On global and pointwise adaptive estimation

SAM EFROMOVICH

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA. e-mail: efrom@math.unm.edu.

Let an estimated function belong to a Lipschitz class of order α. Consider a minimax approach where the infimum is taken over all possible estimators and the supremum is taken over the considered class of estimated functions. It is known that, if the order α is unknown, then the minimax mean squared (pointwise) error convergence slows down from $n^{-2\alpha/(2\alpha+1)}$ for the case of the given α to $[n/\ln(n)]^{-2\alpha/(2\alpha+1)}$. At the same time, the minimax mean integrated squared (global) error convergence is proportional to $n^{-2\alpha/(2\alpha+1)}$ for the cases of known and unknown α. We show that a similar phenomenon holds for analytic functions where the lack of knowledge of the maximal set to which the function can be analytically continued leads to the loss of a sharp constant. Surprisingly, for the more general adaptive minimax setting where we consider the union of a range of Lipschitz and a range of analytic functions neither pointwise error convergence nor global error convergence suffers an additional slowing down.

Keywords: analytic and Lipschitz functions; efficiency; mean integrated squared error; mean squared error

1. Introduction

It is known that for the problems of filtering, density estimation or nonparametric regression, when the estimated function belongs to a Lipschitz class of a given order or a class of analytic functions admitting an analytic continuation onto a set of the complex plane, the minimax mean squared (pointwise) error and the minimax mean integrated squared (global) error have the same rate of convergence as the sample size tends to infinity. Moreover, for the case of analytic functions, even sharp constants are the same (Ibragimov and Khasminskii 1981, 1984; Golubev and Levit 1996; Efromovich 1994).

However, it is also well known that for the case of a range of Lipschitz classes the convergence of the minimax pointwise error slows down up to a logarithmic factor of the sample size while the convergence of the minimax global error remains the same (Efromovich and Pinsker 1984; Lepskii 1990; Brown and Low 1996a; Birgé and Massart 1996).

The focus of this paper is twofold. First, we would like to explore the same adaptive phenomenon for analytic functions. Second, we would like to understand what kind of

penalties we are to pay for the natural case where the estimated functions belong to the union of Lipschitz and analytic functions.

We show that the former problem resembles the case of Lipschitz functions but here we lose only a sharp constant (and not rate!) of the minimax pointwise error convergence. The solution of the latter problem reveals that there are no additional penalties for consideration of this union. This is slightly surprising because the known adaptive lower bounds of Lepskii and Spokoiny (1994) indicate such a possibility. The suggested estimator is also of practical interest because a unique and very promising Lepskii adaptive procedure is based on the assumption that a range of the orders α of Lipschitz classes is finite and given (Lepskii 1990; Lepskii and Spokoiny 1994). Therefore, the suggested adaptive procedure shows how to relax this assumption.

The scope of the paper is as follows. The statistical framework and notation are given in Section 2. Section 3 is devoted to adaptive estimation of analytic functions. The case of the union of Lipschitz and analytic classes is explored in Section 4. All proofs are deferred to the Appendix.

2. Statistical framework and notation

The following model of filtering a signal from Gaussian white noise can be considered as a canonical example for many other statistical models including nonparametric regression and density estimation (Brown and Low 1996b; Nussbaum 1996). On the interval [0, 1] we observe the random processes $Y_n(t)$ having stochastic differential

$$dY_n(t) = f(t)\,dt + n^{-1/2}\,dW(t), \qquad 0 \leqslant t \leqslant 1, \qquad (2.1)$$

where $W(t)$ is a standard Brownian motion. Thus for brevity we restrict our attention only to this model.

The following notation will be used throughout the paper. The inner product and norm in $L_2([0, 1])$ we denote by $\langle f, g \rangle$ and $\|f\|$, i.e., $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$ and $\|f\| = \langle f, f \rangle^{1/2}$.

Here we use an orthogonal series approach and therefore, to make our approach concrete, we employ the classical trigonometric basis in $L_2[0, 1]$: $\varphi_0(x) = 1$, $\varphi_{2j-1}(x) = 2^{1/2}\sin(2\pi jx)$ and $\varphi_{2j}(x) = 2^{1/2}\cos(2\pi jx)$, $j = 1, 2, \ldots$. We refer to $\theta_j = \langle f, \varphi_j \rangle$ as the jth Fourier coefficient of f. The parametric estimator $\hat{\theta}_j = \int_0^1 \varphi_j(t)\,dY_n(t) = \theta_j + n^{-1/2}\xi_j$, where $\xi_0, \xi_1, \ldots$ are independently and identically distributed (i.i.d.) standard normal random variables, will be used for estimating θ_j. By N_γ we denote the minimal even positive integer which is larger than $\gamma^{-1}\ln(n)$.

We consider two classical families of functions. The former is a Lipschitz class Lip_α of m-fold continuously differentiable and 1-periodic functions satisfying the inequality $\max_{u,v}|f^{(m)}(u) - f^{(m)}(v)| \leqslant L|u - v|^{\alpha^*}$ where $L < \infty$. Here $\alpha = m + \alpha^*$ where m is integer and $\alpha^* \in (0, 1]$, $f^{(k)}$ denotes the kth derivative and $f^{(0)} = f$. We assume that $\alpha > 1/2$.

The latter is a class A_γ of analytic functions. Following Bary (1964), we define this class via Fourier coefficients by

$$A_\gamma = \{f : |\theta_{2j-s}| \leqslant M \exp(-\gamma j), \quad s = 0, 1, \quad j = 0, 1, \quad \ldots, \quad 0 < M < \infty\}. \quad (2.2)$$

Analytic functions have been intensively studied in both approximation and statistical literature. For instance, the case where f has an analytic continuation onto a strip is considered by Timan (1963) and Golubev and Levit (1996). In this case, γ is defined by the width of that strip. On the other hand, DeVore and Lorentz (1993), following the famous results of Bernstein, explored the best polynomial approximation of functions that have an analytic extension onto some elliptic discs; Ibragimov and Khasminskii (1984) suggested an asymptotically minimax pointwise estimator of such f. Note that different bases may be employed. The approach of Bary (1964) employed here to define an analytic function is especially convenient for exploring adaptive settings where neither smoothness of f nor a set to which f can be analytically continued is given.

We shall use the following estimators: truncated orthogonal series estimator $\hat{f}_n(J, t) = \sum_{0 \leqslant j \leqslant J} \hat{\theta}_j \varphi_j(t)$ which is employed for pointwise estimation of analytic functions; the de La Vallée Poussin sum $\check{f}_n(J, t) = \hat{\theta}_0 + \sum_{j=1}^{J} \sum_{i=0}^{1} \hat{\theta}_{2j-i} \varphi_{2j-i}(t) + \sum_{j=J+1}^{2J-1}(2 - j/J) \sum_{i=0}^{1} \hat{\theta}_{2j-i} \varphi_{2j-i}(t)$, which is employed for pointwise estimation of Lipschitz functions; the smoothed orthogonal series estimator $\check{f}_n(S, t) = \hat{\theta}_0 + \sum_{1 \leqslant k \leqslant S} \hat{\Lambda}(k, n) \sum_{j=(k-1)k+1}^{k(k+1)} \hat{\theta}_j \varphi_j(x)$, used for global estimation. Here $\hat{\Lambda}(k, n) = \hat{\Theta}(k, n)[\hat{\Theta}(k, n) + n^{-1}]^{-1} I[\hat{\Theta}(k, n) > 1/n \ln(k + 3)]$, $\hat{\Theta}(k, n) = (2k)^{-1} \sum_{j=(k-1)k+1}^{k(k+1)} (\hat{\theta}_j^2 - n^{-1})$ and $I(\cdot)$ is the indicator function.

The data-driven cut-off sequence $\hat{J} = \hat{J}(\{Y_n(t)\})$ for the adaptive pointwise estimator $\check{f}_n(\hat{J}, t)$ is defined as follows. Set J_s^* to be the integer part of $[\ln(n)]^2 d^s$ where $d > 2$ is a fixed constant and $s = 0, 1, \ldots$; set K to be the maximum integer satisfying the inequality $J_K^* < n^{1/2}[\ln(n)]^{-3}$; set β_s to be a solution of equation $[n/\ln(n)]^{1/(2\beta+1)} = J_s^*$; set J_s to be equal to J_m^* which is closest to $J_s^*[\ln(n)]^{1/(2\beta_s+1)}$; set $\tilde{I}(i, j) = \check{f}_n(j, 0) - \check{f}_n(i, 0)$, $\tilde{k} = \min\{l : |\tilde{I}(J_l^*, J_s^*)|^2 \leqslant 6 \ln(n) J_s^* n^{-1}, \ l \leqslant s \leqslant K; \ 0 \leqslant l \leqslant K\}$. Then we denote $\hat{J} = J_{\tilde{k}}$.

The following procedure defines an adaptive pointwise estimator for the case of analytic functions. Let Q_n be a minimal integer larger than $\ln(n)$ and let q be a positive real number. Set $b_k = 1$ for $k = 1, 2, \ldots, Q_n$ and $b_{Q_n+1} = \infty$, $T(r) = \{j : rq \ln(n) \leqslant j < (r+1)q \ln(n)\}$ and define

$$\hat{k} = \min\left(s : \max_{s \leqslant r \leqslant Q_n} \left|\sum_{j \in T(r)} \hat{\theta}_j\right|^2 \leqslant b_{r+1} rq \ln(n) n^{-1}, \quad s = 1, 2, \ldots, Q_n\right). \quad (2.3)$$

Then $\hat{f}_n(q\hat{k} \ln(n), t)$ is the adaptive pointwise estimator for analytic functions.

For the case of the union of Lipschitz and analytic functions we employ the estimator $\tilde{f}_n(t)$; it is equal to $\check{f}_n(J_{\tilde{k}}, t)$ if $\tilde{k} > 0$ and to $\hat{f}_n(q\hat{k} \ln(n), t)$ if $\tilde{k} = 0$.

We also use generic notation C and $o(1)$ for postitive constants and vanishing sequences in n, respectively. Both C and $o(1)$ may depend on α, L, γ or M.

3. Adaptive estimation of analytic functions

The following efficient (sharp) lower bounds are known for some particular settings:

$$\inf \sup_{f \in A_\gamma} \mathrm{E}_f\{[f_n(0) - f(0)]^2\} \geqslant \gamma^{-1} \ln(n)\, n^{-1}[1 + o(1)], \tag{3.1}$$

$$\inf \sup_{f \in A_\gamma} \mathrm{E}_f[\|f_n - f\|^2] \geqslant \gamma^{-1} \ln(n)\, n^{-1}[1 + o(1)], \tag{3.2}$$

where the infimum is taken over all estimators based on the observation $\{Y_n(t),\, 0 \leqslant t \leqslant 1\}$, n, M and γ (Ibragimov and Khasminskii 1984; Efromovich 1994). We give a simple proof of these lower bounds in the Appendix and also show that they are attained by the same estimate $\hat{f}_n(N_\gamma, t)$ (see Lemma A.1).

If the parameter γ is unknown, then the global lower bound (3.2) remains sharp and it is attained by the adaptive estimator $\check{f}_n(\ln(n), t)$ (see Lemma A.1 in the Appendix). At the same time, the following theorem establishes that there is no adaptive (data-driven) estimator which attains the pointwise lower bound (3.1).

Theorem 3.1. *Let b be a real positive number such that $\frac{1}{8}\ln(b) > 1$ and let $\hat{f}_n$ be an adaptive estimator which is based only on data $\{Y_n(t),\, 0 \leqslant t \leqslant 1\}$ and which is efficient for the class $A_{b\gamma}$, i.e.,*

$$\sup_{f \in A_{b\gamma}} \mathrm{E}_f\{[\hat{f}_n(0) - f(0)]^2\} \leqslant (b\gamma)^{-1} \ln(n)\, n^{-1}[1 + o(1)]. \tag{3.3}$$

Then, this adaptive estimator is not efficient for the class A_γ. Moreover,

$$\sup_{f \in A_\gamma} \mathrm{E}_f\{[\hat{f}_n(0) - f(0)]^2 \geqslant \tfrac{1}{8}\ln(b)\, \gamma^{-1} \ln(n)\, n^{-1}[1 + o(1)]. \tag{3.4}$$

Thus, we lose the sharp constant γ^{-1} for the case of unknown γ. At the same time, the following assertion shows that we do not lose the optimal rate $\ln(n)\, n^{-1}$. Recall that owing to Lepskii (1990) we lose the optimal rate $n^{-2\alpha/(2\alpha+1)}$ for the case of the Lipschitz functions and unknown α.

Theorem 3.2. *The adaptive estimator $\hat{f}_n(\hat{k}q\ln(n), t)$ is pointwise rate optimal, i.e.,*

$$\sup_{f \in A_\gamma} \mathrm{E}_f\{[\hat{f}_n(\hat{k}q\ln(n), 0) - f(0)]^2\} \leqslant C \ln(n)\, n^{-1}. \tag{3.5}$$

4. Union of Lipschitz and analytic functions

Let the only given information about an estimated f be that it belongs to the union $\{Lip_\alpha,\ \alpha > \frac{1}{2}\} \cup \{A_\gamma,\ \gamma > 0\}$. The following assertion shows that we can suggest adaptive estimators which perform as well as optimal adaptive estimators for each of these classes.

Theorem 4.1. *The adaptive estimator $\check{f}_n(n^{1/4}\ln^{-1}(n), t)$ is globally rate optimal for the case of Lipschitz functions and globally efficient for the case of analytic functions, i.e.,*

$$\sup_{f \in Lip_\alpha} E_f\{\|\check{f}_n(n^{1/4}\ln^{-1}(n), \cdot) - f(\cdot)\|^2\} \leqslant Cn^{-2\alpha/(2\alpha+1)}, \tag{4.1}$$

$$\sup_{f \in A_\gamma} E_f\{\|\check{f}_n(n^{1/4}\ln^{-1}(n), \cdot) - f(\cdot)\|^2\} = \gamma^{-1}\ln(n)\, n^{-1}[1 + o(1)]. \tag{4.2}$$

The adaptive estimator $\tilde{f}_n(t)$ is pointwise rate optimal, among all data-driven estimators, for both Lipschitz and analytic functions, i.e.,

$$\sup_{f \in Lip_\alpha} E_f\{[\tilde{f}_n(0) - f(0)]^2\} \leqslant C\left(\frac{n}{\ln(n)}\right)^{-2\alpha/(2\alpha+1)}, \tag{4.3}$$

$$\sup_{f \in A_\gamma} E_f\{[\tilde{f}_n(0) - (0)]^2\} \leqslant C\ln(n)\, n^{-1}. \tag{4.4}$$

Appendix

Lemma A.1. *The lower bounds* (3.1) *and* (3.2) *hold and are attained by* $\hat{f}_n(N_\gamma, t)$. *Moreover,* (3.2) *is attained by the data-driven estimate* $\check{f}_n(\ln(n), t)$.

Proof. First, we establish the pointwise lower bound (3.1). Let J be the minimal positive even integer which is greater than $\gamma^{-1}\ln(n)\{1 - 1/\ln[\ln(n)]\}$ and set $D = \{f : f(t) = \sum_{j=1}^J \theta_j \varphi_j(t),\ |\theta_j| \leqslant \ln(n)\, n^{-1/2}\}$. Then straightforward algebra shows that $D \subset A_\gamma$ for sufficiently large n.

Now we employ the traditional method of estimating a minimax risk via a Bayes risk and Rao–Blackwell theorem. Let Θ_j, $j = 1, 2, \ldots$ be i.i.d. uniform random variables on the interval $[-\ln(n)\, n^{-1/2},\ \ln(n)\, n^{-1/2}]$. We write

$$\inf_{f_n} \sup_{f \in A_\gamma} E_f\{[f_n(0) - f(0)]^2\} \geqslant \inf_{f_n} \sup_{f \in D} E_f\{[f_n(0) - f(0)]^2\}$$

$$\geqslant E\left[\left(\sum_{j=1}^J (E\{\Theta_j|\zeta_j\} - \Theta_j)\varphi_j(0)\right)^2\right]$$

$$= E\{[E\{\Theta_1|\zeta_1\} - \Theta_1]^2\} \sum_{j=1}^J [\varphi_j(0)]^2$$

$$= J E\{[E\{\Theta_1|\zeta_1\} - \Theta_1]^2\}.$$

Here $\zeta_j = \int_0^1 \varphi_j(t)\, d\tilde{Y}_n(t) = \Theta_j + n^{-1/2}\xi_j$, $d\tilde{Y}_n(t) = [\sum_{j=1}^J \Theta_j \varphi_j(t)]\, dt + n^{-1/2}\, dW(t)$, and $\xi_1, \xi_2, \ldots$ are i.i.d. standard normal variables which are independent of $\Theta_1, \Theta_2 \ldots$.

Note that $E\{[E\{\Theta_1|\zeta_1\} - \Theta_1]^2\} = n^{-1}[1 + o(1)]$ owing to Lemma 2 of Pinsker (1980). This yields (3.1).

The global lower bound (3.2) is established similarly and even more simply:

$$\inf_{f_n} \sup_{f \in A_\gamma} \mathrm{E}_f\{\|f_n - f\|^2\} \geq \inf_{f_n} \sup_{f \in D} \mathrm{E}_f\{\|f_n - f\|^2\}$$

$$\geq \sum_{j=1}^{J} \mathrm{E}\{[\mathrm{E}\{\Theta_j|\zeta_j\} - \Theta_j]^2\}$$

$$= J\,\mathrm{E}\{[\mathrm{E}\{\Theta_1|\zeta_1\} - \Theta_1]^2\}$$

$$= \gamma^{-1}\ln(n)\,n^{-1}[1 + o(1)].$$

Now we are in a position to show that these lower bounds are sharp and attained by the same estimate $\hat{f}_n(N_\gamma, t)$. We write $\hat{f}_n(N_\gamma, t) - f(t) = n^{-1/2}\sum_{j=0}^{N_\gamma}\xi_j\varphi_j(t) - \sum_{j>N_\gamma}\theta_j\varphi_j(t)$. Owing to (2.2) we get $|\sum_{j>N_\gamma}\theta_j\varphi_j(t)| < 4M\gamma^{-1}n^{-1/2} = o(1)[\ln(n)/n]^{1/2}$ and therefore

$$\sup_{f \in A_\gamma} \mathrm{E}_f\{[\hat{f}_n(N_\gamma, 0) - f(0)]^2\} = \sup_{f \in A_\gamma} \mathrm{E}_f\{\|\hat{f}_n(N_\gamma, \cdot) - f(\cdot)\|^2\}[1 + o(1)]$$

$$= \gamma^{-1}\ln(n)\,n^{-1}[1 + o(1)].$$

This yields the sharpness of (3.1) and (3.2). The fact that the adaptive estimate $\check{f}_n(\ln(n), t)$ attains the lower bound (3.2) follows from (2.2) and Lemmas 1–3 of Efromovich and Pinsker (1984). Lemma A.1 is proved.   □

Note that similarly to Tsybakov (1997) the case of general loss functions may be considered as well.

Proof of Theorem 3.1. The proof is based on the following known parametric result.

Proposition (Brown and Low (1996a). *Let Z be distributed according to the density p_{θ_1} or the density p_{θ_2} with respect to a measure μ. For any estimator $\delta(Z)$ define its risk by $R(\theta, \delta) = \int(\theta - \delta(z))^2 p_\theta(z)\mu\,(\mathrm{d}z)$. Also denote $I = E_{\theta_1}\{[p_{\theta_2}(z)/p_{\theta_1}(z)]^2\}$ and $\theta = \theta_2 - \theta_1$. Assume that $R(\theta_1, \delta) \leq \epsilon^2$ and $0 < \epsilon < |\theta|I^{-1/2}$. Then, $R(\theta_2, \delta) \geq |\theta|^2(1 - \epsilon I^{1/2}|\theta|^{-1})^2$.*

To apply this result to our nonparametric setting we set $\theta_1 = \theta_1(0)$ and $\theta_2 = \theta_2(0)$ where $\theta_1(t) = 0$ and $\theta_2(t) = \kappa_n\sum_{j=0}^{J}\varphi_j(t)$. Here J is the minimal even integer which is greater than $\gamma^{-1}\ln(n)$, $\kappa_n = c\{2\gamma/n\ln(n)\}^{1/2}$ and $c = [\ln(b)/2]^{1/2}$.

Denote by $P_{\theta_1}^n$ and $P_{\theta_2}^n$ the probability measures associated with the processes $\mathrm{d}Y_n(t) = n^{-1/2}\,\mathrm{d}W(t)$ and $\mathrm{d}Y_n(t) = \theta_2(t)\mathrm{d}t + n^{-1/2}\,\mathrm{d}W(t)$, $0 \leq t \leq 1$, respectively. The sufficient statistic for the family of probability measures $(P_{\theta_1}^n, P_{\theta_2}^n)$ is the random variable $Z = \ln[(\mathrm{d}P_{\theta_2}^n/\mathrm{d}P_{\theta_1}^n)(Y_n)]$ that is well defined (see Appendix II of Ibragimov and Khasminskii (1981)). Moreover, using the results of that Appendix we get that the distribution of Z is normal $N(-a_n^2/2, a_n^2)$ or $N(a_n^2/2, a_n^2)$ under $P_{\theta_1}^n$ and $P_{\theta_2}^n$, respectively. Here $a_n^2 = n\kappa_n^2 J$.

Note that the function $f(t) = \theta_1(t)$ belongs to $A_{b\gamma}$ and therefore the given inequality (3.3) may be rewritten as $R(\theta_1, \delta) \leq \epsilon_n^2$ where $\epsilon_n^2 = (b\gamma)^{-1}\ln(n)\,n^{-1}[1 + o(1)]$.

Now we can calculate I and θ:

$$I = e^{a_n^2} = \exp\{nc^2 2\gamma[n\ln(n)]^{-1}\gamma^{-1}\ln(n)[(1+o(1)]\} = \exp\{2c^2[1+o(1)]\},$$

$$\theta = \theta_2 - \theta_1 = \kappa_n 2^{1/2}\frac{J}{2} = c\gamma^{-1/2}\left(\frac{\ln(n)}{n}\right)^{1/2}[1+o(1)].$$

We also note that $\theta_2(t)$ belongs to A_γ for sufficiently large n and therefore, applying the Proposition, we obtain

$$\sup_{f\in A_\gamma} \mathrm{E}_f\{[\hat{f}_n(0) - f(0)]^2\}$$

$$\geqslant R(\theta_2, \hat{f}_n(0))$$

$$\geqslant c^2\gamma^{-1}\frac{\ln(n)}{n}\left[1 - \left(\frac{\ln(n)}{n}\right)^{1/2}\exp(c^2)\frac{\gamma^{1/2}}{(b\gamma)^{1/2}c[\ln(n)/n]^{1/2}}\right]^2[1+o(1)]$$

$$= \gamma^{-1}\ln(n)\,n^{-1}c^2[1 - \exp(c^2)b^{-1/2}c^{-1}]^2[1+o(1)]$$

$$= \gamma^{-1}\ln(n)\,n^{-1}\frac{\ln(b)}{2}\left[1 - \left(\frac{2}{\ln(b)}\right)^{1/2}\right]^2. \tag{5.1}$$

By assumption $\ln(b) > 8$ and this yields (3.4). Theorem 3.1 is proved. $\qquad\square$

Proof of Theorem 3.2. Let $k = k(\gamma, q)$ be the minimal positive integer which is greater than $(\gamma q)^{-1}$. Note that $k < Q_n$ for sufficiently large n and therefore we assume that the last inequality holds.

The inequality (3.5) obviously holds for the cases $\hat{k} = k$ and $\hat{k} = k+1$. For the case $\hat{k} > k+1$, where the variance term is increased, the following relations hold: $\sup_{f\in A_\gamma}|\sum_{j\in T(\hat{k}-1)}\theta_j|^2 < o(1)kq\ln(n)\,n^{-1}$ and $|\sum_{j\in T(\hat{k}-1)}\hat{\theta}_j|^2 > (\hat{k}-1)q\ln(n)\,n^{-1}$. They follow at once from (2.2) and (2.3), respectively. Also recall the elementary equality

$$\left|\sum_{j\in T(r)}\hat{\theta}_j\right|^2 = \left|\sum_{j\in T(r)}\theta_j\right|^2 + 2\left(\sum_{j\in T(r)}\theta_j\right)n^{-1/2}\left(\sum_{j\in T(r)}\xi_j\right) + n^{-1}\left|\sum_{j\in T(r)}\xi_j\right|^2. \tag{5.2}$$

Thus, the inequality $\hat{k} > k+1$ implies the inequality $|\sum_{j\in T(\hat{k}-1)}\xi_j|^2 > C(\hat{k}-1)q\ln(n)$. Using this, the well-known inequality

$$P(\xi_1^2 > a^2) < C\exp\left(-\frac{a^2}{2}\right), \tag{5.3}$$

and denoting by $I(\cdot)$ the indicator function, we write

$$\sup_{f\in A_\gamma} \mathrm{E}_f\{I(\hat{k} > k+1)|\hat{f}_n(\hat{k}q\ln(n), 0) - f(0)|^2\}$$

$$\leqslant C\sum_{r=k+1}^{Q_n-1}\mathrm{E}_f\left\{I(\hat{k} = r+1)rq\ln(n)\,n^{-1}I\left(\left|\sum_{j\in T(r)}\xi_j\right|^2[q\ln(n)]^{-1} > Cr\right)\right\}$$

$$\leq C[kq\ln(n)\,n^{-1}]\sum_{r=k+1}^{Q_n-1}\frac{rq}{kq}\exp(-Cr)$$

$$< C\gamma^{-1}\ln(n)\,n^{-1}.\tag{5.4}$$

Consider the case $\hat{k}<k$ where the bias term is increased. Using the inequality $|\sum_{l=1}^{k}a_l|^2\leq k\sum_{l=1}^{k}a_l^2$ we get

$$\sup_{f\in A_\gamma}\mathrm{E}_f\{I(\hat{k}<k)|\hat{f}_n(\hat{k}q\ln(n),0)-f(0)|^2\}$$

$$\leq kq\ln(n)\,n^{-1}+k\sum_{r=1}^{k-1}\left|\sum_{j\in T(r)}\theta_j\right|^2 I\left(\left|\sum_{j\in T(r)}\theta_j\right|^2\leq 4rq\ln(n)\,n^{-1}\right)+k\sum_{r=1}^{k-1}\left|\sum_{j\in T(r)}\theta_j\right|^2$$

$$\times I\left(\left|\sum_{j\in T(r)}\theta_j\right|^2>4rq\ln(n)\,n^{-1}\right)\mathrm{E}_f\left\{I(\hat{k}=r)I\left(\left|\sum_{j\in T(r)}\hat\theta_j\right|^2\leq rq\ln(n)\,n^{-1}\right)\right\}.\tag{5.5}$$

Note that the inequalities $|a|>2c$ and $|a+b|<c$ imply that $|b|>\frac{3}{8}|a|$. Thus

$$I\left(\left|\sum_{j\in T(r)}\theta_j\right|^2>4rq\ln(n)\,n^{-1}\right)\mathrm{E}_f\left\{I\left(\left|\sum_{j\in T(r)}\hat\theta_j\right|^2\leq rq\ln(n)\,n^{-1}\right)\right\}$$

$$\leq\mathrm{E}_f\left\{I\left(\left|\sum_{j\in T(r)}\xi_j\right|^2\Big/[q\ln(n)]>C\left|\sum_{j\in T(r)}\theta_j\right|^2\Big/[q\ln(n)\,n^{-1}]\right)\right\}.$$

Applying the last relation and (5.3) to the right-hand side of (5.5) we obtain

$$\sup_{f\in A_\gamma}\mathrm{E}_f\{I(\hat{k}<k)|\hat{f}_n(\hat{k}q\ln(n),0)-f(0)|^2\}\leq C(\gamma^{-1}+\gamma^{-3})\ln(n)\,n^{-1}+k\sum_{r=1}^{k-1}\left|\sum_{j\in T(r)}\theta_j\right|^2$$

$$\times I\left(\left|\sum_{j\in T(r)}\theta_j\right|^2>4rq\ln(n)\,n^{-1}\right)\exp\left(-C\sum_{j\in T(r)}\left|\theta_j\right|^2\Big/[q\ln(n)\,n^{-1}]\right)$$

$$\leq C(\gamma^{-1}+\gamma^{-3})\ln(n)\,n^{-1}.$$

This together with (5.4) yields (3.5). Theorem 3.2 is proved. □

Proof of Theorem 4.1. We begin with establishing (4.1) and (4.2). Consider a truncated projection estimate $\hat{f}_n(n^{1(2\alpha+1)})$. This plainly satisfies (4.1), which follows from Parseval identity and Lemma 12 in Chapter 12 of Devroye and Györfi (1985). Thus, specific truncated projection estimators are optimal for Lipschitz and analytic functions. It is well known that a smoothing pseudo-estimator $\check{f}_n^*$ that is defined as $\check{f}_n$ only in place of estimates $\hat\Theta(k,n)$ (we use $\Theta(k,n)=(2k)^{-1}\sum_{j=(k-1)k+1}^{k(k+1)}\theta_j^2$) has a smaller mean integrated squared error (Efromovich and Pinsker 1984). Note that in the modern statistical literature $\check{f}_n^*$ is referred to as oracle with the smoothing weights $\Lambda_k=\Theta_k/(\Theta_k+n^{-1})$ that are applied to Fourier coefficients θ_j, $(k-1)k+1\leq j\leq k(k+1)$.

Note that $\mathrm{E}\{\|\check{f}_n^*(n^{1/4}/\ln(n,\cdot)-f(\cdot)\|^2\}=n^{-1}\sum_{1\leq k\leq n^{1/4}/\ln(n)}2k\Lambda_k+\sum_{k>n^{1/4}/\ln(n)}2k\Theta_k.$

According to Lemma 1 of Efromovich and Pinsker (1984),

$$\mathrm{E}\left\{\left\|\check{f}_n\left(\frac{n^{1/4}}{\ln(n)},\cdot\right)-f(\cdot)\right\|^2\right\} \leq \mathrm{E}\left\{\left\|\check{f}_n^*\left(\frac{n^{1/4}}{\ln(n)},\cdot\right)-f(\cdot)\right\|^2\right\}+R_n, \qquad (5.6)$$

where, using the inequality

$$\mathrm{E}_f\left\{\left((2k)^{-1}\sum_{j=(k-1)k+1}^{k(k+1)}(\hat{\theta}_j^2-n^{-1}-\theta_j^2)\right)^8\right\} \leq Ck^{-4}n^{-4}\left(k^{-1}\sum_{j=(k-1)k+1}^{k(k+1)}\theta_j^2+n^{-1}\right)^4$$

in the place of the last inequality in the Appendix of Efromovich and Pinsker (1984), we obtain, via Lemma 3 of the paper,

$$R_n \leq o(1)\,\mathrm{E}\left\{\left\|\check{f}_n^*\left(\frac{n^{1/4}}{\ln(n)},\cdot\right)-f(\cdot)\right\|^2\right\}+Cn^{-1}\sum_{1\leq k\leq n^{1/4}/\ln(n)}(2k)^{-2}[\ln(k+3)]^6. \qquad (5.7)$$

This yields (4.1) and (4.2).

Now we prove (4.3). The proof is primarily based on results of Efromovich and Low (1994). If $\tilde{k}>0$, then (4.3) is proved in Theorem 4 of that paper. Thus, we are to consider the case $\tilde{k}=0$. Set $k=k(n,\alpha)$ to be the maximum integer satisfying $J_k\leq n^{1/(2\alpha+1)}$. According to (3.29) in that paper we get

$$\mathrm{E}\{I(\tilde{k}=0)[\hat{f}_n(q\hat{k}\ln(n),0)-f(0)]^2\}$$

$$\leq \mathrm{E}\{I[\tilde{I}^2(J_0,J_k^*)\leq 6\ln(n)J_k^*n^{-1}][(\hat{f}_n(q\hat{k}\ln(n),0)-\tilde{f}(J_k,0))+(\tilde{f}(J_k,0)-f(0))]^2\}$$

$$\leq C[\ln(n)J_k^*n^{-1}+J_k^{-2\alpha}]<C\left(\frac{n}{\ln(n)}\right)^{-2\alpha/(2\alpha+1)}.$$

This yields (4.3).

(4.4) for the case $\tilde{k}=0$ follows from Theorem 3.2. Consider the case $\tilde{k}>0$. Then, by definition of $\tilde{k}$, we get $I(\tilde{k}>0)\leq I(|\tilde{I}(J_0^*,J_s^*)|^2>6\ln(n)J_s^*n^{-1})$ for some $s>0$. On the other hand, $\sup_{f\in A_\gamma}\sum_{J>J_0^*}|\theta_j^*|<C\exp^{\{-\gamma\ln^2(n)/2\}}=o(1)n^{-1}$. Using these results and (5.3) we obtain

$$\sup_{f\in A_\gamma}\mathrm{E}\{I(\tilde{k}>0)\} \leq \sup_{f\in A_\gamma}\sum_{s=1}^K\mathrm{E}\left\{I\left(\left|n^{-1/2}\sum_{j=J_0^*}^{J_s^*}\xi_j\varphi_j(0)\right|^2\geq 6\ln(n)J_s^*n^{-1}[1+o(1)]\right)\right\}$$

$$\leq CK\exp\{-3\ln(n)(1+o(1))\}$$

$$= o(1)n^{-2}.$$

Using the last inequality and the Cauchy–Schwarz inquality we get

$$\sup_{f \in A_\gamma} \mathrm{E}\{I(\tilde{k}>0)[\tilde{f}_n(J_{\tilde{k}},0)-f(0)]^2\} \leqslant [\mathrm{E}\{I(\tilde{k}>0)\}]^{1/2}[\sup_{f \in A_\gamma} \mathrm{E}\{|\tilde{f}_n(J_{\tilde{k}},0)-f(0)|^2\}]^{1/2}$$

$$= o(1)n^{-1}(Cn^{-1}J_K K)^{1/2}$$

$$= o(1)n^{-1}.$$

This yields (4.4). Theorem 4.1 is proved.   □

References

Bary, N.K. (1964) *Treatise on Trigonometric Series*. Oxford: Pergamon.

Birgé, L. and Massart, P. (1996) From model selection to adaptive estimation. In D. Polard and G. Yang (eds) *Festschrift for Lucien Le Cam*, pp. 78–114. New York: Springer-Verlag.

Brown, L.D. and Low, M.G. (1996a) A constrained inequality with applications to nonparametric functional estimation. *Ann. Statist.*, **24**, 2524–2535.

Brown, L.D. and Low, M.G. (1996b) Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.

DeVore, R.A. and Lorentz, G.G. (1993) *Constructive Approximation*. New York: Sringer-Verlag.

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation: The L_1 View*. New York: Wiley.

Efromovich, S. (1994) Efficient pointwise and global estimation of nonparametric curves. Technical Report, University of New Mexico.

Efromovich, S. and Low, M. (1994) Adaptive estimates of linear functionals. *Probab. Theory Related Fields*, **98**, 261–275.

Efromovich, S. and Pinsker, M.S. (1984) Learning algorithm for nonparametric filtering. *Autom. Remote Control*, **11**, 1434–1440.

Golubev, Y.K. and Levit, B.Y. (1996) Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist.*, **5**, 357–368.

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

Ibragimov, I.A. and Khasminskii, R.Z. (1984) On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Applic.*, **29**, 18–32.

Lepskii, O.V. (1990) On one problem of adaptive estimation in Gaussian white noise. *Theory Probab. Applic.*, **35**, 459–470.

Lepskii, O.V. and Spokoiny, V.G. (1994) Optimal pointwise adaptive methods in non-parametric regression. Technical Report, Humbolt University.

Nussbaum, M. (1996) Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, **24**, 2399–2430.

Pinsker, M.S. (1980) Optimal filtration of square-integrable signals in Gaussian noise. *Problems Information Transmission*, **16**, 52–68.

Timan, A.F. (1963) *Theory of Approximation of Functions of a Real Variable*. Oxford: Pergamon.

Tsybakov, A.B. (1997) Asymptotically efficient nonparametric estimation in L_2 with general loss. *Problems Information Transmission*, in press.