The Annals of Statistics
1996, Vol. 24, No. 6, 2462–2476

PERFORMANCE OF WAVELET METHODS FOR FUNCTIONS WITH MANY DISCONTINUITIES

BY PETER HALL, IAN MCKAY[1] AND BERWIN A. TURLACH

Australian National University

Compared to traditional approaches to curve estimation, such as those based on kernels, wavelet methods are relatively unaffected by discontinuities and similar aberrations. In particular, the mean square convergence rate of a wavelet estimator of a fixed, piecewise-smooth curve is not influenced by discontinuities. Nevertheless, it is clear that as the estimation problem becomes more complex the limitations of wavelet methods must eventually be apparent. By allowing the number of discontinuities to increase and their size to decrease as the sample grows, we study the limits to which wavelet methods can be pushed and still exhibit good performance. We determine the effect of these changes on rates of convergence. For example, we derive necessary and sufficient conditions for wavelet methods applied to increasingly complex, discontinuous functions to achieve convergence rates normally associated only with fixed, smooth functions, and we determine necessary conditions for mean square consistency.

1. Introduction. A major virtue of wavelet methods in curve estimation is their very high degree of adaptability to different levels of smoothness. They capture discontinuous episodes in a curve far better than do many alternative approaches, such as kernel methods, and achieve this end without requiring subsidiary adjustments. For example, the formula for mean integrated squared error of a "standard" rth-order wavelet estimator of a density or regression function is unaffected, to first order, by a bounded number of jump discontinuities. To achieve the same level of performance when used to estimate a discontinuous curve, a kernel estimator first requires identification of the places where jumps occur and then careful manipulation in the neighborhood of the jumps.

On reflection, however, it is clear that these virtuous properties of wavelet estimators must depend in some way on the complexity of the target function. In the present paper we examine the performance of a wavelet estimator for targets where the number of jumps diverges as sample size increases. Our aim is to determine factors which limit performance of wavelet estimators in very difficult problems. We provide a concise description of the rate at which the number of jump discontinuities may increase if performance of the estimator is not to be affected, to first order, by those jumps. We also describe the relationship between the number of jumps and the sizes of jumps in this context—if the jumps are of smaller size, then more jumps are allowed be-

Received June 1995; revised April 1996.

[1]Dr McKay passed away in Canberra on 25 April 1995.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62G20.

Key words and phrases. Density estimation, discontinuity, mean integrated squared error, jump, nonparametric regression, threshold, wavelet.

fore performance deteriorates. We address issues as fundamental as mean square consistency of wavelet estimators, showing that a certain relationship between the number of jumps and their sizes is sufficient for convergence in mean square.

We shall now summarize our main conclusions. Roughly speaking, the wavelet estimator is consistent if the number of jumps (m, say) multiplied by the size of the jumps (η, say) is of smaller order than the sample size (multiplied by a logarithmic factor). If the jumps are relatively large [specifically, if η is of size $(n^{-1}\log n)^{1/2}$ or larger], then the wavelet estimator will converge at the rate for a smooth function if and only if $m = O(n^{1/(2r+1)})$ (again, multiplied by a suitable logarithmic factor). For smaller jump sizes the necessary and sufficient condition is $m\eta^2 = O(n^{-(2r-1)/(2r+1)})$.

We achieve these results by developing explicit formulas for terms which, in the case of fixed f, are of second order, and monitoring the effect that changing f (by allowing an unboundedly large number of jumps) has on those terms. The convergence rate starts to deteriorate, relative to that for a fixed, piecewise-smooth f, when second-order terms begin to dominate first-order ones. We shall develop these ideas only in the case of nonparametric density estimation, but a longer version of the paper [Hall, McKay and Turlach (1995)] addresses the context of nonparametric regression, and discusses numerical performance. In particular, our theoretical result that mean integrated squared error increases very nearly linearly in number of jumps, and has slope that is virtually independent of noise variance (in the context of regression), is well supported by numerical work. So also is our theoretical conclusion that the optimal value of the truncation point (q, in the analysis in Section 2) increases with number of jumps.

The advantages of wavelet thresholding in the context of curve estimation were first appreciated and analyzed by Donoho, Johnstone, Kerkyacharian and Picard in an important series of papers [e.g., Donoho, Johnstone, Kerkyacharian and Picard (1995, 1996), Kerkyacharian and Picard (1993) and Donoho and Johnstone (1994, 1995)]. Their contributions built upon earlier work on wavelet methods in statistics, for example, that of Doukhan (1988). Concise first-order mean squared error formulas for fixed, piecewise-smooth functions were developed by Hall and Patil (1995).

2. Methodology and theoretical results.

2.1. *Model for the target function.* Our model is of piecewise-continuous densities f whose complexity increases as the sample size becomes larger, so that the estimation problem becomes more difficult as the amount of information grows. Specifically, we assume that f has an unboundedly large number of jumps of unboundedly small size. For technical convenience we ask that for a sample of size n all the jumps be of size $\eta = \eta_n$, this quantity being assumed bounded but possibly converging to 0. The number of jumps within the support of f is denoted by $m-1$, where $m = m_n \geq 1$ may increase without bound. Variants of this model may be treated using similar arguments. The

simplest variant is that where the lth jump is of size $y_{nl}\,\eta$, say, and the y_{nl}'s are bounded away from 0 and ∞. Results in this context differ in only relatively minor, rather obvious ways from those where the jumps are all equal, and so do not require specific attention.

Likewise, the case where jump sizes may be grouped into a finite number of different classes, corresponding to a finite number of pairs (m, η) that vary in different ways, is effectively subsumed by our model. In that context the overall convergence rate is the worst of the convergence rates for the different pairs.

We further assume that consecutive jumps in f are spaced $1/m$ apart. This condition may also be relaxed, for example, by asking that spacings vary within the range $[\varepsilon/m, 1/\varepsilon m]$ for some $\varepsilon \in (0, 1)$, but, again, such generalizations produce only minor, rather obvious alterations to our results.

There is a wide variety of ways of specifying f between jumps. To minimize the notational burden, we suppose that f is constructed by vertical shifts of a single smooth function g, although many other approaches could be adopted. Details are sketched in Remark 2.4.

Bearing these goals in mind, our class of functions f may be defined as follows. Let $\alpha, \beta > 1$ denote fixed constants, and let g be a function on $\mathscr{I} = [0, 1]$ enjoying $r \geq 1$ bounded, uniformly continuous derivatives and such that $\alpha^{-1} \leq g \leq \alpha$. Divide $\mathscr{I} = (0, 1]$ into $m \geq 1$ equal parts $\mathscr{I}_l \equiv ((l-1)/m, l/m]$, and let $\eta \in (0, \alpha]$. Let f denote a density supported on $\mathscr{I}$ such that, for each $1 \leq l \leq m$, f restricted to $\mathscr{I}_l$ is identical to $g + k_l\eta$ for some integer k_l, where $|k_{l+1} - k_l| = 1$ for $1 \leq l \leq m-1$, and $\beta^{-1} \leq f \leq \beta$. Write $\mathscr{F} = \mathscr{F}(\alpha, \beta, g; m, \eta)$ for the class of all f's defined in this way.

For each value n of sample size we select a pair (m_n, η_n) and a candidate $f_n \in \mathscr{F}_n = \mathscr{F}(\alpha, \beta, g; m_n, \eta_n)$, and consider estimating f_n from data drawn from this distribution. We suppress the subscript n in our notation, particularly from f_n, (m_n, η_n) and $\mathscr{F}_n$.

2.2. *Wavelet expansion.* Let ϕ (the "father wavelet" or scale function) and ψ (the "mother wavelet") be bounded, Hölder continuous and compactly supported. Normalize ϕ so that it integrates to unity, and suppose that integer translates of ϕ are orthogonal. We call these conditions (W). If they hold, then, for any $i \geq 0$, $-\infty < j < \infty$ and $p > 0$ with $p_i = 2^i p$, the functions

$$\phi_j(x) \equiv p^{1/2}\,\phi(px - j), \qquad \psi_{ij}(x) \equiv p_i^{1/2}\,\psi(p_i x - j)$$

are orthonormal: $\int \phi_{j_1}\,\phi_{j_2} = \delta_{j_1, j_2}$, $\int \psi_{i_1 j_1}\,\psi_{i_2 j_2} = \delta_{i_1 i_2}\,\delta_{j_1 j_2}$ and $\int \phi_{j_1}\,\psi_{i j_2} = 0$, where δ_{ij} denotes the Dirac delta function. An arbitrary square-integrable function f may be expanded as a generalized Fourier series

$$f(x) = \sum_j b_j \phi_j(x) + \sum_{i=1}^{\infty} \sum_j b_{ij}\,\psi_{ij}(x),$$

which converges in mean square. In this formula $b_j = \int \phi_j f$ and $b_{ij} = \int \psi_{ij} f$.

The scale function is determined by a dilation equation, $\phi(x) = \sum c_j \phi(2x - j)$, for positive constants c_j satisfying $\sum c_j = 2$. Since ϕ is compactly supported, only a finite number of the c_j's are nonzero. Let $r \geq 1$ denote the maximal integer such that

$$\sum_j (-1)^j j^k c_j = 0 \quad \text{for all } 0 \leq k \leq r-1,$$

and define

$$\kappa = (r!)^{-1} \int y^r \psi(y)\, dy = (r!)^{-1} (-\tfrac{1}{2})^{r+1} \sum (-1)^j j^r c_j.$$

These properties and features of wavelets are discussed in much greater detail by Strang (1989, 1993), Meyer (1990) and Daubechies (1992).

The quantity p is sometimes called the "primary resolution level" of the wavelet expansion of f. In the contexts of both density estimation and nonparametric regression, where the wavelet expansion is approximated by an empirical version, p plays the role of the inverse of bandwidth. Details are given in Remark 2.1.

2.3. *Estimation.* Let $X_1, \ldots, X_n$ denote independent random variables having a density $f \in \mathscr{F}$. Put $\hat{b}_j = n^{-1} \sum_k \phi_j(X_k)$ and $\hat{b}_{ij} = n^{-1} \sum_k \psi_{ij}(X_k)$, being unbiased estimators of b_j and b_{ij}, respectively. Our wavelet estimator of f is a thresholded, truncated form of the Fourier expansion of f:

$$hf(x) = \sum_j \hat{b}_j \phi_j(x) + \sum_{i=1}^{q-1} \sum_j \hat{b}_{ij}\, I(|\hat{b}_{ij}| > \delta)\, \psi_{ij}(x),$$

where $\delta > 0$ is the threshold and $q \geq 1$ is an appropriate truncation point for the infinite series over different resolution levels.

2.4. *Main theorem.* We begin with notation. Write $\log n$ to denote a natural logarithm, and $\log_2 n$ for the logarithm function base 2. Define

$$\Psi(x) \equiv \int_{u \leq x} \psi(u)\, du.$$

We assume that the variable parameters $m = m(n)$, $p = p(n)$, $q = q(n)$ and $\eta = \eta(n)$ satisfy

$$\begin{aligned} &m \geq 1, \qquad 0 < \eta = O(1), \qquad \log n = O(q), \\ &n/(p^{2r+1} \log n) \to 0, \qquad p_q (\log n)^6 / n \to 0. \end{aligned} \tag{2.1}$$

We note that by suitably rescaling we can always guarantee that

$$\sup_{-\infty < j < \infty} \inf_{0 \leq \varepsilon < 1} |\Psi(\varepsilon + j)| > 0. \tag{2.2}$$

On occasion we shall also ask that either

$$m/p = O(1) \tag{2.3}$$

or

$$m/p \to 0. \tag{2.4}$$

If we write $\xi_{il} \in [0, 1)$ for the fractional part of $p_i l m^{-1}$, then, as a result of (2.2), the quantities

$$s_{il}^2 \equiv \sum_j \Psi(\xi_{il} + j)^2, \qquad s^2 \equiv p_q (m-1)^{-1} \sum_{i=q}^{\infty} \sum_{l=1}^{m-1} p_i^{-1} s_{il}^2,$$

$$s_0^2 = p_q^{-1} \sum_{i=q}^{\infty} p_i^{-1} s_{i0}^2, \qquad s_1^2 = p_q^{-1} \sum_{i=q}^{\infty} p_i^{-1} s_{im}^2$$

are bounded away from 0 and ∞ as $n \to \infty$. (We define $s^2 = 1$ if $m = 1$.) Let the threshold be

$$\delta = C(2n^{-1} \log n)^{1/2}, \tag{2.5}$$

where $C > 0$ is a constant.

Put $\eta_1 = \min(\eta, 1)$, and define $\rho \equiv n\eta^2/(p \log n)$, $A \equiv 2C\pi^{-1/2}$,

$$B \equiv \left(1 - 2^{-2r}\right)^{-1} \kappa^2 \int \left(g^{(r)}\right)^2,$$

$$a_n \equiv n^{-1} (\log n)^{1/2} p_q \int f^{1/2} n^{-C^2/f},$$

$$b_n \equiv \begin{cases} m\eta^2/p_q, & \text{if } m \le p_q, \\ (m\eta/p_q)^2, & \text{if } p_q < m \le p_q/\eta_1, \\ 1, & \text{if } m > p_q/\eta_1. \end{cases}$$

Let $B_1 = 0$ if g restricted to $(0, 1)$ is a polynomial of degree $r - 1$, and $B_1 = 1$ otherwise.

The theorem below presents an expansion of the dominant contributions to mean integrated squared error, or MISE; the latter is defined as $\int E(\hat{f} - f)^2$. Results (2.6), (2.7) and (2.8) are successively more refined, and hold under successively stronger assumptions. We adopt the convention that, for arbitrary sequences $\{u_n\}$ and $\{v_n\}$ of positive numbers, $u_n \sim v_n$ means that the ratio u_n/v_n converges to 1 as n increases, and $u_n \asymp v_n$ means that $u_n = O(v_n)$ and $v_n = O(u_n)$.

THEOREM. *Under conditions* (W) *and* (2.1), *and with the threshold defined by* (2.5),

$$\int E(\hat{f} - f)^2 = O\left\{n^{-1} p + p^{-2r} + a_n + b_n + p_q^{-1} + mn^{-1} \min(1, \rho) \log n\right\}. \tag{2.6}$$

If, in addition, (2.2) *and* (2.3) *hold, then*

$$\begin{aligned} \int E(\hat{f} - f)^2 \asymp{}& n^{-1} p + p^{-2r} B_1 + p_q^{-1}(m\eta^2 + 1) + a_n \\ &+ mn^{-1} \min(1, \rho) \log n + o(p^{-2r}); \end{aligned} \tag{2.7}$$

and if (2.2) *and* (2.4) *hold, then there exists a positive sequence* $\{\zeta_n\}$, *bounded away from* 0 *and* ∞ *uniformly in* $f \in \mathscr{F}$, *such that*

$$\begin{aligned}\int E(\hat{f}-f)^2 \sim{}& n^{-1}p + p^{-2r}B + (m-1)p_q^{-1}\eta^2 s^2 \\ &+ p_q^{-1}\{f(0+)^2 s_0^2 + f(1-)^2 s_1^2\} \\ &+ a_n A + \zeta_n m n^{-1}\min(1,\rho)\log n + o(p^{-2r}).\end{aligned} \tag{2.8}$$

A proof is outlined in Section 3. The arguments there are sufficient to prove that the bounds and asymptotic equivalences claimed in the theorem are available uniformly in $f \in \mathscr{F}$. Below, in a sequence of remarks, we discuss the theorem's main consequences. There, to avoid trivialities, we shall assume that g restricted to $(0,1)$ is not simply a polynomial of degree $r-1$. We shall say that u_n and v_n are "of the same size" if $u_n \asymp v_n$, and that u_n is "of order v_n" if $u_n = O(v_n)$.

REMARK 2.1 (Necessary and sufficient conditions for achieving specified convergence rates). If $C \geq (\sup f)^{1/2}$, then the terms in a_n in (2.6)–(2.8) are negligibly small, and may be dropped. We shall assume in this remark that this has been done. If m and η are fixed, so that f is smooth in a piecewise sense, then the optimal convergence rate of MISE is $O(n^{-2r/(2r+1)})$, and is achieved when p is of size $n^{1/(2r+1)}$. See Hall and Patil (1995). We observe from (2.7) that if (2.1) and (2.2) hold, but m and η are allowed to vary, then the same convergence rate is achieved if and only if p is of size $n^{1/(2r+1)}$, $m\eta^2 = O(n^{-2r/(2r+1)}p_q)$ and

$$m\min\bigl(1, n^{2r/(2r+1)}\eta^2/\log n\bigr) = O\bigl(n^{1/(2r+1)}/\log n\bigr).$$

Additionally, using (2.6), we may readily derive sufficient conditions for consistency. Indeed, if $\eta = O(n^{-\gamma})$ for some $\gamma \geq 0$, and if p_q is of size $n(\log n)^{-\delta}$ for some $\delta > 6$, then the wavelet estimator is mean-square consistent if (a) $m = o\{n^{1+\gamma}(\log n)^{-\delta}\}$ and (b) $p \to \infty$ more rapidly than $(n/\log n)^{1/(2r+1)}$.

REMARK 2.2 [Importance of conditions (2.3) and (2.4)]. If m/p diverges as n increases, then all the wavelet coefficients b_j, and also the b_{ij}'s for sufficiently small $i \geq 1$, are affected by an unboundedly large number of jump discontinuities. In this case a concise analysis can depend on positions of jumps relative to features of the mother and father wavelet functions, and results such as (2.8) do not appear possible. Similar difficulties arise in the case of (2.7) if m/p converges to a finite, positive limit.

REMARK 2.3 [Sizes of a_n and $p_q^{-1}(m\eta^2+1)$]. Suppose $f(x) = f_n(x) \leq C^2\xi^{-1}$ for all n and x, where $\xi > 0$ and C is as in (2.5). Then, if $q \leq (\xi-\varepsilon)\log_2 n + O(1)$ for some $\varepsilon > 0$, a_n is negligible to first order: $a_n = o(n^{-1}p + p^{-2r})$. Similarly, it may be proved that if f is smooth in a neighborhood of its global maximum, the condition $q \leq \xi\log_2 n + O(1)$ is necessary for first-order negligibility of a_n.

If p is of the size (i.e., $n^{1/(2r+1)}$; see Remark 2.1) which minimizes MISE in the case of a smooth f, then, in order for $p_q^{-1}(m\eta^2+1)$ to be of smaller order than $n^{-1}p + p^{-2r}$, uniformly in pairs (m, η) satisfying (2.1) and (2.4), it is necessary and sufficient that $q \geq \{2r/(2r+1)\}\log_2 n + O(1)$. From this result and that in the previous paragraph we see that, in a wide range of circumstances, in order for a_n and $p_q^{-1}(m\eta^2+1)$ to both equal $o(n^{-1}p + p^{-2r})$ for all (m, η) satisfying (2.1) and (2.4), it is necessary that C be chosen so large that $\sigma \equiv \sup f \leq \frac{1}{2}C^2(2+r^{-1})$ and that q satisfy

$$2(2+r^{-1})^{-1}\log_2 n + O(1) \leq q \leq C^2\sigma^{-1}\log_2 n + O(1),$$

and sufficient that $\sigma < C^2(2+r^{-1})$ and

$$2(2+r^{-1})^{-1}\log_2 n + O(1) \leq q \leq (C^2\sigma^{-1} - \varepsilon)\log_2 n + O(1)$$

for some $\varepsilon > 0$. For these choices of C and q, and for all (m, η) satisfying (2.1) and (2.4), mean integrated squared error in the presence of many discontinuities admits the same asymptotic formula that it does for a smooth function.

REMARK 2.4 (Densities f with variation). The total variation of f equals $m\eta + \int |g'|$, which is bounded if and only if $m\eta$ is. In this case $m\eta^2 = O(1)$, and equals $o(1)$ if m diverges. Therefore, if f is of bounded variation, then second-order terms in an expansion of MISE are of the same size they would be for fixed f. In particular, first-order theory such as that described by Hall and Patil (1995) remains valid.

REMARK 2.5 (Different ways of aligning f at discontinuities). Our definition of f ensures that all derivatives except the zeroth match perfectly at discontinuities. Identical results may be obtained under the assumption that f be bounded away from 0 on $\mathscr{I}$; that between discontinuities, f and its derivatives be uniformly bounded; and that at discontinuities (where f jumps by an amount η) kth derivatives jump by an amount of smaller order than $\eta p^{k/2}$ for each $1 \leq k \leq r$. Of course, the latter restriction is only significant when η is relatively small. It is necessary to replace $\int (g^{(r)})^2$ by $\int (f^{(r)})^2$, where $f^{(r)}$ is defined arbitrarily at jumps, in the definition of B.

3. Proof of the theorem. We derive only (2.6) and (2.8), since a proof of (2.7) is similar.

STEP 1. Preliminaries. In view of the orthogonality properties of wavelets and the wavelet expansion of f,

$$\int (\hat{f} - f)^2 = S_1 + \cdots + S_4, \tag{3.1}$$

where

$$S_1 = \sum_j (\hat{b}_j - b_j)^2, \qquad S_2 = \sum_{i=1}^{q-1} \sum_j (\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta),$$

$$S_3 = \sum_{i=1}^{q-1} \sum_j b_{ij}^2 I(|\hat{b}_{ij}| \le \delta), \qquad S_4 = \sum_{i=q}^{\infty} \sum_j b_{ij}^2.$$

Choose $v > 0$ so large that the support of ϕ and ψ is contained in $(-v, v)$. For each $0 \le l \le m$ and $i \ge 1$, put $\mathscr{K}_{il} = \{j\colon j \in (p_i l m^{-1} - v, p_i l m^{-1} + v)\}$. Note particularly that each set $\mathscr{K}_{il}$ contains no more than $2v + 1$ elements. Unless $j \in \mathscr{K}_{il}$ for some l, the wavelet coefficients b_{ij} are determined entirely from integrals over regions where f is continuous, and likewise the empirical wavelet coefficients $\hat{b}_{ij}$ are constructed entirely from data within such regions. Put $\mathscr{K}_{(i)} = \bigcup_l \mathscr{K}_{il}$ and let $\widetilde{\mathscr{K}}_{(i)}$ denote the complement of $\mathscr{K}_{(i)}$ in the set of all integers. In this notation we may for sufficiently large n write $S_j = S'_j + S''_j$ for $j = 2, 3$, where

$$S'_2 = \sum_{i=1}^{q-1} \sum_{j \in \widetilde{\mathscr{K}}_{(i)}} (\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta),$$

$$S'_3 = \sum_{i=1}^{q-1} \sum_{j \in \widetilde{\mathscr{K}}_{(i)}} b_{ij}^2 I(|\hat{b}_{ij}| \le \delta),$$

$$S_5 = \sum_{i=1}^{q-1} \sum_{j \in \mathscr{K}_{(i)}} \{(\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta) + b_{ij}^2 I(|\hat{b}_{ij}| \le \delta)\}.$$

Therefore, by (3.1),

$$\int (\hat{f} - f)^2 = S_1 + S'_2 + S'_3 + S_4 + S_5. \tag{3.2}$$

STEP 2. Approximation to $E(S_1)$. Analogously to $\mathscr{K}_{il}$, define $\mathscr{K}_l = \{j\colon j \in (plm^{-1} - v, plm^{-1} + v)\}$. Let $\widetilde{\mathscr{K}}$ denote the complement of $\bigcup_l \mathscr{K}_l$ in the set of all integers, and put

$$\sigma_j^2 \equiv n \operatorname{var}(\hat{b}_j) = \int \phi(y)^2 f\{(y + j)/p\}\, dy - b_j^2.$$

If $j \in \widetilde{\mathscr{K}}$, then, by Taylor expansion of $f\{(y + j)/p\}$ about $f(j/p)$,

$$b_j = p^{-1/2} \int \phi(y) f\{(y + j)/p\}\, dy = p^{-1/2} f(j/p) + O(p^{-3/2}),$$

$$\int \phi(y)^2 f\{(y + j)/p\}\, dy = f(j/p) + O(p^{-1}),$$

where the "$O(\cdot)$" terms are of the stated orders uniformly in $j \in \widetilde{\mathscr{K}}$ and $f \in \mathscr{F}$. More simply, $|b_j| = O(p^{-1/2})$ uniformly in $-\infty < j < \infty$ and $f \in \mathscr{F}$. Note, too,

that σ_j^2 is nonzero only for $O(p)$ values of j. Hence, (a), without assuming (2.4),

$$nE(S_1) = \sum_j \sigma_j^2 = O(p) \tag{3.3a}$$

uniformly in $f \in \mathscr{F}$; and (b) a slightly finer analysis gives $p\{1 + O(\eta m/p) + o(1)\}$ in place of $O(p)$ so that assuming (2.4) gives

$$nE(S_1) = \sum_j \sigma_j^2 = \sum_{j \in \widetilde{\mathscr{K}}} f(j/p) + O(m) = \{1 + o(1)\}p, \tag{3.3b}$$

again uniformly in $f \in \mathscr{F}$.

STEP 3. Approximation to $E(S_3')$. Observe that

$$\begin{aligned} b_{ij} &= p_i^{-1/2} \int \psi(y) f\{(y+j)/p_i\}\, dy \\ &= \kappa p_i^{-\{r+(1/2)\}} f^{(r)}(j/p_i) + o\big(p_i^{-\{r+(1/2)\}}\big) \end{aligned} \tag{3.4}$$

uniformly in $j \in \widetilde{\mathscr{K}}_{(i)}$ and $f \in \mathscr{F}$. Furthermore,

$$\sigma_{ij}^2 \equiv n \operatorname{var}(\hat{b}_{ij}) = \int \psi(y)^2 f\{(y+j)/p_i\}\, dy - b_{ij}^2, \tag{3.5}$$

and so $n \operatorname{var}(\hat{b}_{ij})$ is bounded uniformly in j. It follows from these results and the fact that $n/(p^{2r+1} \log n) \to 0$ that $E(\hat{b}_{ij}^2) = o(n^{-1} \log n)$ uniformly in pairs (i, j) such that $j \in \widetilde{\mathscr{K}}_{(i)}$. Hence, by Chebyshev's inequality, $P(|\hat{b}_{ij}| \le \delta) \to 1$ uniformly in such (i, j). It follows that (a), without assuming (2.4),

$$E(S_3') = \sum_{i=1}^{q-1} \sum_{j \in \widetilde{\mathscr{K}}_{(i)}} b_{ij}^2 P(|\hat{b}_{ij}| \le \delta) = O\big(p^{-2r}\big), \tag{3.6a}$$

and (b), assuming (2.4),

$$E(S_3') = \{1 + o(1)\}\big(1 - 2^{-2r}\big)^{-1} \kappa^2 p^{-2r} \int \big(g^{(r)}\big)^2, \tag{3.6b}$$

both uniformly in $f \in \mathscr{F}$.

STEP 4. Approximation to S_4. When (2.4) holds,

$$b_{ij} = p_i^{-1/2} \eta \Psi(p_i l m^{-1} - j) + O\big(\eta p_i^{-3/2} + p_i^{-\{r+(1/2)\}}\big) \tag{3.7}$$

uniformly in (i, j) such that $j \in \mathscr{K}_{il}$ and $1 \le l \le m - 1$, and in $f \in \mathscr{F}$. The analogous results for $l = 0$ and $l = m$ have η replaced by $f(0+)$ and $f(1-)$,

respectively. Therefore,

$$
\begin{aligned}
&\sum_{i=q}^{\infty}\sum_{l=0}^{m}\sum_{j\in\mathscr{K}_{il}} b_{ij}^2 \\
&\quad = \eta^2\sum_{i=q}^{\infty}\sum_{l=1}^{m-1}\sum_{j\in\mathscr{K}_{il}} p_i^{-1}\Psi(p_i l m^{-1}-j)^2 \\
&\qquad + \sum_{i=q}^{\infty} p_i^{-1}\{\Psi(-j)^2 f(0+)^2 + \Psi(p_i-j)^2 f(1-)^2\} \\
&\qquad + O\Bigl\{\sum_{i=q}^{\infty}\sum_{l=1}^{m-1}\sum_{j\in\mathscr{K}_{il}} \bigl(\eta^2 p_i^{-2} + \eta p_i^{-(r+1)} + p_i^{-(2r+1)}\bigr)\Bigr\} + o\bigl(p_q^{-1}\bigr) \\
&\quad = s^2(m-1)\eta^2 p_q^{-1} + p_q^{-1}\{f(0+)^2 s_0^2 + f(1-)^2 s_1^2\} \\
&\qquad + O\bigl\{m\bigl(\eta^2 p_q^{-2} + \eta p_q^{-(r+1)} + p_q^{-(2r+1)}\bigr)\bigr\} + o\bigl(p_q^{-1}\bigr).
\end{aligned}
\tag{3.8}
$$

Next, we treat the terms b_{ij} not covered by these bounds. By (3.4), $|b_{ij}| = O(p_i^{-\{r+(1/2)\}})$ uniformly in (i, j) such that $j \in \widetilde{\mathscr{K}_{(i)}}$, and uniformly in $f \in \mathscr{F}$. For each i, b_{ij} is nonzero only for $O(p_i)$ values of j, and so the contribution made to S_4 by the sum of b_{ij}^2 over $j \in \widetilde{\mathscr{K}_{(i)}}$ is of order p_i^{-2r} uniformly in i and $f \in \mathscr{F}$. Furthermore, the sum of p_i^{-2r} over $i \geq q$ is of order p_q^{-2r}. This result and (3.8) imply that

$$
\begin{aligned}
S_4 = {} & s^2(m+1)\,\eta^2 p_q^{-1} + p_q^{-1}\{f(0+)^2 s_0^2 + f(1-)^2 s_1^2\} \\
& + o\bigl(m\eta^2 p_q^{-1} + p^{-2r}\bigr)
\end{aligned}
\tag{3.9b}
$$

uniformly in $f \in \mathscr{F}$.

Finally, we treat the case where (2.4) is not assumed. Write N_{ij} for the number of discontinuities of f in the interval $((j-v)/p_i, (j+v)/p_i)$, where $j \in \mathscr{K}_{(i)}$. Then $1 \leq N_{ij} \leq \text{const.}\max(1, m/p_i)$. Furthermore, using essentially the argument at (3.7), $|b_{ij}| = O\{\min(1, N_{ij}\eta)p_i^{-1/2} + p_i^{-\{r+(1/2)\}}\}$ for $j \in \mathscr{K}_{il}$ and $1 \leq l \leq m-1$. Also, the number of elements of $\mathscr{K}_{(i)}$ equals $O\{\min(m, p_i)\}$. Therefore,

$$
\begin{aligned}
S &\equiv \sum_{i=q}^{\infty}\sum_{j\in\mathscr{K}_{(i)}} b_{ij}^2 \\
&= O\Bigl[\sum_{i=q}^{\infty}\sum_{j\in\mathscr{K}_{(i)}} \bigl\{\min\bigl(1, N_{ij}^2\eta^2\bigr)\, p_i^{-1} + p_i^{-(2r+1)}\bigr\} + p_q^{-1}\Bigr] \\
&= O\bigl(\eta^2[(m/p_q)\, I(m \leq p_q) + \{1+\log(m/p_q)\}\, I(m > p_q)] + p_q^{-1} + T\bigr),
\end{aligned}
$$

where

$$T \equiv \sum_{i \geq q:\ p_i \leq m} p_i \min\{1, (\eta m/p_i)^2\}\, p_i^{-1}$$
$$= O\Big[\sum_{i=q}^{\infty} \min\{1, (\eta m/p_i)^2\}\, I(p_i \leq m)\Big].$$

If $\eta m/p_q \leq 1$, then $T \leq \text{const.}(\eta m/p_q)^2$; otherwise, we use the bound $S \leq \int f^2$. From these results and the fact that the sum of b_{ij}^2 over $j \in \widetilde{\mathscr{K}}_{(i)}$ and $i \geq q$ is of order p_q^{-2r}, we deduce that

$$S_4 = O\big\{(\eta^2 m/p_q)\, I(m \leq p_q) + (\eta m/p_q)^2\, I\big(p_q < m \leq \eta_1^{-1} p_q\big) + I\big(m > \eta_1^{-1} p_q\big) + p_q^{-1}\big\}. \tag{3.9a}$$

In the next three steps we develop approximations to $E(S_2')$ and $E(S_5)$. Our technique is to work first under the fictitious assumption, which we denote by (N), that $\hat{b}_{ij}$ is exactly normally distributed with the same mean and variance as $\hat{b}_{ij}$. Then we develop bounds to the error committed by making this assumption.

STEP 5. Approximation to $E(S_2')$. Let Z denote a random variable with the standard normal distribution, and write $t = C_1(2\log n)^{1/2}$ and $u = C_2(2\log n)^{1/2}$, where $C_1,\ C_2 \geq 0$. Put $e_n(C_1, C_2) \equiv E\{Z^2\, I(|Z + t| > u)\}$. Integration by parts shows that

$$e_n(C_1, C_2) = \pi^{-1/2} (\log n)^{1/2} n^{-C_2^2} n^{-C_1^2} \times \big[(C_2 + C_1) n^{-2C_1C_2} + (C_2 - C_1) n^{2C_1C_2}\big]\big(1 + O(\log n)^{-1}\big)$$

uniformly over all $C_2 - C_1 \geq \varepsilon$ for an arbitrary $\varepsilon > 0$. Letting C_1 tend to 0, we see that, for a fixed but arbitrary sequence $\varrho = \varrho(n) \to 0$,

$$e_n(C_1, C_2) = \pi^{-1/2}\, C_2\, n^{-C_2^2} (\log n)^{1/2} \{2 + o(1)\} \tag{3.10}$$

uniformly in values of $C_1 = C_1(n)$ and $C_2 = C_2(n)$ satisfying $|C_1| \leq \varrho(n)/\log n$ and $C_2 = o(1/\varrho(n))$. We will choose $\varrho(n) = [n/(p^{2r+1}\log n)]^{1/2}$; by assumption, this tends to 0.

Under assumption (N) we may write $e_{ij} \equiv E\{(\hat{b}_{ij} - b_{ij})^2\, I(|\hat{b}_{ij}| > \delta)\}$ as $e_{ij} = n^{-1}\sigma_{ij}^2\, e_n(C_{1ij}, C_{2ij})$, where σ_{ij}^2 is given by (3.5), $C_{1ij} = (n/2\log n)^{1/2}\, |b_{ij}|/\sigma_{ij}$ and $C_{2ij} = C/\sigma_{ij}$. Since f is bounded away from 0 and ∞, and $|b_{ij}| \to 0$ uniformly in pairs (i, j) such that $j \in \widetilde{\mathscr{K}}_{(i)}$, then σ_{ij} is bounded away from 0 and ∞ uniformly in such i and j for all sufficiently large n. From this result and (3.4), noting too that $n/(p^{2r+1}\log n) \to 0$, we may deduce that

$$C_{1ij} = O\big[\{n/(p_i^{2r+1}\log n)\}^{1/2}\big] = o(1) \tag{3.11}$$

uniformly in $i \geq 1$ and $j \in \widetilde{\mathscr{K}}_{(i)}$. Similarly, if

$$i_0 \equiv 2(1 + \varepsilon)\{(2r + 1)\log 2\}^{-1} \log\log n$$

and

$$\mathscr{T} = \mathscr{T}(\varepsilon) = \{(i, j)\colon i \geq i_0,\ \ j \in \widetilde{\mathscr{K}_{(i)}}\}$$

for $\varepsilon > 0$, then

$$\begin{aligned}\sup_{(i,j)\in\mathscr{T}} C_{1ij} &= O\Big[\sup_{(i,j)\in\mathscr{T}} \big\{np^{-(2r+1)}(\log n)^{-1}2^{-(2r+1)i}\big\}^{1/2}\Big] \\ &= o\big\{(\log n)^{-(1+\varepsilon)}\big\}.\end{aligned}$$

Hence, by (3.10) and (3.11),

$$E(S_2') = \sum_{i=1}^{q-1} \sum_{j\in\widetilde{\mathscr{K}_{(i)}}} e_{ij} = \{2 + o(1)\}\, T_1 + T_2, \tag{3.12}$$

where, for a sequence $\xi = \xi(n) \downarrow 0$,

$$\begin{aligned} T_1 &\equiv \pi^{-1/2} n^{-1} (\log n)^{1/2} \sum_{i=i_0}^{q-1} \sum_{j\in\widetilde{\mathscr{K}_{(i)}}} \sigma_{ij}^2 C_{2ij} n^{-C_{2ij}^2}, \\ T_2 &= O\Big[n^{-1}(\log n)^{1/2} \sum_{i=1}^{i_0-1} \sum_{j\in\widetilde{\mathscr{K}_{(i)}}} \exp\big\{\xi(\log n)/2^{(2r+1)i}\big\} n^{-C_{2ij}^2}\Big], \end{aligned} \tag{3.13}$$

and the "$o(1)$" and "$O(\cdot)$" terms in (3.12) and (3.13) are of the indicated orders uniformly in $f \in \mathscr{F}$.

Define

$$T_3 \equiv \sum_{i=1}^{q-1} \sum_{j\in\widetilde{\mathscr{K}_{(i)}}} \sigma_{ij}^2 C_{2ij} n^{-C_{2ij}^2}, \qquad T_4 \equiv n^{-1} (\log n)^{1/2}\, p_q \int n^{-C^2/f}$$

and $T_5 \equiv \pi^{-1/2} n^{-1} (\log n)^{1/2} T_3$. Noting that the sum over $1 \leq i < i_0$ includes only $O(\log\log n)$ terms, while that over $i \geq i_0$ includes at least a constant multiple of $\log n$ terms, we may show that $T_1 = T_5 + o(T_4)$ and $T_2 = o(T_4)$. (A detailed proof of these results is arguably simpler if one notes that, without essential loss of generality, it may be assumed that $m/p \to l$, where $0 \leq l \leq \infty$, and treats separately the cases $l = 0$, $0 < l < \infty$ and $l = \infty$.) The remainders in both these relations are of the stated orders uniformly in $f \in \mathscr{F}$. Therefore, by (3.12),

$$E(S_2') = T_5 + o(T_4) \tag{3.14}$$

uniformly in $f \in \mathscr{F}$.

Observe next that $\sigma_{ij}^2 = f(j/p_i) + O(p_i^{-1})$ uniformly in $i \geq 1$, $j \in \widetilde{\mathscr{K}_{(i)}}$ and $f \in \mathscr{F}$. Therefore, $C_{2ij} = Cf(j/p_i)^{-1/2} + O(p_i^{-1})$ uniformly in such i, j

and f. Hence, since $p_i^{-1} \le p^{-1} = O(n^{-\varepsilon})$ for some $\varepsilon > 0$, then $T_3 = T_6 + o\{n(\log n)^{-1/2} T_4\}$, where

$$T_6 \equiv \sum_{i=1}^{q-1} \sum_{j \in \widetilde{\mathscr{K}}_{(i)}} C f(j/p_i)^{1/2} n^{-C^2/f(j/p_i)}.$$

It follows from these results, (3.14) and the argument that we shall give below in our more detailed treatment of the case where (2.4) holds, that, without needing to assume (2.4),

$$\begin{aligned} E(S_2') &= O\left\{ n^{-1} (\log n)^{1/2} \sum_{i=1}^{q-1} \sum_j n^{-C^2/f(j/p_i)} \right\} \\ &= O(T_4). \end{aligned} \tag{3.15a}$$

To conclude, we treat the case where (2.4) holds. Recall the definition $\mathscr{I}_l \equiv ((l-1)/m, l/m]$, let $\mathscr{J}_{il}$ denote the set of indices $j \in \widetilde{\mathscr{K}}_{(i)}$ such that $j/p_i \in \mathscr{I}_l$, and put

$$T_{1il} \equiv \sum_{j \in \mathscr{J}_{il}} C f(j/p_i)^{1/2} n^{-C^2/f(j/p_i)}.$$

In this notation $T_6 = \sum_{i \le q} \sum_{l \le m} T_{1il}$. Noting that $\mathscr{J}_{il}$ is a sequence of consecutive integers contained within the interval $[p_i(l-1)m^{-1} - v, p_i l m^{-1} + v]$ and containing all integers in the interval $(p_i(l-1)m^{-1} + v, p_i l m^{-1} - v)$ and that $\sup_i p_i^{-1} = O(n^{-\varepsilon})$ for some $\varepsilon > 0$, we may deduce that $T_{1il}/T_{2il} = 1 + o(1)$ uniformly in $1 \le i \le q-1$ and $1 \le l \le m$, where

$$T_{2il} \equiv p_i \int_{\mathscr{I}_l} C f(x)^{1/2} n^{-C^2/f(x)} \, dx.$$

Now, $\sum_{1 \le i \le q-1} p_i \sim p_q$ as $p, q \to \infty$, and so

$$\begin{aligned} T_3 &\sim \sum_{i=1}^{q-1} \sum_{l=1}^{m} T_{1i} \sim \sum_{i=1}^{q-1} \sum_{l=1}^{m} T_{2i} \\ &\sim p_q \int_0^1 C f(x)^{-1/2} n^{-C^2/f(x)} \, dx. \end{aligned}$$

Therefore, by (3.14),

$$E(S_2') = \{1 + o(1)\} \pi^{-1/2} n^{-1} (\log n)^{1/2} p_q \int C f^{1/2} n^{-C^2/f}. \tag{3.15b}$$

STEP 6. Approximating $E(S_5)$ under assumption (N). Again, let $t = C_1(2 \log n)^{1/2}$ and $u = C_2(2 \log n)^{1/2}$, where $C_1, C_2 > 0$, let C_{1ij} and C_{2ij} be as in Step 5, let Z denote a standard normal variable, and define $d_{ij} \equiv d_n(C_{1ij}, C_{2ij})$ and

$$d_n(C_1, C_2) \equiv E\{Z^2 I(|Z + t| > u)\} + 2C_1^2 \log n \, P(|Z + t| \le u).$$

In this notation and assuming condition (N),

$$E\{(\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta) + b_{ij}^2 I(|\hat{b}_{ij}| \le \delta)\} = n^{-1}\sigma_{ij}^2 d_{ij}.$$

Hence, assuming condition (N),

$$\begin{aligned} nE(S_5) &= \sum_{i=1}^{q-1} \sum_{j \in \mathscr{K}_{(i)}} \sigma_{ij}^2 d_{ij} \\ &= O\Bigg\{\sum_{i=1}^{q-1} \sum_{j \in \mathscr{K}_{(i)}} \Big(E\big[Z^2 I\{Z > (C_{2ij} - C_{1ij})(2\log n)^{1/2}\}\big] \\ &\qquad + C_{1ij}^2 (\log n) P\{Z > (C_{1ij} - C_{2ij})(2\log n)^{1/2}\}\Big)\Bigg\}. \end{aligned} \tag{3.16}$$

Now, $C_{1ij} \le c_1\{(\rho 2^{-i})^{1/2} + (\rho_1 p_i^{-(2r+1)})^{1/2}\}$ and $C_{2ij} \ge c_2$, where c_1, $c_2 > 0$ and $\rho_1 = n/\log n$. From this result, noting that the number of elements of $\mathscr{K}_{(i)}$ equals $O\{\min(m, p_i)\}$ uniformly in i and that $q = O(\log n)$, we may prove that

$$nE(S_5) = O\{m \min(1, \rho) \log n + n \min(m/p, 1)\, p^{-2r}\}. \tag{3.17a}$$

To derive a lower bound, we assume both (2.2) and (2.4). Define

$$C_{ij} = \int \psi(y)^2 f\{(y+j)/p_i\}\, dy, \qquad D_{ij} \equiv \big|\Psi\big(p_i l m^{-1} - j\big)\big|, \qquad C_{0ij} \equiv D_{ij}/C_{ij}.$$

In this notation, for constants $A_1, A_2, \ldots$ not depending on n and satisfying $0 < A_3 \le A_2 < \infty$ and $0 < A_6 \le A_5 < \infty$, we see from (3.16) that

$$\begin{aligned} &nE(S_5) + o(np^{-2r}) \\ &\quad \ge A_1 \sum_{i=1}^{q-1} \sum_{l=0}^{m} \sum_{j \in \mathscr{K}_{il}} \big\{I(\rho 2^{-i} C_{0ij} > A_2) \\ &\qquad\qquad + \rho 2^{-i} C_{0ij}^2 (\log n)\, I(\rho 2^{-i} C_{0ij} \le A_3)\big\} \\ &\quad \ge A_4 m \sum_{i=1}^{q-1} \big\{I(\rho 2^{-i} > A_5) + \rho 2^{-i} (\log n)\, I(\rho 2^{-i} \le A_6)\big\} \\ &\quad \ge A_7\, m \min(1, \rho) \log n. \end{aligned} \tag{3.17b}$$

The theorem, assuming condition (N), follows from (3.2), (3.3), (3.6), (3.9), (3.15) and (3.17). [Except in the case of (3.17) we employ version (a) of these respective results when establishing (2.6), and version (b) for (2.8). When using (3.17), both versions are needed to derive the last term in (2.8).]

STEP 7. Completion. It remains to show that the error committed by making assumption (N) is of no larger order than the remainder terms in (2.6) and (2.8). To this end, let X_{kij}, $k \ge 1$, denote independent random variables, independent also of the data $X_1, \ldots, X_n$, whose distribution equals that of

X_1 conditional on $X_1 \in \mathscr{I}_{ij} \equiv ((j-v)/p_i, (j+v)/p_i)$, and let M_{ij} equal the number of data values X_t, $1 \le t \le n$, that lie in $\mathscr{I}_{ij}$. Write Z for a standard normal random variable, and define $Z_{ij} \equiv (\hat{b}_{ij} - b_{ij})/(\operatorname{var} \hat{b}_{ij})^{1/2}$. Using Bernstein's inequality and classical methods for estimating large deviations [see, e.g., Petrov (1975), Chapter 8], it may be proved that there exist positive constants D_n, diverging to ∞, such that

$$\left|1 - \frac{P(Z_{ij} > z)}{P(Z > z)}\right| + \left|1 - \frac{P(Z_{ij} \le -z)}{P(Z > z)}\right| = o(1) \tag{3.18}$$

uniformly in $0 \le z \le D_n \log n$, and integers i, j such that $1 \le i \le q$ and $\mathscr{I}_{ij} \subseteq \mathscr{I} = (0, 1]$. For each i there are no more than two values of j such that $\mathscr{I}_{ij} \cap \mathscr{I}$ is nonempty. This result may be employed to show that the error arising from assumption (N) is negligibly small. □

Acknowledgments. Constructive comments of three referees, an Editor and an Associate Editor are gratefully acknowledged.

REFERENCES

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

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Minimax estimation via wavelet shrinkage. Unpublished manuscript.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B* **57** 301–369.

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539.

DOUKHAN, P. (1988). Formes de Toeplitz associées à une analyse multi-échele. *C. R. Acad. Sci. Paris Sér. I Math.* **306** 663–668.

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

HALL, P., MCKAY, I. and TURLACH, B. (1995). Performance of wavelet methods for functions with many discontinuities. Technical Report SRR 029-95, Centre for Mathematics and its Applications, Australian National Univ.

KERKYACHARIAN, G. and PICARD, D. (1993). Introduction aux ondelettes et estimation de densité, 1: introduction aux ondelettes et à l'analyse multiresolution. Lecture notes, Univ. Paris XII.

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

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

STRANG, G. (1989). Wavelets and dilation equations: a brief introduction. *SIAM Rev.* **31** 614–627.

STRANG, G. (1993). Wavelet transforms versus Fourier transforms. *Bull. Amer. Math. Soc.* **28** 288–305.

CENTRE FOR MATHEMATICS AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA ACT 0200
AUSTRALIA
E-MAIL: halpstat@fac.anu.edu.au
berwin@alphasun.anu.edu.au