FUNCTION ESTIMATION VIA WAVELET SHRINKAGE FOR LONG-MEMORY DATA[1]

By Yazhen Wang

University of Missouri−Columbia

In this article we study function estimation via wavelet shrinkage for data with long-range dependence. We propose a fractional Gaussian noise model to approximate nonparametric regression with long-range dependence and establish asymptotics for minimax risks. Because of long-range dependence, the minimax risk and the minimax linear risk converge to 0 at rates that differ from those for data with independence or short-range dependence. Wavelet estimates with best selection of resolution level-dependent threshold achieve minimax rates over a wide range of spaces. Cross-validation for dependent data is proposed to select the optimal threshold. The wavelet estimates significantly outperform linear estimates. The key to proving the asymptotic results is a wavelet−vaguelette decomposition which decorrelates fractional Gaussian noise. Such wavelet−vaguelette decomposition is also very useful in fractal signal processing.

1. Introduction. Consider nonparametric regression

$$(1) \qquad y_i = f(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $x_i = i/n \in [0,1]$, $\varepsilon_1, \ldots, \varepsilon_n$ are observational errors with mean 0 and f is an unknown function to be estimated. Common assumptions on $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. errors or stationary processes with short-range dependence such as classic ARMA processes. However, real data do not always meet these assumptions. Scientists in diverse fields have observed empirically that correlations between observations that are far apart decay to 0 at a slower rate than we would expect from independent data or short-range dependent data. These fields include agronomy, astronomy, chemistry, economics, engineering, environmental sciences, geosciences, hydrology, physics and signal and image processing. For example, in the study of trends in global temperature, observational errors exhibit slow decay in correlation [see Beran (1994), Section 1.4]. In fact, due to the vast number of examples from hydrology and geophysics, slowly decaying correlations are recognized by most hydrologists and geophysicists to be the rule rather than the exception [see, e.g., Beran

---

(1992) or Beran (1994), Chapter 1]. Slow decay in correlation is often referred to as long-range dependence or long memory. Suppose $(\varepsilon_i)_{1 \le i \le n}$ is a stationary Gaussian process with mean 0. Then (ε_i) is said to have long-range dependence (or long memory) if there exists $\alpha \in (0, 1)$ such that

$$(2) \qquad R(j - i) = \mathrm{Corr}(\varepsilon_i, \varepsilon_j) \asymp |j - i|^{-\alpha}, \qquad j - i \to \infty,$$

where $a \asymp b$ means that a/b is asymptotically bounded away from 0 and ∞. An alternative definition of long-range dependence is that the spectral density of (ε_i) has a pole at 0 [see, e.g., Beran (1994), Chapter 2, for the equivalence as well as justifications of the definition]. This article will study minimax estimation of f via wavelet shrinkage for long-range dependent data.

Long-range dependence had been observed empirically long before suitable mathematical models were known. Mandelbrot and his co-workers [Mandelbrot and van Ness (1968) and Mandelbrot and Wallis (1968, 1969)] first introduced fractional Gaussian noise and fractional Brownian motion to model long-range dependence. Since then a considerable body of literature has evolved which explores these so-called $1/f$ processes as a study in self-similarity and long-range dependence [Barton and Poor (1988) and Keshner (1982)]. These processes have much more persistent long-term correlation structure than the well-studied short-range processes such as ARMA processes and mixing processes. Traditionally, they have been mathematically awkward to manipulate. Recently wavelets have been used advantageously to analyze them and study their statistical self-similarity and long-range dependence. Wornell (1990) used a wavelet basis to provide an approximate Karhunen–Loève expansion for fractional Brownian motion. Studies of the correlation structure of the wavelet transform of fractional Brownian motion were accomplished in Flandrin (1989, 1992), Ramanathan and Zeitouni (1991) and Tewfik and Kim (1992). The auto and cross-correlation functions of the discrete wavelet coefficients of fractional Brownian motion at different scales decay hyperbolically fast at a rate much faster than that of the fractional Brownian motion itself. Wornell and Oppenheim (1992) considered recovering fractal signals corrupted by white noise using wavelets. Benassi and Jaffard (1993) investigated wavelet decompositions of Gaussian processes including fractional Brownian motion.

The development of a theory of statistical inference for long-memory data has become a very active field of research in the past decade. Estimation for data with long-range dependence is quite different from that for observations with independence or short-range dependence, and many basic problems are still unsolved [see, e.g., Beran (1992) and Beran (1994)]. For example, for estimating a regression function or a density, Hall and Hart (1990a, b) showed that convergence rates differ from those with independence or short-range dependence.

The approach in this paper is as follows. Inspired by the seminal work of Donoho (1995), Donoho and Johnstone (1994, 1995a, b) and Donoho, Johnstone, Kerkyacharian and Picard (1995), we propose a fractional Gauss-

ian noise model to approximate nonparametric regression with long-range dependence and establish asymptotic results for minimax risks. Because of long-range dependence, the minimax risk and the minimax linear risk converge to 0 at rates that differ from those for data with independence or short-range dependence. Wavelet shrinkage estimates with resolution level-dependent threshold can be "tuned" to achieve minimax rates over the entire scale of Besov spaces $B_{p,q}^{\sigma}$. Linear estimates cannot achieve even the minimax rates over Besov classes when $p < 2$, so the wavelet estimates significantly outperform linear estimates.

The wavelet estimates with best selection of level-dependent threshold achieve the minimax risk over a wide range of spaces, so selection of the optimal threshold is very important. Nason (1996) reported that the universal and sure methods and Nason's cross-validation method (removing half of the data each time) do not work well for correlated data. We propose a cross-validation method to select thresholds for dependent data. To reduce correlation, the cross-validation method deletes more than half of the data each time.

The difficulty in proving asymptotic results is that the wavelet coefficients of fractional Gaussian noise are correlated and hence do not form a white noise process. We are, however, able to show that the wavelet coefficients are nearly independent by using a wavelet–vaguelette decomposition (WVD) of fractional Gaussian noise and then deriving minimax risks by methods developed by Donoho (1995). Here we utilize the idea of simultaneous diagonalization through WVD described in Donoho (1995) and the fact that fractional Gaussian noise is linked to dilation-homogeneous operators which are almost diagonal in a wavelet basis.

Decorrelation of fractional Gaussian noise and fractional Brownian motion via WVD has its own interest. In fractal signal processing, it is very desirable to decorrelate fractional Gaussian noise and fractional Brownian motion [Wornell and Oppenheim (1992)]. Although wavelets reduce the dependence of fractional Gaussian noise, the wavelet coefficients of fractional Gaussian noise and fractional Brownian motion are correlated. Hence wavelets themselves do not decorrelate fractional Gaussian noise and fractional Brownian motion.

The rest of this paper is organized as follows. Section 2 introduces a fractional Gaussian noise model and derives asymptotics for the minimax risks. Section 3 relates the fractional Gaussian noise model to nonparametric regression with long-range dependence. Sections 4 to 6 consist of derivations of the asymptotic results. The WVD's to decorrelate fractional Gaussian noise and fractional Brownian motion are given in Section 5. Section 7 discusses wavelet estimates and threshold selection by cross-validation for dependent data.

2. The fractional Gaussian noise model. Suppose we observe the stochastic process $Y(x)$, $x \in [0, 1]$, from the fractional Gaussian noise model

$$(3) \qquad\qquad Y(dx) = f(x)\, dx + \epsilon^{2-2H} B_H(dx),$$

where f is an unknown function, ϵ is the noise level which we think is small and $B_H(dx)$ is a fractional Gaussian noise which is the formal derivative of a standard fractional Brownian motion defined below. Fractional Gaussian noise and fractional Brownian motion are often used to model phenomena exhibiting long-range dependence [see, e.g., Beran (1992, 1994) and Mandelbrot and van Ness (1968)]. Section 3 will show that the fractional Gaussian noise model (3) is an approximation of nonparametric regression with long-range dependence. A standard fractional Brownian motion $B_H(t)$, $t \in [0, 1]$, with parameter H, $1/2 < H < 1$, can be written as

$$B_H(t) = \frac{1}{\Gamma(H + 1/2)} \int_0^t (t - u)^{H - 1/2} B(du), \qquad t \in [0, 1],$$

where B is a standard Brownian motion. B_H has covariance function

(4) $$\mathrm{Cov}\{B_H(s), B_H(t)\} = V_H\{|s|^{2H} + |t|^{2H} - |t - s|^{2H}\}/2,$$

where

(5) $$V_H = \mathrm{Var}\{B_H(1)\} = \frac{\cos(\pi H)\Gamma(1 - 2H)}{\pi H}.$$

Our goal is to estimate f on the basis of the data $Y(x)$, $x \in [0, 1]$, and a priori information $f \in \mathcal{F}$. We use squared-error loss and are interested in the minimax risk

$$R(\epsilon; \mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} E\|\hat{f} - f\|_2^2.$$

2.1. *Wavelets.* Let ψ be a Daubechies wavelet of compact support having M vanishing moments, M continuous derivatives and unit L^2 norm. Wavelets are derived from ψ by a process of dyadic dilations and translations of ψ. Let $\lambda = (j, k)$ with j and k integers, and define

$$\psi_\lambda(x) = 2^{j/2}\psi(2^j x - k), \qquad j, k \in \mathbb{Z}.$$

Daubechies' construction guarantees that $(\psi_\lambda)_{\lambda \in \mathbb{Z}^2}$ is a complete orthonormal basis for the space $L^2(\mathbb{R})$. Hence, for a function $f \in L^2(\mathbb{R})$, the wavelet decomposition of f is written as

(6) $$f(x) = \sum_\lambda \theta_\lambda \psi_\lambda(x),$$

with wavelet coefficients of f:

(7) $$\theta_\lambda = \int f(x)\psi_\lambda(x)\, dx.$$

For functions supported in a fixed unit interval, say $[0, 1]$, we can select an index set $\Lambda \subset \mathbb{Z}$ and modify some of the ψ_λ, $\lambda \in \Lambda$, such that $(\psi_\lambda)_{\lambda \in \Lambda}$ forms a complete orthonormal basis for $L^2[0, 1]$ [see Cohen, Daubechies, Jawerth and Vial (1993), Daubechies (1992, 1994), Donoho and Johnstone (1995a) and Meyer (1992)].

2.2. *Wavelet estimates.*   Let

$$(8) \qquad y_\lambda = \int \psi_\lambda(x) Y(dx)$$

be the empirical wavelet coefficients of the data Y. Wavelet estimates of f are defined to be

$$(9) \qquad \hat{f}_{(t_j)} = \sum_\lambda \delta_{t_j}(y_\lambda) \psi_\lambda,$$

with $\delta_{t_j}(y_\lambda) = \text{sign}(y_\lambda)(|y_\lambda| - t_j)_+$ for the soft-threshold rule, $\delta_{t_j}(y_\lambda) = y_\lambda 1_{\{|y_\lambda| \geq t_j\}}$ for the hard-threshold rule and (t_j) the resolution level-dependent threshold. [See Donoho (1993, 1995), Donoho and Johnstone (1994, 1995a, b) and Donoho, Johnstone, Kerkyacharian and Picard (1993, 1995).]

We define the minimax-wavelet threshold risk:

$$R_W(\epsilon; \mathscr{F}) = \inf_{(t_j)} \sup_{f \in \mathscr{F}} E \| \hat{f}_{(t_j)} - f \|_2^2.$$

This is the best that we can do for recovering objects in $\mathscr{F}$ by the wavelet shrinkage method with optimal selection of (t_j).

2.3. *Asymptotics of minimax risk.*   Let $\mathscr{F}$ be a ball of Besov space $B_{p,q}^\sigma$ such that the wavelet coefficients of functions f in $\mathscr{F}$ belong to the set $\Theta_{p,q}^s$ of sequences $(\theta_\lambda)_\Lambda$ obeying

$$(10) \qquad \left(\sum_j 2^{jsq} \left(\sum_k |\theta_\lambda|^p \right)^{q/p} \right)^{1/q} \leq C,$$

where $s = \sigma + 1/2 - 1/p$ [see Donoho (1995), Donoho and Johnstone (1995a) and Meyer (1992)]. For comparison purposes, we define the minimax linear risk:

$$R_L(\epsilon, \mathscr{F}) = \inf_{\hat{f} \text{ linear}} \sup_{\mathscr{F}} E \| \hat{f} - f \|_2^2.$$

The following theorem, which is proved in Sections 4 to 6, gives asymptotics of the minimax risks.

THEOREM 1.   *Suppose $\sigma > (1 - H)(2 - p)/p$. Then*

$$R(\epsilon; \mathscr{F}) \asymp \epsilon^{2r}, \qquad \epsilon \to 0,$$

with rate exponent

$$r = \frac{2\sigma(1 - H)}{\sigma + 1 - H};$$

and

$$R_W(\epsilon; \mathscr{F}) \leq \text{const } R(\epsilon; \mathscr{F}), \qquad \epsilon \to 0.$$

However, the minimax linear risk tends to 0 *at the rate*

$$R_L(\epsilon; \mathscr{F}) \asymp \epsilon^{2r'}, \qquad \epsilon \to 0,$$

with rate exponent

$$r' = \frac{2(\sigma + 1/2 - 1/p_-)(1 - H)}{\sigma + 3/2 - 1/p_- - H},$$

where $p_- = \min(2, p)$. *The minimax linear risk converges to* 0 *slower than the minimax risk for* $p < 2$.

Comparing the convergence rates in Theorem 1 with those in Corollary 1 of Donoho and Johnstone (1995a), Section 1, we see that minimax risks converge to 0 slower than those for the white noise model, which corresponds to the case with independence or short-range dependence. The convergence rates in Theorem 1 are also related to those in Theorem 4 of Donoho (1995), Section 6.2, for a linear inverse problem with white noise and in Theorem 1 of Wang (1994) for a linear inverse problem with long-range dependence (see Section 5 for details).

3. Nonparametric regression and fractional white noise. We now relate the fractional Gaussian noise model to nonparametric regression with long-range dependence. As the white noise model approximates nonparametric regression with i.i.d. errors [Brown and Low (1996) and Donoho and Johnstone (1995a), Section 9], the fractional Gaussian noise model is an approximation of the nonparametric regression model (1) with long-range dependence, and results established for the fractional Gaussian noise model can then be applied to nonparametric regression.

For estimating $f \in \mathscr{F}$ based on the data $y_1, \ldots, y_n$ from model (1), we denote by $R(n, \mathscr{F})$ and $R_L(n, \mathscr{F})$ the minimax risk and the minimax linear risk, respectively, that is

$$R(n, \mathscr{F}) = \inf_{\hat{f}} \sup_{f} E\|\hat{f} - f\|_2^2 \quad \text{and} \quad R_L(n, \mathscr{F}) = \inf_{\hat{f} \text{ linear}} \sup_{f} E\|\hat{f} - f\|_2^2.$$

Let

$$L_n^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} R(i - j), \qquad H = 1 - \alpha/2 \quad \text{and} \quad \epsilon_n = (L_n/n)^{1/(2 - 2H)}.$$

From (2) we have

$$(11) \qquad\qquad L_n^2 \asymp n^{2H} \quad \text{and} \quad \epsilon_n \asymp n^{-1/2}.$$

The definition of the rate exponent r (see Theorem 1) implies that $r < 2 - 2H$ and hence $(\epsilon_n^{2 - 2H})^2 = o(n^{-r})$.

Theorem 5.1 of Taqqu (1975) implies that there exist fractional Brownian motions B_H such that, for all x_i,

$$(12) \qquad\qquad \frac{1}{L_n} \sum_{j=1}^{i} \varepsilon_j = B_H(x_i) + c_{n,i},$$

where $\max\{c_{n,i}, \ 1 \le i \le n\} \to 0$ as $n \to \infty$. [It may be necessary to change probability spaces such that $\varepsilon_1, \ldots, \varepsilon_n$ and B_H live on common probability spaces. Also see Lemma 5.1 and its proof in Taqqu (1975).] Define the regression process $\{Y_n(x), \ x \in [0,1]\}$ via

$$Y_n(0) = 0, \qquad Y_n(x_i) = \frac{1}{n}\sum_{j=1}^{i} y_j, \qquad i = 1, \ldots, n,$$

with interpolation between the x_i by B_H: for $x_i < x < x_{i+1}$ set

$$\begin{aligned}
Y_n(x) = \ & Y_n(x_i) + (x - x_i)y_{i+1} \\
& + \epsilon_n^{2-2H}\big[B_H(x) - n(x - x_i)B_H(x_{i+1}) \\
& \qquad + \{n(x - x_i) - 1\}B_H(x_i) - n(x - x_i)c_{n,i+1} \\
& \qquad\qquad + \{n(x - x_i) - 1\}c_{n,i}\big].
\end{aligned}$$

Let

$$(13) \qquad\qquad f_n(x) = \sum_{i=1}^{n} f(x_i)\mathbf{1}_{\{x_{i-1} \le x < x_i\}}.$$

Then by the definition of Y_n, (12) and (13), we can easily check that Y_n satisfies

$$(14) \qquad\qquad Y_n(dx) = f_n(x)\,dx + \epsilon_n^{2-2H}B_H(dx).$$

Therefore, on a common probability space, we have that Y_n differs from Y (with $\epsilon = \epsilon_n$) defined by (3) exactly by the difference between f_n and f.

For a class $\mathscr{F}$ of functions, define $D_n(\mathscr{F}) = \|f - f_n\|_2^2$. If $D_n(\mathscr{F}) = o(n^{-r})$, where r is the rate exponent defined in Theorem 1, then minimax risks based on Y_n and based on Y (with $\epsilon = \epsilon_n$) are equivalent. Similar to Theorems 15 and 16 of Donoho and Johnstone (1995a), Section 9, we have that for $\sigma > \alpha(2 - p)/(2p)$ the minimax risks based on $Y_n(x)$, $x \in [0,1]$, are equivalent to those based on $Y(x)$, $x \in [0,1]$ (with $\epsilon = \epsilon_n$), and hence from Theorem 1 with $\epsilon_n \asymp n^{-1/2}$ [by (11)] and $H = 1 - \alpha/2$, we conclude that

$$R(n;\mathscr{F}) \asymp n^{-r}, \qquad n \to \infty,$$

with rate exponent

$$r = \frac{2\alpha\sigma}{2\sigma + \alpha};$$

the minimax-wavelet threshold risk based on Y_n satisfies

$$R_W(n,\mathscr{F}) = \inf_{(t_j)}\sup_{f} E\|\hat{f}_{(t_j)} - f\|_2^2 \le \mathrm{const}\, R(n;\mathscr{F}), \qquad n \to \infty,$$

where $\hat{f}_{(t_j)}$ are wavelet estimates based on Y_n; and

$$R_L(n;\mathscr{F}) \asymp n^{-r'}, \qquad n \to \infty,$$

with rate exponent

$$r' = \frac{2\alpha(\sigma + 1/2 - 1/p_-)}{2\sigma + 1 - 2/p_- + \alpha},$$

where $p_- = \min(2, p)$. The minimax linear risk converges to 0 slower than the minimax risk for $p < 2$.

Hall and Hart (1990a) considered nonparametric regression (1) with ε_i following an infinite-order Gaussian autoregression, which satisfies the long-range dependence relation (2) with $0 < \alpha < 1$. They showed that, for a class of functions with bounded second derivative, the risk of the best kernel estimate has a convergence rate $n^{-4\alpha/(4+\alpha)}$. This convergence rate corresponds to the minimax rate with $\sigma = 2$ and $p = q = \infty$. Since $p \geq 2$, linear estimates can achieve the minimax rate. However, linear estimates such as kernel estimates cannot achieve the minimax risk when $p < 2$. For example, for estimating $f \in \hat{\mathscr{F}} = \{f, \int |f''(x)|^p \, dx \leq C\}$ with $p < 2$, the best kernel estimate can achieve only the linear rate $n^{-\alpha(5-2/p)/(5-2/p+\alpha)}$, which is slower than the optimal rate $n^{-4\alpha/(4+\alpha)}$.

4. Estimation in sequence space.

4.1. *Wavelet coefficients of fractional Gaussian noise.*

LEMMA 1.

$$E\left\{\int g_1(s)\, dB_H(s) \int g_2(t)\, dB_H(t)\right\}$$

$$= V_H H(2H-1) \int\int g_1(s) g_2(t)|s-t|^{2H-2}\, ds\, dt,$$

where V_H is defined in (5).

PROOF. The lemma is easily proved by use of integration by parts. □

LEMMA 2. *Let*

$$\tau = \left\{V_H H(2H-1) \int\int \psi(u)\psi(v)|u-v|^{2H-2}\, du\, dv\right\}^{1/2}.$$

Then

$$\mathrm{Var}\left\{\int \psi_\lambda(s)\, dB_H(s)\right\} = \tau^2 2^{j_\lambda(1-2H)}.$$

PROOF.   By Lemma 1, we have

$$\mathrm{Var}\Big\{\int \psi_\lambda(s)\, dB_H(s)\Big\}$$

$$= E\Big\{\int \psi_\lambda(s)\, dB_H(s) \int \psi_\lambda(t)\, dB_H(t)\Big\}$$

$$= E\Big\{\iint \psi_\lambda(s)\psi_\lambda(t)\, dB_H(s)\, dB_H(t)\Big\}$$

$$= V_H H(2H-1)\iint \psi_\lambda(s)\psi_\lambda(t)|s-t|^{2H-2}\, ds\, dt$$

$$= \Big\{V_H H(2H-1)\iint \psi(u)\psi(v)|u-v|^{2H-2}\, du\, dv\Big\}2^{j_\lambda(1-2H)}. \qquad \square$$

Let

$$(15)\qquad
\begin{aligned}
z_\lambda &= \Big(\mathrm{Var}\Big\{\int \psi_\lambda(s)\, dB_H(s)\Big\}\Big)^{-1/2}\int \psi_\lambda(u)\, dB_H(u)\\
&= \tau^{-1}2^{j_\lambda(H-1/2)}\int \psi_\lambda(u)\, dB_H(u).
\end{aligned}$$

LEMMA 3.

$$\mathrm{Corr}(z_\lambda, z_\mu) \asymp \Big\{2^{(j_\lambda+j_\mu)/2}|2^{-j_\mu}k_\mu - 2^{-j_\lambda}k_\lambda|\Big\}^{2(H-M)}$$

holds uniformly in $\lambda = (j_\lambda, k_\lambda)$ *and* $\mu = (j_\mu, k_\mu)$. *In particular,* $(z_\lambda)_\lambda$ *at each level is a stationary Gaussian process with zero mean, unit variance and*

$$\mathrm{Corr}(z_\lambda, z_\mu) \asymp |k_\mu - k_\lambda|^{2(H-M)},$$

uniformly for $\lambda = (j, k_\lambda)$ *and* $\mu = (j, k_\mu)$.

PROOF.   Using Lemma 1 and arguments in proving Theorem 1 of Tewfik and Kim (1992), we obtain

$$\begin{aligned}
E(z_\lambda z_\mu) &= \tau^{-2}2^{(j_\lambda+j_\mu)(H-1/2)}\iint \psi_\lambda(s)\psi_\mu(t)|s-t|^{2H-2}\, ds\, dt\\
&= \Big\{\iint \psi(u)\psi(v)|u-v|^{2H-2}\, du\, dv\Big\}^{-1}2^{(H-1)(j_\lambda+j_\mu)}\\
&\quad \times \iint \psi(u)\psi(v)\big|(u+k_\lambda)2^{-j_\lambda} - (v+k_\mu)2^{-j_\mu}\big|^{2H-2}\, du\, dv\\
&\asymp 2^{(H-M)(j_\lambda+j_\mu)}|2^{-j_\mu}k_\mu - 2^{-j_\lambda}k_\lambda|^{2(H-M)},
\end{aligned}$$

uniformly in $\lambda = (j_\lambda, k_\lambda)$ *and* $\mu = (j_\mu, k_\mu)$. $\square$

4.2. *Regression on the wavelet domain.*   The empirical wavelet coefficients y_λ of the data Y defined in (8) satisfy

$$(16) \qquad\qquad y_\lambda = \theta_\lambda + \tau \epsilon^{2-2H} 2^{j_\lambda(1/2-H)} z_\lambda,$$

where θ_λ and z_λ are defined in (7) and (15), respectively.

Now model (3) is equivalent to model (16). The problem of estimating f is equivalent to recovering θ_λ. Because wavelets provide a complete orthonormal system, $\|\hat{f} - f\|_2^2$ is essentially the same as $\|\hat{\theta} - \theta\|_{l^2(\Lambda)}^2$. We therefore work with the sequence of observations from model (16). We assume that $(\theta_\lambda) \in \Theta_{p,q}^s$ and define the minimax risk

$$R_z(\epsilon) = \inf_{\hat{\theta}} \sup_{\Theta_{p,q}^s} E \|\hat{\theta}(y) - \theta\|_{l^2(\Lambda)}^2.$$

The noise (z_λ) in (16) is not a white noise, but we will show that it satisfies

$$(17) \qquad\qquad \gamma_0^2 \le \mathrm{Var}\big(z_\lambda | (z_\lambda)_{\lambda \ne \mu}\big) \le \mathrm{Var}(z_\lambda) \le \gamma_1^2$$

for each $\lambda \in \Lambda$, where $0 < \gamma_0 < \gamma_1 < \infty$ are positive generic constants. Such a noise process is called a nearly independent noise in Donoho (1995).

For observations from the model (16) with nearly independent noise, Donoho (1995), Section 7.3, showed that the minimax risk $R_z(\epsilon)$ is equivalent to the minimax risk $R_w(\epsilon)$ for observations from the discrete white noise model

$$(18) \qquad\qquad x_\lambda = \theta_\lambda + \tau \epsilon^{2-2H} 2^{j_\lambda(1/2-H)} w_\lambda, \qquad \lambda \in \Lambda,$$

where (w_λ) is a Gaussian white noise, and

$$R_w(\epsilon) = \inf_{\hat{\theta}} \sup_{\Theta_{p,q}^s} E \|\hat{\theta}(x) - \theta\|_{l^2(\Lambda)}^2.$$

For the white noise model (18), $R_w(\epsilon)$ can be explicitly given by the renormalization arguments in Sections 7.4 to 7.6 of Donoho (1995). Also, the minimax-wavelet risk and the minimax linear risk can be derived by the methods developed in Sections 8 and 9 of Donoho (1995). Hence it is enough to prove (17). In the next two sections we will decorrelate fractional Gaussian noise by WVD and then use the WVD to prove (17).

5. Decorrelation of fractional Gaussian noise via WVD.   In the study of a linear inverse problem, Donoho (1995) introduced a WVD to simultaneously quasidiagonalize both an operator K and a priori information $f \in \mathcal{F}$ such that f can be recovered through a reproducing formula by use of information about Kf. The WVD uses wavelets and vaguelettes instead of eigenfunctions used by the singular value decomposition (SVD) and hence is better at representing spatially variable functions than the SVD [see also Kolaczyk (1994) and Wang (1995b)]. A WVD is defined as follows:

1. three sets of functions $(\psi_\lambda)_\lambda$—an orthonormal wavelet basis—and $(u_\lambda)_\lambda$ and $(v_\lambda)_\lambda$—near-orthogonal vaguelettes with unit L^2-norm;

2. quasisingular value relations $K\psi_\lambda = \kappa_j v_\lambda$ and $K^* u_\lambda = \kappa_j \psi_\lambda$ for $\lambda = (j, k)$ with quasisingular values (κ_j), depending on resolution index j but not spatial index k;

3. biorthogonality relations $[u_\lambda, v_\mu] = \delta_{\lambda, \mu}$;

4. near-orthogonality relations

$$(19) \qquad \left\| \sum_\lambda a_\lambda u_\lambda \right\|_2 \asymp \|(a_\lambda)\|_{l^2} \quad \text{and} \quad \left\| \sum_\lambda a_\lambda v_\lambda \right\|_2 \asymp \|(a_\lambda)\|_{l^2}.$$

We will use the WVD to decorrelate fractional Brownian motion and fractional Gaussian noise below. Let $A_H = (-\Delta)^{H+1/2}$, where $\Delta = d^2/dx^2$ is an elliptic operator. For $f, g \in L^2(\mathbb{R})$, define

$$[f, g] = \int f(x) g(x)\, dx, \qquad [f, g]_H = \int A_H f(x) g(x)\, dx,$$

$$\|f\|_H^2 = [f, f]_H$$

and

$$K_H = \{ f \in L^2(\mathbb{R}), \|f\|_H < \infty \}.$$

Then fractional Brownian motion B_H is linked to the operator A_H and K_H is the space of the reproducing kernel A_H of B_H. Let ψ_λ be a usual wavelet orthonormal basis of $L^2(\mathbb{R})$, and define

$$(20) \qquad u_\lambda = (-\Delta)^{H/2 - 1/4} \psi_\lambda, \qquad v_\lambda = (-\Delta)^{1/4 - H/2} \psi_\lambda.$$

THEOREM 2. *Assume ψ_λ has vanishing moments of order less than or equal to M and continuous derivatives of order less than or equal to M. If $M > H + 3/2$, then*:

(i) $(\psi_\lambda)_\lambda$, $(2^{-j_\lambda(H+1/2)} A_H^{1/2} \psi_\lambda)_\lambda$ *and* $(2^{j_\lambda(H+1/2)} A_H^{-1/2} \psi_\lambda)_\lambda$ *form a WVD of $A_H^{1/2}$, and hence the WVD decorrelates B_H; that is, B_H has a representation*

$$(21) \qquad B_H(x) = \sum_\lambda \xi_\lambda A_H^{-1/2} \psi_\lambda,$$

where (ξ_λ) is a white noise;

(ii) $(\psi_\lambda)_\lambda$, $(2^{j_\lambda(1/2 - H)} u_\lambda)_\lambda$ *and* $(2^{j_\lambda(H - 1/2)} v_\lambda)_\lambda$ *form a WVD of $(-\Delta)^{H/2 - 1/4}$, and hence the WVD decorrelates fractional Gaussian noise $B_H(dx)$ and nearly decorrelates white noise $B(dx)$; that is, $B_H(dx)$ has a representation*

$$(22) \qquad B_H(dx) = \sum_\lambda \xi_\lambda v_\lambda(x)\, dx,$$

where (ξ_λ) is a white noise and $[u_\lambda, B(dx)]$ and $[v_\lambda, B(dx)]$ are nearly independent Gaussian noises.

PROOF. The proofs of the two results are similar, so we give arguments only for the first result. It is easy to show that K_H is the reproducing kernel space of the process defined on the right-hand side of (21) and hence (21) holds [see also Benassi and Jaffard (1993)]. The near independence of the noise processes $[u_\lambda, B(dx)]$ and $[v_\lambda, B(dx)]$ is a consequence of the near orthogonality of vaguelettes [Donoho (1995), Section 7.2]. So it is enough to show ψ_λ, u_λ and v_λ form a WVD of $A_H^{1/2}$. The biorthogonality and quasisingu-

lar value relations can be easily checked. The near orthogonality of u_λ and v_λ is proved as follows.

The Fourier transformations of $A_H^{1/2}\psi$ and $A_H^{-1/2}\psi$ are $\hat{\psi}(\omega)|\omega|^{H+1/2}$ and $\hat{\psi}(\omega)|\omega|^{-H-1/2}$, respectively, and hence they are vaguelettes by Definition 1 and Lemma 4 of Donoho (1995). Therefore, the near orthogonality of $A_H^{1/2}\psi_\lambda$ and $A_H^{-1/2}\psi_\lambda$ is a consequence of Theorem 2 of Donoho (1995). □

It is of interest to point out that the two problems, a linear inverse problem and estimation with long-range dependence, can be formally converted from one to another by WVD. By (22), model (3) can be written as

$$(23) \qquad Y(dx) = f(x)\,dx + \epsilon^{2-2H}\sum_\lambda \xi_\lambda v_\lambda(x)\,dx.$$

Let $y_\lambda^* = [Y(dx), u_\lambda]$. By (6) we have

$$(24) \qquad [f(x)\,dx, u_\lambda] = \sum_\mu \theta_\mu [\psi_\mu dx, u_\lambda] = (K\theta)_\lambda,$$

where $\theta = (\theta_\lambda)$, θ_λ are the wavelet coefficients of f defined in (7) and K is a transformation on θ which is represented as an infinite matrix. By (23) and (24), we formally convert model (3) as a linear inverse problem with white noise

$$(25) \qquad y_\lambda^* = (K\theta)_\lambda + \epsilon^{2-2H}\xi_\lambda, \qquad \lambda \in \Lambda.$$

For an ill-posed inverse problem, minimax rates are slower than those for direct observations [see Donoho (1995) and Donoho and Low (1992)], so from this perspective we also naturally see slow convergence of minimax risks for data with long-range dependence. The linear inverse problem observes a transformed function contaminated by white noise, while the long-range dependence problem has a transformation in white noise. These noninvertible transformations cause the convergence rate losses. Moreover, there is a corresponding relationship between the minimax rates in Theorem 1 and those in Theorem 4 of Donoho (1995) for a linear inverse problem with white noise.

It is also worth mentioning that Wang (1994) considered the following linear inverse problem with long-range dependence:

$$(26) \qquad Y(dx) = (Kf)(x)\,dx + \epsilon^{2-2H}B_H(dx), \qquad x \in [0,1],$$

where ϵ and $B_H(dx)$ are the same as those in model (3), K is a linear transformation and f is an unknown function to be recovered. Model (3) is the special case where $K = I$ and $H > 1/2$, and linear inverse problems with white noise considered in Donoho (1995) correspond to (26) with $H = 1/2$ (white noise). By using two WVD's (one decorrelates fractional Gaussian noise and one simultaneously quasidiagonalizes both the operator K and the prior information $f \in \mathcal{F}$), Wang (1994) derived minimax risks for model (26) and showed that estimates constructed by the WVD's can achieve minimax rates over a wide range of spaces. For example, suppose K has a WVD with quasisingular values $\kappa_j \asymp 2^{-j\alpha}$, then for estimating $f \in \mathcal{F}$ [a ball in Besov space $B_{p,q}^\sigma$ defined by (10)], the minimax risk has a convergence rate $2\sigma(1-H)/(\sigma + 1 + \alpha - H)$.

6. Near independence of the noise $(z_\lambda)_\lambda$. This section proves (17) by WVD. The upper bound γ_1 in (17) can be taken as $\gamma_1 = \text{Var}(z_\lambda) = 1$. Now we prove the lower bound. Let $\hat{z}_\lambda = E(z_\lambda|(z_\mu)_{\lambda \neq \mu})$. Then $\text{Var}(z_\lambda|(z_\mu)_{\lambda \neq \mu}) = \text{Var}(z_\lambda - \hat{z}_\lambda)$. As $(z_\lambda)_{\lambda \in \Lambda}$ is a zero-mean Gaussian random field, $\hat{z}_\lambda$ and $z_\lambda - \hat{z}_\lambda$ are independent, and there exists a sequence $(a_\mu)_{\mu \neq \lambda}$ such that

$$(27) \qquad \hat{z}_\lambda = \sum_{\lambda \neq \mu} a_\mu z_\mu.$$

For convenience, let $a_\lambda = -1$, and for $\lambda = (j_\lambda, k_\lambda)$ let $\sigma_\lambda = \tau 2^{j_\lambda(H-1/2)}$ with τ defined in Lemma 2. On the other hand, by (22) we have, for any μ,

$$(28) \qquad z_\mu = \sigma_\mu\big[B_H(dx), \psi_\mu\big] = \sigma_\mu \sum_{\lambda'} [v_{\lambda'}, \psi_\mu]\,\xi_{\lambda'}.$$

By (27), (28) and the biorthogonality of (u_μ) and (v_μ), we obtain

$$z_\lambda - \hat{z}_\lambda = z_\lambda - \sum_{\lambda \neq \mu} a_\mu z_\mu$$

$$= \sigma_\lambda \sum_{\lambda'} [v_{\lambda'}, \psi_\lambda]\,\xi_{\lambda'} - \sum_{\lambda \neq \mu} \sigma_\mu a_\mu \sum_{\lambda'} [v_{\lambda'}, \psi_\mu]\,\xi_{\lambda'}$$

$$(29) \qquad = \sum_{\lambda'} \left[v_{\lambda'}, \sigma_\lambda \psi_\lambda - \sum_{\lambda \neq \mu} \sigma_\mu a_\mu \psi_\mu\right]\xi_{\lambda'}$$

$$= -\sum_{\lambda'}\left[v_{\lambda'}, \sum_\mu \sigma_\mu a_\mu \psi_\mu\right]\xi_{\lambda'}$$

and

$$(30) \qquad [v_{\lambda'}, \psi_\mu] = \big[(-\Delta)^{1/4-H/2}\psi_{\lambda'}, \psi_\mu\big] = \big[\psi_{\lambda'}, (-\Delta)^{1/4-H/2}\psi_\mu\big] = [\psi_{\lambda'}, v_\mu].$$

By (29) and (30), we get

$$z_\lambda - \hat{z}_\lambda = -\sum_{\lambda'}\left[\psi_{\lambda'}, \sum_\mu \tau_\mu a_\mu v_\mu\right]\xi_{\lambda'},$$

and hence by complete orthogonormal bases of wavelets, the near orthogonality of the vaguelette $(\sigma_\mu v_\mu)$ proved in Theorem 2 and $\|(\alpha_\mu)\|_{l^2} \geq |a_\lambda| = 1$, we have

$$\text{Var}(z_\lambda - \hat{z}_\lambda) = \sum_{\lambda'}\left[\psi_{\lambda'}, \sum_\mu \sigma_\mu a_\mu v_\mu\right]^2$$

$$= \left\|\sum_\mu \sigma_\mu a_\mu v_\mu\right\|^2$$

$$= \left\|\sum_\mu (a_\mu)(\sigma_\mu v_\mu)\right\|^2$$

$$\geq C_0\|(\alpha_\mu)\|_{l^2}$$

$$\geq C_0 = \gamma_0 > 0.$$

This proves the lower bound in (17).

7. Discussion. Data from nonparametric regression are discrete, and, consequently, a discrete version of wavelet transformation must be performed to find wavelet coefficients and compute wavelet estimates. Suppose we have observations $y_1, \ldots, y_n$, $n = 2^J$, from model (1).

7.1. *Discrete wavelet transformation.* The discrete wavelet transformation (DWT) is a discretized version of the continuous wavelet transformation and can be written as linear transformation involving an $n \times n$ orthogonal matrix $\mathscr{W}$ which depends on parameters M (number of vanishing moments), S (support width), j_0 (low-resolution cutoff) and boundary adjustments [see Cohen, Daubechies, Jawerth and Vial (1993), Daubechies (1994), Donoho and Johnstone (1994) and Wang (1995a)]. The rows of $\mathscr{W}$ correspond to a discretized version of the wavelets ψ_λ. Indeed, if we denote by $\mathbf{W}_{j,k}(i)$ the ith element of the (j, k)th row of $\mathscr{W}$, then

$$(31) \qquad \sum_{i=1}^{n-1} i^l \mathbf{W}_{j,k}(i) = 0, \qquad l = 0, \ldots, M, j \geq j_0, k = 0, \ldots, 2^j - 1,$$

and

$$(32) \qquad \sqrt{n}\, \mathbf{W}_{j,k}(i) \approx 2^{j/2} \psi(2^j x), \qquad x = i/n - k 2^{-j}.$$

[See Donoho and Johnstone (1994).] Let $\mathbf{y} = (y_1, \ldots, y_n)$. The DWT of the data $\mathbf{y}$ is given by $\mathbf{w} = \mathscr{W}\mathbf{y}$. The elements of $\mathbf{w}$ are indexed dyadically as follows: $w_{j,k}$, $j = 0, \ldots, J - 1$, $k = 0, \ldots, 2^j - 1$, and $w_{-1,0}$. $w_{j,k}$ is called the empirical wavelet coefficient at level j and position $k 2^{-j}$, $k = 0, \ldots, 2^j - 1$, $j = 0, \ldots, J - 1$. Because $\mathscr{W}$ is orthogonal, the inverse DWT is easy and $\mathbf{y}$ is recovered from the DWT $\mathbf{w}$ by

$$\mathbf{y} = \mathscr{W}^T \mathbf{w} \quad \text{or} \quad y_i = \sum_{j,k} w_{j,k} \mathbf{W}_{j,k}(i).$$

Mallat's pyramidal algorithm [see, e.g., Mallat (1989) and Chui (1992), pages 20 and 21] requires only $O(n)$ operations for computing the DWT and reconstruction of the DWT.

7.2. *Wavelet estimates.* The wavelet estimates $\hat{f}_{(t_j)}$ of f are computed as follows. First, find the DWT $w_{j,k}$ and then threshold $w_{j,k}$ by the soft-threshold rule

$$\delta_{t_j}(w_{j,k}) = \text{sign}(w_{j,k})\big(|w_{j,k}| - t_j\big)_+,$$

or the hard-threshold rule

$$\delta_{t_j}(w_{j,k}) = w_{j,k} \mathbf{1}_{\{|w_{j,k}| \geq t_j\}},$$

with level-dependent threshold (t_j). [In practice, the wavelet coefficients $w_{j,k}$ at the first three levels are often left untouched. See Donoho and Johnstone (1994) and Nason (1996).] The wavelet estimates are the reconstruction of the thresholded $w_{j,k}$.

7.3. *Threshold selection.* Wavelet estimates with best selection of level-dependent threshold achieve the minimax risk over a wide range of spaces, so it is very crucial for wavelet estimates to select the threshold (t_j). Donoho and Johnstone (1994, 1995b) developed the universal and sure methods to select thresholds. Nason (1996) proposed a threshold selection by cross-validation which removes half of the data each time. However, Nason (1996) reported that for correlated data these methods are unable to find the "true" optimal threshold and noted that level-dependent thresholds must be used [see also Johnstone and Silverman (1994)].

Because of long-range dependence, the wavelet coefficients of the data are correlated. However, by Lemma 3, (31) and (32), the correlations of the wavelet coefficients $w_{j,k}$ decay faster than those of the data (y_i). In particular, at each level the wavelet coefficients $\varepsilon_{j,k}$ of the errors (ε_i) are stationary with mean 0 and correlation

$$\mathrm{Corr}(\varepsilon_{j,k}, \varepsilon_{j,k'}) \asymp |k - k'|^{2(H-M)}.$$

We take the structure of dependence into consideration and propose two level-dependent thresholds.

7.3.1. *Universal threshold.* Lemma 2 implies that the variance of the wavelet coefficients $w_{j,k}$ at level j is approximately equal to that of $w_{j,k}$ at the finest level multiplying by $2^{(2H-1)(J-j)}$, so we select the threshold (t_j) in the following way:

$$(33) \qquad\qquad t_j^U = \sqrt{2\log(n)}\, 2^{(H-1/2)(J-j)}\hat{\sigma},$$

where $\hat{\sigma}$ is the median of absolute values of $w_{j,k}$ at the finest level.

The threshold (33) depends on H, so a simple method is proposed to estimate H as follows: since the wavelet coefficients $w_{j,k}$ at level j are stationary with variance approximately proportional to $2^{j(1-2H)}$, we regress the logarithm of the sample variance of $w_{j,k}$ at level j on j for $3 \le j \le J$, and denote by β the estimated slope. Then H is estimated by $1/2 - \beta/(2\log 2)$.

7.3.2. *Selection by cross-validation.* Nason's cross-validation method removes half the data each time and hence deletes only adjacent observations. As in applying cross-validation to density estimation and nonparametric regression for dependent data [see Hart and Vieu (1990)], in order to reduce correlation we propose a cross-validation method to select the threshold (t_j) by removing more than half the data each time. The algorithm is described as follows.

1. Select an integer l_n and divide the data into 2^{J-l_n} groups as follows: the first 2^{l_n} observations $y_1, \ldots, y_{2^{l_n}}$ are the first group, the next 2^{l_n} observations $y_{2^{l_n}+1}, \ldots, y_{2^{l_n+1}}$ are the second group, so on and so forth.
2. We take the first observation and the $(2^{l_n-1} + 1)$th observation out from each group. All the first observations form a data set of size 2^{J-l_n}, while the $(2^{l_n-1} + 1)$th observations consist of a data set of the same size. A

wavelet estimate $\bar{f}_1^t = \{\bar{f}_{1,b}^t, 1 \le b \le 2^{l_n}\}$ for the former data set is constructed using the threshold $t_j = t 2^{(H-1/2)(J-j)}$, and the latter data set is used to interpolate the values $\bar{g}_1 = \{\bar{g}_{1,b}^t, 1 \le b \le 2^{l_n}\}$ of the former data set. Similarly, we take the second observation and the $(2^{l_n-1} + 2)$th observation, the third observation and the $(2^{l_n-1} + 3)$th observation,..., and the (2^{l_n-1})th observation and the (2^{l_n})th observation out from each group, and obtain $(\bar{f}_a^t, \bar{g}_a)$, $a = 2,\ldots,2^{l_n-1}$, respectively.

3. Switching the data sets used to construct wavelet estimates and the data sets used to do interpolation in step (2), we repeat step (2) and get $(\bar{f}_a^t, \bar{g}_a)$, $a = 2^{l_n-1} + 1,\ldots,2^{l_n}$.

4. Define

$$M(t) = \sum_{a=1}^{2^{l_n}} \sum_b \left[\bar{f}_{a,b}^t - \bar{g}_{a,b} \right]^2,$$

where $\bar{f}_{a,b}^t$ ($\bar{g}_{a,b}$, resp.) is the bth coordinate of $\bar{f}_a^t$ ($\bar{g}_a$, resp.). Let t^{CV} be the value of t at which $M(t)$ takes its minimum. Then the cross-validation threshold is defined to be

$$(34) \qquad t_j^{\mathrm{CV}} = \left(1 - \frac{l_n \log 2}{\log n} \right)^{-1/2} 2^{(H-1/2)(J-j)} t^{\mathrm{CV}}.$$

The factor $(1 - l_n \log 2/\log n)^{-1/2}$ is an adjustment constant because $\bar{f}_a^t$ are based on 2^{J-l_n} data [see Nason (1996)]. Nason's cross-validation method corresponds to $l_n = 1$.

The cross-validation method not only provides a decrease in correlation within the observations used to construct wavelet estimates, but it also reduces the dependence between the samples used to calculate wavelet estimates and the samples used to assess the wavelet estimates. In fact, the correlation between the samples for construction of $\bar{f}_a$ and for assessment of $\bar{f}_a$ is $2^{-(l_n-1)\alpha}$ of the correlation between two adjacent original observations y_i and y_{i+1}. The dependence within the observations used to construct $\bar{f}_a$ is approximately reduced to $2^{-l_n\alpha}$ of the correlation of the original data. For example, if $\alpha = 0.8$ and the correlation of two adjacent original observations y_i and y_{i+1} is 0.8, then the correlation of two adjacent observations in each selected data set will be $0.264, 0.152, 0.087$ for $l_n = 2, 3, 4$, respectively. Much of the dependence is reduced, and hence the selected threshold will be close to the optimal threshold. Generally, the larger l_n is, the more dependence is reduced. As l_n increases, however, the sample sizes of the selected data sets decrease very fast. For large l_n, we may not have enough data to obtain accurate estimates $\bar{f}_a$. So, in practice, we select a moderate l_n such as 2, 3 or 4. For short-range dependence, cross-validation with $l_n \ge 2$ decreases correlation dramatically. For example, for stationary errors (ε_i) with $\mathrm{Corr}(\varepsilon_i, \varepsilon_j) = \rho^{|i-j|}$, $|\rho| < 1$, the correlation of the selected data will be $(\rho')^{|i-j|}$ with $\rho' = \rho^{2^{l_n}}$.

7.3.3. *Simulations*.   A small-scale Monte Carlo study was conducted to check performance of the methods. We consider the piecewise polynomial example in Nason (1996):

$$
f(x) = \begin{cases} 4x^2(3 - 4x), & x \in [0, 1/2], \\ 4x(4x^2 - 10x + 7)/3 - 3/2, & x \in (1/2, 3/4], \\ 16x(x - 1)^2/3, & x \in (3/4, 1]. \end{cases}
$$

The function was sampled 1024 times evenly on the interval $[0, 1]$, and then simulated errors ε_i were added, where (ε_i) is a stationary Gaussian process with mean 0, variance 0.1 and $\alpha = 0.5$ [see (3)] or $H = 0.75$. As indicated in Nason (1996), these correlated data derail the universal, sure and Nason's cross-validation thresholds.

We applied the proposed methods to this example and repeated 100 times. Based on 100 replications, we found that the estimate of H has the MSE $= 0.003$. As in Nason (1996), for each simulated data set we computed the thresholds t_j^U and t_j^{CV} (with $l_n = 3$) and the "true" optimal threshold $t_j^O = 2^{(H-1/2)(J-j)}t^O$, where t^O was chosen to minimize the integrated squared error between the true function and the thresholded noisy function. Since t_j^U, t_j^{CV} and t_j^O are proportional to $2^{(H-1/2)(J-j)}$, we plot the two thresholds against the "true" optimal threshold in Figure 1 as follows: $t^O = t_j^O 2^{(H-1/2)(j-J)}$ is plotted on the x axis, while on the y axis we plot $t_j^U 2^{(H-1/2)(j-J)}$ and $t_j^{CV} 2^{(H-1/2)(j-J)}$. The line $x = y$ is also plotted to give some idea of whether the method underfits (above the line: too many wavelet coefficients are removed because of the large threshold) or overfits (below the
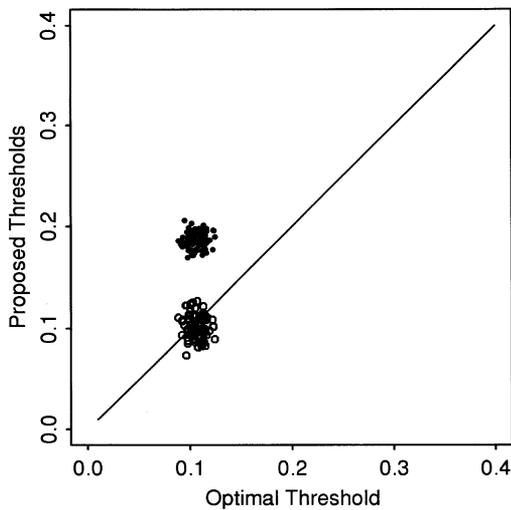

FIG. 1.   *Plot of the thresholds* (t_j^U) *and* (t_j^{CV}) *against the "true" optimal threshold* (t_j^O): $\bullet$ *and* $\bigcirc$ *correspond to* (t_j^U) *and* (t_j^{CV}), *respectively, and the line is* $x = y$.

line: too many wavelet coefficients are retained because the chosen threshold is too small). Thresholds that are clustered around the line $y = x$ indicate that the corresponding method can find the "true" optimal threshold. Figure 1 shows that the cross-validation threshold (t_j^{CV}) performs very well, while the threshold (t_j^U) often tends to underfit. The finding is very similar to that of the universal threshold and Nason's cross-validation threshold for i.i.d. data.

REFERENCES

BARTON, R. J. and POOR, H. V. (1988). Signal detection in fractional Gaussian noise. *IEEE Trans. Inform. Theory* **34** 943–959.

BENASSI, A. and JAFFARD, S. (1993). Wavelet decomposition of one and several dimensional Gaussian processes. In *Recent Advances in Wavelet Analysis* (L. L. Schumaker and G. Webb, eds.) 119–154. Academic, New York.

BERAN, J. (1992). Long-range dependence (with discussion). *Statist. Sci.* **7** 404–427.

BERAN, J. (1994). *Statistics for Long Memory Processes*. Chapman and Hall, New York.

BROWN, L. and LOW, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* To appear.

CHUI, C. K. (1992). *An Introduction to Wavelets*. Academic, New York.

COHEN, A., DAUBECHIES, I., JAWERTH, B. and VIAL, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. *C. R. Acad. Sci. Paris Sér. I Math.* **316** 417–421.

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

DAUBECHIES, I. (1994). Two recent results on wavelets: wavelet bases for the interval, and biorthogonal wavelet diagonalizing the derivative operator. In *Recent Advances in Wavelet Analysis* (L. L. Schumaker and G. Webb, eds.) 237–257. Academic, New York.

DONOHO, D. L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In *Different Perspectives on Wavelet* (I. Daubechies, ed.) 173–205. Amer. Math. Soc., Providence, RI.

DONOHO, D. L. (1995). Nonlinear solutions of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126.

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. (1995a). Minimax estimation via wavelet shrinkage. Unpublished manuscript.

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

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1993). Density estimation by wavelet thresholding. Technical Report 419, Dept. Statistics, Stanford Univ.

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. and LOW, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944–970.

FLANDRIN, P. (1989). On the spectrum of fractional Brownian motion. *IEEE Trans. Inform. Theory* **35** 197–199.

FLANDRIN, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theory* **38** 910–917.

HALL, P. and HART, J. D. (1990a). Nonparametric regression with long-range dependence. *Stochastic Process. Appl.* **36** 339–351.

HALL, P. and HART, J. D. (1990b). Convergence rates in density estimation for data from infinite-order moving average processes. *Probab. Theory Related Fields* **87** 253–274.

HART, J. D. and VIEU, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* **18** 873–890.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1994). Wavelet threshold estimators for data with correlated noise. Technical report, Dept. Statistics, Stanford Univ.

KESHNER, M. S. (1982). $1/f$ noise. *Proc. IEEE* **70** 212–218.

KOLACZYK, E. (1994). Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noise data. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

MALLAT, (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11** 674–693.

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. B. and WALLIS, J. R. (1968). Noah, Joseph and operational hydrology. *Water Resource Research* **4** 909–918.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Computer experiments with fractional Gaussian noises. *Water Resource Research* **5** 228–267.

MEYER, Y. (1992). *Wavelets and Operators*. Oxford Univ. Press.

NASON, G. P. (1996). Wavelet shrinkage using cross-validation. *J. Roy. Statist. Soc. Ser. B.* **58** 463–479.

RAMANATHAN, J. and ZEITOUNI, O. (1991). On the wavelet transform of fractional Brownian motion. *IEEE Trans. Inform. Theory* **37** 1156–1158.

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302.

TEWFIK, A. H. and KIM, M. (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inform. Theory* **38** 904–909.

WANG, Y. (1994). Nonlinear solutions of linear inverse problems via wavelet–vaguelette decomposition for data with long-range dependence. Technical report, Dept. Statistics, Univ. Missouri–Columbia.

WANG, Y. (1995a). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397.

WANG, Y. (1995b). Change-point analysis via wavelet–vaguelette decomposition for indirect data. Technical report, Dept. Statistics, Univ. Missouri–Columbia.

WORNELL, G. W. (1990). A Karhunen–Loève-like expansion for $1/f$ processes via wavelets. *IEEE Trans. Inform. Theory* **36** 859–861.

WORNELL, G. W. and OPPENHEIM, A. V. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing* **40** 611–623.

DEPARTMENT OF STATISTICS
222 MATHEMATICAL SCIENCES BUILDING
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65211