APPLICATIONES MATHEMATICAE
27,4 (2000), pp. 445–454

W. POPIŃSKI (Warszawa)

CONVERGENCE RATES OF ORTHOGONAL SERIES REGRESSION ESTIMATORS

Abstract. General conditions for convergence rates of nonparametric orthogonal series estimators of the regression function $f(x) = E(Y \mid X = x)$ are considered. The estimators are obtained by the least squares method on the basis of a random observation sample (Y_i, X_i), $i = 1, \ldots, n$, where $X_i \in A \subset \mathbb{R}^d$ have marginal distribution with density $\varrho \in L^1(A)$ and $\mathrm{Var}(Y \mid X = x)$ is bounded on A. Convergence rates of the errors $E_X(f(X) - \widehat{f}_N(X))^2$ and $\|f - \widehat{f}_N\|_\infty$ for the estimator $\widehat{f}_N(x) = \sum_{k=1}^N \widehat{c}_k e_k(x)$, constructed using an orthonormal system e_k, $k = 1, 2, \ldots$, in $L^2(A)$, are obtained.

1. Introduction. Let (Y_i, X_i), $i = 1, \ldots, n$, be a random sample of size n from the distribution of (X, Y), where X represents a predictor variable and Y a real-valued response variable. We assume that X ranges over a compact subset A of some euclidean space $\mathbb{R}^d$, $d \geq 1$, and has absolutely continuous distribution with density $\varrho \in L^1(A)$. Set $f(x) = E(Y \mid X = x)$, $\sigma^2(x) = \mathrm{Var}(Y \mid X = x)$ and assume that $\sigma^2(x) \leq C$ for $x \in A$ and the function f can be uniformly approximated on this set by finite linear combinations of functions e_k, $k = 1, 2, \ldots$, forming a complete orthonormal system in $L^2(A)$. We consider the problem of estimating the regression function f using series estimators of the form

$$\widehat{f}_N(x) = \sum_{k=1}^{N} \widehat{c}_{kN} e_k(x),$$

where the vector of coefficient estimators $\widehat{c}_N = (\widehat{c}_{1N}, \ldots, \widehat{c}_{NN})^T$ is, for a fixed N, obtained by the least squares method, i.e.

2000 *Mathematics Subject Classification*: 62G08, 62G20.
Key words and phrases: nonparametric series regression, least squares method, orthonormal system, convergence rate.

$$\widehat{c}_N = \arg\min_{a\in\mathbb{R}^N} \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle a, e^N(X_i)\rangle)^2,$$

where $e^N(x) = (e_1(x), \ldots, e_N(x))^T$.

The vector $\widehat{c}_N$ can be obtained as a solution of the normal equations

$$G_n\widehat{c}_N = g_n, \tag{1}$$

where

$$G_n = \frac{1}{n}\sum_{i=1}^{n} e^N(X_i)e^N(X_i)^T, \qquad g_n = \frac{1}{n}\sum_{i=1}^{n} Y_i e^N(X_i),$$

and when $\det G_n \neq 0$ it is uniquely determined. Set $G(N) = EG_n$ and observe that for any vector $v = (v_1, \ldots, v_N)^T \in \mathbb{R}^N$ the following equality is true:

$$\langle G(N)v, v\rangle = \langle Ee^N(X)e^N(X)^T v, v\rangle = \sum_{k=1}^{N}\sum_{l=1}^{N} v_k v_l \int_A e_k(x)e_l(x)\varrho(x)\,dx$$
$$= \int_A \Big(\sum_{k=1}^{N} v_k e_k(x)\Big)^2 \varrho(x)\,dx.$$

If we assume that $\varrho \geq c > 0$, then in view of orthogonality of the functions e_k, $k = 1, 2, \ldots$, the above equality implies that $\lambda(N) \geq c$, where $\lambda(N)$ denotes the minimal eigenvalue of the matrix $G(N)$, and consequently $G(N)$ is nonsingular.

Let us also note that in the case when $\det G_n \neq 0$ the estimator $\widehat{f}_N$ is invariant under nonsingular linear transformations of $e^N(x)$, i.e. it does not change when we use the vector function $h^N(x) = Be^N(x)$, where B is a nonsingular matrix, instead of $e^N(x)$ for constructing it. In consequence, in the case when $\varrho \geq c > 0$ and $\det G_n \neq 0$ the series estimator considered can be represented in the form $\widehat{f}_N(x) = \langle h^N(x), \widehat{b}_N\rangle$, where $h^N(x) = G(N)^{-1/2}e^N(x)$ and the vector of coefficient estimators $\widehat{b}_N = G(N)^{1/2}\widehat{c}_N$ is determined by the least squares method. Such a representation of $\widehat{f}_N$ is convenient, since then $Eh^N(X)h^N(X)^T = I_N$ and consequently if $h^N(x) = (h_{1N}(x), \ldots, h_{NN}(x))^T$, then the functions $h_{kN}(x), k = 1, \ldots, N$, are orthonormal with weight ϱ, i.e. $\int_A h_{kN}(x)h_{lN}(x)\varrho(x)\,dx = \delta_{kl}$.

Moreover, if $H_n = (1/n)\sum_{i=1}^{n} h^N(X_i)h^N(X_i)^T$ is the matrix of normal equations corresponding to the vector function $h^N(x)$, then

$$E\|H_n - I\|^2 = \sum_{k=1}^{N}\sum_{l=1}^{N} E\Big(\frac{1}{n}\sum_{i=1}^{n} h_{kN}(X_i)h_{lN}(X_i) - \delta_{kl}\Big)^2$$
$$\leq \sum_{k=1}^{N}\sum_{l=1}^{N} \frac{1}{n} Eh_{kN}^2(X)h_{lN}^2(X) = \frac{1}{n}E\|h^N(X)\|^2\|h^N(X)\|^2$$

$$\le \frac{1}{n}\|h^N\|_\infty^2 E\|h^N(X)\|^2 = \frac{N}{n}\|h^N\|_\infty^2,$$

where $\|h^N\|_\infty = \sup_{s\in A}\|h^N(s)\|$, and since $h^N(x) = G(N)^{-1/2}e^N(x)$ we also have

$$\|h^N\|_\infty^2 \le \lambda(N)^{-1}\|e^N\|_\infty^2 \le \|e^N\|_\infty^2 c^{-1}, \tag{2}$$

so finally

$$E\|H_n - I\|^2 \le \frac{N}{cn}\|e^N\|_\infty^2.$$

If we put $M_N = \|e^N\|_\infty$ and λ_n denotes the smallest eigenvalue of the matrix H_n, then since $|\lambda_n - 1| \le \|H_n - I\|$, we also have

$$E|\lambda_n - 1|^2 \le \frac{NM_N^2}{cn}, \tag{3}$$

which implies that $\lambda_n \to 1$ in probability on condition that $NM_N^2/n \to 0$ as $n \to \infty$. This fact will be used to prove the results presented below. Let us also note that for $\varrho \ge c > 0$, by (3) we have $P(\det G_n = 0) = P(\det H_n = 0) \le P(\lambda_n < 1/2) \le 4NM_N^2 c^{-1}/n$. Thus, the conditions $\varrho \ge c > 0$ and $NM_N^2/n \to 0$ as $n \to \infty$ assure that $P(\det G_n = 0) \to 0$, i.e. the estimator is uniquely determined with growing probability.

In this work we continue the investigations of [8]–[10] on asymptotic properties of series regression estimators by considering convergence rates of the errors $E_X(f(X) - \widehat{f}_N(X))^2$ and $\|f - \widehat{f}_N\|_\infty$. We give sufficient conditions for the convergence rates and extend the results of [9], [10] where only convergence in probability of such errors for trigonometric and polynomial estimators was investigated. Huang [4] has recently obtained general conditions for convergence rates of $E_X(f(X) - \widehat{f}_N(X))^2$ for the relevant series estimators, under the assumption that $D \ge \varrho \ge c > 0$, but the estimator measurability and uniqueness conditions are not discussed there. In the present work it is shown that the boundedness condition imposed on the density ϱ can be relaxed.

Asymptotic properties of other nonparametric series regression estimators for similar observation models were investigated in the works of Lugosi and Zeger [6] and Györfi and Walk [3] but the results obtained there concern the universal consistency of their estimators and the problem of convergence rates is only briefly discussed in [6]. The series estimator considered in [6] is obtained via constrained empirical risk minimization (whereas our estimators are obtained by unconstrained empirical risk minimization) and the one considered in [3] is based on stochastic approximation procedure in a function space. Convergence rates of $E_X(f(X) - \widetilde{f}_N(X))^2$ for estimators $\widetilde{f}_N$ constructed using radial basis functions and neural networks are investigated in [7]. Properties of series estimators in the case of other observation models are investigated in [13].

2. Convergence rates of the L^2-error. Let χ_n denote the indicator function of $\{\lambda_n \geq 1/2\}$. According to (3) we have $P(\lambda_n < 1/2) \leq 4NM_N^2 c^{-1}/n$ and the assumption $NM_N^2/n \to 0$ as $n \to \infty$ implies that $P(\chi_n \neq 1) \to 0$.

We need the following lemmas. Only the proof of the second lemma is given since the first one is proved in [11].

Lemma 2.1. *Let* $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ *and*

$$H_n = \frac{1}{n}\sum_{i=1}^n h^N(X_i)h^N(X_i)^T.$$

Then

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n v_i v_j h^N(X_i)^T H_n^{-1} h^N(X_j) \leq \frac{1}{n}\sum_{i=1}^n v_i^2.$$

Lemma 2.2. *Assume that* $\varrho \geq c > 0$ *and for* $N > 0$ *there exist* $f_N \in \operatorname{span}\{e_1, \ldots, e_N\}$ *such that* $\|f - f_N\|_\infty = O(N^{-\alpha})$ *as* $N \to \infty$, *where* $\alpha > 0$. *If* $f_N(x) = \langle h^N(x), b_N\rangle$ *and* $\widehat{b}_N = H_n^{-1} h_n$, *where* $h_n = (1/n)\sum_{i=1}^n Y_i h^N(X_i)$, *then*

$$E\chi_n\|\widehat{b}_N - b_N\|^2 = O(N/n + N^{-2\alpha}).$$

Proof. Putting $\eta_i = Y_i - f(X_i)$, $i = 1, \ldots, n$, and $f(x) = f_N(x) + r_N(x)$, we have

$$\widehat{b}_N = b_N + H_n^{-1}\Big(\frac{1}{n}\sum_{i=1}^n r_N(X_i)h^N(X_i)\Big) + H_n^{-1}\Big(\frac{1}{n}\sum_{i=1}^n \eta_i h^N(X_i)\Big). \tag{4}$$

Now, putting $D_n = (X_1, \ldots, X_n)$, since $\operatorname{Var}(Y \mid X = X_i) \leq C$, $i = 1, \ldots, n$, we easily obtain

$$\begin{aligned} E\Big[\chi_n\Big\|H_n^{-1/2}\frac{1}{n}\sum_{i=1}^n \eta_i h^N(X_i)\Big\|^2 \;\Big|\; D_n\Big] &= \chi_n E\Big[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \eta_i\eta_j h^N(X_i)^T H_n^{-1} h^N(X_j) \;\Big|\; D_n\Big] \\ &\leq \chi_n \frac{C}{n^2}\sum_{i=1}^n h^N(X_i)^T H_n^{-1} h^N(X_i) \\ &= \chi_n \frac{C}{n^2}\operatorname{Tr}\Big(\sum_{i=1}^n h^N(X_i)h^N(X_i)^T H_n^{-1}\Big) \leq \frac{CN}{n}. \end{aligned}$$

The last inequality implies

$$E\chi_n\Big\|H_n^{-1/2}\frac{1}{n}\sum_{i=1}^n \eta_i h^N(X_i)\Big\|^2 \leq CN/n,$$

which further yields

$$E\chi_n\left\|H_n^{-1}\frac{1}{n}\sum_{i=1}^n \eta_i h^N(X_i)\right\|^2$$

$$= E\chi_n\left(\frac{1}{n^2}\sum_{i=1}^n \eta_i h^N(X_i)^T H_n^{-1/2}H_n^{-1}H_n^{-1/2}\sum_{j=1}^n \eta_j h^N(X_j)\right)$$

$$\le E\chi_n\lambda_n^{-1}\left\|H_n^{-1/2}\frac{1}{n}\sum_{i=1}^n \eta_i h^N(X_i)\right\|^2 \le 2CN/n,$$

since $\chi_n\lambda_n^{-1} \le 2$. In view of Lemma 2.1 it is also easy to see that

$$E\chi_n\left\|H_n^{-1}\frac{1}{n}\sum_{i=1}^n r_N(X_i)h^N(X_i)\right\|^2$$

$$= E\chi_n\left(\frac{1}{n^2}\sum_{i=1}^n r_N(X_i)h^N(X_i)^T H_n^{-1/2}H_n^{-1}H_n^{-1/2}\sum_{j=1}^n r_N(X_j)h^N(X_j)\right)$$

$$\le E\chi_n\lambda_n^{-1}\left(\frac{1}{n^2}\sum_{i=1}^n r_N(X_i)h^N(X_i)^T H_n^{-1}\sum_{j=1}^n r_N(X_j)h^N(X_j)\right)$$

$$\le E\chi_n\lambda_n^{-1}\left(\frac{1}{n}\sum_{i=1}^n r_N^2(X_i)\right) \le E\chi_n\lambda_n^{-1}\max_{1\le i\le n}|r_N^2(X_i)| = O(N^{-2\alpha}).$$

The above bounds together with (4) imply the assertion of the lemma. ∎

In the case when the regression function f is square-integrable and the density ϱ satisfies the additional condition $D \ge \varrho \ge c > 0$ we also have the following lemma.

LEMMA 2.3. *Assume that* $D \ge \varrho \ge c > 0$, $f \in L^2(A)$ *and* $f_N(x) = \langle h^N(x), b_N\rangle$ *is its orthogonal projection on* $\operatorname{span}\{e_1,\ldots,e_N\}$. *If* $\widehat{b}_N = H_n^{-1}h_n$, *then*

$$E\chi_n\|\widehat{b}_N - b_N\|^2 = O(N/n + \|f-f_N\|^2).$$

Proof. First observe that for $r_N = f - f_N$,

$$E\left(\frac{1}{n}\sum_{i=1}^n r_N^2(X_i)\right) = Er_N^2(X) \le D\|f-f_N\|^2,$$

and then follow the proof of Lemma 2.2. ∎

Now, we can prove the following theorem on convergence rates of the estimators considered.

THEOREM 2.1. *Assume that $\varrho \ge c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and for $N > 0$ there exist $f_N \in \operatorname{span}\{e_1, \ldots, e_N\}$ such that $\|f - f_N\|_\infty = O(N^{-\alpha})$ as $N \to \infty$, where $\alpha > 0$. Then the orthogonal series estimator $\widehat{f}_{N(n)}$ satisfies

$$\int_A (f - \widehat{f}_{N(n)})^2 \varrho = O_p(N(n)/n + N(n)^{-2\alpha}).$$

Proof. Putting $f_N(x) = \langle h^N(x), b_N\rangle$, we easily obtain, by the triangle inequality and the equality $Eh^N(X)h^N(X)^T = I_N$,

$$\begin{aligned}\chi_n \int_A (f - \widehat{f}_N)^2 \varrho &= \chi_n \int_A (f(x) - f_N(x) + \langle h^N(x), b_N - \widehat{b}_N\rangle)^2 \varrho(x)\,dx \\ &\le 2\chi_n \int_A (f(x) - f_N(x))^2 \varrho(x)\,dx + 2\chi_n \|b_N - \widehat{b}_N\|^2 \\ &\le O(N^{-2\alpha}) + 2\chi_n \|b_N - \widehat{b}_N\|^2.\end{aligned}$$

As remarked earlier, the assumption $N(n)M_{N(n)}^2/n \to 0$ assures that $P(\chi_n \ne 1) \to 0$ so the assertion follows by applying Lemma 2.2. ■

In the case when the regression function can be approximated in the mean-square sense the following theorem holds.

THEOREM 2.2. *Assume that $D \ge \varrho \ge c > 0$, the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and f_N is the orthogonal projection of the regression function $f \in L^2(A)$ on $\operatorname{span}\{e_1, \ldots, e_N\}$. Then the orthogonal series estimator $\widehat{f}_{N(n)}$ satisfies

$$\int_A (f - \widehat{f}_{N(n)})^2 \varrho = O_p(N(n)/n + \|f - f_{N(n)}\|^2).$$

Proof. Observe that $\int_A (f - f_N)^2 \varrho \le D\|f - f_N\|^2$ and follow the proof of Theorem 2.1 using Lemma 2.3 instead of Lemma 2.2. In fact we can even prove that $E\chi_n \int_A (f - \widehat{f}_{N(n)})^2 \varrho = O(N(n)/n + \|f - f_{N(n)}\|^2)$. ■

Since for $f \in L^2(A)$ we have $\|f - f_N\|^2 \to 0$ as $N \to \infty$, under the assumptions of Theorem 2.2, $E_X(f(X) - \widehat{f}_{N(n)}(X))^2 = o_p(1)$ and consequently also $\|f - \widehat{f}_{N(n)}\|^2 = o_p(1)$. Moreover, if $\|f - f_N\| = O(N^{-\alpha})$, where $\alpha > 0$, Theorem 2.2 allows one to obtain convergence rates of the error

$E_X(f(X)-\widehat{f}_{N(n)}(X))^2$. The term N/n in the above formulae for convergence rates essentially corresponds to a variance term, and $N^{-2\alpha}$ to a bias term. When $N(n)$ is chosen so that these two terms go to zero at the same rate, which occurs when $N(n)\sim n^{1/(1+2\alpha)}$ (i.e. $r_1\ge N(n)n^{-1/(1+2\alpha)}\ge r_2$, $r_1,r_2>0$), the convergence rate of the estimators will be $n^{-2\alpha/(1+2\alpha)}$. Thus, Stone's [12] bound on the best obtainable rate is attained.

Let us also remark that the error $E_X(f(X)-\widehat{f}_N(X))^2$, for which convergence rates were obtained above, is related to the prediction error $E(Y-\widehat{f}_N(X))^2=E_X(f(X)-\widehat{f}_N(X))^2+E\sigma^2(X)$.

It is easy to see that for the estimator $\widetilde{f}_N$ defined by the formula

$$\widetilde{f}_N(x)=\begin{cases}\widehat{f}_N(x) & \text{if } \lambda_n\ge 1/2,\\ 0 & \text{otherwise,}\end{cases}$$

Theorems 2.1 and 2.2 are true, but we also have the following result concerning its IMSE.

THEOREM 2.3. *Assume that* $D\ge\varrho\ge c>0$, *the sequence of natural numbers* $N(n)$, $n=1,2,\ldots$, *satisfies*

$$\lim_{n\to\infty}N(n)=\infty,\qquad \lim_{n\to\infty}\frac{N(n)M^2_{N(n)}}{n}=0,$$

and f_N *is the orthogonal projection of the regression function* $f\in L^2(A)$ *on* $\operatorname{span}\{e_1,\ldots,e_N\}$. *Then the orthogonal series estimator* $\widetilde{f}_{N(n)}$ *satisfies*

$$E\int_A(f-\widetilde{f}_{N(n)})^2=O(N(n)/n+\|f-f_{N(n)}\|^2+N(n)M^2_{N(n)}/n).$$

Proof. According to the definition of $\widetilde{f}_N$, we have

$$\begin{aligned}E\int_A(f-\widetilde{f}_N)^2&=E\chi_n\int_A(f-\widehat{f}_N)^2+E(1-\chi_n)\int_A f^2\\&\le c^{-1}E\chi_n\int_A(f-\widehat{f}_N)^2\varrho+P(\lambda_n<1/2)\|f\|^2.\end{aligned}$$

As remarked in the proof of Theorem 2.2, the first term on the right hand side is $O(N(n)/n+\|f-f_{N(n)}\|^2)$ and the second term is bounded by $4\|f\|^2NM_N^2c^{-1}/n$, which completes the proof. ∎

Since λ_n is the minimal eigenvalue of the matrix H_n which is not used in computations we cannot verify directly whether the condition $\lambda_n\ge 1/2$ is satisfied. However, Theorem 2.3 allows us to learn about IMSE convergence rates of the estimator. Namely, for many orthonormal systems it is possible to obtain a bound of the form

$$M_N^2=\|e^N\|_\infty^2=\sup_{x\in A}\sum_{k=1}^N e_k^2(x)\le KN, \tag{5}$$

where K is a constant [1]. This clearly holds for uniformly bounded systems (e.g. the trigonometric system in $L^2([0,2\pi]^d)$) but also for strongly localized systems (e.g. the splines, piecewise polynomials) and leveled localized systems (e.g. compactly supported wavelets in $L^2[0,1]$) [1], [13]. For such systems the condition $N(n)M_{N(n)}^2/n \to 0$ is satisfied if $N(n)^2/n \to 0$ as $n \to \infty$ and under the assumptions of Theorem 2.3 we have $E\|f - \widetilde{f}_{N(n)}\|^2 = O(\|f - f_{N(n)}\|^2 + N(n)^2/n)$. Assuming further that $\|f - f_N\| = O(N^{-\alpha})$ we see that for $N(n) \sim n^{1/(2+2\alpha)}$ the IMSE convergence rate of $\widetilde{f}_N$ is $n^{-\alpha/(\alpha+1)}$.

3. Uniform convergence rates. In this section a result on uniform convergence rates of orthogonal series estimators is proved. It extends the results of [9] where only uniform pointwise consistency of trigonometric and polynomial estimators was examined.

THEOREM 3.1. *Assume that* $\varrho \geq c > 0$, *the sequence of natural numbers* $N(n)$, $n = 1, 2, \ldots,$ *satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} \frac{N(n)M_{N(n)}^2}{n} = 0,$$

and for $N > 0$ *there exist* $f_N \in \operatorname{span}\{e_1, \ldots, e_N\}$ *such that* $\|f - f_N\|_\infty = O(N^{-\alpha})$ *as* $N \to \infty$, *where* $\alpha > 0$. *Then the orthogonal series estimator* $\widehat{f}_{N(n)}$ *satisfies*

$$\|f - \widehat{f}_{N(n)}\|_\infty = O_p(M_{N(n)}(N(n)^{1/2}/n^{1/2} + N(n)^{-\alpha})).$$

Proof. Putting $f_N(x) = \langle h^N(x), b_N \rangle$, by the triangle and Cauchy inequalities we have

$$\begin{aligned}\chi_n|f(x) - \widehat{f}_N(x)| &\leq \chi_n|f(x) - f_N(x)| + \chi_n|\langle h^N(x), b_N - \widehat{b}_N\rangle| \\ &\leq O(N^{-\alpha}) + \chi_n\|h^N(x)\| \cdot \|b_N - \widehat{b}_N\|\end{aligned}$$

for $x \in A$, and according to (2) we further obtain

$$\chi_n\|f - \widehat{f}_N\|_\infty \leq O(N^{-\alpha}) + \chi_n\|b_N - \widehat{b}_N\|M_N c^{-1/2}.$$

Since the condition $N(n)M_{N(n)}^2/n \to 0$ as $n \to \infty$ implies $P(\chi_n \neq 1) \to 0$ we get the assertion by Lemma 2.2. ∎

The uniform convergence rates for the above series estimators do not attain Stone's [12] bound on the best obtainable rate for the uniform error but they improve on some rates obtained earlier, e.g. on those of Cox [2].

4. Conclusions. As proved in [8], if we use orthogonal systems of analytic functions to construct estimators of regression functions (e.g. trigonometric functions or multivariate polynomials), the normal equations matrix

G_n is almost surely positive definite for $N \le n$ for any density ϱ. Thus, in that case the estimators are uniquely defined with probability one.

Let $\beta = (\beta_1, \ldots, \beta_d)$ be a vector of nonnegative integers, and $|\beta| = \sum_{k=1}^d \beta_k$. For a vector function $f(x) = (f_1(x), \ldots, f_N(x))^T$ define the vector of partial derivatives $\partial^{|\beta|} f(x) = \partial^{|\beta|} f(x)/\partial x_1^{\beta_1} \ldots \partial x_d^{\beta_d}$ and for any nonnegative integer r let $\|f\|_r = \max_{|\beta| \le r} \sup_{x\in A} \|\partial^{|\beta|} f(x)\|$.

Now put $P_r(N) = \max_{|\beta|\le r} \sup_{x\in A} \|\partial^{|\beta|} e^N(x)\|$ and assume that $\varrho \in L^1(A)$, $\varrho \ge c > 0$, and for $N > 0$ there exists a vector b_N such that $\|f - b_N^T e^N\|_r i = O(N^{-\alpha})$, where $\alpha > 0$. Under the above assumptions, using the same technique as in the proof of Theorem 3.1 we can prove that $\|f - \widehat{f}_{N(n)}\|_r = O_p(P_r(N(n))(N(n)^{1/2}/n^{1/2} + N(n)^{-\alpha}))$ if the natural numbers $N(n)$ satisfy $\lim_{n\to\infty} N(n) = \infty$ and $\lim_{n\to\infty} N(n)P_0^2(N(n))/n = 0$.

The exponent α defining the decrease rate of the uniform approximation error of the regression function and its derivatives up to order r is related not only to the smoothness of the regression function but also to the dimensionality of X and the size of r. For example, if f is continuously differentiable of order s on $[-1,1]^d$, then in the case of polynomial approximation and $r = 0$ we have $\alpha = s/d$ according to Lorentz [5]. It is much more difficult to find in the literature a corresponding result for the case when $r > 0$, except in two cases. Namely, when X is univariate it is well known that $\alpha = s - r$ (see [5]), and when f is analytic it is known that for any r the assumption of uniform approximation rate in the norm $\| * \|_r$ will hold with α equal to an arbitrarily large positive number.

Now consider the observation model $Y_i = f(X_i) + \eta_i$, $i = 1, \ldots, n$, where the η_i are realizations of some strictly stationary β mixing process, centered in expectation, with β mixing sequence $(\beta_k)_{k\ge 0}$ satisfying the condition $\sum_{k\ge 0} \beta_k < \infty$ (see [13]). In that case in view of the inequalities $\chi_n \lambda_n^{-1} \le 2$ and $\|h_N\|_\infty \le M_N/\sqrt{c}$ we have

$$\begin{aligned}
E\chi_n \bigg\| H_n^{-1} \frac{1}{n} \sum_{i=1}^n \eta_i h^N(X_i) \bigg\|^2 &\le E_X E_\eta \chi_n \lambda_n^{-2} \bigg\| \frac{1}{n} \sum_{i=1}^n \eta_i h^N(X_i) \bigg\|^2 \\
&\le 4 E_X \chi_n \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E_\eta \eta_i \eta_j \langle h^N(X_i), h^N(X_j) \rangle \\
&\le 4 E_X \chi_n \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |\mathrm{cov}(\eta_i, \eta_j)| \cdot \|h^N(X_i)\| \cdot \|h^N(X_j)\| \\
&\le \frac{4M_N^2}{n^2 c} \sum_{i=1}^n \sum_{j=1}^n |\mathrm{cov}(\eta_i, \eta_j)| \le \frac{8M_N^2}{nc} \sum_{i=1}^n |\mathrm{cov}(\eta_i, \eta_1)| = O\bigg(\frac{M_N^2}{n}\bigg),
\end{aligned}$$

since for β mixing processes satisfying the imposed conditions the sums

$\sum_{i=1}^{n} |\mathrm{cov}(\eta_i, \eta_1)|$ are bounded by a constant (see Theorem 2.2 in [13] for details). Thus, for such observation models we obtain, as in Lemma 2.2, the bound $E\chi_n \|\widehat{b}_N - b_N\|^2 = O(M_N^2/n + N^{-2\alpha})$, which implies that for orthogonal systems satisfying (5) the convergence rates of the relevant errors are the same as in the case of independent observations.

References

[1] L. Birgé and P. Massart, *Minimum contrast estimators on sieves: exponential bounds and rates of convergence*, J. Bernoulli Soc. 4 (1998), 329–375.

[2] D. D. Cox, *Approximation of least squares regression on nested subspaces*, Ann. Statist. 16 (1988), 713–732.

[3] L. Györfi and H. Walk, *On the strong universal consistency of a series type regression estimate*, Math. Methods Statist. 5 (1996), 332–342.

[4] J. Z. Huang, *Projection estimation in multiple regression with application to functional ANOVA models*, Ann. Statist. 26 (1998), 242–272.

[5] G. G. Lorentz, *Approximation of Functions*, Holt, Reinehart & Winston, New York, 1966.

[6] G. Lugosi and K. Zeger, *Nonparametric estimation via empirical risk minimization*, IEEE Trans. Inform. Theory IT-41 (3) (1995), 677–687.

[7] P. Niyogi and F. Girosi, *Generalization bounds for function approximation from scattered noisy data*, Adv. Comput. Math. 10 (1999), 51–80.

[8] W. Popiński, *On least squares estimation of Fourier coefficients, and of the regression function*, Appl. Math. (Warsaw) 22 (1993), 91–102.

[9] —, *Consistency of trigonometric and polynomial regression estimators*, ibid. 25 (1998), 73–83.

[10] —, *A note on orthogonal series regression function estimators*, ibid. 26 (1999), 281–291.

[11] E. Rafajłowicz, *Nonparametric least-squares estimation of a regression function*, Statistics 19 (1988), 349–358.

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

[13] G. Viennet, *Least-square estimation for regression on random design for absolutely regular observations*, Statist. Probab. Lett. 43 (1999), 13–23.

Waldemar Popiński
Department of Survey Organization
Central Statistical Office
Al. Niepodległości 208
00-925 Warszawa, Poland
E-mail: w.popinski@stat.gov.pl

Received on 23.12.1999;
revised version on 31.7.2000