The Annals of Statistics
1998, Vol. 26, No. 5, 1783–1799

WAVELET SHRINKAGE FOR NONEQUISPACED SAMPLES

BY T. TONY CAI AND LAWRENCE D. BROWN

Purdue University and University of Pennsylvania

Standard wavelet shrinkage procedures for nonparametric regression are restricted to equispaced samples. There, data are transformed into empirical wavelet coefficients and threshold rules are applied to the coefficients. The estimators are obtained via the inverse transform of the denoised wavelet coefficients. In many applications, however, the samples are nonequispaced. It can be shown that these procedures would produce suboptimal estimators if they were applied directly to nonequispaced samples.

We propose a wavelet shrinkage procedure for nonequispaced samples. We show that the estimate is adaptive and near optimal. For global estimation, the estimate is within a logarithmic factor of the minimax risk over a wide range of piecewise Hölder classes, indeed with a number of discontinuities that grows polynomially fast with the sample size. For estimating a target function at a point, the estimate is optimally adaptive to unknown degree of smoothness within a constant. In addition, the estimate enjoys a smoothness property: if the target function is the zero function, then with probability tending to 1 the estimate is also the zero function.

1. Introduction. Suppose we are given data:

$$y_i = f(t_i) + \varepsilon z_i, \tag{1.1}$$

$i = 1, 2, \dots, n$, $0 < t_1 < t_2 < \cdots < t_n = 1$, and z_i are independently and identically distributed as $N(0, 1)$.

The function f is an unknown function of interest. We wish to estimate the function f globally or to estimate f at a point. In the case of recovering the entire function f on $[0, 1]$, one can measure the performance of an estimate $\hat{f}$, for example, by the global squared L_2 norm risk:

$$R(\hat{f}, f) = E\int_0^1 \big(\hat{f}(t) - f(t)\big)^2\, dt.$$

The goal is to construct estimates that have "small" risk. In order to have some meaningful estimate according to this criterion, one must assume certain regularity conditions on the unknown function f, such as f belongs to some Hölder classes, Sobolev classes, Besov classes and so forth.

The more traditional approaches to nonparametric regression include fixed-bandwidth kernel methods, orthogonal series methods and linear spline smoothers. These methods are not adaptive. That is, the estimators based on these methods may achieve substantially slower rate of convergence if the

Received July 1997; revised May 1998.

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

Key words and phrases. Wavelets, multiresolution approximation, nonparametric regression, minimax, adaptivity, piecewise Hölder class.

smoothness of the underlying regression functions is misspecified. In recent years, more efforts have been made to develop adaptive procedures. A variety of adaptive methods have been proposed, such as variable-bandwidth kernel methods and variable-knot spline smoothers.

The recent development of wavelet bases based on multiresolution analyses suggests new techniques for nonparametric function estimation. Wavelets offer a degree of localization both in space and in frequency. This gives great advantage over the traditional Fourier basis. In the recent few years, wavelet theory has been widely applied to the fields of signal and image processing, as well as statistical estimation.

The application of wavelet theory to the field of statistical function estimation was pioneered by Donoho and Johnstone. In a series of important papers (see, e.g., [6], [7] and [9]), Donoho and Johnstone and coauthors present a coherent set of procedures that are spatially adaptive and near optimal over a range of function spaces of inhomogeneous smoothness. Wavelet procedures achieve adaptivity through thresholding of the empirical wavelet coefficients. They enjoy excellent mean squared error properties when used to estimate functions that are only piecewise smooth and have near optimal convergence rates over large function classes. In contrast, traditional linear estimators typically achieve good performance only for relatively smooth functions.

Despite their considerable advantages, however, standard wavelet procedures have limitations. One serious limitation is the requirement of equispaced samples. Standard wavelet procedures are restricted to equispaced samples; that is, t_i in (1.1) are equally spaced on $[0, 1]$. In practice, however, there are many interesting applications in statistics where the samples are not equispaced. In some wavelet software packages, nonequispaced samples are currently treated the same as equispaced ones. As we shall explain later, nonequispaced samples should not in general be treated as equispaced. Otherwise the convergence rate could be far below the optimal rate. Different treatments are needed. So how to apply the wavelet shrinkage method to nonequispaced samples is of practical interest.

We formulate the nonequispaced regression model as follows:

$$y_i = f(t_i) + \varepsilon z_i, \tag{1.2}$$

with $i = 1, 2, \dots, n$, $n = 2^J$, $z_i \overset{\text{iid}}{\sim} N(0, 1)$ and $t_i = H^{-1}(i/n)$ for some cumulative density function H on $[0, 1]$. Note that the design points t_i are assumed to be fixed, not randomly drawn from H.

We develop an adaptive wavelet threshold procedure for the nonequispaced model based on multiresolution analysis and projection as well as nonlinear thresholding. The algorithm for implementing the procedure has the following ingredients:

1. Precondition the data by a sparse matrix.
2. Transform the preconditioned data by the discrete wavelet transform.
3. Denoise the noisy wavelet coefficients via thresholding.

The function with the denoised wavelet coefficients $\hat{\theta}_{jk}$ is our estimate of the function f that we intend to recover. If one is interested in estimating the function at the sample points, two more steps are added:

4. Apply the inverse transform to the denoised coefficients.
5. Postcondition the data by a matrix to get the estimate at the sample points.

Both the preconditioning and postconditioning matrices, defined in (5.3) and (5.4), respectively, are sparse matrices containing $O(n)$ nonzero entries. The preconditioning matrix operation is equivalent to a projection in multiresolution analysis to account for the irregular spacing of the sample points. The postconditioning matrix operation is a step to evaluate the estimated function at the given nonequispaced sample points. Compared to Donoho and Johnstone's VisuShrink, this procedure has two additional steps, preconditioning and postconditioning. The procedure agrees with the VisuShrink when the sample is, in fact, equispaced.

The procedure is near optimal and is adaptive up to the smoothness of the wavelets used. We investigate the adaptivity of the estimators over a wide range of piecewise Hölder classes, indeed with a number of discontinuities that increases polynomially fast with the sample size. We show in Section 4 that the rate of convergence for estimating regression function f globally over the function classes is a logarithmic factor away from the minimax risk. Furthermore, for estimating a target function at a point, the estimate is optimally adaptive to unknown degree of smoothness within a constant factor. The estimate also enjoys a smoothness property. If the target function is the zero function, then the estimate will also be the zero function with probability tending to 1. Therefore, the procedure removes pure noise completely with high probability.

The rest of the paper is organized as follows. Section 2 describes the wavelet basis, multiresolution analysis and wavelet approximation. Section 3 introduces the nonequispaced procedure. Optimality of the estimators will be presented in Section 4. Further discussion about the procedure and related topics are given in Section 5. Section 6 contains proofs of the main results.

2. Wavelets and wavelet approximation. We summarize in this section the basics on wavelets and multiresolution analysis that will be needed in later sections. Further details on wavelet theory can be found in Daubechies [5] and Meyer [14].

An orthonormal wavelet basis is generated from dilation and translation of two basic functions, a "father" wavelet ϕ and a "mother" wavelet ψ. The functions ϕ and ψ are assumed to be compactly supported. Assume that $\text{supp}(\phi) = \text{supp}(\psi) = [0, N]$. Also assume that ϕ satisfies $\int \phi = 1$. We call a wavelet ψ *r-regular* if ψ has r vanishing moments and r continuous derivatives.

Let

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \qquad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$$

and denote the periodized wavelets

$$\phi^p_{jk}(t) = \sum_{l\in\mathcal{Z}} \phi_{jk}(t-l), \qquad \psi^p_{jk}(t) = \sum_{l\in\mathcal{Z}} \psi_{jk}(t-l) \quad \text{for } t \in [0, 1].$$

For simplicity in exposition, we use the periodized wavelet bases on [0, 1] in the present paper. The collection $\{\phi^p_{j_0 k},\ k = 1, \ldots, 2^{j_0};\ \psi^p_{jk},\ j \geq j_0,\ k = 1, \ldots, 2^j\}$ constitutes such an orthonormal basis of $L_2[0, 1]$. Note that the basis functions are periodized at the boundary. The superscript "p" will be suppressed from the notation for convenience.

A wavelet basis has an associated multiresolution analysis on [0, 1]. Let V_j and W_j be the closed linear subspaces generated by $\{\phi_{jk},\ k = 1, \ldots, 2^j\}$ and $\{\psi_{jk},\ k = 1, \ldots, 2^j\}$, respectively. Then:

1. $V_{j_0} \subset V_{j_0+1} \subset \cdots \subset V_j \subset \cdots$;
2. $\overline{\bigcup_{j=j_0}^{\infty} V_j} = L_2([0, 1])$;
3. $V_{j+1} = V_j \oplus W_j$.

The nested sequence of closed subspaces $V_{j_0} \subset V_{j_0+1} \subset \cdots$ is called a multiresolution analysis on [0, 1].

An orthonormal wavelet basis has an associated exact orthogonal discrete wavelet transform (DWT) that transforms sampled data into the wavelet coefficient domain. A crucial point is that the transform is not implemented by matrix multiplication but by a sequence of finite-length filtering that produces an order $O(n)$ orthogonal transform. See [5] and [15] for further details about the discrete wavelet transform.

For a given square-integrable function f on [0, 1], denote

$$\xi_{jk} = \langle f, \phi_{jk}\rangle, \qquad \theta_{jk} = \langle f, \psi_{jk}\rangle.$$

So the function f can be expanded into a wavelet series:

$$f(x) = \sum_{k=1}^{2^{j_0}} \xi_{j_0 k}\phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} \theta_{jk}\psi_{jk}(x). \tag{2.1}$$

The wavelet transform decomposes a function into different resolution components. In (2.1), $\xi_{j_0 k}$ are the coefficients at the coarsest level. They represent the gross structure of the function f. And θ_{jk} are the wavelet coefficients. They represent finer and finer structures of the function f as the resolution level j increases.

We note that the DWT is an orthogonal transform, so it transforms i.i.d. Gaussian noise to i.i.d. Gaussian noise and it is norm preserving. This important property of DWT allows us to transform the problem in the function domain into a problem in the sequence domain of the wavelet coefficients with isometry of risks.

Wavelets provide smoothness characterization of function spaces. Many traditional smoothness spaces, for example, Hölder spaces, Sobolev spaces and Besov spaces, can be completely characterized by wavelet coefficients. See

Meyer [14]. In the present paper, we consider the estimation problem over a range of piecewise Hölder classes. A function in a piecewise Hölder class can be regarded as the superposition of a regular smooth function in a Hölder class and an irregular perturbation consisting of jump discontinuities. In our main results, the maximum number of jump discontinuities is allowed to grow polynomially fast with the sample size. This enables the function classes to effectively model functions of significant spatial inhomogeneity.

DEFINITION 1. A piecewise Hölder class $\Lambda^\alpha(M, B, m)$ on $[0, 1]$ with at most m discontinuous jumps consists of functions f satisfying the following conditions:

1. The function f is bounded by B, that is, $|f| \le B$.
2. There exist $l \le m$ points $0 \le a_1 < \cdots < a_l \le 1$ such that, for all $a_i \le x$, $y < a_{i+1}$, $i = 0, 1, \dots, l$ (with $a_0 = 0$ and $a_{l+1} = 1$),

   (i) $|f(x) - f(y)| \le M|x - y|^\alpha$ if $\alpha \le 1$;
   (ii) $|f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \le M|x - y|^{\alpha'}$ and $|f'(x)| \le B$ if $\alpha > 1$

   where $\lfloor\alpha\rfloor$ is the largest integer less than α and $\alpha' = \alpha - \lfloor\alpha\rfloor$.

In words, the function class $\Lambda^\alpha(M, B, m)$ consists of functions that are piecewise Hölder with the number of discontinuities bounded by m. The following are the upper bounds of wavelet coefficients of functions in a piecewise Hölder class $\Lambda^\alpha(M, B, m)$. Throughout, C denotes a generic constant not depending on function f and sample size n, and the standard notation $\langle\ ,\ \rangle$ denotes inner product in L_2 space.

LEMMA 1. *Let $f \in \Lambda^\alpha(M, B, m)$. Suppose that the wavelet function ψ is r-regular with $r \ge \alpha$. Then*:

(i) *If* $\operatorname{supp}(\psi_{jk})$ *does not contain any jump points of f, then*

$$\theta_{jk} \equiv |\langle f, \psi_{jk}\rangle| \le C2^{-j(1/2+\alpha)}. \tag{2.2}$$

(ii) *If* $\operatorname{supp}(\psi_{jk})$ *contains at least one jump point of f, then*

$$\theta_{jk} \equiv |\langle f, \psi_{jk}\rangle| \le C2^{-j/2}. \tag{2.3}$$

Now suppose we have a dyadically sampled function $\{f(k/n)\}_{k=1}^n$ with $n = 2^J$. We can utilize a wavelet basis and the associated multiresolution analysis to get a good approximation of the entire function f. Let us begin with the following result. The proof is straightforward.

LEMMA 2. *Suppose $f \in \Lambda^\alpha(M, B, m)$. Let $\xi_{Jk} \equiv \langle f, \phi_{Jk}\rangle$ and $s(\alpha) = \min(\alpha, 1)$. Then*:

(i) *If* $\operatorname{supp}(\phi_{Jk})$ *does not contain any jump points of f, then*

$$|n^{-1/2} f(k/n) - \xi_{Jk}| \le Cn^{-(1/2+s(\alpha))}. \tag{2.4}$$

(ii) *If* $\text{supp}(\phi_{Jk})$ *contains jump points of the function* f, *then*

$$|n^{-1/2} f(k/n) - \xi_{Jk}| \le C n^{-1/2}. \tag{2.5}$$

According to this result, we may use $n^{-1/2} f(k/n)$ as an approximation of $\xi_{Jk} = \langle f, \phi_{Jk} \rangle$. This means that if a dyadically sampled function is given, we may use a multiresolution analysis to get an approximation of the projection of the function f onto subspace V_J because ξ_{Jk} are the coefficients of the projection. This in turn provides a good approximation of the entire function f. More specifically, we may use $f_n(t) = \sum_{k=1}^n n^{-1/2} f(k/n)\phi_{Jk}(t)$ as an approximation of f. Based on Lemmas 1 and 2, simple calculation shows that the approximation error $\|f_n - f\|_2^2$ is on the order of $n^{-2s(\alpha)}$ for functions in the piecewise Hölder class $\Lambda^\alpha(M, B, m)$ with fixed α, M, B and m.

3. The nonequispaced procedure.

3.1. *The estimator.* Suppose now that we observe the data $\{y_i\}$ as in (1.2) and we wish to recover the regression function f. Our estimation method is based on multiresolution analysis and the projection method. The motivation of the method will be given in Section 3.2 from the approximation point of view.

Let $\tilde{g}(t) = n^{-1/2} \sum_{i=1}^n y_i \phi_{Ji}(t)$ and let

$$\tilde{f}_J(t) = \text{Proj}_{V_J}\, \tilde{g}(H(t)) = n^{-1/2} \sum_{k=1}^{2^{j_0}} \tilde{\xi}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \tilde{\theta}_{jk} \psi_{jk}(t),$$

where

$$\tilde{\xi}_{jk} = n^{-1/2} \sum_{i=1}^n y_i \langle \phi_{Ji} \circ H, \phi_{jk} \rangle, \qquad \tilde{\theta}_{jk} = n^{-1/2} \sum_{i=1}^n y_i \langle \phi_{Ji} \circ H, \psi_{jk} \rangle. \tag{3.1}$$

We can regard $\tilde{\xi}_{j_0 k}$ and $\tilde{\theta}_{jk}$ as noisy observations of the true wavelet coefficients $\xi_{j_0 k}$ and θ_{jk}. Indeed, we estimate θ_{jk} by thresholding $\tilde{\theta}_{jk}$. Let

$$\hat{\xi}_{j_0 k} = \tilde{\xi}_{j_0 k}, \qquad \hat{\theta}_{jk} = \text{sgn}(\tilde{\theta}_{jk})(|\tilde{\theta}_{jk}| - \lambda_{jk})_+ \tag{3.2}$$

be the estimate of the wavelet coefficients of f where the threshold λ_{jk} is derived in Section 3.3. Then a soft-thresholded wavelet estimator of f is given as follows:

$$\hat{f}_n^*(t) = \sum_{k=1}^{2^{j_0}} \hat{\xi}_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \hat{\theta}_{jk} \psi_{jk}(t). \tag{3.3}$$

Similarly, a hard-thresholded estimator can be obtained by setting the coefficients in (3.3) as

$$\hat{\xi}_{j_0 k} = \tilde{\xi}_{j_0 k}, \qquad \hat{\theta}_{jk} = \tilde{\theta}_{jk} I(|\tilde{\theta}_{jk}| > \lambda_{jk}), \tag{3.4}$$

with the same threshold λ_{jk} as in (3.2).

The coefficients $\hat{\xi}_{j_0 k}$ contain the gross structure of the function f and we do not threshold these coefficients. The risk of the estimate (3.3) can be decomposed as approximation error and estimation error. From Theorem 1, it is easy to see that the dominant term is the estimation error. We will show in Section 4 that the estimation error is comparable to the equispaced samples and the estimate enjoys the same convergence rate as the Donoho–Johnstone VisuShrink estimate in the equispaced case.

REMARK. We consider here the case of fixed design variables t_i. The method can be extended to random designs. The case of random designs has also been studied by Hall and Turlach (1996). Their method is based on linear interpolation.

3.2. *Approximation.* Let us see why the estimation method makes sense. We first consider the problem of approximating a whole function based on a noiseless nonequispaced sample. Denote by $\Lambda^1(h)$ the collection of Lipschitz functions f satisfying

$$|f(x) - f(y)| \le h\,|x - y| \quad \text{for } x, y \in [0, 1].$$

Suppose we are given a sampled function $\{f(t_i),\ i = 1, 2, \dots, n(= 2^J)\}$ with $t_i = H^{-1}(i/n)$, where H is a strictly increasing cumulative density function on [0, 1] and $H^{-1} \in \Lambda^1(h)$ for some constant h. How do we approximate the function f via multiresolution analysis?

If t_i are equispaced, it follows from Lemmas 1 and 2 that $f_n(t) = \sum_{k=1}^{n} n^{-1/2} f(t_k)\phi_{Jk}(t)$ is a good approximation. When t_i are nonequispaced, an approximation using multiresolution analysis can be derived by the following consideration. One can first approximate $f(H^{-1}(t))$ by $g_n(t) = \sum_{k=1}^{n} n^{-1/2} f(t_k)\phi_{Jk}(t)$, then use the projection of $g_n(H(t))$ onto the multiresolution space V_J as the approximation of f. To be more specific, let

$$\xi'_{Ji} = n^{-1/2} \sum_{k=1}^{2^J} f(t_k)\,\langle \phi_{Jk} \circ H, \phi_{Ji} \rangle \tag{3.5}$$

and let

$$f_n(t) = \sum_{i=1}^{2^J} \xi'_{Ji} \phi_{Ji}(t) \tag{3.6}$$

be an approximation of the function f. Note that f_n is in the multiresolution approximation space V_J. An upper bound for the approximation error is shown in the following result.

THEOREM 1. *Suppose that a sampled function $\{f(t_i), i = 1, 2, \dots, n(= 2^J)\}$ is given with $t_i = H^{-1}(i/n)$, where H is a strictly increasing cumulative density function on* [0, 1] *with $H^{-1} \in \Lambda^1(h)$. Let the wavelet function ψ be r-regular with $r > \alpha$. Let ξ'_{Ji} and f_n be given as in* (3.5) *and* (3.6), *respectively. Then the* *approximation error* $\|f_n - f\|_2^2$ *satisfies*

$$\sup_{f\in\Lambda^\alpha(M,B,m)} \|f_n - f\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}), \tag{3.7}$$

where the maximum number of jump discontinuities $m = Cn^\gamma$ *with constants* $C > 0$ *and* $0 < \gamma < 1/(1+2\alpha)$.

Theorem 1 shows that the approximation error over function class $\Lambda^\alpha(M, B, m)$ is of higher order than $n^{-2\alpha/(1+2\alpha)}$ even when the number of jump points increases polynomially with the sample size. Because the optimal convergence rate for estimating f over uniform Hölder class $\Lambda^\alpha(M, B, 0)$ under the model (1.2) is $n^{-2\alpha/(1+2\alpha)}$, the approximation error is smaller in order than the minimax risk for statistical estimation.

3.3. *The threshold.* The approximation result (3.7) implies that $\tilde{\xi}_{j_k}$ and $\tilde{\theta}_{jk}$ in (3.1) have the "correct" means. In order to make thresholding work, we need to know the noise level of each coefficient $\tilde{\theta}_{jk}$.

The function H^{-1} is strictly increasing, so H^{-1} is differentiable almost everywhere. Denote by $\tilde{h}(t)$ the derivative of $H^{-1}(t)$. Then

$$0 < \tilde{h}(t) \le h \quad \text{for almost all } t \in [0, 1].$$

It is easy to see from (3.1) that

$$\begin{aligned}\sigma_{jk}^2 \equiv \operatorname{var}(\tilde{\theta}_{jk}) &= n^{-1}\varepsilon^2 \sum_{i=1}^{n} \big(\langle \phi_{Ji} \circ H, \psi_{jk}\rangle\big)^2 \\ &\le n^{-1}\varepsilon^2 \int \psi_{jk}^2(t)\tilde{h}(H(t))\,dt \equiv u_{jk}^2.\end{aligned} \tag{3.8}$$

Note that the inequality in (3.8) is asymptotically sharp, $\sigma_{jk} \to u_{jk}$, as $n \to \infty$. We set the threshold

$$\lambda_{jk} = u_{jk}\,(2\log n)^{1/2}. \tag{3.9}$$

REMARK. This procedure generalizes Donoho and Johnstone's VisuShrink for equispaced samples. When the samples are, in fact, equispaced, that is, when H is the identity function and thus $h = 1$, then $\tilde{\xi}_{j_0 k}$ and $\tilde{\theta}_{jk}$ are discrete wavelet transforms of $\{n^{-1/2} y_i\}$ and $\lambda_{jk} = \varepsilon(2n^{-1}\log n)^{1/2}$. Therefore, the procedure agrees with the VisuShrink when the sample is equispaced.

4. Optimality results. In this section, we discuss the properties of the wavelet estimate (3.3) given in Section 3.1. We begin by showing that the estimate enjoys a smoothness property. If the target function is the zero function, then the estimate $\hat{f}_n^*$ given in (3.3) and (3.9) is also the zero function with high probability. Specifically, we have:

THEOREM 2. *If the regression function is the zero function $f \equiv 0$, then there exists a sequence of constants P_n such that*

$$P(\hat{f}_n^* \equiv 0) \geq P_n \to 1 \quad \text{as } n \to \infty. \tag{4.1}$$

Therefore, with high probability, the estimate removes pure noise completely. We then prove that the estimate enjoys near minimaxity for global estimation and the estimate optimally adapts to unknown degree of local smoothness within a constant factor when used for estimating a function at a point.

4.1. *Global estimation.* We investigate the adaptivity of the wavelet estimate constructed in Section 3.1 over a range of piecewise Hölder classes $\Lambda^\alpha(M, B, m)$, where the maximum number of jump discontinuities is allowed to increase polynomially with the sample size. This enhances the power of the function classes $\Lambda^\alpha(M, B, m)$ for modeling spatially inhomogeneous functions. We show that the estimate (3.3) is near optimal. The convergence rate is within a logarithmic factor of the minimax rate over a range of function classes $\Lambda^\alpha(M, B, m)$.

THEOREM 3. *Suppose we observe $\{(t_i, y_i),\ i = 1, 2, \ldots n(= 2^J)\}$ as in* (1.2) *with $t_i = H^{-1}(i/n)$, where H is a strictly increasing cumulative density function on $[0, 1]$ with $H^{-1} \in \Lambda^1(h)$. Let $\hat{f}_n^*$ be either the soft-thresholded or hard-thresholded wavelet estimator of f given in* (3.3) *and* (3.9). *Suppose that the wavelet function ψ is r-regular. Then the estimator $\hat{f}_n^*$ is near optimal*:

$$\sup_{f \in \Lambda^\alpha(M, B, m)} E\,\|\hat{f}_n^* - f\|_2^2 \leq C(\log n/n)^{2\alpha/(1+2\alpha)}(1 + o(1)) \tag{4.2}$$

for all $0 < \alpha < r$ and all $m \leq Cn^\gamma$ with constants $C > 0$ and $0 < \gamma < 1/(1+2\alpha)$.

4.2. *Estimation at a point.* Theorem 3 gives the convergence rate of global estimation. Now we turn our attention to local estimation. The adaptive estimation in this case is similar to global estimation, but with a very interesting distinction. The adaptive minimax rate for estimation at a point is different from that for estimation of a whole function.

By the results of Brown and Low [2] and Lepski [13], an estimator adaptive to unknown smoothness without loss of efficiency is impossible for pointwise estimation, even when the function is known to belong to one of two Hölder classes. Therefore, local adaptation cannot be achieved "for free." The minimum loss of efficiency is a $(\log n)^{2\alpha/(1+2\alpha)}$ factor for estimating a function of unknown degree of local Hölder smoothness at a point. See [2] and [13]. We call $(\log n/n)^{2\alpha/(1+2\alpha)}$ the adaptive minimax rate. Donoho and Johnstone [8] discuss pointwise performance of the wavelet estimate for equispaced samples. They show that the VisuShrink estimate attains the adaptive minimax rate for estimating functions at a point. See [8] for details.

We will show that the estimator given in Theorem 3 attains the exact adaptive minimax rate for estimating a function in a Hölder class at a fixed point. Therefore, the estimator is optimally adaptive to unknown degree of smoothness within a constant factor. To be more precise, we have the following:

THEOREM 4. *For any fixed* $t_0 \in [0, 1]$, *let* $\hat{f}_n^*(t)$ *be given as in* (3.3) *and* (3.9). *Under the conditions given in Theorem* 3, *we have*

$$\sup_{f \in \Lambda^\alpha(M, B, 0)} E\big(\hat{f}_n^*(t_0) - f(t_0)\big)^2 \le C(\log n/n)^{2\alpha/(1+2\alpha)}(1 + o(1)) \tag{4.3}$$

for all $0 < \alpha < r$.

We have stated here the result in the case of uniform smoothness without jumps for the sake of simplicity. The wavelet procedure is locally adaptive; the result also holds for general piecewise Hölder classes so long as the jump points are away from a fixed neighborhood of t_0.

5. Discussion.

5.1. *Choice of threshold.* In (3.9), we set the threshold $\lambda_{jk} = u_{jk}(2 \log n)^{1/2}$, where $u_{jk} = (n^{-1}\varepsilon^2 \int \psi_{jk}^2(t)\tilde{h}(H(t))\,dt)^{1/2}$. It is clear that

$$u_{jk}^2 \le n^{-1}\varepsilon^2 h_{jk}, \tag{5.1}$$

where $h_{jk} = \text{ess}\,.\sup\{\tilde{h}(t)\colon t \in [H^{-1}(2^{-j}k), H^{-1}(2^{-j}(k+N))]\}$.

We may replace the threshold λ_{jk} by

$$\lambda'_{jk} = \varepsilon(2h_{jk}n^{-1}\log n)^{1/2}. \tag{5.2}$$

The optimality results hold with λ'_{jk} as the threshold. The threshold λ'_{jk} has computational advantage over the threshold λ_{jk}.

5.2. *The function H.* We have modeled the design points as $t_i = H^{-1}(i/n)$, where H is a strictly increasing c.d.f. with $H^{-1} \in \Lambda^1(h)$. In practice, H is usually unknown. In this case, one can use the piecewise linear empirical $\hat{H}_n$ in place of the "true" H. Here $\hat{H}_n$ is the piecewise linear function satisfying $\hat{H}(t_i) = i/n$. All of the theoretical results remain valid if we replace H by $\hat{H}_n$ in the construction of the estimator. This modification is useful for implementing the estimator.

5.3. *Implementation.* In this section, we address the issue of numerical implementation of the procedure we propose in Section 3.1.

Let P_H be a matrix with entries

$$P_H(k, i) = \langle \phi_{Ji} \circ H, \phi_{Jk} \rangle. \tag{5.3}$$

The cascade algorithm (see [5]), which converges exponentially, can be used to compute ϕ. Then $P_H(k, i)$ can be computed numerically. Also, based on

Lemma 2, we may use $n^{-1/2}\phi_{Ji}(H(k/n)) = \phi(n\,H(k/n) - i)$ as an approximation of $P_H(k, i)$.

Let W be the discrete wavelet transform and let

$$\tilde{\Theta} = (\tilde{\xi}_{j_0 1}, \ldots, \tilde{\xi}_{j_0 2^{j_0}}, \tilde{\theta}_{j_0 1}, \ldots, \tilde{\theta}_{j_0 2^{j_0}}, \ldots, \tilde{\theta}_{J-1,1}, \ldots, \tilde{\theta}_{J-1,2^{J-1}})',$$

where $\tilde{\xi}_{j_0 k}$ and $\tilde{\theta}_{jk}$ are given as in (3.1). We can view P_H as a preconditioning matrix because

$$\tilde{\Theta} = W(P_H n^{-1/2} Y).$$

Our algorithm for implementing the procedure has the following steps:

Step 1. Use the cascade algorithm to compute P_H; then precondition the data $n^{-1/2}Y$ by P_H, say $Y_p = P_H n^{-1/2} Y$.

Step 2. Apply the discrete wavelet transform to the preconditioned data to get the noisy wavelet coefficients; let $\tilde{\Theta} = WY_p$.

Step 3. Threshold the noisy wavelet coefficients; denote $\hat{\theta}_{jk} = \eta_{\lambda_{jk}}(\tilde{\theta}_{jk})$, where $\eta_{\lambda_{jk}}$ is either the hard- or the soft-thresholding function.

Then

$$\hat{f}_n(t) = \sum_{k=1}^{2^j} \hat{\xi}_{j_0 k}\, \phi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \hat{\theta}_{jk}\, \psi_{jk}(t)$$

is our estimate of the target function f.

If one is also interested in estimating the function at sample points, then two more steps are needed to get there:

Step 4. Apply the inverse wavelet transform to the denoised wavelet coefficients to get $W^{-1} \cdot \hat{\Theta}$.

Step 5. Compute P^H by using the cascade algorithm, where

$$P^H(k, i) = \phi_{Ji}(t_k); \tag{5.4}$$

then apply this postconditioning transform to $W^{-1}\hat{\Theta}$ to get the estimate of $f(t_i)$:

$$\hat{f}_n = P^H(W^{-1}\hat{\Theta}). \tag{5.5}$$

Combining the five steps together, the estimator is given by

$$\hat{f}_n = P^H W^{-1} T W P_H Y.$$

where T denotes the thresholding operation.

Note that both the preconditioning matrix P_H and the postconditioning matrix P^H are sparse matrices with only $O(n)$ nonzero entries.

5.4. *Why not treat nonequispaced samples the same as equispaced samples?* The nonequispaced model (1.2) is reduced to the equispaced model when H is the identity function. For general H, however, one can still "pretend" the sample is equispaced. Let $g = f \circ H^{-1}$. Then the sample is equispaced in terms

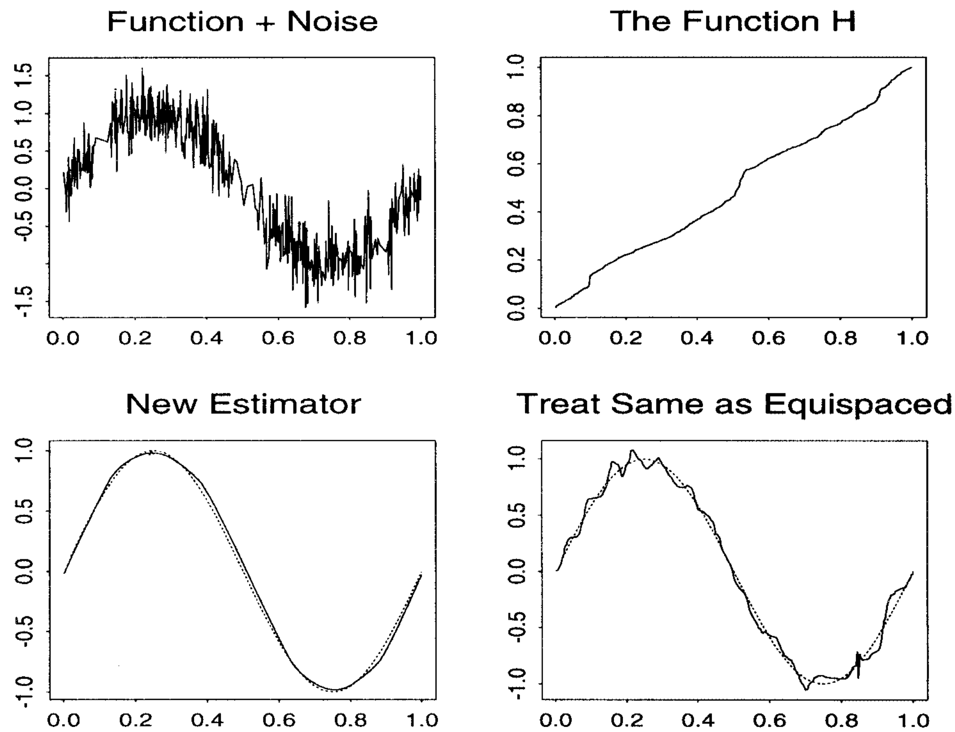

FIG. 1. *Comparison.*

of the function g. One can use the standard wavelet shrinkage procedure to estimate g by $\hat{g}$ and then use $\hat{g} \circ H$ as an estimator of f. This is what we mean by treating nonequispaced samples as equispaced. Here the estimator does not depend on the distribution of t_i.

The estimator is simple and very easy to implement. However, the estimator often does not perform well, especially when the underlying true function f is smoother than the function H. This can be shown by a formal calculation of asymptotic risk. One can show that, in many situations, the convergence rate of the estimator is suboptimal if nonequispaced samples are simply treated as equispaced. See [3] for more details. Another disadvantage is that the estimator is often visually unpleasant. Here is an example. The true function is $\sin(2\pi x)$, which is much smoother than the function H. The new estimator implemented by the previous algorithm is smooth and close to the true function, whereas the estimator treating the nonequispaced sample the same as the equispaced sample looks very rough. See Figure 1.

6. Proofs. This section contains proofs of the main results. We begin with a brief proof of Theorem 1 by using Lemmas 1 and 2.

PROOF OF THEOREM 1. Let $g(t) = f(H^{-1}(t))$. Denote $s(\alpha) = \min(\alpha, 1)$ and $M \vee B = \max(M, B)$. Then it is easy to see that $g \in \Lambda^{s(\alpha)}(h^{s(\alpha)}M \vee B, B, m)$.

Now $f_n = \operatorname{Proj}_{V_J} g_n \circ H$. It follows from Lemmas 1 and 2 that

$$\begin{aligned}\|f_n - f\|_2^2 &\le \|\operatorname{Proj}_{V_J}(g_n \circ H - g \circ H)\|_2^2 + \|\operatorname{Proj}_{V_J} f - f\|_2^2 \\ &\le C\, n^{-2s(\alpha)} + C\, m\, n^{-1} = o(n^{-2\alpha/(1+2\alpha)}). \qquad \square\end{aligned}$$

The proof of Theorem 2 is straightforward. For brevity, we omit the proof of the theorem. Before we prove Theorems 3 and 4, let us consider the problem of estimating a univariate normal mean.

Let $y \sim N(\theta, \sigma^2)$ be a normal variable with known variance σ^2. We are interested in estimating the mean θ with threshold estimator and we wish to assess the risk of the estimator. Let $\lambda = a\,\sigma$ with $a \ge 1$. And let $\hat\theta_\lambda^h = yI(|y| > \lambda)$ be a hard threshold estimator and let

$$\hat\theta_\lambda^s = \operatorname{sgn}(y)(|y| - \lambda)_+$$

be a soft threshold estimator of the mean θ. We recall the following results on the risk upper bound of the threshold estimator $\hat\theta$ from [3].

LEMMA 3. *Suppose $y \sim N(\theta, \sigma^2)$. Let $\hat\theta_\lambda^s$ and $\hat\theta_\lambda^h$ be the soft and hard threshold estimators of θ, respectively. Let $\lambda = a\,\sigma$ with $a \ge 1$. Then*

$$\text{(i)}\quad E(\hat\theta_\lambda^s - \theta)^2 \le (a^2+1)\sigma^2 \wedge (2\theta^2 + \exp(-a^2/2)\sigma^2), \tag{6.1}$$

$$\text{(ii)}\quad E(\hat\theta_\lambda^h - \theta)^2 \le (2a^2+2)\sigma^2 \wedge (2\theta^2 + 2a\exp(-a^2/2)\sigma^2). \tag{6.2}$$

The proofs of Theorems 3 and 4 are given only for soft threshold estimators. The proofs for hard threshold estimators are similar.

PROOF OF THEOREM 3. We follow the notation in Section 3.1. Let $g(t) = f(H^{-1}(t))$ and $\tilde g(t) = n^{-1/2}\sum_{i=1}^n y_i \phi_{Ji}(t)$ and let $\tilde f(t) = \tilde g(H(t))$ Then

$$\begin{aligned}\tilde f(t) &= n^{-1/2}\sum_{i=1}^n f(t_i)\phi_{Ji}(H(t)) + n^{-1/2}\varepsilon\sum_{i=1}^n z_i\phi_{Ji}(H(t)) \\ &= f(t) + \Delta(t) + r(t),\end{aligned}$$

where $\Delta(t) = n^{-1/2}\sum_{i=1}^n f(t_i)\phi_{Ji}(H(t)) - f(t)$ is the approximation error and $r(t) = n^{-1/2}\varepsilon\sum_{i=1}^n z_i\phi_{Ji}(H(t))$. Now project $\tilde f$ onto the multiresolution space V_J and decompose the orthogonal projection $\tilde f_J(t) = \operatorname{Proj}_{V_J}\tilde f(t)$ into three terms:

$$\tilde f_J(t) = f_J(t) + \Delta_J(t) + r_J(t), \tag{6.3}$$

where $f_J = \operatorname{Proj}_{V_J} f$, $\Delta_J = \operatorname{Proj}_{V_J}\Delta$ and $r_J = \operatorname{Proj}_{V_J} r$, respectively. Theorem 1 yields

$$\|\Delta_J\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}). \tag{6.4}$$

Denote $\tilde{\theta}_{jk} = \langle \tilde{f}_J, \psi_{jk} \rangle$. In the same fashion as in (6.3), we decompose $\tilde{\theta}_{jk}$ into three parts:

$$\tilde{\theta}_{jk} = \theta_{jk} + d_{jk} + r_{jk} \quad \text{for } k = 1, \dots, 2^j, \ j = j_0, \dots, J-1,$$

where $\theta_{jk} = \langle f, \psi_{jk} \rangle$ is the true wavelet coefficient of f, $d_{jk} = \langle \Delta_J, \psi_{jk} \rangle$ is the approximation error and $r_{jk} = \langle r_J, \psi_{jk} \rangle$ is the noise. Similarly separate $\tilde{\xi}_{j_0 k} = \langle \tilde{f}_J, \phi_{j_0 k} \rangle$ into three terms:

$$\tilde{\xi}_{j_0 k} = \xi_{j_0 k} + d'_{j_0 k} + r'_{j_0 k} \quad \text{for } k = 1, \dots, 2^{j_0}.$$

Let $\hat{\xi}_{j_0 k}$ and $\hat{\theta}_{jk}$ be given as in (3.2). Note that

$$\sum_{k=1}^{2^{j_0}} (d'_{j_0 k})^2 + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} d_{jk}^2 = \|\Delta_J\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}). \tag{6.5}$$

By the orthonormality of the wavelet basis, we have the isometry between the L_2 function norm and the l_2 wavelet sequence norm:

$$\begin{aligned} E\|\hat{f}_n^* - f\|^2 &= \sum_{k=1}^{2^{j_0}} E(\hat{\xi}_{j_0 k} - \xi_{j_0 k})^2 + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=J}^{\infty} \sum_{k=1}^{2^j} \theta_{jk}^2 \\ &\equiv S_1 + S_2 + S_3. \end{aligned}$$

It is easy to see from (3.9) that

$$S_1 \le 2^{j_0} n^{-1} \varepsilon^2 h + \sum_{k=1}^{2^{j_0}} (d'_{j_0 k})^2 = o(n^{-2\alpha/(1+2\alpha)}). \tag{6.6}$$

At each resolution level j, denote

$$\begin{aligned} G_j \equiv \{k\colon\ & \text{supp}(\psi_{jk}) = [2^{-j}k, 2^{-j}(N+k)] \\ & \text{contains at least one jump point of } f\}. \end{aligned}$$

Then $\text{card}(G_j) \le N(m+2)$ (counting two end points 0 and 1 as jump points as well). Lemma 1 yields

$$|\theta_{jk}| \le C 2^{-j(1/2+\alpha)} \quad \text{for } k \notin G_j, \tag{6.7}$$

$$|\theta_{jk}| \le C 2^{-j/2} \quad \text{for } k \in G_j, \tag{6.8}$$

where C is a constant not depending on f. Therefore,

$$\begin{aligned} S_3 &= \sum_{j=J}^{\infty} \sum_{k \in G_j} \theta_{jk}^2 + \sum_{j=J}^{\infty} \sum_{k \notin G_j} \theta_{jk}^2 \\ &\le \sum_{j=J}^{\infty} N(m+2) C^2 2^{-j} + \sum_{j=J}^{\infty} \sum_{k=1}^{2^j} C^2 2^{-j(1+2\alpha)} \\ &= o(n^{-2\alpha/(1+2\alpha)}). \end{aligned} \tag{6.9}$$

Now we consider S_2. First, note from (3.9) that $\sigma_{jk} \le u_{jk}$ and $\lambda_{jk} = u_{jk}(2\log n)^{1/2}$, so $a_{jk} \equiv \lambda_{jk}/\sigma_{jk} \ge (2\log n)^{1/2}$. It follows from (6.1) that

$$E(\hat{\theta}_{jk} - \theta_{jk})^2 \le (4\log n + 2)h\varepsilon^2 n^{-1} \wedge (8\theta_{jk}^2 + 2h\varepsilon^2 n^{-2}) + 10d_{jk}^2 \tag{6.10}$$

Write

$$\begin{aligned} S_2 &= \sum_{j=j_0}^{J-1} \sum_{k\in G_j} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=j_0}^{J-1} \sum_{k\notin G_j} E(\hat{\theta}_{jk} - \theta_{jk})^2 \\ &\equiv S_{21} + S_{22}. \end{aligned}$$

Since $\operatorname{card}(G_j) \le N(m+2)$, it follows from (6.10) that

$$S_{21} \le \sum_{j=j_0}^{J-1} N(m+2)[(4\log n + 2)h\varepsilon^2 n^{-1} + 10d_{jk}^2] = o(n^{-2\alpha/(1+2\alpha)}). \tag{6.11}$$

Now let J_1 be an integer satisfying $2^{J_1(1+2\alpha)} = n/\log n$. (For simplicity, we assume the existence of such an integer. In general, choose $J_1 = \lfloor 1/(1+2\alpha)\log_2(n/\log n)\rfloor$.) From (6.10), we have

$$E(\hat{\theta}_{jk} - \theta_{jk})^2 \le 5\varepsilon^2 n^{-1}\log n + 10d_{jk}^2 \quad \text{for } j_0 \le j \le J_1 - 1,\ k \notin G_j, \tag{6.12}$$

$$\begin{aligned} E(\hat{\theta}_{jk} - \theta_{jk})^2 &\le 8C^2 2^{-j(1+2\alpha)} + 2h\varepsilon^2 n^{-2} + 10d_{jk}^2 \\ &\qquad \text{for } J_1 \le j \le J-1,\ k \notin G_j. \end{aligned} \tag{6.13}$$

Therefore,

$$\begin{aligned} S_{22} &\le \sum_{j=j_0}^{J_1-1} \sum_{k\notin G_j} 5\varepsilon^2 n^{-1}\log n + \sum_{j=J_1}^{J-1} \sum_{k\notin G_j} (8C^2 2^{-j(1+2\alpha)} + 2h\varepsilon^2 n^{-2}) \\ &\quad + 10\sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} d_{jk}^2 \\ &= C(\log n/n)^{2\alpha/(1+2\alpha)}(1+o(1)). \end{aligned} \tag{6.14}$$

We finish the proof by putting (6.6), (6.9), (6.11) and (6.14) together:

$$E\|\hat{f}_n^* - f\|_2^2 \le C\,(\log n/n)^{2\alpha/(1+2\alpha)}(1+o(1)). \tag{6.15}$$ □

Proof of Theorem 4. First, we recall a simple but useful inequality.

Lemma 4. *Let X_i be random variables, $i = 1, \dots, n$. Then*

$$E\left(\sum_{i=1}^n X_i\right)^2 \le \left(\sum_{i=1}^n (EX_i^2)^{1/2}\right)^2. \tag{6.16}$$

Now, applying the inequality (6.16), we have

$$
\begin{aligned}
E(f_n^*(t_0)-f(t_0))^2 &= E\Bigg[\sum_{k=1}^{2^{j_0}}(\hat{\xi}_{j_0k}-\xi_{j_0k})\phi_{j_0k}(t_0)+\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j}(\hat{\theta}_{jk}-\theta_{jk})\psi_{jk}(t_0)\Bigg]^2\\
&\le \Bigg[\sum_{k=1}^{2^{j_0}}(E(\hat{\xi}_{j_0k}-\xi_{j_0k})^2\phi_{j_0k}^2(t_0))^{1/2}\\
&\quad+\sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j}(E(\hat{\theta}_{jk}-\theta_{jk})^2\psi_{jk}^2(t_0))^{1/2}+\sum_{j=J}^{\infty}\sum_{k=1}^{2^j}|\theta_{jk}\psi_{jk}(t_0)|\Bigg]^2\\
&\equiv (Q_1+Q_2+Q_3)^2.
\end{aligned}
$$

Now consider the three terms separately. Note that at each resolution level j there are at most N basis functions ψ_{jk} that are nonvanishing at t_0, where N is the length of the support of wavelet functions ϕ and ψ. Therefore,

$$
\begin{aligned}
Q_1 &= \sum_{k=1}^{2^{j_0}}\big(E(\hat{\xi}_{j_0k}-\xi_{j_0k})^2\big)^{1/2}|\phi_{j_0k}(t_0)|\\
&\le C\left(Nn^{-1}\varepsilon^2h+\sum_{k=1}^{2^{j_0}}(d'_{jk})^2\right)^{1/2}=o(n^{-\alpha/(1+2\alpha)}).
\end{aligned}
\tag{6.17}
$$

For the third term, it follows from Lemma 1(i) that

$$
Q_3=\sum_{j=J}^{\infty}\sum_{k=1}^{2^j}|\theta_{jk}|\,|\psi_{jk}(t_0)|\le\sum_{j=J}^{\infty}N\|\psi\|_\infty 2^{j/2}C2^{-j(1/2+\alpha)}\le Cn^{-\alpha}.
\tag{6.18}
$$

Now let us consider the term Q_2. First, note that for the function $f\in\Lambda^\alpha(M,B,0)$ the approximation error $\Delta(t)$ satisfies $\sup_t|\Delta(t)|\le Cn^{-s(\alpha)}$. This yields

$$
|d_{jk}|=|\langle\Delta,\psi_{jk}\rangle|\le C_1 2^{-j/2}n^{-s(\alpha)},
$$

where the constant C_1 does not depend on f. Let the integer J_1 be given as in the proof of Theorem 3. Applying (6.12) and (6.14),

$$
\begin{aligned}
Q_2 &\le N\|\psi\|_\infty\sum_{j=j_0}^{J_1-1}\big(5\varepsilon^2n^{-1}\log n+10C_1^2 2^{-j}n^{-2s(\alpha)}\big)^{1/2}\\
&\quad+N\|\psi\|_\infty\sum_{j=J_1}^{J-1}\big(8C^2 2^{-j(1+2\alpha)}+2h\varepsilon^2n^{-2}+10C_1^2 2^{-j}n^{-2s(\alpha)}\big)^{1/2}\\
&=C\left(\frac{\log n}{n}\right)^{\alpha/(1+2\alpha)}(1+o(1)).
\end{aligned}
\tag{6.19}
$$

Combining (6.17), (6.18) and (6.19), we have

$$
E(f_n^*(t_0)-f(t_0))^2\le C(\log n/n)^{2\alpha/(1+2\alpha)}(1+o(1)).
\tag{6.20}
$$

□

Acknowledgments. Part of the research was completed when the first author was a visitor at the Department of Statistics, University of Pennsylvania. The first author would like to thank the second author and the department for their (financial and other) support. The authors thank an Associate Editor and the referees for suggestions which improve the presentation of this work.

REFERENCES

[1] BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

[2] BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimations. *Ann. Statist.* **24** 2524–2535.

[3] CAI, T. (1996). Nonparametric function estimation via wavelets. Ph.D. dissertation, Cornell Univ.

[4] CASELLA, G. and STRAWDERMAN, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–878.

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

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

[7] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.

[8] DONOHO, D. L. and JOHNSTONE, I. M. (1996). Neo-classic minimax problems, thresholding, and adaptation. *Bernoulli* **2** 39–62.

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

[10] HALL, P. and TURLACH, B. A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.* **25** 1912–1925.

[11] HÄRDLE, W. (1990). *Applied Nonparametric Regression.* Cambridge Univ. Press.

[12] JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* **59** 319–351.

[13] LEPSKI, O. V. (1990). On a problem of adaptive estimation on white Gaussian noise. *Theory Probab. Appl.* **35** 454–466.

[14] MEYER, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes*. Hermann, Paris.

[15] STRANG, G. (1989). Wavelet and dilation equations: A brief introduction. *SIAM Rev.* **31** 614–627.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
LAFAYETTE, INDIANA 47907
E-MAIL: tcai@stat.purdue.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302
E-MAIL: lbrown@compstat.wharton.upenn.edu