Bernoulli 2(2), 1996, 167–181

Asymptotically efficient estimation of analytic functions in Gaussian noise

YURI K. GOLUBEV,[1] BORIS Y. LEVIT[2*] and ALEXANDER B. TSYBAKOV[3]

[1]*Institute for Problems of Information Transmission, Ermolovoy Str. 19, 101447 Moscow GSP-4, Russia*
[2]*University of Utrecht, Budapestlaan 6, 3584 CD Utrecht, Netherlands*
[3]*Université de Paris VI, 4 place Jussieu, LSTA, 75252 Paris Cédex 05, France*

The problem of recovery of an unknown regression function $f(x)$, $x \in \mathbb{R}^1$ from noisy data is considered. The function $f(\cdot)$ is assumed to belong to a class of functions analytic in a strip of the complex plane around the real axis. The performance of an estimator is measured either by its deviation at a fixed point, or by its maximal error in the L_∞-norm over a bounded interval. It is shown that in the case of equidistant observations, with an increasing design density, asymptotically minimax estimators of the unknown regression function can be found within the class of linear estimators. Such best linear estimators are explicitly obtained.

Keywords: analytic function; asymptotically minimax estimator; nonparametric regression

1. Introduction

Nonparametric regression models are widely studied in the statistical literature, since they are mathematically attractive and have many useful applications. Recent results by Ibragimov and Hasminskii (1981; 1982a; 1984a) and Stone (1980; 1982) marked a new approach to these models, with the emphasis on optimal (minimax) rates of convergence in estimating an unknown regression function in various functional classes.

In some remarkable cases not only minimax rates of convergence, but also exact asymptotic constants have been found, and the corresponding asymptotically minimax estimators have been derived. Pinsker (1980) was the first to do this in the problem of regression estimation in continuous-time Gaussian white noise. He obtained asymptotically minimax estimators of the regression function in the L_2-norm, with the underlying functional classes defined as ellipsoids in L_2. These classes include as special cases Sobolev's classes as well as the classes of periodic analytic functions. In the ensuing papers of Nussbaum (1985) and Golubev and Nussbaum (1992), this study was extended to regression models in discrete time and Sobolev's classes of functions.

Another example of asymptotically efficient nonparametric regression estimators was given by Ibragimov and Hasminskii (1984b) for the problem of estimating the unknown

* To whom correspondence should be addressed.

regression function at a single fixed point, if the underlying class consists of periodic functions, analytic in the strip $\{(x+iy):|y|<\gamma\}$, $\gamma>0$, of the complex plane.

The third, and last, example is due to Korostelev (1993). It relates to the problem of asymptotically efficient nonparametric regression estimation in the L_∞-norm, on Lipschitz classes C^β. Korostelev (1993) solved this problem for $\beta \le 1$. The extension to $\beta > 1$ was established by Donoho (1994).

In this paper we provide new examples of asymptotically minimax nonparametric regression estimation. We consider the following nonparametric regression model:

$$y_j = f(jh) + \xi_j, \qquad j = 0, \pm 1, \pm 2, \ldots,$$

where $h > 0$, the unknown function $f(x)$, $x \in \mathbb{R}^1$, is assumed to belong to a class $\mathcal{A}_\gamma(L)$ of functions admitting analytic continuation into the strip $\{(x+iy):|y|<\gamma\}$ (for a precise definition, see Section 2) and $\{\xi_j\}$ is a Gaussian white noise sequence. Clearly, this model presents a certain idealization of a real observation process. It approximates the practical situation where the observations are available on a large (but finite) interval. For mathematical convenience we consider here the infinite observation interval.

The classes of functions related to $\mathcal{A}_\gamma(L)$ featured quite prominently in earlier studies of the statistical estimation problems for stationary time series (see Parzen 1958 and references therein) and later in nonparametric density estimation (Watson and Leadbetter 1963). It was shown, by that time, that the rate of convergence $(n^{-1}\log n)^{1/2}$ is achievable for such classes, where n is the length of the observation period. In the paper by Ibragimov and Hasminskii (1983), whose definition of the class $\mathcal{A}_\gamma(L)$ we are using here, it was shown that this rate is optimal in the minimax sense.

The functional classes such as $\mathcal{A}_\gamma(L)$ provide a representative alternative to the classes of functions which are only finitely smooth. They are much broader than, for example, the class of entire functions of an exponential type (cf. Ibragimov and Hasminskii 1982b) and contain curves such as the normal, Cauchy and Student densities, as well as their mixtures.

What makes such classes even more attractive, from the statistical point of view, is that in estimating an unknown function $f(x) \in \mathcal{A}_\gamma(L)$, at any given point x, one can find asymptotically efficient (to be precise, locally asymptotically minimax) estimators (cf. our Theorem 2.1, where the corresponding lower bound actually has a local character). In that sense statistical properties of functionals $\Phi(f) = f(x)$ resemble those of the differentiable functionals, as defined in Koshevnik and Levit (1976) (see also the recent monograph by Bickel *et al.* 1993). However, these functionals are not differentiable in the above sense. The difference manifests itself in our Theorem 2.2. where estimating $f(x)$ in the uniform norm gives rise to another rate of convergence: $(n^{-1}\log n \log\log n)^{1/2}$. Such discrepancy typically does not occur in the case of differentiable functions (cf. Levit 1978; Ibragimov and Hasminskii 1991; or Bickel *et al.* 1993, Chapter 5). The above rate has been shown already to be optimal, for the classes closely related to $\mathcal{A}_\gamma(L)$, in Ibragimov and Hasminskii (1993). Our Theorem 2.2. provides the exact constant for this rate.

Two different kinds of result are discussed below. First, we consider the asymptotically minimax estimation of analytic regression functions at a fixed point x. We present an estimator $\tilde{f}_h(x)$ of $f(x)$ such that, for any x and $m = 0, 1, \ldots$, the derivative $\tilde{f}_h^{(m)}(x)$ is asymptotically minimax in estimating the derivative $f^{(m)}(x)$ for a large variety of loss functions. The property of $\tilde{f}_h(x)$ that in a sense dominates and facilitates this kind of result is the asymptotic negligibility of the bias of $\tilde{f}_h(x)$, as compared to its variance. These results are in the spirit of the earlier work of Golubev and Levit (1994), dealing with asymptotically minimax estimation of analytic probability densities and distribution functions.

Second, we study the problem of estimating an analytic regression function $f(\cdot)$ if the quality of an estimator is measured in $L_\infty[0,1]$. We show that the asymptotically minimax solution to this problem is given by a slightly modified estimator $\tilde{f}_h(x)$. This result represents, within its scope of normally distributed equidistant observations, a noticeable sharpening of the result due to Ibragimov and Hasminskii (1982a) concerning the optimal rates of convergence in L_∞ for the classes of analytic functions.

2. Main results

Suppose that we are given the following observations:

$$y_j = f(jh) + \xi_j, \qquad j = 0, \pm 1, \pm 2, \ldots, \tag{2.1}$$

where ξ_j are independent Gaussian $N(0,\sigma^2)$ random variables.

It is assumed that the unknown real-valued function $f(\cdot)$ belongs to the functional class $\mathcal{A}_\gamma(L)$ of analytic functions. This class has been used in the context of statistical estimation problems by Ibragimov and Hasminskii (1983). The class $\mathcal{A}_\gamma(L)$ consists of all functions $f(x)$ in $\mathbb{R}^1$ admitting analytic continuation into the strip $S_\gamma = \{(x+iy) : |y| \le \gamma\}$ of the complex plane $\mathbb{C}$ such that $f(x+iy)$ is analytic inside S_γ, bounded up to its boundary and

$$\int |f(x+i\gamma)|^2 \,dx \le L \qquad (\gamma > 0,\ L > 0).$$

Equivalently, $\mathcal{A}_\gamma(L)$ can be defined as the set of functions $f(x)$ satisfying the assumption (see, for example, Golubev and Levit 1994)

$$\frac{1}{2\pi}\int \cosh^2(\gamma t)|\hat{f}(t)|^2 \,dt \le L, \tag{2.2}$$

where $\hat{f}(t)$ is the Fourier transform of a real-valued function $f(x) \in L_2(\mathbb{R}^1)$. In particular, if $f(x) \in L_2(\mathbb{R}^1) \cap L_1(\mathbb{R}^1)$,

$$\hat{f}(t) = \int e^{itx} f(x) \,dx.$$

Our goal is to estimate the unknown function $f(x) \in \mathcal{A}_\gamma(L)$, based on the vector $\boldsymbol{y} = (\ldots, y_{-1}, y_0, y_1, \ldots)$ of observations in (2.1), using as the quality criterion one of the two risk functions described below.

Let $\mathcal{W}_\alpha$ be the class of loss functions $w(x) \ge 0$, $x \in \mathbb{R}^1$, such that

$$w(x) = w(-x), \qquad w(x) \ge w(y), \qquad |x| \ge |y|,$$

$$\int e^{-\alpha x^2} w(x) \,dx < \infty.$$

For $w \in \mathcal{W}_\alpha$, $\alpha < 1/2$, denote

$$\mathrm{E}w(\xi) = \frac{1}{(2\pi)^{1/2}} \int \mathrm{e}^{-x^2/2} w(x)\,\mathrm{d}x.$$

For convenience, we also define

$$Q_h = \frac{L}{\sigma^2 h}.$$

Let $\bar{f}_h(x; \boldsymbol{y})$ be an arbitrary estimator of the unknown regression function $f(x)$, based on the observations (2.1), and let P_f, E_f respectively be the distribution of the vector $\boldsymbol{y}$ and the expectation, corresponding to a given $f(\cdot)$ in (2.1).

Let

$$\psi_{h,m} = \left(\frac{\sigma^2 h \log^{2m+1} Q_h}{\pi(2m+1)(2\gamma)^{2m+1}} \right)^{1/2}.$$

Our first risk function is related to estimating the unknown regression function (in the case of $m = 0$), or its derivative of orders m, at a given point $x \in \mathbb{R}^1$:

$$R_{h,m}(\bar{f}_h, f) = \mathrm{E}_f w((\bar{f}_h^{(m)}(x, \boldsymbol{y}) - f^{(m)}(x))/\psi_{h,m}).$$

Below we also consider asymptotically minimax estimators of the unknown f, with respect to the L_∞-norm on a bounded interval which, without loss of generality, we assume to be $[0, 1]$:

$$R_h^\infty(\bar{f}_h, f) = \mathrm{E}_f w\left(\sup_{0 \le x \le 1} |\bar{f}_h(x, \boldsymbol{y}) - f(x)| / \psi_h^\infty \right),$$

where

$$\psi_h^\infty = \left(\frac{\sigma^2 h \log Q_h \log\log Q_h}{\pi\gamma} \right)^{1/2}.$$

Let us describe the estimators, which will be shown to be asymptotically minimax. Define the family of kernels

$$k(x, A) = \frac{\sin(x \cosh^{-1}(2A+1)/(2\gamma))}{2\gamma(1 + A^{-1})^{1/2} \sinh(\pi x/(2\gamma))}, \qquad x \in \mathbb{R}^1.$$

This kernel was derived in Golubev and Levit (1994), where it was shown, in the density estimation set-up, to provide asymptotically minimax estimators of an unknown density $f \in \mathcal{A}_\gamma(L)$, at any given point. The Fourier transform of $k(x, A)$, with respect to x, is given by (see Gradshtein and Ryzhik 1980), formula (3.983.1)):

$$\hat{k}(t, A) = (1 + A^{-1} \cosh^2 \gamma t)^{-1}. \tag{2.3}$$

Our estimator of the unknown function $f(\cdot)$ at a given point x is defined by

$$\tilde{f}_h(x,y) = h \sum_{j=-\infty}^{\infty} k(x - jh, Q_h) y_j.$$

Below $\inf_{\bar{f}}$ denotes infimum over all estimators.

Theorem 2.1. *Let $w \in \mathcal{W}_\alpha$, for some $0 < \alpha < 1/2$. Then for any $x \in \mathbb{R}^1$ and $m = 0, 1, \ldots,$*

$$\lim_{h\to 0} \inf_{\bar{f}} \sup_{f \in \mathcal{A}_\gamma(L)} R_{h,m}(\bar{f}, f) = \lim_{h\to 0} \sup_{f \in \mathcal{A}_\gamma(L)} R_{h,m}(\tilde{f}_h, f) = \mathrm{E}w(\xi).$$

When we deal with the L_∞-norm, the following estimator is used:

$$\tilde{f}_h^\infty(x,y) = h \sum_{j=-\infty}^{\infty} k\left(x - jh, \frac{Q_h}{2\log\log Q_h}\right) y_j.$$

Theorem 2.2. *Let $w(\cdot)$ be continuous at $x = 1$ and $w \in \mathcal{W}_\alpha$, for some $\alpha > 0$. Then*

$$\lim_{h\to 0} \inf_{\bar{f}} \sup_{f \in \mathcal{A}_\gamma(L)} R_h^\infty(\bar{f}, f) = \lim_{h\to 0} \sup_{f \in \mathcal{A}_\gamma(L)} R_h^\infty(\tilde{f}_h^\infty, f) = w(1).$$

Remark 2.1. The definition of $\mathcal{A}_\gamma(L)$ in terms of (2.2) shows that this functional class is closely related to ellipsoids in L_2, with exponential coefficients. In fact, results analogous to Theorems 2.1 and 2.2 can be obtained for the problem of estimating a periodic regression on $[0, 1]$, whose Fourier coefficients belong to such an ellipsoid.

Remark 2.2. As compared to the previous exact asymptotically minimax results, mentioned in Section 1, Theorem 2.2 has a peculiarity: it deals with the case of 'non-matched' norms. While the loss is measured in the L_∞-norm, the class of functions $\mathcal{A}_\gamma(L)$ is defined by the L_2-type restriction (2.2).

Remark 2.3. Theorem 2.1 presents a result of quite a different type than Theorem 2.2. Theorem 2.1 has a form typical of asymptotically minimax statements in parametric estimation, with Gaussian limiting distribution (cf. Ibragimov and Hasminskii 1981). In Theorem 2.2, however, we encounter a type of behaviour characterized by a degenerate limiting distribution. The results of Korostelev (1993) and Donoho (1994) are of this type. It can be shown that Pinsker's (1980) theorem is also a result on degenerate asymptotics (Tsybakov 1994).

Remark 2.4. In proving Theorems 2.1, 2.2 we extensively exploit the fact that ξ_j are assumed Gaussian. Assuming any other form of distribution with a finite second moment would overload the derivation with technicalities, which – while apparently standard – might greatly overshadow the main ideas; cf. Ibragimov and Hasminskii (1983). Another possible

– nonparametric – approach would apparently be much easier to implement. This will be treated elsewhere.

3. Proofs of the theorems

We first state the well-known formula due to Poisson, which will be used below on several occasions.

Lemma 3.1. *For real-valued functions f, g such that $f, f', g^{(m)}, g^{(m+1)} \in L_2(\mathbb{R}^1)$,*

$$h\sum_{j=-\infty}^{\infty} f(jh)g^{(m)}(jh) = \frac{1}{2\pi}\int \hat{f}(t)\hat{g}(-t)(\mathrm{i}t)^m\,\mathrm{d}t + \frac{1}{2\pi}\sum_{j\neq 0}\int \hat{f}\left(t-\frac{2\pi j}{h}\right)\hat{g}(-t)(\mathrm{i}t)^m\,\mathrm{d}t. \tag{3.1}$$

In particular, with $f \in \mathcal{A}_\gamma(L)$ and $g(y) = k(x-y, A)$, $A > 3$,

$$h\sum_{j=-\infty}^{\infty} k^{(m)}(x-jh, A)f(jh) = \frac{1}{2\pi}\int \mathrm{e}^{-\mathrm{i}tx}(\mathrm{i}t)^m\hat{k}(t,A)\hat{f}(t)\,\mathrm{d}t + O(\sqrt{A}\,\mathrm{e}^{-\pi\gamma/h}\log^m A)\left(\int \cosh^2(\gamma t)|\hat{f}(t)|^2\,\mathrm{d}t\right)^{1/2}. \tag{3.2}$$

Proof. Equation (3.1) directly follows from Titchmarsh (1962, Chapter II, Theorem 45). For completeness we provide the reader with a sketch of its proof. Inverting Fourier transforms and using the well-known formula in the theory of distributions (see, for example, Antosik *et al.* 1973, Chapter 9.6),

$$\sum_{j=-\infty}^{\infty} \mathrm{e}^{2\pi\mathrm{i}jx} = \sum_{j=-\infty}^{\infty} \delta(x-j),$$

one easily obtains

$$\begin{aligned} h\sum_{j=-\infty}^{\infty} f(jh)g^{(m)}(jh) &= \frac{h}{(2\pi)^2}\sum_{j=-\infty}^{\infty}\int \mathrm{e}^{-\mathrm{i}tjh}\hat{f}(t)\,\mathrm{d}t\int \mathrm{e}^{-\mathrm{i}sjh}\hat{g}(s)(-\mathrm{i}s)^m\,\mathrm{d}s \\ &= \frac{h}{(2\pi)^2}\int\int \hat{f}(t)\hat{g}(s)(-\mathrm{i}s)^m\sum_{j=-\infty}^{\infty}\delta\left(\frac{h(s+t)}{2\pi}-j\right)\mathrm{d}t\,\mathrm{d}s \\ &= \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\int \hat{f}\left(t-\frac{2\pi j}{h}\right)\hat{g}(-t)(\mathrm{i}t)^m\,\mathrm{d}t. \end{aligned}$$

This proves (3.1).

To prove (3.2), apply the Cauchy–Schwarz inequality to the integrals

$$\int e^{-itx}(it)^m \hat{k}(t,A)\hat{f}\left(t-\frac{2\pi j}{h}\right)dt$$

on the right-hand side of (3.1) and note that, according to (2.3),

$$\int t^{2m}|\hat{k}(t,A)|^2\cosh^{-2}(\gamma(t-2\pi j/h))\,dt = O(e^{-2\pi\gamma|j|/h})\int_0^\infty t^{2m}e^{2\gamma t}(1+(2A)^{-1}e^{2\gamma t})^{-2}\,dt$$

$$= O(A\,e^{-2\pi\gamma|j|/h})\int_1^\infty \frac{\log^{2m}(tA)}{(1+t)^2}\,dt = O(A\,e^{-2\pi\gamma|j|/h}\log^{2m}A). \qquad \square$$

Lemma 3.2. *Let*

$$\chi(x,A)=\begin{cases}1, & |x|\le(2\gamma)^{-1}\log A,\\ 0, & |x|>(2\gamma)^{-1}\log A.\end{cases}$$

Then, for $A\to\infty$,

$$\int t^{2m}|\hat{k}(t,A)-\chi(t,A)|\,dt = O(\log^{2m}A). \tag{3.3}$$

Indeed, (3.3) follows from relations (see also (2.3))

$$\int_{(2\gamma)^{-1}\log A}^{\infty} t^{2m}|\hat{k}(t,A)-\chi(t,A)|\,dt \le 4A\int_{(2\gamma)^{-1}\log A}^{\infty} t^{2m}\,e^{-2\gamma t}\,dt = O(\log^{2m}A),$$

$$\int_0^{(2\gamma)^{-1}\log A} t^{2m}|\hat{k}(t,A)-\chi(t,A)|\,dt \le A^{-1}\int_0^{(2\gamma)^{-1}\log A} t^{2m}\,e^{2\gamma t}\,dt = O(\log^{2m}A).$$

We turn now to the theorems stated in Section 2.

Proof of Theorem 2.1. We deal first with the upper bound on the risk. According to (2.1), one can split $\tilde{f}_h^{(m)}(x)-f^{(m)}(x)$ into two parts:

$$\tilde{f}_h^{(m)}(x)-f^{(m)}(x) = h\sum_{j=-\infty}^{\infty} k^{(m)}(x-jh,Q_h)\xi_j$$

$$+\,h\sum_{j=-\infty}^{\infty} k^{(m)}(x-jh,Q_h)\,f(jh)-f^{(m)}(x) \stackrel{\text{def}}{=} v(x)+b_f(x), \tag{3.4}$$

where $v(x)$ is a zero-mean stochastic term and $b_f(x)$ is the bias term.

Due to Lemma 3.1, for $h\to 0$,

$$b_f(x)=\frac{1}{2\pi}\int e^{-itx}(\hat{k}^{(m)}(t,Q_h)\hat{f}(t)-\hat{f}^{(m)}(t))\,dt + O(e^{-\pi\gamma/h}\sqrt{Q_hL}\log^m Q_h)$$

$$=\frac{1}{2\pi}\int e^{-itx}(-it)^m(\hat{k}(t,Q_h)-1)\hat{f}(t)\,dt + O(e^{-\pi\gamma/h}\sqrt{Q_hL}\log^m Q_h).$$

Applying the Cauchy–Schwarz inequality, Lemma 3.2, (2.2) and (2.3), one can show that the squared absolute value of the leading term on the right-hand side of the above equation is bounded by

$$\begin{aligned}&\frac{L}{2\pi}\int t^{2m}(\cosh(\gamma t))^{-2}(\hat{k}(t,Q_h)-1)^2\,\mathrm{d}t\\&\quad=\frac{L}{2\pi Q_h}\int t^{2m}\hat{k}(t,Q_h)(1-\hat{k}(t,Q_h))\,\mathrm{d}t\\&\quad\le\frac{L}{2\pi Q_h}\int t^{2m}|\hat{k}(t,Q_h)-\chi(t,Q_h)|\,\mathrm{d}t=O\left(\frac{L\log^{2m}Q_h}{Q_h}\right).\end{aligned}$$

Thus we have

$$\sup_{f\in\mathcal{A}_\gamma(L)}\sup_x(b_f(x))^2=O\left(\frac{L\log^{2m}Q_h}{Q_h}\right),\qquad h\to 0. \tag{3.5}$$

To evaluate the variance of the stochastic term

$$\sigma^2(x)\stackrel{\text{def}}{=}\operatorname{var}v(x)=\sigma^2h^2\sum_{j=-\infty}^{\infty}(k^{(m)}(x-jh,Q_h))^2,$$

one can again apply equality (3.2) of Lemma 3.1, with $f(y)=k^{(m)}(x-y,Q_h)$. To evaluate the last term appearing in (3.2), note that

$$\begin{aligned}\int t^{2m}\cosh^2(\gamma t)|\hat{k}(t,Q_h)|^2\,\mathrm{d}t&=O(1)\int_0^\infty t^{2m}\,\mathrm{e}^{2\gamma t}(1+(2Q_h)^{-1}\,\mathrm{e}^{2\gamma t})^{-2}\,\mathrm{d}t\\&=O(Q_h)\int_1^\infty\frac{\log^{2m}(uQ_h)}{(1+u)^2}\,\mathrm{d}u=O(Q_h\log^{2m}Q_h).\end{aligned}$$

Thus, by using Lemmas 3.1, 3.2, one obtains:

$$\begin{aligned}\sigma^2(x)&=\frac{\sigma^2h}{2\pi}\int|\hat{k}(t,Q_h)|^2\,\mathrm{d}t+O(\sigma^2h)\\&=\frac{\sigma^2h}{2\pi}\int t^{2m}|\hat{k}^{(m)}(t,Q_h)|^2\,\mathrm{d}t+O(\sigma^2h)\\&=\frac{h\sigma^2}{\pi(2m+1)}\left(\frac{\log Q_h}{2\gamma}\right)^{2m+1}+O(\sigma^2h\log^{2m}Q_h).\end{aligned} \tag{3.6}$$

According to (3.4)–(3.6), the distribution of $(f_h^{(m)}(x)-f^{(m)}(x))/\psi_{h,m}$ is $N(\mu_h,\sigma_h^2)$, where $\mu_h=o(1)$, $\sigma_h^2=1+o(1)$, and the terms $o(1)$ are small uniformly in $f\in\mathcal{A}_\gamma(L)$. Therefore, by virtue of the dominated convergence term, uniformly in $f\in\mathcal{A}_\gamma(L)$,

$$\lim_{h\to\infty}R_{h,m}(\hat{f}_h,f)=\lim_{h\to\infty}\frac{1}{(2\pi)^{1/2}}\int\exp\left(-\frac{(x-\mu_h)^2}{2\sigma_h^2}\right)w(x)\,\mathrm{d}x=\mathrm{E}w(\xi).$$

Turning now to the lower bound on the risk, denote

$$\mathrm{sinc}(x) = \frac{\sin x}{x} = \frac{1}{2}\int_{-1}^{1} \mathrm{e}^{-\mathrm{i}tx}\,\mathrm{d}t.$$

Consider the family of functions denoted by

$$f_c(z) = cg(z - x),$$

where

$$g(z) = \frac{(-1)^m(2m+1)}{a^m}\mathrm{sinc}^{(m)}(az), \qquad a = \frac{\log Q_h}{2\gamma}.$$

Note that

$$f_c^{(m)}(x) = cg^{(m)}(0) = \frac{(-1)^m(2m+1)c}{2}\int_{-1}^{1}(-\mathrm{i}t)^{2m}\,\mathrm{d}t = c \tag{3.7}$$

and

$$\int |\hat{g}(t)|^2 \cosh^2(\gamma t)\,\mathrm{d}t \le \text{const.}\, Q_h a^{-(2m+2)}.$$

Therefore $f_c(\cdot) \in \mathcal{A}_\gamma(L)$, for all c such that $|c| \le c(h)$, where $c^2(h) = \rho Q_h^{-1}\log^{2m+2} Q_h$, for some sufficiently small $\rho > 0$.

Note that (see, for example, Kuo 1975, Section II.2)

$$\frac{\mathrm{d}P_c}{\mathrm{d}P_0}(y) = \exp\left\{\frac{1}{2\sigma^2}\sum_{j=-\infty}^{\infty}(2cy_j g(jh) - c^2 g^2(jh))\right\},$$

where $P_c(\cdot) = P_{f_c}(\cdot)$. Clearly the statistic

$$T = \frac{\sum_{j=-\infty}^{\infty} y_j g(jh)}{\sum_{j=-\infty}^{\infty} g^2(jh)}$$

is sufficient for the parameter c of the family of distributions $P_c(\cdot)$ and is normally distributed,

$$T \sim N\left(c, \frac{\sigma^2}{\sum_{j=-\infty}^{\infty} g^2(jh)}\right),$$

where, according to Lemma 3.1, for all sufficiently small h,

$$\sigma^{-2}\sum_{j=-\infty}^{\infty} g^2(jh) = \frac{\pi\sigma^{-2}(2m+1)^2}{2ha^{2(2m+1)}}\int_{-a}^{a} t^{2m}\,\mathrm{d}t = (\psi_{h,m})^{-2}. \tag{3.8}$$

Let

$$\lambda(c) = \frac{1}{c(h)} \cos^2 \frac{\pi c}{2c(h)}$$

be the prior density on the interval $(-c(h), c(h))$.

Due to the sufficiency of T and (3.7), (3.8), the initial problem is reduced to estimation of the parameter c based on observation

$$T = c + \xi \psi_{h,m},$$

where $\xi \sim N(0,1)$. More precisely, we have

$$\begin{aligned}\inf_{\bar{f}} \sup_{f \in \mathcal{A}_\gamma(L)} R_{h,m}(\bar{f}, f) &\geq \inf_{\bar{f}} \sup_{|c| \leq c(h)} R_{h,m}(\bar{f}, f_c) \geq \inf_{\bar{c}} \sup_{|c| \leq c(h)} \mathrm{E}_c w((\bar{c} - c)/\psi_{h,m}) \\ &\geq \inf_{\bar{c}} \int_{-c(h)}^{c(h)} \mathrm{E}_c w((\bar{c} - c)/\psi_{h,m}) \lambda(c)\, \mathrm{d}c \\ &= \inf_{\bar{c}(T)} \int_{-c(h)}^{c(h)} \mathrm{E}_c w((\bar{c}(T) - c)/\psi_{h,m}) \lambda(c)\, \mathrm{d}c.\end{aligned}$$

Finally, according to Levit (1980, p. 565),

$$\begin{aligned}\inf_{\bar{c}(T)} \int_{-c(h)}^{c(h)} \mathrm{E}_c w((\bar{c}(T) - c)/\psi_{h,m})) \lambda(c)\, \mathrm{d}c &= \mathrm{E} w(\xi) + O(\psi_{h,m}^2 / c^2(h)) \\ &= \mathrm{E} w(\xi) + o(1), \qquad h \to 0.\end{aligned}$$

This completes the proof of the lower bound. □

Proof of Theorem 2.2. Again we look first at the upper bound on the risk. Similar to (3.4)–(3.6), one obtains

$$\tilde{f}_h^\infty(x) - f(x) \stackrel{\text{def}}{=} v(x) + b_f(x),$$

where

$$\sup_{f \in \mathcal{A}_\gamma(L)} \sup_x (b_f(x))^2 = O(2L \log\log(Q_h)/Q_h), \tag{3.9}$$

$$\sigma^2(x) \stackrel{\text{def}}{=} \operatorname{var} v(x) = \frac{h\sigma^2}{2\pi\gamma} \log\left(\frac{Q_h}{2\log\log Q_h}\right) + O(\sigma^2 h). \tag{3.10}$$

By using Lemma 3.2, one obtains similarly:

$$\sigma_a^2(x) \stackrel{\text{def}}{=} \operatorname{var} v'(x) = (1 + o(1)) \frac{h\sigma^2}{3\pi(2\gamma)^3} \log^3 Q_h. \tag{3.11}$$

To evaluate the supremum of $v(x)$ we use a well-known result describing the behaviour of the extremes of a differentiable Gaussian process (see Cramér and Leadbetter 1967, Chapter 13.5). This result and (3.9)–(3.11) show that for any $\delta > 0$, uniformly in

$f \in \mathcal{A}_\gamma(L)$, and for $z \geq 0$,

$$\begin{aligned} P_f\left\{(\psi_h^\infty)^{-1} \sup_{0\leq x\leq 1} |\tilde{f}_h^\infty(x,y) - f(x)| > 1+\delta+z\right\} \\ = P_f\left\{\sup_{0\leq x\leq 1} \frac{|v(x)|}{\sigma(x)} > (1+o(1))(1+\delta+z)(2\log\log Q_h)^{1/2}\right\} \\ \leq \frac{(1+o(1))}{\pi} \exp(-(1+o(1))(1+\delta+z)^2 \log\log Q_h) \int_0^1 \frac{\sigma_1(x)}{\sigma(x)}\,\mathrm{d}x \\ \leq \frac{(1+o(1))}{2(3)^{1/2}\pi\gamma} (\log Q_h)^{1-(1+o(1))(1+\delta)^2} \exp(-(1+o(1))z^2 \log\log Q_h) \\ \leq O(\log^{-\delta} Q_h) \exp(-z^2 \log\log Q_h). \end{aligned} \tag{3.12}$$

Denote

$$\mu = \sup_{0\leq x\leq 1} |\tilde{f}_h^\infty(x,y) - f(x)|/\psi_h^\infty.$$

Integrating by parts, one obtains

$$R_h^\infty(\tilde{f}_h^\infty, f) \leq w(1+\delta) + \int_{1+\delta}^\infty P_f\{\mu \geq z\}\,\mathrm{d}w(z). \tag{3.13}$$

Using (3.12) and again integrating by parts gives

$$\begin{aligned} \int_{1+\delta}^\infty P_f\{\mu \geq z\}\,\mathrm{d}w(z) = \int_0^\infty P_f\{\mu \geq z+1+\delta\}\,\mathrm{d}w(z+1+\delta) \\ \leq O(\log^{-\delta} Q_h) \int_0^\infty \mathrm{e}^{-z^2 \log\log Q_h}\,\mathrm{d}w(z+\delta+1) = o(1)w(1+\delta) \\ + O(\log^{-\delta} Q_h) \log\log Q_h \int_0^\infty z\,\mathrm{e}^{-z^2 \log\log Q_h} w(z+1+\delta)\,\mathrm{d}z. \end{aligned} \tag{3.14}$$

Finally note that, since $w \in W_\alpha$ for some $\alpha > 0$, we have

$$\begin{aligned} \int_0^\infty z\,\mathrm{e}^{-z^2 \log\log Q_h} w(z+1+\delta)\,\mathrm{d}z \leq \max_{x>0} x \exp(-x^2 \log\log Q_h + \alpha(1+\delta+x)^2) \\ \times \int_0^\infty w(z)\,\mathrm{e}^{-\alpha z^2}\,\mathrm{d}z \leq \text{const.} (\log\log Q_h)^{-1/2}. \end{aligned}$$

This inequality, together with (3.13) and (3.14), completes the proof of the upper bound.

Turning now to the lower bound on the risk, consider the family of functions denoted by

$$f_c(x) = \sum_{k=1}^N c_k g_k(x), \qquad c = (c_1, \ldots, c_N),$$

where

$$g_k(x) = \psi_h^\infty \operatorname{sinc}(W(Q_h)(x - x_k)), \qquad W(Q_h) = (\log Q_h - 2\log\log Q_h)/(2\gamma),$$

and

$$x_k = \pi k/W(Q_h), \qquad N = \lfloor W(Q_h)/\pi \rfloor, \qquad |c_k| \le 1, \qquad k = 1, \ldots, N.$$

Denote for brevity $P_c(\cdot) = P_{fc}(\cdot)$, $\mathrm{E}_c(\cdot) = \mathrm{E}_{fc}(\cdot)$.

Note that

$$f_c(x_k) = \psi_h^\infty c_k$$

and

$$\begin{aligned}\int |\hat{f}_c(t)|^2 \cosh^2(\gamma t)\,\mathrm{d}t &= (\psi_h^\infty)^2 \int \left| \sum_{k=1}^N \mathrm{e}^{\mathrm{i}tx_k} c_k \widehat{\mathrm{sinc}}(W(Q_h)t) \right|^2 \cosh^2(\gamma t)\,\mathrm{d}t \\ &\le \frac{\pi^2 N^2 (\psi_h^\infty)^2}{W^2(Q_h)} \int_{-W(Q_h)}^{W(Q_h)} \cosh^2(\gamma t)\,\mathrm{d}t \\ &\le \frac{\pi^2 N^2 (\psi_h^\infty)^2}{2W^2(Q_h)} \exp(\log Q_h - 2\log\log Q_h) \\ &= O(L \log\log Q_h \log^{-1} Q_h) = o(1).\end{aligned}$$

Therefore $f_c(x) \in \mathcal{A}_\gamma(L)$, for all sufficiently small h.

Next, according to Lemma 3.1, for all sufficiently small h and $k, l = 1, \ldots, N$,

$$\begin{aligned}\sum_{j=-\infty}^{\infty} g_k(jh) g_l(jh) &= \frac{(\pi\psi_h^\infty)^2}{2\pi h W^2(Q_h)} \int_{-W(Q_h)}^{W(Q_h)} \mathrm{e}^{\mathrm{i}t(x_k - x_l)}\,\mathrm{d}t \\ &= \frac{\pi(\psi_j^\infty)^2}{hW^2(Q_h)} \delta_{kl} = 2\delta_{kl}\sigma^2(\log\log Q_h)(1 + o(1)),\end{aligned} \tag{3.15}$$

where δ_{kl} is the Kronecker delta.

Due to (3.15), the likelihood ratio associated with the observations y can be written as (cf. Kuo 1975, Section II.2)

$$\frac{\mathrm{d}P_c}{\mathrm{d}P_0}(y) = \exp\left\{ \frac{1}{2\sigma^2} \sum_{k=1}^N \sum_{j=-\infty}^{\infty} g_k^2(jh)(2T_k c_k - c_k^2) \right\}, \tag{3.16}$$

where

$$T_k = \frac{\sum_{j=-\infty}^{\infty} y_j g_k(jh)}{\sum_{j=-\infty}^{\infty} g_k^2(jh)}.$$

Note that due to (3.15) and (3.16), for all $k = 1, \ldots, N$ the statistic T_k is sufficient for the corresponding parameter c_k of the family $P_c(\cdot)$. Moreover $T_1, \ldots, T_N$ are independent and normally distributed, $T_k \sim N(c_k, S^2)$, where

$$S^2 = \frac{1 + o(1)}{2\log\log Q_h}.$$

For all sufficiently small h and arbitrary $\delta > 0$, one obtains:

$$\inf_{\bar f} \sup_{f \in \mathcal{A}_\gamma(L)} R_h^\infty(\bar f, f) \geq \inf_{\bar f} \sup_{c: |c_k| \leq 1} R_h^\infty(\bar f, f_c)$$

$$\geq \inf_{\bar c} \sup_{c: |c_k| \leq 1} \mathrm{E}_c w\left(\max_{1 \leq k \leq N} |\bar c_k - c_k| \right)$$

$$\geq w(1-\delta) \inf_{\bar c} \sup_{c: |c_k| \leq 1} P_c \left\{ \max_{1 \leq k \leq N} |\bar c_k - c_k| \geq 1 - \delta \right\},$$

where $\bar c_k = \bar f(x_k)/\psi_h^\infty$, $\bar c = (\bar c_1, \ldots, \bar c_N)$.

To evaluate from below the last expression, we use the method of Korostelev (1993), with suitable modifications. Denote by $\phi_\sigma(x)$ the density of normal distribution $N(0, \sigma^2)$. It is not difficult to see (cf. Korostelev 1993), that for any $\sigma^2 > 0$, $\delta \in (0, 1)$,

$$\max_{z \in \mathbb{R}^1} \int_{-1}^1 \phi_\sigma(x-c) \chi\{|z-c| < 1-\delta\}\, \mathrm{d}c = \int_{-1}^1 \phi_\sigma(x-c)\chi\{|z^*(x) - c| < 1 - \delta\}\, \mathrm{d}c, \quad (3.18)$$

where

$$z^*(x) = \begin{cases} x, & |x| \leq \delta, \\ \delta\,\mathrm{sign}(x), & |x| > \delta. \end{cases}$$

Let $C = \{c : |c_k| \leq 1,\ k = 1, \ldots, N\}$. Then

$$\inf_{\bar c} \sup_{c: |c_k| \leq 1} P_c \left\{ \max_{1 \leq k \leq N} |\bar c_k - c_k| \geq 1 - \delta \right\}$$

$$\geq 2^{-N} \inf_{\bar c} \int_C P_c \left\{ \max_{1 \leq k \leq N} |\bar c_k - c_k| \geq 1 - \delta \right\} \mathrm{d}c$$

$$\leq 1 - 2^{-N} \sup_{\bar c} \int_C P_c \left\{ \max_{1 \leq k \leq N} |\bar c_k - c_k| < 1 - \delta \right\} \mathrm{d}c$$

$$= 1 - 2^{-N} \sup_{\bar c} \int_C \mathrm{E}_c \prod_{k=1}^N \chi\{|\bar c_k - c_k| < 1 - \delta\}\, \mathrm{d}c. \quad (3.19)$$

Using sufficiency of T_k, $k = 1, \ldots, N$, one obtains, according to (3.16) and (3.18):

$$2^{-N} \sup_{\bar c} \int_C \mathrm{E}_c \prod_{k=1}^N \chi\{|\bar c_k - c_k| < 1 - \delta\}\, \mathrm{d}c$$

$$= 2^{-N} \sup_{\bar c} \mathrm{E}_0 \int_C \prod_{k=1}^N \frac{\phi_S(T_k - c_k)}{\phi_S(T_k)} \chi\{|\bar c_k - c_k| < 1 - \delta\}\, \mathrm{d}c$$

$$\leq 2^{-N} \mathrm{E}_0 \prod_{k=1}^N \sup_{\bar c} \int_{-1}^1 \frac{\phi_S(T_k - c_k)}{\phi_S(T_k)} \chi\{|\bar c_k - c_k| < 1 - \delta\}\, \mathrm{d}c_k$$

$$= 2^{-N} \prod_{k=1}^{N} \int \sup_{z_k} \int_{-1}^{1} \phi_S(T_k - c_k)\chi\{|z_k - c_k| < 1 - \delta\}\, \mathrm{d}c_k\, \mathrm{d}T_k$$
$$= \left(\frac{1}{2}\int\int_{-1}^{1} \phi_S(T - c)\chi\{|z^*(T) - c| < 1 - \delta\}\, \mathrm{d}c\, \mathrm{d}T\right)^N$$
$$= \left(1 - \frac{1}{2}\int\int_{-1}^{1} \phi_S(T - c)\chi\{|z^*(T) - c| \geq 1 - \delta\}\, \mathrm{d}c\, \mathrm{d}T\right)^N, \qquad (3.20)$$

where

$$\frac{1}{2}\int\int_{-1}^{1} \phi_S(T - c)\chi\{|z^*(T) - c| \geq 1 - \delta\}\, \mathrm{d}c\, \mathrm{d}T$$
$$\geq \frac{1}{2}\int_{-\delta/2}^{\delta/2} \int_{1-\delta \leq |T-c| \leq 1-\delta/2} \phi_S(T - c)\, \mathrm{d}c\, \mathrm{d}T$$
$$\geq \delta \int_{1-\delta}^{1-\delta/2} \phi_S(x)\, \mathrm{d}x = \frac{\delta S(1 + o(1))}{\sqrt{2\pi}(1 - \delta)} \exp\left\{-\frac{(1 - \delta)^2}{2S^2}\right\}.$$

Noting that $N = O(\log Q_h)$, this, together with (3.17) and (3.19), (3.20), completes the proof. □

Acknowledgement

The research of the first author was supported by the Netherlands Organization for Scientific Research (NWO) while the author was visiting the University of Utrecht in the autumn of 1994.

References

Antosik, P., Mikusiński, J. and Sikorski, R. (1973) *Theory of Distributions. The Sequential Approach.* Amsterdam: Elsevier.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore, MD: Johns Hopkins University Press.

Cramér, H. and Leadbetter, M.R. (1967) *Stationary and Related Stochastic Processes.* New York: Wiley.

Donoho, D.L. (1994) Asymptotic minimax risk for sup-norm loss: solution via optimal recovery. *Probab. Theory Related Fields*, **99**, 145–170.

Golubev, Y.K. and Levit, B.Y. (1994) Asymptotically efficient estimation with analytic distributions. Technical Report No. 860, Department of Mathematics, University of Utrecht.

Golubev, G.K. and Nussbaum, M. (1992) Adaptive spline estimates for nonparametric regression models. *Theory Probab. Appl.*, **37**, 521–529.

Gradshtein, I.S. and Ryzhik, I.M. (1980) *Table of Integrals, Series, and Products.* New York: Academic Press.

Ibragimov, I.A. and Hasminskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag. (Originally published in Russian 1979.)

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

Ibragimov, I.A. and Hesminskii, R.Z. (1982b) Estimation of distribution density belonging to a class of entire functions. *Theory Probab. Appl.*, **27**, 551–562.

Ibragimov, I.A. and Hasminskii, R.Z. (1983) Estimation of distribution density. *J. Soviet Math.*, **25**, 40–57. (Originally published in Russian in 1980.)

Ibragimov, I.A. and Hasminskii, R.Z. (1984a) Asymptotic bounds on the quality of the nonparametric regression estimation in L_p. *J. Soviet Math.*, **25**, 540–549.

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

Ibragimov, I.A. and Hasminskii, R.Z. (1991) Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.*, **19**, 1681–1724.

Korostelev, A.P. (1993) Exact asymptotical minimax estimate of a nonparametric regression in the uniform norm. *Theory Probab. Appl.*, **38**, 875–882.

Koshevnik, Yu.A. and Levit B.Y. (1976) On a non-parametric analogue of the information matrix. *Theory. Probab. Appl.*, **21**, 738–753.

Kuo Hui-Hsiung (1975) *Gaussian Measures in Banach Spaces*. Lecture Notes in Math. 463. Berlin: Springer-Verlag.

Levit, B.Y. (1978) Infinite-dimensional informational inequalities. *Theory Probab. Appl.*, **23**, 371–377.

Levit, B.Y. (1980) On the asymptotic minimax estimates of the second order. *Theory Probab. Appl.*, **25**, 552–568.

Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.*, **13**, 984–997.

Parzen, E. (1958) On asymptotically efficient consistent estimates of the spectral density function of a stationary time series. *J. Roy. Statist. Soc. Ser. B*, **20**, 303–322.

Pinsker, M.S. (1980) Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission*, **16**, 120–133.

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

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

Titchmarsh, E.C. (1962) *Introduction to the Theory of Fourier Integrals*. London: Oxford University Press.

Tsybakov, A.B. (1994) Asymptotically efficient nonparametric estimation in L_2 with general loss. Technical report, Université Paris VI.

Watson, G.S. and Leadbetter, M.R. (1963) On estimation of the probability density. *Ann. Math. Statist.*, **34**, 480–491.

Received January 1995 and revised February 1996