Statistica Sinica **9**(1999), 33-49

ON THE MINIMAX OPTIMALITY OF BLOCK THRESHOLDED WAVELET ESTIMATORS

Peter Hall, Gérard Kerkyacharian and Dominique Picard

Australian National University, Université de Picardie and Université de Paris VII

Abstract: Block thresholding methods have been proposed by Hall, Kerkyacharian and Picard (1995) as a means of obtaining increased adaptivity when estimating a function using wavelet methods. For example, it has been shown that block thresholding reduces mean squared error by rendering the estimator more adaptive to relatively subtle, local changes in curvature, of the type that local bandwidth choice is designed to accommodate in traditional kernel methods. In this paper we show that block thresholding also provides extensive adaptivity to many varieties of aberration, including those of chirp and Doppler type. Indeed, in a wide variety of function classes, block thresholding methods possess minimax-optimal convergence rates, and in particular enjoy those rates without the extraneous logarithmic penalties that are usually suffered by term-by-term thresholding methods.

Key words and phrases: Besov space, chirp function, convergence rate, Doppler function, mean squared error, nonparametric regression, smoothing parameter.

1. Introduction

Wavelet methods for nonparametric function estimation are renown for their extraordinary adaptivity across a wide range of function classes. Indeed, nonlinear wavelet estimators based on term-by-term thresholding enjoy almost-optimal convergence rates across many of these classes, in that they achieve the minimax-optimal speed of convergence up to a power of the logarithm of sample size. See for example Donoho, Johnstone, Kerkyacharian and Picard (1995). It has recently been shown that so-called "block thresholding" allows the extraneous logarithmic factor to be removed in the case of estimating piecewise-continuous densities (see Hall, Kerkyacharian and Picard (1995)).

In the present paper we show that the minimax optimality of block thresholding is available substantially more generally, and in the context of nonparametric regression, across function classes that involve unboundedly many irregularities of a wide variety of types, including chirp and Doppler functions, and jump discontinuities. Moreover, the optimality is exact in the sense of convergence rates, since the ratio of the uniform upper bound to the minimax lower bound is close to neither zero nor infinity as sample size increases. In particular, the lower bound is not inferior to the upper bound by a power of the logarithm of sample size.

Our results also provide information about the level of irregularity that is allowable before it completely determines the overall convergence rate. This feature is perhaps best explained via an example, as follows. In the case of a Doppler, such as $g_1(x) = |x-x_0|^\beta \cos(|x-x_0|^{-\alpha})$ for $\alpha, \beta \geq 0$, frequency increases without bound in the vicinity of the potential discontinuity at x_0. (The case $\alpha = 0$ corresponds to a chirp, and there, x_0 does indeed produce a discontinuity.) One consequence of our results is that if such a Doppler is added to a smooth function g_2 of regularity s (meaning that it is in a Besov space B_{spq}; see Example 3.1 for a definition), then the overall convergence rate is still that of an optimal estimator of g_2, provided that $s < (2\beta+1)/(2\alpha)$. (This result requires the order, or Daubechies number, N, of the wavelet to be chosen sufficiently large.)

Block thresholding has the practical advantage of providing spatial adaptivity to relatively subtle changes in the target function. When applied to smooth curves it produces a degree of graduated smoothing, in much the same manner that a well-chosen, locally-varying bandwidth does in kernel estimation. By way of contrast, term-by-term thresholding applied to smooth curves is analogous to using a single, global bandwidth in kernel methods.

One way of viewing the performance of block thresholding is to note that, in the terminology of Donoho et al. (1995), it interprets the "oracle" perfectly to first order. That is to say, it includes an estimated wavelet coefficient if and only if the standard deviation of that quantity is less than its absolute value, except for terms of second order. Term-by-term thresholding can only interpret the "oracle" correctly up to a constant factor, even for terms of first order. These results are derived and further discussed by Hall, Kerkyacharian and Picard (1995).

In practice, block length for block thresholding must be chosen empirically, but its selection is not as critical as that of the primary resolution level (2^{j_0} in notation of Donoho et al. (1995), or p in notation of Hall and Patil (1995)). Indeed, if chosen within an appropriate range of values the block length does not have any first-order effect on mean squared error properties, and so it would not usually be considered a smoothing parameter. These issues will not be addressed in the present paper, however, since here we are principally interested in the performance of block thresholding when the target function is highly non-smooth. In effect we are concerned with adaptivity to high-frequency changes in the target curve, not to relatively subtle, low-frequency changes.

Section 2 will suggest a nonparametric regression model for the mechanism generating the data, and introduce block thresholding estimators of the regression mean. To ensure as much generality as possible our methods will be based on projections of estimators produced by the empirical wavelet transform method of Donoho and Johnstone (1995), although we shall discuss more direct approaches that are available in some circumstances. Section 3 will introduce a class $\mathcal{H}$ of regression means that may often be represented as the superposition of a function g_1

that has a possibly unbounded number of irregularities, and a relatively smooth function g_2. There we shall discuss examples, such as piecewise-continuous functions, chirps and Dopplers, that belong to the class. Section 4 will describe convergence rates achieved by the estimators in Section 2, uniformly over functions in $\mathcal{H}$; and it will point out that those rates are identical to minimax lower bounds for the regular components, g_2, of functions in $\mathcal{H}$. In this way we shall show at once that our block thresholding estimators achieve optimal rates of convergence over classes of irregular functions, and that in a wide range of settings these rates are identical to those of particularly regular functions. All proofs will be deferred to Section 5.

2. Model and Estimators

Assume that data $\{Y_m\}$ are generated by the model

$$Y_m = g(x_m) + \epsilon_m\,, \quad 1 \le m \le n, \tag{2.1}$$

where $x_m = m/n$ and the variables ϵ_m are independent and normally distributed with zero mean and variance $\sigma^2 > 0$. (In this setting, g is restricted to the interval $\mathcal{I} = [0,1]$, and the integrals in the definitions of α_j and β_{ij} are taken over $\mathcal{I}$.) The assumption of normality may be relaxed, as we shall discuss in Remark 4.5.

Let ϕ be a scale function and ψ its associated wavelet, and define $\phi_{ij}(x) = 2^{i/2}\,\phi(2^i x - j)$ and $\psi_{ij}(x) = 2^{i/2}\,\psi(2^i x - j)$. Given a square-integrable g, put $\alpha_{ij} = \int g\,\phi_{ij}$ and $\beta_{ij} = \int g\,\psi_{ij}$. An empirical wavelet expansion based on term-by-term thresholding is given by

$$\bar g = \sum_{-\infty<j<\infty} \bar\alpha_{i_0 j}\,\phi_{i_0 j} + \sum_{i=i_0}^{i_1-1} \sum_{-\infty<j<\infty} \bar\beta_{ij}\,\psi_{ij}\,I(\bar\beta_{ij}^2 > cn^{-1}\log n), \tag{2.2}$$

where $\bar\alpha_{ij} = n^{-1}\sum_m Y_m\,\phi_{ij}(x_m)$, $\bar\beta_{ij} = n^{-1}\sum_m Y_m\,\psi_{ij}(x_m)$, c is an appropriate threshold constant, and $i_1 > i_0$ is a truncation point. Note that here, a thresholding decision is made about each term in ψ_{ij}. Here and below, while we consider the regression model only on the compact interval $\mathcal{I}$, we simplify our notation and analysis by not employing special boundary wavelets. The block method is sufficiently adaptive to deal with edge effects without requiring subsidiary adjustments. Indeed, this is precisely one of its advantages. Some practical improvements (offering only second-order theoretical advantages) may be expected through using special boundary wavelets, but they can be difficult to implement.

In block thresholding, the integers j are divided among consecutive, nonoverlapping blocks of length l_i, say $\mathcal{B}_{ik} = \{j : (k-1)l_i + \nu + 1 \le j \le kl_i + \nu\}$,

$-\infty < k < \infty$, where ν is an arbitrary integer. (It simplifies notation a little if we take $\nu = 0$, which we shall do.) In this approach, all terms involving the functions ψ_{ij} for $j \in \mathcal{B}_{ik}$ are included in or excluded from the empirical wavelet transform. This leads to the estimator,

$$\tilde{g} = \sum_{-\infty<j<\infty} \bar{\alpha}_{i_0 j}\, \phi_{i_0 j} + \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} \Bigl(\sum\nolimits_{(ik)} \bar{\beta}_{ij}\, \psi_{ij} \Bigr) I(\widehat{B}_{ik} > cn^{-1}), \quad (2.3)$$

where $\sum_{(ik)}$ denotes summation over $j \in \mathcal{B}_{ik}$, and $\widehat{B}_{ik}$ is an estimator of the "average" value of β_{ij}^2 for $j \in \mathcal{B}_{ik}$.

In general, block length is an increasing function of the weight of the tails of the error distribution. To appreciate why, observe that estimates of terms β_{ij} will exhibit more tendency towards large-deviation fluctuations if the error distribution has heavier tails. Therefore, with heavier-tailed errors it is necessary to pool more spatial levels, j, into each block, if this effect is to be countered. Under the assumption of normality we shall use blocks whose length is only a power of $\log n$, but considerably longer blocks are necessary for error distributions such as Student's t. (Further discussion of this phenomenon appears in Remark 4.6.)

One of the ways in which the estimators at (2.2) and (2.3) differ is choice of the threshold level, which is proportional to $n^{-1} \log n$ in the former case and n^{-1} in the latter (for thresholding *squares* of empirical wavelet coefficients). This is a reflection of the greater accuracy with which β_{ij}^2 may be estimated if (as in the context of block thresholding) we pool information about neighbouring coefficients, and is discussed in detail by Hall, Kerkyacharian and Picard (1995).

The estimator at (2.3) is precisely the one employed by Hall, Kerkyacharian and Picard (1995), with an appropriate definition of $\widehat{B}_{ik}$. In the present work it is suitable if the "irregular" parts of the functions $g \in \mathcal{H}$ (to be defined in Section 3) are sufficiently smooth. In particular, they should be in a Besov space $\mathcal{B}_{s2\infty}$ (see Kerkyacharian and Picard (1993), or Example 3.1 below) for some $s > \frac{1}{2}$. (Remark 4.6 will provide further details.) For more general g, however, an alternative construction, based on the empirical wavelet transform of Donoho and Johnstone (1995), seems to be required.

In this construction we require that the scaling function ϕ be a Coiflet, with an associated wavelet ψ. Specifically, we suppose that ϕ and ψ are orthonormal and compactly supported, on $[0, v]$ say, and that the integral of ψ against any polynomial of degree no more than $N - 1$ vanish. We call N the Daubechies number of the Coiflet/wavelet pair.

Let V_i and W_i be the spaces spanned by $\{\phi_{ij},\ -\infty < j < \infty\}$ and $\{\psi_{ij},\ -\infty < j < \infty\}$, respectively, and let $\mathrm{Proj}_{V_i}(\cdot)$ and $\mathrm{Proj}_{W_i}(\cdot)$ be the projection operators

on these spaces. If $i < i_1$ and $f \in V_{i_1}$ then the coefficients of $\mathrm{Proj}_{V_i}(f)$ and $\mathrm{Proj}_{W_i}(f)$ may be computed from the values of $\int f\phi_{i_1 j}$, $-\infty < j < \infty$, using "subband filtering schemes" discussed by Daubechies (1992), Chapter 5. Define

$$\widehat{G}_{i_1} = n^{-1/2} \sum_{m=1}^{n} Y_m \phi_{i_1 m} \,.$$

Let the coefficients $\hat{\alpha}_{ij}$ and $\hat{\beta}_{ij}$ be given by

$$\mathrm{Proj}_{W_i}(\widehat{G}_{i_1}) = \sum_{-\infty<j<\infty} \hat{\beta}_{ij}\psi_{ij} \quad \text{and} \quad \mathrm{Proj}_{V_{i_0}}(\widehat{G}_{i_1}) = \sum_{-\infty<j<\infty} \hat{\alpha}_{i_0 j}\phi_{i_0 j} \,,$$

and put $\widehat{B}_{ik} = l_i^{-1} \sum_{(ik)} \hat{\beta}_{ij}^2$. In this notation our wavelet estimator of g is

$$\hat{g} = \sum_{-\infty<j<\infty} \hat{\alpha}_{i_0 j}\, \phi_{i_0 j} + \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} \Big(\sum\nolimits_{(ik)} \hat{\beta}_{ij}\, \psi_{ij} \Big)\, I(\widehat{B}_{ik} > n^{-1}\, c). \tag{2.4}$$

Choice of i_0, i_1, l_i and c will be discussed in Section 4.

3. The Class of Functions, $\mathcal{H}$

Given $0 < s_1 < s_2 < N$ and $\gamma, C_1, C_2, C_3, v \geq 0$, we shall define a class of functions $\mathcal{H} = \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$. In motivating the class we consider each of its elements to be the superposition of a regular function g_2 from the Besov space $B_{s_2\infty\infty}$, with a function g_1 that may possess irregularities of a variety of different types — for example, discontinuities or transients of unboundedly high frequency. The parameter s_i describes the minimum allowed smoothness of g_i (greater s_i corresponds to greater minimum smoothness), γ is a measure of the maximum allowed range of each irregularity of g_1 (that component of an irregularity which is of dyadic frequency i must involve fluctuations over a band of width no more than $2^{i(\gamma-1)}$), C_3 represents the number of irregularities in g_1, and $v/2^i$ denotes the distance over which we may expect to approximate aberrations of dyadic frequency i by a polynomial of degree $N-1$ (hence our reason for taking ϕ and ψ to have support of length v). For $i = 1, 2$ the constant C_i accommodates the amplitude of fluctuations in g_i when describing its smoothness by s_i. We hold $s_1, s_2, \gamma, C_1, C_2, N$ and v fixed, but allow C_3 to depend on n. Convergence rates are determined by the smooth component g_2, reflecting the adaptivity of wavelet methods to irregularities such as those in g_1.

More broadly, one may say that the irregularity of functions in $\mathcal{H}$ is described "macroscopically" by a "bad" function g_2, and "microscopically" by the number, $C_3 2^{\gamma i}$, of "bad" wavelet coefficients at level i. (See Proposition 3.2.) Specifically

$\mathcal{H}$ is the class of functions g such that for any $i \geq 0$ there exists a set of integers S_i for which the following is true: $\text{card}(S_i) \leq C_3 2^{i\gamma}$ and

> for each $j \in S_i$ there exist constants $a_0 = g(j/2^i), a_1, \ldots, a_{N-1}$ such that $\left| g(x) - \sum_{l=0}^{N-1} a_l (x - 2^{-i} j)^l \right| \leq C_1 \, 2^{-i s_1}$ for all $x \in [j/2^i, (j+v)/2^i]$; and
>
> for each $j \notin S_i$ there exist constants $a_0 = g(j/2^i), a_1, \ldots, a_{N-1}$ such that $\left| g(x) - \sum_{l=0}^{N-1} a_l (x - 2^{-i} j)^l \right| \leq C_2 \, 2^{-i s_2}$ for all $x \in [j/2^i, (j+v)/2^i]$.

If s_2 is not an integer then it follows by Taylor expansion that a ball of radius C_2 in a Besov space $B_{s_2 \infty \infty}$ is a subspace of $\mathcal{H}(s_1, s_2, \gamma, C_1, C_2, 0, N, v)$, for arbitrary $s_1 < s_2$ and $\gamma > 0$, and with $C_1 > 0$ depending on choice of the other constants. We may take $\gamma = 0$ in the case of jump discontinuities, and

$$\gamma = \frac{(s - \beta + \alpha N) \vee 0}{(\alpha + 1) N - \beta}$$

in the case of Doppler functions of the type $|x - x_0|^\beta \cos(|x - x_0|^{-\alpha})$. The next two examples describe more general forms of these respective cases.

Example 3.1. *Functions with discontinuities.* We begin by introducing the space B_{spq} and its norm $\|g\|_{spq}$. Given a sequence of real numbers $\{u_{ij}, -\infty < j < \infty\}$, put $\|u_{i.}\|_p = (\sum_{-\infty < j < \infty} |u_{ij}|^p)^{1/p}$ for $1 \leq p < \infty$, and $\|u_{i.}\|_\infty = \sup_j |u_{ij}|$. For functions g whose wavelet coefficients α_{ij} and β_{ij} are as defined in Section 2, and assuming $s < N$, let

$$\|g\|_{spq} = \|\alpha_{0.}\|_p + \Big\{ \sum_{i \geq 0} (2^{i\{s + (1/2) - (1/p)\}} \, \|\beta_{i.}\|_p)^q \Big\}^{1/q}$$

for $1 \leq q \leq \infty$, with the obvious change when q is replaced by ∞. Then B_{spq} is the set of functions g such that $\|g\|_{spq} < \infty$. (For definitions and properties of Besov spaces, and their relationships to wavelets, the reader is referred to Peetre (1975), Bergh and Löfström (1976), Meyer (1990), Kerkyacharian and Picard (1993) and Triebel (1992).)

Let $P_{d\tau A}$ be the set of piecewise polynomials of degree $d \leq N - 1$, with support contained in $[0, 1]$, such that the number of discontinuities is less than τ and the supremum norm less than A. Put

$$F_{s\infty\infty}(M) = \{ g \in B_{s\infty\infty} : \text{ supp } g \subseteq [0, 1], \ \|g\|_{s\infty\infty} \leq M \}.$$

Then it may be shown that the set $V_{d\tau A}\{F_{s\infty\infty}(M)\}$ of all functions of the form $g_1 + g_2$, with $g_1 \in P_{d\tau A}$ and $g_2 \in F_{s\infty\infty}(M)$, is a subset of $\mathcal{H}(0, s, 0, A, M, C\tau, N, v)$ if $C = C(s, A, M, N, v) > 0$ is chosen sufficiently large.

Example 3.2. *Chirp and Doppler functions.* Let $D(\alpha, \beta, \tau, A)$ denote the set of restrictions to $[0,1]$ of functions of the form

$$f(x) = \sum_{m=1}^{\tau} A_m (x - x_m)^{\beta_m} \cos\{(x - x_m)^{-\alpha_m}\},$$

with $\beta_m \geq \beta$, $\alpha_m \leq \alpha$ and $|A_m| \leq A$. Let $DF(\alpha, \beta, \tau, A, s, M)$ denote the set of functions $g_1 + g_2$ with $g_1 \in D(\alpha, \beta, \tau, A)$ and $g_2 \in F_{s\infty\infty}(M)$. The following proposition shows that all such functions lie in the class $\mathcal{H}$. It will be proved in Section 5.

Proposition 3.1. *If* $\max(s, \beta) < N$ *then for each* $v > 0$ *there exists* $C = C(\alpha, \beta, A, s, M, v) > 0$*, chosen sufficiently large, such that*

$$DF(\alpha, \beta, \tau, A, s, M) \subseteq \mathcal{H}\Big(\frac{(N-s)\beta}{(\alpha+1)N - \beta}, s, \frac{(s - \beta + \alpha N) \vee 0}{(\alpha+1)N - \beta}, 2NA, C, C\tau, N, v\Big).$$

More generally still, the function g may involve a mixture of jump discontinuities and Chirp and Doppler irregularities. Indeed, the convergence rates that we shall describe are available over somewhat larger classes than $\mathcal{H}$. It is difficult, however, to provide clear motivation for such larger classes, and to relate them to practically-occurring signals to which wavelet methods might be applied.

The conditions defining $\mathcal{H}$ have direct implications for the wavelet expansion of a function $g \in \mathcal{H}$, described in the next lemma. Let α_{ij} and β_{ij} denote the wavelet coefficients of g, defined in Section 2.

Proposition 3.2. *For every* $g \in \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$*,*

$$\begin{aligned}
|\beta_{ij}| &\leq \|\psi\|_1 \, C_1 \, 2^{-i(s_1+1/2)} \ \textit{if } j \in S_i\,, \\
|\beta_{ij}| &\leq \|\psi\|_1 \, C_2 \, 2^{-i(s_2+1/2)} \ \textit{if } j \notin S_i\,, \\
|\alpha_{ij} - g(j/2^i)| &\leq \|\phi\|_1 \, C_1 \, 2^{-i(s_1+1/2)} \ \textit{if } j \in S_i\,, \\
|\alpha_{ij} - g(j/2^i)| &\leq \|\phi\|_1 \, C_2 \, 2^{-i(s_2+1/2)} \ \textit{if } j \notin S_i\,.
\end{aligned}$$

4. Main Results

Our main theorem provides an upper bound to convergence rates uniformly over functions in $\mathcal{H}$. Since the bound is of the same size as the minimax lower bound (see Remark 4.1), then it is optimal. As a prelude to stating the bound, we introduce regularity conditions.

Let ϕ be a Coiflet, and ψ the associated wavelet, with Daubechies number N and support contained in the interval $[0, v]$ for some $0 < v < \infty$. Define the indices i_0 and i_1 in terms of N by $2^{i_0 - 1} \leq n^{1/(2N+1)} \leq 2^{i_0}$ and $2^{i_1 - 1} \leq n \leq 2^{i_1}$. Assume that the errors ϵ_m in the model at (2.1) are independent and identically

distributed as normal $\mathrm{N}(0,\sigma^2)$. Put $l_i = l = (\log n)^2$ for each i, and assume that $c \geq 48\sigma^2$, $0 \leq s_1 \leq s_2 < N$ and $0 \leq \gamma < (2s_1+1)/(2s_2+1)$; and that for all $\delta > 0$,

$$C_3 = O\Big(n^{\{1/(2s_2+1)\}-\{\gamma/(2s_1+1)\}+\delta}\Big)\,.$$

(Recall that c is the threshold constant in the formula for $\hat{g}$ (see (2.4)).) We call these conditions (C). It will be convenient to write $c = 48V^2$ where $V \geq \sigma$.

The condition $\gamma < (2s_1+1)/(2s_2+1)$ is crucial. It effects a balance between the number of "bad" wavelet coefficients at level i ($C_3 2^{\gamma i}$) and the difference between the irregularities of the components g_1 and g_2 of g.

Theorem 4.1. *If conditions* (C) *hold, and if the estimator $\hat{g}$ is as defined at* (2.4)*, then for each $C_1, C_2 > 0$ there exists a constant $K = K(s_1, s_2, \gamma, C_1, C_2, V, N, v) > 0$ such that*

$$\sup_{f \in \mathcal{H}(s_1,s_2,\gamma,C_1,C_2,C_3,N,v)} \int E(\hat{g}-g)^2 \leq n^{-2s_2/(2s_2+1)}\,\{K + o(1)\}\,.$$

Remark 4.1. *Achievement of minimax convergence rate.* Minimax theory (see e.g. Kerkyacharian and Picard (1993)) declares that the convergence rate over $F_{s\infty\infty}(M)$ is at best $n^{-2s/(1+2s)}$. Now, $F_{s_2\infty\infty}(C_2) \subseteq \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$. Therefore, the estimator $\hat{g}$ attains the minimax lower bound exactly, without any extraneous logarithmic factors.

Remark 4.2. *Adaptivity to different levels of regularity.* Let $\mathcal{C}(a)$, $a \in \mathcal{A}$, denote a sequence of classes of functions f, where $\mathcal{A}$ is a general index set. We have particularly in mind $\mathcal{C}(a) = V_{d\tau A}\{F_{s\infty\infty}(M)\}$, where $a = (d, A, s, M)$ and $\tau = \tau(n)$ does not increase too quickly (see Proposition 4.1); or $\mathcal{C}(A) = DF(\alpha, \beta, \tau, A, s, M)$, where $a = (A, s, M)$, α and β are fixed, and again $\tau(n)$ does not diverge too rapidly (see Proposition 4.2). Minimax optimality may be interpreted as an expression of adaptivity, as follows. An estimator $f^\star$ is said to be adaptive for the class $\{\mathcal{C}(\alpha),\ \alpha \in \mathcal{A}\}$ if for each $\alpha \in \mathcal{A}$ there exists $K(\alpha) > 0$ such that for all n, $R_n\{f^\star; \mathcal{C}(\alpha)\} \leq K(\alpha) \inf_{\hat{f}} R_n\{\hat{f}; \mathcal{C}(\alpha)\}$, where $R_n(\hat{f}; \mathcal{F}) = \sup_{f \in \mathcal{F}} \int E(\hat{f} - f)^2$. Propositions 4.1 and 4.2 express adaptation properties of the estimator $\hat{g}$ in the contexts of Examples 3.1 and 3.2, respectively.

Proposition 4.1. (*Discontinuities.*) *The estimator $\hat{g}$ defined at* (2.4) *is adaptive for the class of functions $V_{d\tau A}\{F_{s\infty\infty}(M)\}$, defined in Example* 3.1*, provided $d \leq N-1$, $0 < A < \infty$, $0 < s < N$, $0 < M < \infty$, $\sigma^2 \leq V^2$ and $\tau = O(n^{\{1/(1+2s)\}+\delta})$ for all $\delta > 0$.*

Proposition 4.2. (*Chirps and Dopplers.*) *The estimator $\hat{g}$ defined at* (2.4) *is adaptive for the class of functions $DF(\alpha, \beta, \tau, A, s, M)$, defined in Example* 3.2*,*

provided that $0 < A < \infty$, $0 < M < \infty$, $\sigma^2 \le V^2$,

$$0 < s < \tfrac{1}{4}\Big[\{(1+2\alpha N)^2 + 8N(2\beta+1)\}^{1/2} - (1+2\alpha N)\Big] \tag{4.1}$$

and $\tau = O(n^{\{1/(1+2s)\}-[\{(s+N\alpha-\beta)\vee 0\}/\{N(1+\alpha)+\beta\{2(N-s)-1\}]+\delta})$ *for all* $\delta > 0$.

Proposition 4.1 follows from Theorem 4.1, and Proposition 4.2 follows from Theorem 4.1 and Proposition 3.1. Note that in Proposition 4.1 we allow discontinuities at the edges of the interval $\mathcal{I}$, and do not require the pair (ϕ, ψ) to be adapted to edges or discontinuities. In Proposition 4.2, observe that if N is sufficiently large then condition (4.1) is implied by $0 < s < (1+2\beta)/(2\alpha)$, which is the condition noted in the third paragraph of Section 1.

Remark 4.3. *Choice of block length.* Theorem 4.1 remains valid, although for a different constant K, if the sequence l_i is nondecreasing and satisfies $C(\log n)^{1+\eta} \le l_i = O(i^2)$ for some $C, \eta > 0$.

Remark 4.4. *Versions of the theorem for L_p risk.* The theorem remains valid, albeit with a different constant, if the L_2 norm is replaced by the L_p norm for $1 \le p < \infty$, if the definition of $\widehat{B}_{ik}$ is changed to $l_i^{-1} \sum_{j \in \mathcal{B}_{ik}} |\hat{\beta}_{ij}|^p$, and if the threshold c/n is changed to $c/n^{p/2}$.

Remark 4.5. *Assumption of normal errors.* The assumption of normally distributed errors ϵ_m in the model at (2.1) may be replaced by the condition that for some $C > 0$, $P(|\epsilon_m| \le C) = 1$, without affecting the validity of the theorem. In that case, at the point of the proof where we bound large deviations, we employ an inequality due to Talagrand (1994) instead of one due to Cirel'son, Ibragimov and Sudakov (1976).

Remark 4.6. *Use of Coiflets.* If we may restrict the value of s_1 to $(\frac{1}{2}, \infty)$ then there is no need to assume that ϕ is a Coiflet, and the estimator $\hat{g}$ may be constructed without using a projection argument. In this case the only assumptions necessary for the error distribution are that $E|\epsilon_m|^{C_1} < \infty$ for some $C_1 > 1$ sufficiently large, and $E(\epsilon_m) = 0$. Block length, however, must now increase more rapidly than the rate $(\log n)^2$ assumed in Theorem 4.1, and in fact we should ask that $l = l_i = n^{C_2}$, where $C_2 = C_2(C_1) > 0$ is a decreasing function of C_1.

Remark 4.7. *Extension to heteroscedastic errors and irregular design.* Generalization of our results to the case of heteroscedastic errors is straightforward, provided we may write $\epsilon_m = \sigma_m \epsilon'_m$ where the σ_m's are bounded positive constants and the ϵ'_m's are independent normal $N(0,1)$ random variables. To achieve the generalization, simply choose the threshold constant c sufficiently large; it suffices to have $c \ge 48 \sup_m \sigma_m^2$, instead of $c \ge 48\sigma^2$ in conditions (C). In practice,

noting the possibility for heteroscedasticity and choosing the threshold accordingly can be important. Arguably the most appropriate approach is to estimate the variance function, graph it, and select an appropriate upper bound.

The non-normal case is similar, provided the ϵ'_m's are independent and identically distributed and (as noted in Remark 4.6) block length is chosen appropriately with respect to the number of the finite moments of the distribution of ϵ'_m.

Extension to the case of irregular design, where the observed data have the form (x_i, Y_i) with $x_1 < \cdots < x_n$, instead of $(i/n, Y_i)$, may often by achieved by an interpolation argument, as follows. First, fit a piecewise-linear function $Y(x)$ through the points (x_i, Y_i) by simply joining (x_i, Y_i) to (x_{i+1}, Y_{i+1}) for each i. (Horizontal extrapolation is appropriate for defining $Y(x)$ for $x < x_1$ or $x > x_n$.) Then, define functions $\hat\alpha_i(t)$ and $\hat\beta_i(t)$ (analogues of $\hat\alpha_{ij}$ and $\hat\beta_{ij}$, respectively) through integration rather than summation. For example, with $\psi_{it}(x) = 2^{i/2}\psi(2^i x - t)$ let

$$\mathrm{Proj}_{W_i}(\widehat G_{i_1})(x) = \int_{\infty}^{-\infty} \hat\beta_i(t)\, \psi_{it}(x)\, dt\,.$$

Finally, with integration over t in $\hat\alpha_i(t)$ and $\hat\beta_i(t)$ replacing summation over j in $\hat\alpha_{ij}$ and $\hat\beta_{ij}$, define first $\widehat B_i(u)$ (the analogue of $\widehat B_{ik}$) and then $\hat g$ as in Section 2. Versions of all our results may be established for this $\hat g$, provided the design points x_i may be expressed either as $F^{-1}(i/n)$ or $\widehat F^{-1}(i/n)$, where F is a distribution function whose density is smooth, supported on $\mathcal{I}$ and bounded away from zero there, and $\widehat F$ is the empiric of a random sample drawn from F.

5. Proofs of Theorem 4.1 and Proposition 3.1

Proof of Theorem 4.1. We shall consider only the setting where n is dyadic, in which case $n = 2^{i_1}$. Other contexts may be treated similarly.
Part (a): Properties of the projection operator. Observe that

$$\alpha_{i_1 m} \approx n^{-1/2} g(m/2^{i_1}) \int \phi = n^{-1/2} g(m/n)\,.$$

Hence, for a small number $r_{i_1 m}$ we have $n^{-1/2} g(m/n) = \alpha_{i_1 m} + r_{i_1 m}$. In this notation,

$$\widehat G_{i_1}(x) = \sum_{m=1}^{n} (\alpha_{i_1 m} + r_{i_1 m})\, \phi_{i_1 m}(x) + n^{-1/2} \sum_{m=1}^{n} \epsilon_m\, \phi_{i_1 m}(x)\,.$$

Similarly we may write

$$\mathrm{Proj}_{W_i}(\widehat G_{i_1}) = \sum_{-\infty < j < \infty} (\beta_{ij} + u_{ij} + U_{ij})\, \psi_{ij}(x)\,,$$

$$\mathrm{Proj}_{V_{i_0}}(\widehat{G}_{i_1}) = \sum_{-\infty<j<\infty} (\alpha_{i_0 j} + v_{i_0 j} + V_{i_0 j})\, \phi_{i_0 j}(x)\,,$$

where u_{ij} and $v_{i_0 j}$ denote real numbers, and U_{ij} and $V_{i_0 j}$ are normally distributed random variables with zero mean. By Parseval's identity,

$$\sum_{i_0 \le i < i_1} \sum_{-\infty<j<\infty} u_{ij}^2 + \sum_{-\infty<j<\infty} v_{i_0 j}^2 = \sum_m r_{i_1 m}^2\,.$$

The bounds noted in Proposition 3.2 may be employed to show that the last-written sum does not exceed $C_1 C_3 n^{-(2s_1+1-\gamma)} + C_2 n^{-2s_2}$, which in turn equals $C(n)\, n^{-2s_2/(2s_2+1)}$, where $C(n) = C_1 C_3 n^{-\{4s_1 s_2/(2s_2+1)\}-\eta} + C_2 n^{-4s_2^2/(2s_2+1)}$ and $\eta = \{(2s_1+1)/(2s_2+1)\} - \gamma$. Therefore,

$$\sum_{i_0 \le i < i_1} \sum_{-\infty<j<\infty} u_{ij}^2 + \sum_{-\infty<j<\infty} v_{i_0 j}^2 \le C(n)\, n^{-2s_2/(2s_2+1)}. \tag{5.1}$$

Define $\langle p, q\rangle = \int pq$, the usual inner product on the space of square-integrable functions on $\mathcal{I}$. In this notation, $u_{ij} = \sum_l r_{i_1 l}\, \langle \phi_{i_1 l}, \psi_{i_1 j}\rangle$, and also, $|\langle \phi_{i_1 l} \psi_{i_1 j}\rangle| \le 2^{i/2} \|\psi\|_\infty\, 2^{-i_1/2} \|\phi\|_1$ and $|r_{i_1 l}| \le (C_1 \vee C_2)\, 2^{-i(s_1+1/2)}$, whence it follows that

$$|u_{ij}| \le a(C_1 \vee C_2)\, 2^{-i(s_1+1/2)}, \tag{5.2}$$

where the constant a depends only in the pair (ϕ, ψ). In similar fashion,

$$U_{ij} = n^{-1/2} \sum_{m=1}^{n} \epsilon_m\, \langle \phi_{i_1 m}, \psi_{ij}\rangle\,, \quad V_{i_0 j} = n^{-1/2} \sum_{m=1}^{n} \epsilon_m \langle \phi_{i_1 m}, \phi_{i_0 j}\rangle\,,$$

whence it follows that

$$\begin{aligned} &\text{the variables } U_{ij} \text{ and } V_{i_0 j} \text{ are both normally distributed} \\ &\text{with zero means with variances not exceeding } \sigma^2/n. \end{aligned} \tag{5.3}$$

Part (b): Decomposition of the quadratic risk. We may decompose the quadratic risk as

$$E\|\hat{g} - g\|_2^2 = T_1 + T_2 + T_3 + T_4\,, \tag{5.4}$$

where, recalling that $\sum_{(ik)}$ denotes summation over $j \in \mathcal{B}_{ik}$,

$$T_1 = \sum_{i=i_1}^{\infty} \sum_{-\infty<j<\infty} \beta_{ij}^2\,, \quad T_2 = E\|\mathrm{Proj}_{V_{i_0}}(\widehat{G}_{i_1} - g)\|_2^2\,,$$

$$T_3 = \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} E\Big\{ I(\widehat{B}_{ik} > n^{-1} c) \sum\nolimits_{(ik)} (\hat{\beta}_{ij} - \beta_{ij})^2 \Big\}$$

$$= \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} E\Big\{ I(\widehat{B}_{ik} > n^{-1}\, c) \sum\nolimits_{(ik)} (u_{ij} + U_{ij})^2 \Big\}\,,$$

$$T_4 = \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} P(\widehat{B}_{ik} \le n^{-1}\, c) \sum\nolimits_{(ik)} \beta_{ij}^2\,.$$

The remainder of the proof consists of bounding $T_1, \ldots, T_4$.

Bound for T_1. By Proposition 3.2,

$$T_1 \le n^{-2s_2/(2s_2+1)} \Big[\{C_1 C_3/(1 - 2^{2s_1+1-\gamma})\}\, n^{-\{(4s_1 s_2/(2s_2+1)\}-\eta} + \{C_2/(1 - 2^{-2s_2})\}\, n^{-1/(2s_2+1)} \Big]\,. \tag{5.5}$$

Bound for T_2. Using (5.1) we obtain,

$$T_2 \le C(n)\, n^{-2s_2/(2s_2+1)} + n^{-2N/(2N+1)} \sigma^2\,. \tag{5.6}$$

Bounds for T_3 *and* T_4. We begin with T_3. Again by (5.1), writing

$$T_3' = \sum_{i=i_0}^{i_1-1} \sum_{-\infty<k<\infty} E\Big\{ I(\widehat{B}_{ik} > n^{-1}\, c) \sum\nolimits_{(ik)} U_{ij}^2 \Big\}\,,$$

we have

$$\tfrac{1}{2}\, T_3 \le \sum_{i=i_0}^{i_1-1} \sum_{-\infty<j<\infty} u_{ij}^2 + T_3' \le C(n)\, n^{-2s_2/(2s_2+1)} + T_3'\,. \tag{5.7}$$

Let $i' - 1$ denote the integer part of the base-2 logarithm of $n^{1/(2s_2+1)}$; thus, $2^{-i'}$ is of the optimal order for a bandwidth in kernel estimation of a function of known smoothness s_2. Put $B_{ik} = l^{-1} \sum_{(k)} (\beta_{ij} + u_{ij})^2$, where $l = l_i$ denotes block length, and write

$$T_3' = T_{31} + T_{32} + T_{33} + T_{34}\,, \tag{5.8}$$

where

$$T_{31} = \sum_{i=i_0}^{i'} \sum_{-\infty<k<\infty} E\Big\{ I(\widehat{B}_{ik} > n^{-1}\, c) \sum\nolimits_{(ik)} U_{ij}^2 \Big\}\,,$$

$$T_{32} = \sum_{i=i'+1}^{i_1-1} \sum_{k \in S_i} E\Big[I(\widehat{B}_{ik} > n^{-1}\, c)\, I\{B_{ik} > (2n)^{-1}\, c\} \sum\nolimits_{(ik)} U_{ij}^2 \Big]\,,$$

$$T_{33} = \sum_{i=i'+1}^{i_1-1} \sum_{k \notin S_i} E\Big[I(\widehat{B}_{ik} > n^{-1}\, c)\, I\{B_{ik} > (2n)^{-1}\, c\} \sum\nolimits_{(ik)} U_{ij}^2 \Big]\,,$$

$$T_{34} = \sum_{i=i'+1}^{i_1-1} \sum_{-\infty<k<\infty} E\Big[I(\widehat{B}_{ik} > n^{-1}\, c)\, I\{B_{ik} \le (2n)^{-1}\, c\} \sum\nolimits_{(ik)} U_{ij}^2 \Big]\,.$$

By (5.3),

$$T_{31} \le n^{-1}\sigma^2 \sum_{i=i_0}^{i'} 2^i \le 2\sigma^2\, n^{-2s_2/(2s_2+1)}\,. \tag{5.9}$$

By (5.2), (5.3) and Proposition 4.2, for each $r \ge 1$,

$$T_{32} \le (2^r\sigma^2 l/c^r n^{1-r}) \sum_{i=i'+1}^{i_1-1} C_3\, 2^{i\gamma}\{(a+1)^2(C_1^2+C_2^2)\,2^{-i(2s_1+1)}\}^r\,.$$

If $r \ge \gamma/(2s_1+1)$ then the right-hand side does not exceed

$$\left[C_3\{(a+1)^2(C_1^2+C_2^2)\}^r 2^r\sigma^2 l\Big/c^r(1-2^{(2s_1+1)r-\gamma})\right] n^{\zeta}\,,$$

where

$$\zeta = \frac{(2s_1+1)\,r-\gamma}{2s_2+1} + 1 - r = -\frac{2s_2}{2s_2+1} - \frac{1-\gamma-2r\,(s_2-s_1)}{2s_2+1}\,.$$

Recall the definition of η just above (5.1), and take $r = \gamma/(2s_1+1)$, to obtain the identity $\zeta = -2s_2/(2s_2+1) - \eta/(2s_1+1)$. Combining the results between (5.9) and here we deduce finally the bound,

$$T_{32} \le \left[C_3\{(a+1)^2(C_1^2+C_2^2)\}^r 2^r\sigma^2 l\Big/c^r\right] n^{-2s_2/(2s_2+1)-\eta/(2s_1+1)}\,. \tag{5.10}$$

Again by (5.1), (5.2) and Proposition 4.2,

$$T_{33} \le (4\sigma^2/c)\,\{C_2^2 + C(n)\}\,n^{-2s_2/(2s_2+1)}\,. \tag{5.11}$$

In Parts (c) and (d) of the proof we shall show that

$$T_{34} = O(n^{-\lambda}) \quad \text{for all} \quad \lambda > 0\,, \tag{5.12}$$

$$T_4 \le \{2c + C_2 + o(1)\}\,n^{-2s_2/(2s_2+1)}\,. \tag{5.13}$$

Results (5.7)–(5.12) imply that

$$T_3 \le 2\left\{2\sigma^2 + (4\sigma^2/c)\,C_2^2 + o(1)\right\} n^{-2s_2/(2s_2+1)}\,,$$

which in conjunction with (5.4)–(5.6) and (5.13) gives Theorem 4.1.

Part (c): Derivation of (5.12). First we prove:

Lemma 5.1. *If*

$$\textstyle\sum_{(ik)} (\beta_{ij}+u_{ij})^2 \le \frac{1}{2}n^{-1}cl \tag{5.14}$$

then

$$\left\{\textstyle\sum_{(ik)} (\beta_{ij}+u_{ij}+U_{ij})^2 \ge n^{-1}cl\right\} \subseteq \left\{\textstyle\sum_{(ik)} U_{ij}^2 \ge \frac{1}{12}n^{-1}cl\right\}.$$

Proof. Since $|\|u\|^2 - \|v\|^2| \le \|u-v\|^2 + 2\|u-v\|\,\|v\|$ then if (5.14) holds,

$$\begin{aligned}\tfrac{1}{2}n^{-1}cl &\le \textstyle\sum_{(ik)} (\beta_{ij} + u_{ij} + U_{ij})^2 - \sum_{(ik)} (\beta_{ij} + u_{ij})^2 \\ &\le \textstyle\sum_{(ik)} U_{ij}^2 + 2\Big(\sum_{(ik)} U_{ij}^2\Big)^{1/2} (\tfrac{1}{2}n^{-1}cl)^{1/2}\,.\end{aligned}$$

This implies that $\sum_{(ik)} U_{ij}^2 \ge (\frac{1}{2}n^{-1}cl)\,(\sqrt{2}-1)^2 \ge \frac{1}{12}n^{-1}cl$.

Lemma 5.2. *For each positive integer i, each integer k, and all $t > 0$,*

$$P\Big\{\textstyle\sum_{(ik)} U_{ij}^2 \ge \sigma^2 l n^{-1}(1+t)^2\Big\} \le e^{-lt^2/2}\,.$$

Proof. Let $\mathcal{A}_{ik}$ denote the set of sequences $\{a_j : j \in \mathcal{B}_{ik}\}$ such that $\sum_j a_j^2 = 1$. By (5.3),

$$E\Big(\textstyle\sum_{(ik)} U_{ij}^2\Big)^{1/2} \le \Big(E\sum_{(ik)} U_{ij}^2\Big)^{1/2} \le (l\sigma^2/n)^{1/2}\,,$$

and also,

$$\textstyle\sum_{(ik)} U_{ij}^2 = \sup'\Big(\sum_{(ik)} a_j U_{ij}\Big)^2\,,$$

where $\sup'$ denotes the supremum over all sequences $\{a_j\} \in \mathcal{A}_{ik}$. Hence, noting Cirel'son, Ibragimov and Sudakov (1976),

$$P\Big\{\Big(\textstyle\sum_{(ik)} U_{ij}^2\Big)^{1/2} \ge (l\sigma^2/n)^{1/2} + \lambda\Big\} \le \exp\{-n\lambda^2/(2\sigma^2)\}\,.$$

The lemma follows on taking $\lambda^2 = lt^2\sigma^2/n$.

Using Lemmas 5.1 and 5.2, and defining t by $\sigma^2(1+t)^2 = \frac{1}{12}c$ in the latter, we may prove that $T_{34} = O(n^{-\lambda})$ for all $\lambda > 0$, which establishes (5.12).

Part (d). Derivation of (5.13). We divide T_4 into five parts,

$$T_4 = T_{41} + T_{42} + T_{43} + T_{44} + T_{45}\,, \tag{5.15}$$

where

$$\begin{aligned}T_{41} &= \sum_{i=i_0}^{i_1-1} \sum_{k\in S_i} P(\widehat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} \ge 2n^{-1}c) \textstyle\sum_{(ik)} \beta_{ij}^2\,,\\ T_{42} &= \sum_{i=i_0}^{i'} \sum_{k\notin S_i} P(\widehat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} \ge 2n^{-1}c) \textstyle\sum_{(ik)} \beta_{ij}^2\,,\\ T_{43} &= \sum_{i=i_0}^{i'} \sum_{-\infty<k<\infty} P(\widehat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} < 2n^{-1}c) \textstyle\sum_{(ik)} \beta_{ij}^2\,,\end{aligned}$$

$$T_{44} = \sum_{i=i'+1}^{i_1-1} \sum_{k \in S_i} P(\widehat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} < 2n^{-1}c) \sum_{(ik)} \beta_{ij}^2 \,,$$

$$T_{45} = \sum_{i=i'+1}^{i_1-1} \sum_{k \notin S_i} P(\widehat{B}_{ik} \le n^{-1}c) \sum_{(ik)} \beta_{ij}^2 \,.$$

Shortly we shall show that

$$T_{41} \le C_1^2 C_3 \, n^{-(2s_1+1-\gamma)/(2N+1)} \, e^{-lt^2/2} = o(n^{-2s_2/(2s_2+1)}) \,, \tag{5.16}$$

$$T_{42} \le C_2^2 \, n^{-2s_2/(2N+1)} \, e^{-lt^2/2} = o(n^{-2s_2/(2s_2+1)}) \,, \tag{5.17}$$

$$T_{43} \le 2\{c + C(n)\} \, n^{-2s_2/(2s_2+1)} \,, \tag{5.18}$$

$$T_{44} = o(n^{-2s_2/(2s_2+1)}) \,, \tag{5.19}$$

$$T_{45} \le C_2 \, n^{-2s_2/(2s_2+1)} \,. \tag{5.20}$$

Combining (5.15)–(5.20) we deduce that, under the conditions of Theorem 4.1, (5.13) holds.

Results (5.16) and (5.17) follow from Lemma 5.2 and the following analogue of Lemma 5.1: if $\sum_{(ik)} (\beta_{ij} + u_{ij})^2 \ge 2n^{-1}cl$ then

$$\left\{ \sum_{(ik)} (\beta_{ij} + u_{ij} + U_{ij})^2 \le n^{-1}cl \right\} \subseteq \left\{ \sum_{(ik)} U_{ij}^2 \ge \tfrac{1}{12} n^{-1}cl \right\} .$$

For example, to derive (5.16), observe that

$$T_{41} \le C_1^2 C_3 \sum_{i=i_0}^{i_1-1} 2^{i\gamma} \, 2^{-i(2s_1+1)} \, e^{-lt^2/2} \le C_1^2 C_3 n^{-(2s_1+1-\gamma)/(1+2N)} \, e^{-lt^2/2} \,.$$

To obtain (5.18), note that since

$$\tfrac{1}{2} \sum_{(ik)} \beta_{ij}^2 \le \sum_{(ik)} (\beta_{ij} + u_{ij})^2 + \sum_{(ik)} u_{ij}^2$$

then

$$T_{43} \le 2 \sum_{i=i_0}^{i'} (2^i/l) \, (n^{-1}cl) + 2 \sum_{i=i_0}^{i'} \sum_{-\infty<j<\infty} u_{ij}^2 \le 2\{c + C(n)\} \, n^{-2s_2/(2s_2+1)} \,.$$

Similarly,

$$T_{44} \le 2 \sum_{i=i'+1}^{i_1-1} \sum_{k \in S_i} P(\widehat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} < 2n^{-1}c) \times \sum_{(ik)} (\beta_{ij} + u_{ij})^2 + 2 \sum_{i=i'+1}^{i_1-1} \sum_{-\infty<j<\infty} u_{ij}^2$$

$$
\begin{aligned}
&\le 2 \sum_{i=i'+1}^{i_1-1} \sum_{k \in S_i} (2n^{-1}cl)^r \Big\{ \textstyle\sum_{(ik)} (\beta_{ij} + u_{ij})^2 \Big\}^{1-r} + 2C(n)\, n^{-2s_2/(2s_2+1)} \\
&\le 2(2n^{-1}cl)^r \sum_{i=i'+1}^{i_1-1} C_3\, 2^{i\gamma} \Big\{ (a+1)^2 \, (C_1^2 + C_2^2)\, 2^{-i(2s_1+1)} \Big\}^{1-r} \\
&\qquad + 2C(n)\, n^{-2s_2/(2s_2+1)} \\
&= o(n^{-2s_2/(2s_2+1)}) \,,
\end{aligned}
$$

which establishes (5.19). Finally, to derive (5.20) note that

$$
T_{45} \le \sum_{i=i'+1}^{i_1-1} 2^i \, C_2 \, 2^{-i(2s_2+1)} \le C_2 \, n^{-2s_2/(2s_2+1)} \,.
$$

Proof of Proposition 3.1. We treat only the function $f(x) = x^\beta \cos(x^{-\alpha})$, showing that it is an element of the appropriate function class $\mathcal{H}$. Let $\mathcal{I}_{ij}$ denote the interval $[j/2^i, (j+v)/2^i]$. Using Leibnitz' formula for the differential of a product of two functions we may show that

$$
\sup_{x \in I_{ij}} \left| \frac{d^m}{dx^m} \, x^\beta \cos\left(x^{-\alpha}\right) \right| \le C(m, \alpha, \beta) \sum_{l=0}^{m} (j/2^i)^{\beta - l} \, (j/2^i)^{-(m-l)(\alpha+1)} \,.
$$

Hence, by Taylor expansion,

$$
\begin{aligned}
&\sup_{x \in I_{ij}} \left| x^\beta \cos\left(x^{-\alpha}\right) - \sum_{l=0}^{m-1} a_k \, (x - 2^{-i} j)^k \right| \\
&\qquad\qquad \le 2^{-im} C(m, \alpha, \beta) \sum_{l=0}^{m} (j/2^i)^{\beta - l} \, (j/2^i)^{-(m-l)(\alpha+1)} \,.
\end{aligned}
$$

Taking $m = N$ and $j \ge 2^{i\gamma}$, where $\gamma = (s - \beta + \alpha N)/\{(\alpha+1)N - \beta\}$, we see that the right-hand side equals $O(2^{-is})$ as $i \to \infty$. This establishes the appropriate Taylor expansion in the definition of $\mathcal{H}$, in the case where $j \notin S_i = \{j : 0 \le j \le 2^{i\gamma}\}$. When $j \in S_i$ we have, more simply,

$$
\sup_{j \in S_i} \sup_{x \in I_{ij}} \left| x^\beta \cos\left(x^{-\alpha}\right) - (j/2^i)^\beta \, \cos\left\{(j/2^i)^{-\alpha}\right\} \right| \le 2 \sup_{j \in S_i} (j/2^i)^\beta \le 2^{-is_1+1} \,,
$$

where $s_1 = (N - s)\beta/\{(\alpha+1)N - \beta\}$. This gives the Taylor expansion in the definition of $\mathcal{H}$ for the case $j \in S_i$.

Acknowledgements

We are grateful to two referees and an associate editor for their helpful comments.

References

Bergh, J. and Löfström, J. (1976). *Interpolation Spaces — An Introduction.* Springer, New York.

Cirel'son, B. S., Ibragimov, I. A. and Sudakov, V. N. (1976). Norm of Gaussian sample functions. In *Proc. 3rd Japan—USSR Symp. Probab. Thy.* **550** (Edited by G. Maruy-ama and J. V. Prohorov), 20-41. Springer Lecture Notes in Mathematics.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Donoho, D. and Johnstone, I. M. (1995). Minimax estimation by wavelet shrinkage. *Statist. Probab. Lett.* To appear.

Donoho, D., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (With discussion.) *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Hall, P., Kerkyacharian, G. and Picard, D. (1995). Note on the wavelet oracle. Manuscript.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23**, 905-928.

Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelet methods, optimality of Besov Spaces. *Statist. Probab. Lett.* **18**, 327-336.

Meyer, Y. (1990). *Ondelettes.* Hermann, Paris.

Peetre, J. (1975). *New Thoughts on Besov Spaces.* Duke Univ. Math. Ser. **1**.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22**, 28-76.

Triebel, H. (1992). *Theory of Function Spaces II.* Birkhäuser Verlag, Basel.

Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia.

E-mail: halpstat@pretty.anu.edu.au

Faculté Mathematiques et Informatiques, Université de Picardie, 33 rue Saint-Leu, 80039 Amiens, Cedex 01, France.

Département de Mathematiques, Université de Paris VII, 75251 Paris, Cedex 05, France.

E-mail: dominique.loicard@gauss.math.jussieu.fr

(Received January 1996; accepted June 1997)