1/(2k+1)}\, n^{1/(2k+1)}\, (\ln n)^{-1/(2k+1)}\,.$$

Therefore, by Lemma A.1 we obtain

$$\lim_{n\to\infty} \sup_{S\in W^k_r} n^{2k/(2k+1)}\, \Delta_1(n) = 0\,.$$

Let us consider now the next term $\Delta_2(n)$. We have

$$|\Delta_2(n)| = \left| \frac{1}{n^2} \sum_{d=1}^{n} \sigma^2_d \sum_{j=1}^{n} \tilde{\lambda}^2_j \,\overline{\phi}_j(x_d) \right| \le \frac{\sigma_*}{n} \sup_{0\le x\le 1} \left| \sum_{j=1}^{n} \tilde{\lambda}^2_j \,\overline{\phi}_j(x) \right| \,,$$

where $\overline{\phi}_j(x) = \phi^2_j(x) - 1$. Now by Lemma A.2 and definition (2.5) we obtain directly the same property for $\Delta_2(n)$, i.e.

$$\lim_{n\to\infty} \sup_{S\in W^k_r} n^{2k/(2k+1)}\, |\Delta_2(n)| = 0\,.$$

Setting

$$\hat{\gamma}_{k,n}(S) = n^{2k/(2k+1)} \sum_{j=\tilde{j}_0}^{\tilde{j}_1-1} (1-\tilde{\lambda}_j)^2 \vartheta^2_j + \varsigma_n n^{-1/(2k+1)} \sum_{j=1}^{n} \tilde{\lambda}^2_j$$

and applying the well-known inequality

$$(a+b)^2 \le (1+\delta)a^2 + (1+1/\delta)b^2$$

to the first term in the right-hand side of inequality (5.5) we obtain that, for any $\delta > 0$ and for any $p \in \mathcal{P}_*$,

$$\mathbf{E}_{S,p}\,\|\tilde{S} - S\|_n^2 \le (1+\delta)\,\hat{\gamma}_{k,n}(S)\,n^{-2k/(2k+1)}$$
$$+\,\Delta_1(n) + \Delta_2(n) + (1+1/\delta)\,\Delta_3(n)\,, \qquad (5.6)$$

where

$$\Delta_3(n) = \sum_{j=\tilde{j}_0+1}^{\tilde{j}_1-1} (\vartheta_{j,n} - \vartheta_j)^2\,.$$

Taking into account that $k \ge 1$ and that

$$\tilde{j}_1 \le (A_k)^{1/(2k+1)}\,n^{1/(2k+1)}(\ln n)^{(2k+2)/(2k+1)}\,,$$

we can show through Lemma A.3 that

$$\lim_{n\to\infty} \sup_{S\in W_r^k} n^{2k/(2k+1)}\,\Delta_3(n) = 0\,.$$

Therefore inequality (5.6) yields

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n(\tilde{S},S)/\gamma_k(S) \le \limsup_{n\to\infty} \sup_{S\in W_r^k} \hat{\gamma}_{k,n}(S)/\gamma_k(S)$$

and to prove (5.3) it suffices to show that

$$\limsup_{n\to\infty} \sup_{S\in W_r^k} \hat{\gamma}_{k,n}(S)/\gamma_k(S) \le 1\,. \qquad (5.7)$$

First it should be noted that definition (5.1) and inequalities (3.7)-(3.8) imply directly

$$\lim_{n\to\infty} \sup_{S\in W_r^k} \left|\tilde{t}_n/\overline{r}(S) - 1\right| = 0\,.$$

Moreover, by the definition of $(\tilde{\lambda}_j)_{1\le j\le n}$ for sufficiently large n for which $\tilde{t}_n \ge \overline{r}(S)$ we can calculate the following supremum

$$\sup_{j\ge 1} n^{2k/(2k+1)}(1-\tilde{\lambda}_j)^2/(\pi j)^{2k} = \pi^{-2k}(A_k\tilde{t}_n)^{-2k/(2k+1)}$$
$$\le \pi^{-2k}(A_k\overline{r}(S))^{-2k/(2k+1)}\,.$$

Therefore, taking into account the definition of the coefficients $(a_j)_{j\ge1}$ in (3.5) we obtain that

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_r^k}\sup_{j\ge\tilde{j}_0} \pi^{2k}(A_k\overline{r}(S))^{2k/(2k+1)}(1-\tilde{\lambda}_j)^2/a_j \le 1\,.$$

Moreover, by definition (2.5) we get that

$$\lim_{n\to\infty}\sup_{S\in W_r^k}\left| n^{-1/(2k+1)}\sum_{j=1}^n \tilde{\lambda}_j^2 - (A_k\overline{r}(S))^{1/(2k+1)}\int_0^1 (1-z^k)^2\mathrm{d}z\right| = 0\,.$$

Taking into account definition of W_r^k in (3.3) and condition (3.6) we obtain inequality (5.7). Hence Theorem 5.1. $\square$

Now Theorem 4.1 and Theorem 5.1 imply Theorem 4.2.

6 Lower bound for parametric heteroscedastic regression models

Let $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbf{P}_\vartheta, \vartheta\in\Theta\subseteq\mathbb{R}^l)$ be a statistical model relative to the observations $(y_j)_{1\le j\le n}$ governed by the regression equation

$$y_j \,=\, S_\vartheta(x_j) \,+\, \sigma_j(\vartheta)\,\xi_j\,, \tag{6.1}$$

where $\xi_1,\ldots,\xi_n$ are i.i.d. $\mathcal{N}(0,1)$ random variables, $\vartheta=(\vartheta_1,\ldots,\vartheta_l)'$ is a unknown parameter vector, $S_\vartheta(x)$ is a unknown (or known) function and

$\sigma_j(\vartheta) = g(x_j, S_\vartheta)$, with the function $g(x,S)$ defined in condition $\mathbf{H}_1$). Assume that a prior distribution μ_ϑ of the parameter ϑ in $\mathbb{R}^l$ is defined by the density $\Phi(\vartheta)$ of the following form

$$\Phi(\vartheta) \;=\; \Phi(\vartheta_1,\ldots,\vartheta_l) = \prod_{i=1}^{l} \varphi_i(\vartheta_i)\,,$$

where φ_i is a continuously differentiable bounded density on $\mathbb{R}$ with

$$\mathcal{I}_i = \int_{\mathbb{R}} \frac{(\dot{\varphi}_i(z))^2}{\varphi_i(z)} \mathrm{d}z < \infty\,.$$

Let $\lambda(\cdot)$ be a continuously differentiable $\mathbb{R}^l \to \mathbb{R}$ function such that, for any $1 \le i \le l$,

$$\lim_{|\theta_i|\to\infty} \lambda(\vartheta)\,\varphi_i(\vartheta_i) = 0 \quad \text{and} \quad \int_{\mathbb{R}^l} \left|\lambda_i'(\vartheta)\right|\,\Phi(\vartheta)\mathrm{d}\vartheta < \infty\,, \tag{6.2}$$

where

$$\lambda_i'(\vartheta) = (\partial/\partial\vartheta_i)\,\lambda(\vartheta)\,.$$

Let $\hat{\lambda}_n$ be an estimator of $\lambda(\vartheta)$ based on observations $(y_j)_{1\le j\le n}$. For any $\mathcal{B}(\mathbb{R}^n \times \mathbb{R}^l)$ - mesurable integrable function $G(x,\vartheta), x \in \mathbb{R}^n, \vartheta \in \mathbb{R}^l$, we set

$$\tilde{\mathbf{E}}\,G(Y,\vartheta) \;=\; \int_{\mathbb{R}^l} \mathbf{E}_\vartheta\,G(Y,\vartheta)\,\Phi(\vartheta)\,\mathrm{d}\vartheta\,,$$

where $\mathbf{E}_\vartheta$ is the expectation with respect to the distribution $\mathbf{P}_\vartheta$ of the vector $Y = (y_1,\ldots,y_n)$. Note that in this case

$$\mathbf{E}_\vartheta\,G(Y,\vartheta) \;=\; \int_{\mathbb{R}^n} G(v,\vartheta)\,f(v,\vartheta)\,\mathrm{d}v\,,$$

where

$$f(v,\vartheta) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j(\vartheta)} \exp\left\{-\,\frac{(v_j - S_\vartheta(x_j))^2}{2\sigma_j^2(\vartheta)}\right\}\,. \tag{6.3}$$

We prove the following result.

Theorem 6.1. *Assume that conditions* $\mathbf{H}_1) - \mathbf{H}_2)$ *hold. Moreover, assume that the function* $S_\vartheta(\cdot)$ *is uniformly over* $0 \le x \le 1$ *differentiable in* $\mathcal{C}[0,1]$ *with respect to* ϑ_i, $1 \le i \le l$, *i.e. for any* $1 \le i \le l$ *there exists a function* $S'_{\vartheta,i} \in \mathcal{C}[0,1]$ *such that*

$$\lim_{h\to 0} \max_{0\le x\le 1} \left| \left(S_{\vartheta+h\mathbf{e}_i}(x) - S_\vartheta(x) - S'_{\vartheta,i}(x)h \right) /h \right| = 0\,, \qquad (6.4)$$

where $\mathbf{e}_i = (0, ..., 1, ..., 0)'$, *all coordinates are* 0, *except the ith equals to* 1 . *Then for any square integrable estimator* $\hat{\lambda}_n$ *of* $\lambda(\vartheta)$ *and any* $1 \le i \le l$,

$$\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda)^2 \ge \Lambda_i^2/(F_i \,+\, B_i \,+\, \mathcal{I}_i)\,, \qquad (6.5)$$

where $\Lambda_i = \int_{\mathbb{R}^l} \lambda'_i(\vartheta)\,\Phi(\vartheta)\mathrm{d}\vartheta$, $F_i = \sum_{j=1}^n \int_{\mathbb{R}^l} (S'_{\vartheta,i}(x_j)/\sigma_j(\vartheta))^2\,\Phi(\vartheta)\mathrm{d}\vartheta$ *and*

$$B_i = \frac{1}{2} \sum_{j=1}^n \int_{\mathbb{R}^l} \frac{\tilde{\mathbf{L}}_i^2(x_j, S_\vartheta)}{\sigma_j^4(S_\vartheta)} \Phi(\vartheta)\mathrm{d}\vartheta\,,$$

$\tilde{\mathbf{L}}_i(x,\vartheta) = \mathbf{L}_{x,S_\vartheta}(S'_{\vartheta,i})$, *the operator* $\mathbf{L}_{x,S}$ *is defined in the condition* $\mathbf{H}_2)$.

Proof. We put

$$\varrho_i(v,\vartheta) = \frac{1}{f(v,\vartheta)\Phi(\vartheta)} \frac{\partial}{\partial\vartheta_i} \left(f(v,\vartheta)\Phi(\vartheta) \right)\,.$$

Note that due to condition (3.7) the density (6.3) is bounded, i.e.

$$f(v,\vartheta) \le (2\pi g_*)^{-n/2}\,.$$

So through (6.2) we obtain that

$$\lim_{|\vartheta_i|\to\infty} \lambda(\vartheta)\, f(v,\vartheta)\varphi_i(\vartheta_i) = 0\,.$$

Therefore, integrating by parts yields

$$\begin{aligned}\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda)\varrho_i &= \int_{\mathbb{R}^{n+l}} (\hat{\lambda}_n(v) - \lambda(\vartheta)) \frac{\partial}{\partial\vartheta_i} \left(f(v,\vartheta)\Phi(\vartheta) \right) \mathrm{d}\vartheta \mathrm{d}v \\ &= \int_{\mathbb{R}^l} \left(\frac{\partial}{\partial\vartheta_i} \lambda(\vartheta) \right) \Phi(\vartheta) \left(\int_{\mathbb{R}^n} f(v,\vartheta)\mathrm{d}v \right) \mathrm{d}\vartheta = \Lambda_i\,.\end{aligned}$$

Now the Bouniakovskii-Cauchy-Schwarz inequality gives the following lower bound

$$\tilde{\mathbf{E}}(\hat{\lambda}_n - \lambda)^2 \geq \Lambda_i^2 / \tilde{\mathbf{E}} \varrho_i^2 \,.$$

To estimate the denominator in the last ratio, note that

$$\begin{aligned} \varrho_i(v, \vartheta) &= \frac{1}{f(v,\vartheta)} \frac{\partial}{\partial \vartheta_i} f(v,\vartheta) + \frac{\dot{\varphi}_i(\vartheta_i)}{\varphi_i(\vartheta_i)} \\ &= \tilde{f}_i(v,\vartheta) + \frac{\dot{\varphi}_i(\vartheta_i)}{\varphi_i(\vartheta_i)} \,, \end{aligned}$$

where

$$\tilde{f}_i(v,\vartheta) = (\partial/\partial\vartheta_i) \ln f(v,\vartheta) \,.$$

From (6.1) it follows that

$$\tilde{f}_i(v,\vartheta) = \sum_{j=1}^n (\xi_j^2 - 1) \, \frac{1}{2\sigma_j^2(\vartheta)} \frac{\partial}{\partial \vartheta_i} \, \sigma_j^2(\vartheta) + \sum_{j=1}^n \, \xi_j \, \frac{S_i'(x_j)}{\sigma_j(\vartheta)} \,.$$

Moreover, conditions $\mathbf{H}_2$) and (6.4) imply

$$(\partial/\partial\vartheta_i) \, \sigma_j^2(\vartheta) \; = \; \partial/\partial\vartheta_i) \, g^2(x_j, S_\vartheta) \; = \; \tilde{\mathbf{L}}_i(x_j, \vartheta)$$

from which it follows

$$\tilde{\mathbf{E}} \, \left(\tilde{f}_i(Y,\vartheta) \right)^2 = F_i + B_i \,.$$

This implies inequality (6.5). Hence Theorem 6.1. □

7 Parametric kernel function family

In this section we define and study some special parametric kernel functions family which will be used to prove the sharp lower bound (4.6).

Let us begin by kernel functions. We fix $\eta > 0$ and we set

$$I_\eta(x) = \eta^{-1} \int_{\mathbb{R}} \mathbf{1}_{(|u| \le 1-\eta)} \, V\left(\frac{u-x}{\eta}\right) \mathrm{d}u \,, \tag{7.1}$$

where $\mathbf{1}_A$ is the indicator of a set A, the kernel $V \in \mathbf{C}^\infty(\mathbb{R})$ is such that

$$V(u) = 0 \quad \text{for} \quad |u| \ge 1 \quad \text{and} \quad \int_{-1}^{1} V(u) \, \mathrm{d}u = 1 \,.$$

It is easy to see that the function $I_\eta(x)$ possesses the properties :

$$0 \le I_\eta \le 1 \,, \quad I_\eta(x) = 1 \quad \text{for} \quad |x| \le 1 - 2\eta \quad \text{and}$$
$$I_\eta(x) = 0 \quad \text{for} \quad |x| \ge 1 \,.$$

Moreover, for any $c > 0$ and $m \ge 1$

$$\lim_{\eta \to 0} \sup_{f \,:\, |f|_* \le c} \left| \int_{\mathbb{R}} f(x) I_\eta^m(x) \mathrm{d}x - \int_{-1}^{1} f(x) \mathrm{d}x \right| = 0 \,, \tag{7.2}$$

where $|f|_* = \sup_{-1 \le x \le 1} |f(x)|$.

We divide the interval $[0,1]$ into M equal parts of length $2h$ and on each of them we construct a kernel-type function that was used in Ibragimov, Hasminskii, 1981, to obtain the lower bound for estimation at a fixed point. A such constructed on each interval function equals to zero at the extremities together with all derivatives. It means that Fourier partial sums with respect to the trigonometric basis in $\mathcal{L}_2[-1,1]$ give a natural parametric approximation to the function on each interval.

Let $(e_j)_{j \ge 1}$ be the trigonometric basis in $L_2[-1,1]$, i.e.

$$e_1 = 1/\sqrt{2} \,, \quad e_j(x) = Tr_j \left(\pi [j/2] x\right) \,, \; j \ge 2 \,, \tag{7.3}$$

where $Tr_j(x) = \cos(x)$ for even j and $Tr_j(x) = \sin(x)$ for odd j.

Now, for any array $z = \{(z_{m,j})_{1 \le m \le M_n\,,\,1 \le j \le N_n}\}$ we define the following function

$$S_{z,n}(x) = \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} z_{m,j}\, D_{m,j}(x)\,, \tag{7.4}$$

where $D_{m,j}(x) = e_j\,(v_m(x))\, I_\eta\,(v_m(x))$,

$$v_m(x) = (x - \tilde{x}_m)/h_n\,, \quad \tilde{x}_m = 2mh_n \quad \text{and} \quad M_n = [1/(2h_n)] - 1\,.$$

We assume that the sequences $(N_n)_{n\ge1}$ and $(h_n)_{n\ge1}$, satisfy the following conditions.

$\mathbf{A}_1$) *The sequence* $N_n \to \infty$ *as* $n \to \infty$ *and for any* $p > 0$

$$\lim_{n\to\infty} N_n^p/n \,=\, 0\,.$$

Moreover, there exist $0 < \delta_1 < 1$ *and* $\delta_2 > 0$ *such that*

$$h_n \,=\, \mathrm{O}(n^{-\delta_1}) \quad \textit{and} \quad h_n^{-1} \,=\, \mathrm{O}(n^{\delta_2}) \quad \textit{as} \quad n \to \infty\,.$$

To define a prior distribution on the family of arrays, we choose the following random array $\vartheta = \{(\vartheta_{m,j})_{1 \le m \le M_n\,,\,1 \le j \le N_n}\}$ with

$$\vartheta_{m,j} \,=\, t_{m,j}\, \zeta_{m,j}\,, \tag{7.5}$$

where $(\zeta_{m,j})$ are i.i.d. $\mathcal{N}(0,1)$ random variables and $(t_{m,j})_{1 \le m \le M_n\,,\,1 \le j \le N_n}$ are some nonrandom positive coefficients. We make use of gaussian variables since they possess the minimal Fisher information and therefore maximize the lower bound (6.5). We set

$$t_n^* = \max_{1 \le m \le M_n} \sum_{j=1}^{N_n} t_{m,j}\,. \tag{7.6}$$

We assume that the coefficients $(t_{m,j})_{1\le m\le M_n\,,\,1\le j\le N_n}$ satisfy the following conditions.

$\mathbf{A}_2)$ *There exists a sequence of positive numbers* $(d_n)_{n\ge 1}$ *such that*

$$\lim_{n\to\infty}\frac{d_n}{h_n^{2k-1}}\sum_{m=1}^{M_n}\sum_{j=1}^{N_n}t_{m,j}^2\,j^{2(k-1)}=0\,,\qquad \lim_{n\to\infty}\sqrt{d_n}\,t_n^*=0\,,\tag{7.7}$$

moreover, for any $p>0$,

$$\lim_{n\to\infty}n^p\,\exp\{-d_n/2\}=0\,.$$

$\mathbf{A}_3)$ *For some* $0<\varepsilon<1$

$$\limsup_{n\to\infty}\frac{1}{h_n^{2k-1}}\sum_{m=1}^{M_n}\sum_{j=1}^{N_n}t_{m,j}^2\,j^{2k}\ \le(1-\varepsilon)r\left(\frac{2}{\pi}\right)^{2k}\,.$$

$\mathbf{A}_4)$ *There exists* $\epsilon_0>0$ *such that*

$$\lim_{n\to\infty}\frac{1}{h_n^{4k-2+\epsilon_0}}\sum_{m=1}^{M_n}\sum_{j=1}^{N_n}t_{m,j}^4\,j^{4k}\ =0\,.$$

Proposition 7.1. *Let conditions* $\mathbf{A}_1)$–$\mathbf{A}_2)$. *Then, for any* $p>0$ *and for any* $\delta>0$,

$$\lim_{n\to\infty}n^p\max_{0\le l\le k-1}\mathbf{P}\left(\|S^{(l)}_{\vartheta,n}\|>\delta\right)=0\,.$$

Proof. First note that for $0\le x\le 1$ we can represent the lth derivative as

$$S^{(l)}_{\vartheta,n}(x)=\frac{1}{h^l}\sum_{m=1}^{M_n}\sum_{i=0}^{l}\binom{l}{i}\,I_\eta^{(l-i)}(v_m(x))\,Q_{i,m}(v_m(x))\,,\tag{7.8}$$

where

$$Q_{i,m}(v) = \sum_{j=1}^{N_n} \vartheta_{m,j} e_j^{(i)}(v)\,.$$

Therefore

$$\|S_{\vartheta,n}^{(l)}\|^2 = \frac{1}{h_n^{2l-1}} \sum_{m=1}^{M_n} \int_{-1}^{1} \left(\sum_{i=0}^{l} \binom{l}{i} \, I_\eta^{(l-i)}(v)\, Q_{i,m}(v) \right)^2 \mathrm{d}v$$

and by the Bounyakovskii-Cauchy-Schwarz inequality we obtain that

$$\|S_{\vartheta,n}^{(l)}\|^2 \le \frac{C^*(l,\eta)}{h_n^{2l-1}} \sum_{i=0}^{l} \overline{Q}_{i,m} \tag{7.9}$$

with $C^*(l,\eta) = \max_{-1\le v\le 1} \sum_{i=0}^{l} \left(\binom{l}{i} \, I_\eta^{(l-i)}(v) \right)^2$ and

$$\overline{Q}_{i,m} = \sum_{m=1}^{M_n} \int_{-1}^{1} Q_{i,m}^2(v)\,\mathrm{d}v\,.$$

Now we show that for any $0 \le i \le k-1$ and $\delta > 0$

$$\lim_{n\to\infty} n^p\, \mathbf{P}\left(\overline{Q}_{i,m} > \delta h_n^{2k-1} \right) = 0\,. \tag{7.10}$$

To that end we introduce the following set

$$\Xi_n = \{ \max_{1\le m\le M_n} \max_{1\le j\le N} \zeta_{m,j}^2 \le d_n \}\,, \tag{7.11}$$

where the sequence $(d_n)_{n\ge 1}$ is given in condition $\mathbf{A}_2$). Therefore, taking into account that

$$\int_{-1}^{1} Q_{i,m}^2(v)\,\mathrm{d}v = \sum_{j=1}^{N_n} \vartheta_{m,j}^2 \int_{-1}^{1} (e_j^{(i)}(v))^2\,\mathrm{d}v$$
$$\le \left(\frac{\pi}{2}\right)^{2i} \sum_{j=1}^{N_n} t_{m,j}^2\, j^{2i} \zeta_{m,j}^2\,,$$

the function $\overline{Q}_{i,m}$ can be estimated on the set Ξ_n as

$$\overline{Q}_{i,m} \le \left(\frac{\pi}{2}\right)^{2i} d_n \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t^2_{m,j}\, j^{2i}$$

and by (7.7) we get, for any $\delta > 0$ and sufficiently large n,

$$\mathbf{P}\left(\overline{Q}_{i,m} > \delta h_n^{2k-1}\right) \le \mathbf{P}\left(\Xi_n^c\right) .$$

Moreover, for sufficiently large n

$$\mathbf{P}\left(\Xi_n^c\right) \le M_n\, N_n\, e^{-d_n/2} .$$

Therefore, conditions $\mathbf{A}_1$) and (7.7) imply

$$\limsup_{n\to\infty} n^p\, \mathbf{P}\left(\Xi_n^c\right) = 0 , \tag{7.12}$$

for any $p > 0$. Hence Proposition 7.1.

□

Proposition 7.2. *Let conditions* $\mathbf{A}_1$)–$\mathbf{A}_4$)*. Then, for any* $p > 0$*,*

$$\lim_{n\to\infty} n^p\, \mathbf{P}(S_{\vartheta,n} \notin W_r^k) = 0 .$$

Proof. First of all we prove that for ε from condition $\mathbf{A}_3$)

$$\lim_{n\to\infty} n^p\, \mathbf{P}\left(\|S^{(k)}_{\vartheta,n}\| > \sqrt{(1-\varepsilon/4)r}\right) = 0 . \tag{7.13}$$

Indeed, putting in (7.8) $l = k$ we can represent the kth derivative of $S_{\vartheta,n}$ as follows

$$S^{(k)}_{\vartheta,n}(x) = \hat{S}_k(x) + \overline{S}_k(x) \tag{7.14}$$

with

$$\hat{S}_k(x) = \frac{1}{h^k} \sum_{m=1}^{M_n} \sum_{i=0}^{k-1} \binom{k}{i} I_\eta^{(k-i)}(v_m(x))\, Q_{i,m}(v_m(x))$$

and

$$\overline{S}_k(x) = \frac{1}{h^k} \sum_{m=1}^{M_n} I_\eta(v_m(x))\, Q_{k,m}(v_m(x))\,.$$

First, note that, we can estimate the norm of $\hat{S}_k(x)$ by the same way as in inequality (7.9), i.e.

$$\|\hat{S}_k\|^2 \le \frac{C^*(k,\eta)}{h_n^{2k-1}} \sum_{i=0}^{k-1} \overline{Q}_{i,m}\,.$$

By making use of (7.10) we obtain that, for any $p > 0$ and for any $\delta > 0$,

$$\lim_{n\to\infty} n^p\, \mathbf{P}\left(\|\hat{S}_k\| > \delta\right) = 0\,. \tag{7.15}$$

Let us consider now the last term in (7.14). Taking into account that $0 \le I_\eta(v) \le 1$ we get

$$\begin{aligned}\|\overline{S}_k\|^2 &= \frac{1}{h_n^{2k-1}} \sum_{m=1}^{M_n} \int_{-1}^{1} I_\eta^2(v) Q_{k,m}^2(v)\mathrm{d}v \\ &\le \left(\frac{\pi}{2}\right)^{2k} \frac{1}{h_n^{2k-1}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2\, j^{2k}\, \zeta_{m,j}^2\,.\end{aligned}$$

Therefore from condition $\mathbf{A}_3$) we get for sufficiently large n

$$\|\overline{S}_k\|^2 \le (1-\varepsilon/2)r + \left(\frac{\pi}{2}\right)^{2k} \sum_{m=1}^{M_n} \overline{\zeta}_m := (1-\varepsilon/2)r + \left(\frac{\pi}{2}\right)^{2k} Y_n$$

with

$$\overline{\zeta}_m = \frac{1}{h_n^{2k-1}} \sum_{j=1}^{N_n} t_{m,j}^2\, j^{2k} \tilde{\zeta}_{m,j} \quad \text{and} \quad \tilde{\zeta}_{m,j} = \zeta_{m,j}^2 - 1\,.$$

We show that for any $p > 0$ and for any $\delta > 0$

$$\lim_{n\to\infty} n^p\, \mathbf{P}\left(|Y_n| > \delta\right) = 0\,. \tag{7.16}$$

Indeed, by the Chebyshev inequality for any $\iota > 0$

$$\mathbf{P}\left(|Y_n| > \delta\right) \le \mathbf{E}\,\left(Y_n\right)^{2\iota} / \delta^{2\iota}\,. \tag{7.17}$$

Note now that according to the Burkholder-Davis-Gundy inequality for any $\iota > 1$ there exists a constant $B^*(\iota) > 0$ such that

$$\mathbf{E}\,(Y_n)^{2\iota} \le B^*(\iota)\,\mathbf{E}\,\left(\sum_{m=1}^{M_n} \overline{\zeta}^2_m\right)^{\iota}\,.$$

Moreover, by putting

$$\tilde{\zeta}_* = \max_{1\le m\le M_n}\,\max_{1\le j\le N_n}\,\tilde{\zeta}^2_{m,j}$$

we obtain that

$$\overline{\zeta}^2_m \,\le\, \frac{N_n}{h^{4k-2}_n}\sum_{j=1}^{N_n} t^4_{m,j}\,j^{4k}\,\tilde{\zeta}_*\,.$$

Therefore, by condition $\mathbf{A}_4$) for sufficiently large n

$$\begin{aligned}\mathbf{E}\,(Y_n)^{2\iota} &\le B^*(\iota)\,N^{\iota}_n\,h^{\iota\epsilon_0}_n\,\mathbf{E}\,\tilde{\zeta}^{\iota}_*\\ &\le B^*(\iota)\,\mathbf{E}\,(\zeta^2-1)^{2\iota}\,M_n N^{\iota+1}_n h^{\iota\epsilon_0}_n\,,\end{aligned}$$

where $\zeta \sim \mathcal{N}(0,1)$. Taking into account here condition $\mathbf{A}_1$) we obtain for sufficiently large n

$$\mathbf{E}\,(Y_n)^{2\iota} \le n^{-\delta_1\,(\iota\epsilon_0-2)}\,.$$

Thus, choosing in (7.17)

$$\iota > p/(\epsilon_0\delta_1) + 2/\epsilon_0$$

we obtain limiting equality (7.16) which together with (7.14)-(7.15) implies (7.13). Now it is easy to deduce that Proposition 7.1 yields Proposition 7.2.

□

Proposition 7.3. *Let conditions* $\mathbf{A}_1$)–$\mathbf{A}_4$). *Then, for any* $p > 0$,

$$\lim_{n\to\infty} n^p\,\mathbf{E}\,\|S_{\vartheta,n}\|^2\,\left(\mathbf{1}_{\{S_{\vartheta,n}\notin W^k_r\}} \,+\, \mathbf{1}_{\Xi^c_n}\right) = 0\,.$$

Proof. First of all, we remind that due to condition $\mathbf{A}_2$)

$$\lim_{n\to\infty} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 \le \lim_{n\to\infty} \frac{d_n}{h_n^{2k-1}} \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 \, j^{2(k-1)} = 0 \,.$$

Therefore, taking into account that

$$\|S_{\vartheta,n}\|^2 \le h_n \sum_{m=1}^{M_n} \sum_{j=1}^{N_n} t_{m,j}^2 \zeta_{m,j}^2 \tag{7.18}$$

we obtain, for sufficiently large n,

$$\mathbf{E}\, \|S_{\vartheta,n}\|^2 \, \left(\mathbf{1}_{\{S_{\vartheta,n} \notin W_r^k\}} + \mathbf{1}_{\Xi_n^c} \right) \le \max_{m,j} \mathbf{E}\, \zeta_{m,j}^2 \left(\mathbf{1}_{\{S_{\vartheta,n} \notin W_r^k\}} + \mathbf{1}_{\Xi_n^c} \right) \,.$$

Moreover, for any $1 \le m \le M_n$ and $1 \le j \le N_n$, we estimate the last term as

$$\mathbf{E}\, \zeta_{m,j}^2 \left(\mathbf{1}_{\{S_{\vartheta,n} \notin W_r^k\}} + \mathbf{1}_{\Xi_n^c} \right) \le n\, \mathbf{P}(S_{\vartheta,n} \notin W_r^k) + n\, \mathbf{P}(\Xi_n^c) + 2\mathbf{E}\, \zeta^2 \, \mathbf{1}_{\{\zeta^2 \ge n\}} \,,$$

where $\zeta \sim \mathcal{N}(0,1)$. By applying now Proposition 7.2 and limit (7.12) we obtain Proposition 7.3. $\square$

Proposition 7.4. *Let conditions* $\mathbf{A}_1$)–$\mathbf{A}_4$)*. Then for any function g satisfying conditions* (3.7) *and* $\mathbf{H}_4$)

$$\lim_{n\to\infty} \sup_{0\le x\le 1} \mathbf{E} \, \left| g^{-2}(x, S_{\vartheta,n}) - g_0^{-2}(x) \right| = 0 \,.$$

Proof. First, note that on the set Ξ the random function $S_{\vartheta,n}$ is uniformly bounded, i.e.

$$|S_{\vartheta,n}|_* = \sup_{0\le x\le 1} |S_{\vartheta,n}(x)| \le \sqrt{d_n}\, t_n^* \,, \tag{7.19}$$

where the coefficient t^*_n is defined in (7.6). Therefore by condition $\mathbf{H}_1$) we obtain

$$\mathbf{E}\,\left|g^{-2}(x,S_{\vartheta,n})-g_0^{-2}(x)\right|\le \max_{|S|_*\le\sqrt{d_n}\,t^*_n}|g^{-2}(x,S)-g_0^{-2}(x)|+(2/g_*)\,\mathbf{P}\left(\Xi^c_n\right)\,.$$

Conditions $\mathbf{A}_2$) and $\mathbf{H}_4$) together with the limit relation (7.12) imply Proposition 7.4. □

8 Lower bound

In this section we prove Theorem 4.3. To that end we establish the following auxiliary result.

Lemma 8.1. *For any $0<\delta<1$ and any estimate $\hat{S}_n$ of $S\in W^k_r$,*

$$\|\hat{S}_n-S\|^2_n\ge(1-\delta)\|T_n(\hat{S})-S\|^2\,-\,(\delta^{-1}\,-\,1)\,r/n^2\,,$$

where $T_n(\hat{S})(x)=\sum_{k=1}^n\,\hat{S}_n(x_k)\mathbf{1}_{(x_{k-1},x_k]}(x)$.

Proof of this Lemma is given in Appendix A.2.

This Lemma implies that to prove (4.6), it suffices to show the same asymptotic inequality for the integral risk, i.e.

$$\liminf_{n\to\infty}\,\inf_{\hat{S}_n}\,n^{2k/(2k+1)}\,\mathcal{R}_0(\hat{S}_n)\,\ge\,1\,,\tag{8.1}$$

where

$$\mathcal{R}_0(\hat{S}_n)\;=\;\sup_{S\in W^k_r}\mathbf{E}_{S,q}\,\|\hat{S}_n-S\|^2/\gamma_k(S)\,,$$

q is the gaussian $(0,1)$ density of the noise (ξ_j) and $\|S\|^2=\int_0^1 S^2(x)\mathrm{d}x$.

To show (8.1) we will make use of the sequence of random functions $(S_{\vartheta,n})_{n\ge1}$ defined in (7.4)-(7.5) with the coefficients $(t_{m,j})$ satisfying conditions $\mathbf{A}_1)$–$\mathbf{A}_4)$ which will be chosen later.

For any estimator $\hat{S}_n$, we denote by $\hat{S}^0_n$ its projection onto W^k_r, i.e. $\hat{S}^0_n = \mathrm{Pr}_{W^k_r}(\hat{S}_n)$. Since W^k_r is a convex set, we get that

$$\|\hat{S}_n - S\|^2 \ge \|\hat{S}^0_n - S\|^2 \,.$$

Therefore, we can write that

$$\mathcal{R}_0(\hat{S}_n) \ge \int_{\{z:S_{z,n}\in W^k_r\}\cap\Xi_n} \frac{\mathbf{E}_{S_{z,n},q}\|\hat{S}^0_n - S_{z,n}\|^2}{\gamma_k(S_{z,n})}\,\mu_\vartheta(\mathrm{d}z)\,.$$

Here μ_ϑ denotes the distribution of ϑ in $\mathbb{R}^l$ with $l = M_n N_n$. We recall also that the set Ξ_n is defined in (7.11). Moreover, taking into account here inequality (7.19) we estimate the risk $\mathcal{R}_0(\hat{S}_n)$ from below as

$$\mathcal{R}_0(\hat{S}_n) \ge \frac{1}{\gamma^*_n}\int_{\{z:S_{z,n}\in W^k_r\}\cap\Xi_n} \mathbf{E}_{S_{z,n},q}\|\hat{S}^0_n - S_{z,n}\|^2\,\mu_\vartheta(\mathrm{d}z)$$

with

$$\gamma^*_n = \sup_{|S|_*\le\sqrt{d_n}\,t^*_n} \gamma_k(S)\,. \tag{8.2}$$

Let us introduce now the corresponding Bayes risk

$$\tilde{\mathcal{R}}_0(\hat{S}^0_n) = \int_{\mathbb{R}^l} \mathbf{E}_{S_{z,n},q}\|\hat{S}^0_n - S_{z,n}\|^2\,\mu_\vartheta(\mathrm{d}z)\,. \tag{8.3}$$

Now through this risk we rewrite the lower bound for $\mathcal{R}_0(\hat{S}_n)$ as

$$\mathcal{R}_0(\hat{S}^0_n) \ge \tilde{\mathcal{R}}_0(\hat{S}^0_n)/\gamma^*_n - 2\,\varpi_n/\gamma^*_n \tag{8.4}$$

with

$$\varpi_n = \mathbf{E}(\mathbf{1}_{\{S_{\vartheta,n}\notin W^k_r\}} + \mathbf{1}_{\Xi^c_n})(r + \|S_{\vartheta,n}\|^2)\,.$$

First of all, we reduce the nonparametric problem to parametric one. For this we replace the functions $\hat{S}_n^0$ and S by their Fourier series with respect to the basis

$$\tilde{e}_{m,i}(x) = (1/\sqrt{h})\, e_i\,(v_m(x))\,\mathbf{1}_{(|v_m(x)|\le 1)}\,.$$

By making use of this basis we can estimate the norm $\|\hat{S}_n^0 - S_{z,n}\|^2$ from below as

$$\|\hat{S}_n^0 - S_{z,n}\|^2 \ge \sum_{m=1}^{M_n}\sum_{j=1}^{N_n}(\hat{\lambda}_{m,j} \,-\, \lambda_{m,j}(z))^2\,,$$

where

$$\hat{\lambda}_{m,j} = \int_0^1 \hat{S}_n^0(x)\tilde{e}_{m,j}(x)\mathrm{d}x \quad \text{and} \quad \lambda_{m,j}(z) = \int_0^1 S_{z,n}(x)\tilde{e}_{m,j}(x)\,\mathrm{d}x\,.$$

Moreover, from definition (7.4) one gets

$$\lambda_{m,j}(z) = \sqrt{h}\sum_{i=1}^{N_n} z_{m,i}\int_{-1}^1 e_i(u)e_j(u)I_\eta(u)\,\mathrm{d}u\,.$$

It is easy to see that the functions $\lambda_{m,j}(\cdot)$ satisfy condition (6.2) for gaussian prior densities. In this case (see the definition in (6.5)) we have

$$\Lambda_{m,j} = (\partial/\partial z_{m,j})\lambda_{m,j}(z) = \sqrt{h}\overline{e}_j(I_\eta)\,,$$

where

$$\overline{e}_j(f) = \int_{-1}^1 e_j^2(v)\,f(v)\,\mathrm{d}v\,. \tag{8.5}$$

Now to obtain a lower bound for the Bayes risk $\tilde{\mathcal{R}}_0(\hat{S}_n^0)$ we make use of Theorem 6.1 which implies that

$$\tilde{\mathcal{R}}_0(\hat{S}_n^0)\,\ge \sum_{m=1}^{M_n}\sum_{j=1}^{N_n}\frac{h\overline{e}_j^2(I_\eta)}{F_{m,j}+B_{m,j}\,+t_{m,j}^{-2}}\,, \tag{8.6}$$

where $F_{m,j}=\sum_{i=1}^{n}\, D^2_{m,j}(x_i)\,\mathbf{E}\,g^{-2}(x_i,S_{\vartheta,n})$ and

$$B_{m,j}=\frac{1}{2}\sum_{i=1}^{n}\mathbf{E}\,\left(\frac{\tilde{\mathbf{L}}_{m,j}(x_i,S_{\vartheta,n})}{g^2(x_i,S_{\vartheta,n})}\right)^2$$

with $\tilde{\mathbf{L}}_{m,j}(x,S)=\mathbf{L}_{x,S}\left(D_{m,j}\right)$. In the appendix we show that

$$\lim_{n\to\infty}\sup_{1\le m\le M_n}\sup_{1\le j\le N_n}\left|F_{m,j}/(nh)\,-\,\overline{e}_j(I^2_\eta)\,g_0^{-2}(\tilde{x}_m)\right|=0 \tag{8.7}$$

and

$$\lim_{n\to\infty}\sup_{1\le m\le M_n}\sup_{1\le j\le N_n}\left|B_{m,j}/(nh)\right|\,=\,0\,. \tag{8.8}$$

This means that, for any $\nu>0$ and for sufficiently large n,

$$\sup_{1\le m\le M_n}\sup_{1\le j\le N_n}\frac{F_{m,j}+B_{m,j}+t^{-2}_{m,j}}{nh\overline{e}_j(I^2_\eta)g_0^{-2}(\tilde{x}_m)+t^{-2}_{m,j}}\le 1+\nu\,.$$

Therefore, if we denote in (8.6)

$$\kappa^2_{m,j}=\,nh\,g_0^{-2}(\tilde{x}_m)\,t^2_{m,j}\quad\text{and}\quad \tau_j(\eta,y)=\overline{e}^2_j(I_\eta)\,y/(\overline{e}^2_j(I^2_\eta)y+1)$$

we obtain that, for sufficiently large n,

$$n^{2k/(2k+1)}\tilde{\mathcal{R}}_0(\hat{S}^0_n)\,\ge\,\frac{1}{1+\nu}\,n^{-1/(2k+1)}\,\sum_{m=1}^{M_n}\,g_0^2(\tilde{x}_m)\,\sum_{j=1}^{N_n}\,\tau_j(\eta,\kappa^2_{m,j})\,.$$

In the appendix we show that

$$\lim_{\eta\to 0}\sup_{N\ge 1}\sup_{(y_1,\ldots,y_N)\in\mathbb{R}^N_+}\left|\sum_{j=1}^{N}\,\tau_j(\eta,y_j)/\sum_{j=1}^{N}\,\overline{\tau}(y_j)\,-\,1\right|\,=\,0\,, \tag{8.9}$$

where

$$\overline{\tau}(y)=y/(y+1)\,.$$

Therefore we can write that, for sufficiently large n,

$$n^{\frac{2k}{2k+1}}\tilde{\mathcal{R}}_0(\hat{S}^0_n)\,\ge\,\frac{1-\nu}{1+\nu}\,n^{-\frac{1}{2k+1}}\,\sum_{m=1}^{M_n}\,g_0^2(\tilde{x}_m)\,J_{N_n}(\kappa^2_{m,1},\ldots,\kappa^2_{m,N_n})\,, \tag{8.10}$$

where

$$J_N(y_1,\ldots,y_N)=\sum_{j=1}^{N}\overline{\tau}(y_j)\,.$$

Obviously, to obtain a "good" lower bound for the risk $\tilde{\mathcal{R}}_0(\hat{S}^0_n)$ one needs to maximize the right-hand side of inequality (8.10). Hence we choose the coefficients $(\kappa^2_{m,j})$ by maximization of the function J_N, i.e.

$$\max_{y_1,\ldots,y_N} J_N(y_1,\ldots,y_N)\quad\text{subject to}\quad \sum_{j=1}^{N} y_j j^{2k}\le R\,.$$

The parameter $R>0$ will be chosen later to satisfy condition $\mathbf{A}_3$). By the Lagrange multipliers method it is easy to find that the solution of this problem is

$$y^*_j(R)=(R+\sum_{j=1}^{N} j^{2k})\,j^{-k}/\sum_{j=1}^{N} j^{k}-1\quad\text{for}\quad 1\le j\le N\,. \tag{8.11}$$

To obtain a positive solution in (8.11) we need to impose the following condition

$$R\ge\ N^k\sum_{j=1}^{N} j^{k}-\sum_{j=1}^{N} j^{2k}\,. \tag{8.12}$$

Moreover, from condition $\mathbf{A}_3$) we obtain that

$$R\le 2^{2k+1}(1-\varepsilon)r\,n\,h^{2k+1}_n/(\pi^{2k}\hat{g}_0):=R^*_n\,, \tag{8.13}$$

where

$$\hat{g}_0=2h_n\sum_{m=1}^{M_n} g^2_0(\tilde{x}_m)\,.$$

Note that by condition $\mathbf{H}_4$) the function $g_0(\cdot)=g(\cdot,S_0)$ is continuous on the interval $[0,1]$, therefore

$$\lim_{n\to\infty}\hat{g}_0=\int_0^1 g^2(x,S_0)\mathrm{d}x=\varsigma(S_0) \tag{8.14}$$

with $S_0 \equiv 0$.

Now we have to choose the sequence (h_n). Note that if we put in (7.5)

$$t_{m,j} = (g_0(\tilde{x}_m)/\sqrt{nh_n})\,\sqrt{y_j^*(R)} \quad \text{i.e.} \quad \kappa_{m,j} \;=\; y_j^*(R)\,, \tag{8.15}$$

we can rewrite inequality (8.10) as

$$n^{\frac{2k}{2k+1}}\tilde{\mathcal{R}}_0(\hat{S}_n^0) \;\geq\; \frac{(1-\nu)}{(1+\nu)}\,\frac{\hat{g}_0 J^*_{N_n}(R)}{2h_n n^{\frac{1}{2k+1}}}\,, \tag{8.16}$$

where

$$J_N^*(R) = N - \left(\sum_{j=1}^{N} j^k\right)^2 /(R + \sum_{j=1}^{N} j^{2k})\,.$$

It is clear that

$$k^2/(k+1)^2 \leq \liminf_{N\to\infty}\,\inf_{R>0}\,J_N^*(R)/N \leq \limsup_{N\to\infty}\,\sup_{R>0}\,J_N^*(R)/N \leq 1\,.$$

Therefore to obtain a positive finite asymptotic lower bound in (8.16) we have to take the parameter h_n as

$$h_n = h_* n^{-1/(2k+1)} N_n \tag{8.17}$$

with some positive coefficient h_*. Moreover, conditions (8.12)-(8.13) imply that

$$(1-\varepsilon)r\,\frac{2^{2k+1}}{\pi^{2k}}\,\frac{1}{\hat{g}_0}\,h_*^{2k+1} \geq \frac{1}{N_n^{k+1}}\sum_{j=1}^{N_n} j^k - \frac{1}{N_n^{2k+1}}\sum_{j=1}^{N_n} j^{2k}\,.$$

Taking here limit as $n \to \infty$ thanks to asymptotic equality (8.14), we obtain the following condition on h_*

$$h_* \geq (\upsilon_\varepsilon^*)^{1/(2k+1)}\,, \tag{8.18}$$

where

$$\upsilon_\varepsilon^* = \frac{k}{c_\varepsilon^*(k+1)(2k+1)} \quad \text{and} \quad c_\varepsilon^* = \frac{2^{2k+1}(1-\varepsilon)r}{\pi^{2k}\varsigma(S_0)}\,.$$

To maximize the function $J^*_{N_n}(R)$ at the right-hand side of inequality (8.16) we take $R = R^*_n$ defined in (8.13). Therefore we obtain that

$$\liminf_{n\to\infty} \inf_{\hat{S}^0_n} n^{2k/(2k+1)} \tilde{\mathcal{R}}_0(\hat{S}^0_n) \geq (\varsigma(S_0)/2)\, F(h_*)\,, \qquad (8.19)$$

where

$$F(x) = \frac{1}{x} - \frac{2k+1}{(k+1)^2(c^*_\varepsilon(2k+1)x^{2k+2} + x)}\,.$$

Taking into account that

$$F'(x) = -\frac{(c^*_\varepsilon(2k+1)(k+1)x^{2k+1} - k)^2}{(k+1)^2(c^*_\varepsilon(2k+1)x^{2k+2} + x)^2} \leq 0$$

we find that

$$\max_{h_* \geq (\upsilon^*_\varepsilon)^{1/(2k+1)}} F(h_*) = F((\upsilon^*_\varepsilon)^{1/(2k+1)}) = (k/(k+1))(\upsilon^*_\varepsilon)^{-1/(2k+1)}\,.$$

This means that to obtain in (8.19) the maximal lower bound we have to take in (8.17)

$$h_* = (\upsilon^*_\varepsilon)^{1/(2k+1)}\,. \qquad (8.20)$$

Therefore, inequality (8.19) implies

$$\liminf_{n\to\infty} \inf_{\hat{S}^0_n} n^{2k/(2k+1)} \tilde{\mathcal{R}}_0(\hat{S}^0_n) \geq (1-\varepsilon)^{1/(2k+1)}\, \gamma_k(S_0)\,, \qquad (8.21)$$

where the function $\gamma_k(S_0)$ is defined in (4.4) for $S_0 \equiv 0$.

Now to end the definition of the sequence of the random functions $(S_{\vartheta,n})$ defined by (7.4) and (7.5) we have to define the sequence (N_n). We remind that we make use of the sequence $(S_{\vartheta,n})$ with the coefficients $(t_{m,j})$ constructed in (8.15) for $R = R^*_n$ given in (8.13) and for the sequence h_n given by (8.17) and (8.20) for some fixed arbitrary $0 < \varepsilon < 1$.

We will choose the sequence (N_n) to satisfy conditions $\mathbf{A}_1)$–$\mathbf{A}_4)$. We can take, for example $N_n = [\ln^4 n] + 1$. Then condition $\mathbf{A}_1)$ is trivial. Moreover, taking into account that in this case

$$R_n^* = \frac{2^{2k+1}(1-\varepsilon)r}{\pi^{2k}\hat{g}_0}\upsilon_\varepsilon^* N_n^{2k+1} = \frac{\varsigma(S_0)}{\hat{g}_0}\frac{k}{(k+1)(2k+1)}N_n^{2k+1}$$

we find thanks to convergence (8.14)

$$\lim_{n\to\infty}(R_n^* + \sum_{j=1}^{N_n} j^{2k})/(N_n^k \sum_{j=1}^{N_n} j^k) = 1\,.$$

Therefore, solution (8.11) for sufficiently large n satifies the following inequality

$$\max_{1\le j\le N_n} y_j^*(R_n^*)\, j^k \le 2N_n^k\,.$$

Now it is easy to check conditions $\mathbf{A}_2)$ with $d_n = \sqrt{N_n}$ and $\mathbf{A}_4)$ for arbitrary $0 < \epsilon_0 < 1$. As to condition $\mathbf{A}_3)$, note that by definition of $t_{m,j}$ in (8.15) we have

$$\frac{1}{h_n^{2k-1}}\sum_{m=1}^{M_n}\sum_{j=1}^{N_n} t_{m,j}^2\, j^{2k} = \frac{1}{2nh_n^{2k+1}}\,\hat{g}_0\sum_{j=1}^{N_n} y_j^*(R_n^*)\, j^{2k}$$
$$= \frac{R_n^*\hat{g}_0}{N_n^{2k+1}2\upsilon_\varepsilon^*} = (1-\varepsilon)r\left(\frac{2}{\pi}\right)^{2k}\,.$$

Hence condition $\mathbf{A}_3)$.

Therefore Propositions 7.2-7.3 and limit (7.12) imply that for any $p>0$

$$\lim_{n\to\infty} n^p\,\varpi_n = 0\,.$$

Moreover, by condition $\mathbf{H}_4)$ the sequence γ_n^* goes to $\gamma_k(S_0)$ as $n\to\infty$. Therefore, from this, (8.21) and (8.4) we get for any $0<\varepsilon<1$

$$\liminf_{n\to\infty}\inf_{\hat{S}_n} n^{2k/(2k+1)}\,\mathcal{R}_0(\hat{S}_n)\ \ge\ (1-\varepsilon)^{1/(2k+1)}\,.$$

Limiting here $\varepsilon\to 0$ implies inequality (8.1). Hence Theorem 4.3. $\square$

9 Appendix

A.1 Properties of trigonometric basis

Lemma A.1. *For any function $S \in W_r^k$,*

$$\sup_{n\ge 1}\sup_{1\le m\le n-1} m^{2k}\left(\sum_{j=m+1}^{n}\theta_{j,n}^2\right) \le \frac{4r}{\pi^{2(k-1)}}\,. \qquad \text{(A.1)}$$

Lemma A.2. *For any $m \ge 0$,*

$$\sup_{N\ge 2}\sup_{x\in[0,1]} N^{-m}\left|\sum_{l=2}^{N} l^m \overline{\phi}_l(x)\right| \le 2^m\,, \qquad \text{(A.2)}$$

where $\overline{\phi}_l(x) = \phi_l^2(x) - 1$.

Proofs of Lemma A.1 and Lemma A.2 are given in [6].

Lemma A.3. *Let $\theta_{j,n}$ and θ_j be the Fourier coefficients defined in* (2.2) *and* (3.4) *respectively. Then, for $1 \le j \le n$ and $n \ge 2$,*

$$\sup_{S\in W_r^1} |\theta_{j,n} - \theta_j| \le 2\pi\sqrt{r}\, j/n\,. \qquad \text{(A.3)}$$

Proof. Indeed, we have

$$\begin{aligned}
|\theta_{j,n} - \theta_j| &= \left|\sum_{l=1}^{n}\int_{x_{l-1}}^{x_l} \left(S(x_l)\phi_j(x_l) - S(x)\phi_j(x)\right) \mathrm{d}x\right| \\
&\le n^{-1}\sum_{l=1}^{n}\int_{x_{l-1}}^{x_l}\left(|\dot{S}(z)\phi_j(z)| + |S(z)\dot{\phi}_j(z)|\right)\mathrm{d}z \\
&= n^{-1}\int_0^1 \left(|\dot{S}(z)|\,|\phi_j(z)| + |S(z)|\,|\dot{\phi}_j(z)|\right)\mathrm{d}z\,.
\end{aligned}$$

By making use of the Bounyakovskii-Cauchy-Schwarz inequality we get

$$\begin{aligned}
|\theta_{j,n} - \theta_j| &\le n^{-1}\left(\|\dot{S}\|\,\|\phi\| + \|\dot{\phi}\|\,\|S\|\right) \\
&\le n^{-1}\left(\|\dot{S}\| + \pi j\,\|S\|\right)\,.
\end{aligned}$$

The definition of class W_r^1 implies (A.3). Hence Lemma A.1. $\square$

A.2 Proof of Lemma 8.1

First notice that, for any $S \in W_r^k$, one has

$$\|\hat{S}_n - S\|_n^2 = \|T_n(\hat{S}) - S\|^2 + \Psi_n + \Delta_n \,,$$

where

$$\Psi_n = 2\sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} (\hat{S}_n(x_j) - S(x))(S(x) - S(x_j)) \mathrm{d}x$$

and

$$\Delta_n = \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} (S(x) - S(x_j))^2 \mathrm{d}x \,.$$

For any $0 < \delta < 1$, by making use of the elementary inequality

$$2ab \le \delta a^2 + \delta^{-1} b^2 \,,$$

one gets

$$\Psi_n \, \le \, \delta\|T_n(\hat{S}) - S\|^2 + \delta^{-1}\Delta_n \,.$$

Moreover, for any $S \in W_r^k$ with $k \ge 1$, by the Bounyakovskii-Cauchy-Schwarz inequality we obtain that

$$\Delta_n \le \frac{1}{n} \sum_{j=1}^{n} \int_{x_{j-1}}^{x_j} \dot{S}^2(t) \, \mathrm{d}t = \frac{1}{n^2} \|\dot{S}\|^2 \le \frac{r}{n^2} \,.$$

Hence Lemma 8.1. $\square$

A.3 Proof of (8.7)

First of all, note that Proposition 7.4, condition (3.7) and condition $\mathbf{H}_4$) imply that

$$\lim_{n\to\infty} \max_{1\le m\le M_n} \sup_{0\le x\le 1} \mathbf{1}_{\{|v_m(x)|\le 1\}} \mathbf{E} \left| g^{-2}(x, S_{\vartheta,n}) \, - \, g_0^{-2}(\tilde{x}_m) \right| \, = \, 0 \,. \qquad \text{(A.4)}$$

Let us show now that for any continuously differentiable function f on $[-1,1]$

$$\lim_{n\to\infty}\sup_{1\le m\le M_n}\left|\frac{1}{nh}\sum_{i=1}^{n} f(v_m(x_i))\mathbf{1}_{\{|v_m(x_i)|\le 1\}}-\int_{-1}^{1} f(v)\mathrm{d}v\right| = 0\,. \quad \text{(A.5)}$$

Indeed, setting

$$\Delta_{n,m}=\frac{1}{nh}\sum_{i=1}^{n} f(v_m(x_i))\mathbf{1}_{\{|v_m(x_i)|\le 1\}}-\int_{-1}^{1} f(v)\mathrm{d}v$$

we obtain that

$$\begin{aligned}\left|\Delta_{n,m}\right| &= \left|\frac{1}{nh}\sum_{i=i_*}^{i^*} f(v_m(x_i))-\int_{-1}^{1} f(v)\mathrm{d}v\right| \\ &\le \sum_{i=i_*}^{i^*}\int_{v_m(x_{i-1})}^{v_m(x_i)} |f(v_m(x_i))-f(z)|\mathrm{d}z+\max_{|z|\le 1}|f(z)|(2-v^*+v_*)\,.\end{aligned}$$

where $i_*=[n\tilde{x}_m-nh]+1$, $i_*=[n\tilde{x}_m+nh]$,

$$v_*=([n\tilde{x}_m-nh]+1-n\tilde{x}_m)/(nh) \quad \text{and} \quad v^*=([n\tilde{x}_m+nh]-n\tilde{x}_m)/(nh)\,.$$

Therefore, taking into accout that the derivative of the function f is bounded on the interval $[-1,1]$ we obtain that

$$\left|\Delta_{n,m}\right|\le 3\max_{|z|\le 1}|\dot{f}(z)|/(nh_n)+2\max_{|z|\le 1}|f(z)|/(nh_n)\,.$$

Taking into account the conditions on the sequence $(h_n)_{n\ge 1}$ given in $\mathbf{A}_1$) we obtain limiting equality (A.5) which together with (A.4) implies (8.7). □

A.4 Proof of (8.8)

Now we study the behaviour of $B_{m,j}$. Due to inequality (3.9) we obtain that

$$|\tilde{\mathbf{L}}_{m,j}(x,S_{\vartheta,n})|\le C^*\left(|S_{\vartheta,n}(x)D_{m,j}(x)|+|D_{m,j}|_1+\|S_{\vartheta,n}\|\,\|D_{m,j}\|\right)\,.$$

Note that

$$\mathbf{E}(S_{\vartheta,n}(x)D_{m,j}(x))^2 = \mathbf{E}\left(\sum_{l=1}^{N_n}\vartheta_{m,l}e_l(v_m(x))\right)^2 e_j^2(v_m(x))I_\eta^4(v_m(x))$$

$$\leq \sum_{l=1}^{N_n} t_{m,l}^2\,\mathbf{1}_{\{|v_m(x)|\leq 1\}} \leq (t_n^*)^2\mathbf{1}_{\{|v_m(x)|\leq 1\}}\,.$$

We remind that the sequence t_n^* is defined in (7.6). Therefore, property (A.5) implies

$$\max_{1\leq m\leq M_n}\ \max_{1\leq j\leq N_n}\ \frac{1}{nh}\sum_{i=1}^{n}\mathbf{E}(S_{\vartheta,n}(x_i)D_{m,j}(x_i))^2 = \mathrm{O}((t_n^*)^2)\,.$$

Moreover, as to the function $D_{m,j}(\cdot)$ we find that

$$|D_{m,j}|_1 = \int_0^1 |e_j(v_m(x))\,I_\eta(v_m(x))|\mathrm{d}x = h\int_{-1}^1 |e_j(v)\,I_\eta(v)|\mathrm{d}v \leq 2h\,.$$

Similarly we obtain $\|D_{m,j}\|^2 \leq h$.

Finally, by (7.18)we obtain that

$$\mathbf{E}\|S_{\vartheta,n})\|^2 \leq h\sum_{m=1}^{M_n}\sum_{j=1}^{N_n} t_{m,j}^2 \leq (t_n^*)^2\,.$$

Therefore,

$$\max_{1\leq m\leq M_n}\ \max_{1\leq j\leq N_n}\ B_{m,j}/(nh)\ =\ \mathrm{O}((t_n^*)^2 + h_n)$$

and condition $\mathbf{A}_1$) implies (8.8).

□

A.5 Proof of (8.9)

Indeed, by the direct calculation it easy to see that for any $N\geq 1$ and for any vector $(y_1,\ldots,y_N)'\in\mathbb{R}_+^N$

$$\left|\frac{\sum_{j=1}^N\ \tau_j(\eta,y_j)}{\sum_{j=1}^N\ \overline{\tau}(y_j)}\ -\ 1\right|\ \leq\ \frac{\max_{j\geq 1}\left(|\overline{e}_j^2(I_\eta)-\overline{e}_j(I_\eta^2)|+|\overline{e}_j^2(I_\eta)-1|\right)}{\min_{j\geq 1}\ \overline{e}_j(I_\eta^2)}\,,$$

where the operator $\overline{e}_j(f)$ is defined in in (8.5). Moreover, we remind that $\int_{-1}^{1} e_j^2(v)\mathrm{d}v = 1$. Therefore, taking into account property (7.2) we obtain (8.9). $\square$

References

[1] Brua, J.-Y. Asymptotic efficient estimators for non-parametric heteroscedastic model. *Preprint, University of Strasbourg* (2007) available at *http://hal.archives-ouvertes.fr/hal/-00178536/fr/*

[2] Efromovich, S. Sequential design and estimation in heteroscedastic nonparametric regression. *Sequential Analysis*, **26** 3-25 (2007)

[3] Efromovich, S. and Pinsker, M. Sharp - optimal and adaptive estimation for heteroscedastic nonparametric regression. *Statistica Sinica*, **6** 925-942 (1996)

[4] Galtchouk, L. and Pergamenshchikov, S. Asymptotically efficient estimates for nonparametric regression models. *Statistics & Probability Letters.* **76** 852-860 (2006)

[5] Galtchouk, L. and Pergamenshchikov, S. Efficient adaptive nonparametric estimation in heteroscedastic regression models. Preprint of the Strasbourg Louis Pasteur University, IRMA, 2005/020 (2005) available at *http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchtchikov*

[6] Galtchouk, L. and Pergamenshchikov, S. Adaptive nonparametric estimation in heteroscedastic regression models. Part 1. Sharp non-asymptotic oracle inequalities. Preprint of the Strasbourg

Louis Pasteur University, IRMA, 2007/09 (2007), available at *http://hal.archives-ouvertes.fr/hal/-00179856/fr/*

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

[8] Goldfeld, S.M. and Quandt, R.E. *Nonlinear Methods in Econometrics.* North-Holland, London (1972)

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

[10] Nussbaum, M. Spline smothing in regression models and asymptotic efficiency in $\mathbf{L}_2$. *Ann. Statist.* **13** 984-997 (1985)

[11] Rohde, A. On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statistics & Decisions* **22** 235-243 (2004)

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

L.Galtchouk
Department of Mathematics
Strasbourg University
7, rue Rene Descartes
67084, Strasbourg, France
e-mail: galtchou@math.u-strasbg.fr

S. Pergamenshchikov
Laboratoire de Mathématiques Raphael Salem,
Avenue de l'Université, BP. 12,
Université de Rouen,
F76801, Saint Etienne du Rouvray, Cedex France.
Serge.Pergamenchtchikov@univ-rouen.fr