On the Minimax Optimality of Block Thresholded Wavelet Estimators with Long Memory Data

Linyuan Li[*]

Department of Mathematics and Statistics, University of New Hampshire, USA,

Yimin Xiao [†]

Department of Statistics and Probability, Michigan State University, USA

October 2, 2006

Abstract

We consider the estimation of nonparametric regression function with long memory data and investigate the asymptotic rates of convergence of wavelet estimators based on block thresholding. We show that the estimators achieve optimal minimax convergence rates over a large class of functions that involve many irregularities of a wide variety of types, including chirp and Doppler functions, and jump discontinuities. Therefore, in the presence of long memory noise, wavelet estimators still provide extensive adaptivity to many irregularities of large function classes.

Short title: Wavelet Estimators with Long Memory Data
2000 Mathematics Subject Classification: Primary: 62G07; Secondary: 62C20
Keywords: Adaptive estimation; block thresholding; long range dependent data; minimax estimation; nonlinear wavelet-based estimator; rate of convergence

1 Introduction

Hall, et al. (1999) considered the nonparametric regression

$$Y_m = g(x_m) + \varepsilon_m, \quad m = 1, 2, \cdots, n, \tag{1.1}$$

where $x_m = m/n \in [0,1]$, $\varepsilon_1, \cdots, \varepsilon_n$ are independent, identically distributed (i.i.d.) normal random variables with mean 0 and variance σ^2, and g belongs to a large function class $\mathcal{H}$ (its definition will be given in the next section). In order to improve adaptivity when estimating g using wavelet methods, Hall, et al. (1999) introduced a local block thresholding estimator which thresholds empirical wavelet coefficients in groups rather than individually and showed that the estimator achieves optimal minimax convergence rates over a large class of functions $\mathcal{H}$ that involve many irregularities of a wide variety of types, including chirp and Doppler functions, and jump discontinuities. Therefore, wavelet estimators provide extensive adaptivity to many irregularities of large function classes. Cai (2002) considered the asymptotic and numerical properties of a class of block thresholding estimators for model (1.1) with i.i.d. Gaussian errors. He investigated the block size and the thresholding constant such that the corresponding block thresholding estimators achieve optimal convergence rates for both global and local estimation over a large class of functions as in Hall, et al. (1999). However, in many fields which include agronomy, astronomy, economics, environmental sciences, geosciences, hydrology, signal and image processing, it is sometimes unrealistic to assume that the observational errors are independent or short-range dependent. Instead, these observational errors exhibit slow decay in correlation which is often referred to as long-range dependence or long memory (formal definition will be given in next section).

In this paper, we consider the estimation of nonparametric regression function in model (1.1) with long memory error and investigate the asymptotic convergence rates of the block thresholding wavelet estimators. We are particularly interested in describing the effect of long memory on the performance of the wavelet estimators in terms of the Hurst index α. We show that these estimators achieve optimal minimax convergence rates over a large class of functions as in Hall, et al. (1999). Hence, our results extend those of Hall, et al. (1999) from i.i.d. Gaussian errors to long memory stationary Gaussian processes.

The literature on long-range dependence is very extensive, see, e.g., the monograph of Beran (1994) and the references cited therein. Estimation for data with long-range dependence is quite different from that for observations with independence or short-range dependence. For example, Hall and Hart (1990) showed that the convergence rates of mean regression function estimators differ from those under the assumptions of independence or short-range dependence. Robinson and Hidalgo (1997) considered a general multiple linear regression model with long range dependent errors and showed that the weighted least squares estimator has $\sqrt{n}$-asymptotic normality. As to the non-parametric model (1.1), Csörgö and Mielniczuk (1995) and Robinson (1997) proposed kernel estimators of mean regression function and proved central limit theorems when the errors are long range dependent Gaussian sequences and stationary martingale difference sequences, respectively. They all assumed that the regression function g is a fixed continuously differentiable function. In this

paper, we follow the framework of Hall, et al. (1999) and consider regression functions g which belong to a large function class $\mathcal{H}$ as in that paper (see Section 2 for its definition).

Wavelet methods in nonparametric curve estimation has become a well-known technique. For a systematic discussion of wavelets and their applications in statistics, see the recent monograph by Härdle, et al. (1998). The major advantage of wavelet method is its adaptability to the degree of smoothness of the underlying unknown curve. These wavelet estimators typically achieve optimal convergence rates over exceptionally large function spaces. For reference, see Donoho, et al. (1995, 1996) and Donoho and Johnstone (1998). Hall and Patil (1995, 1996) also demonstrated explicitly the extraordinary local adaptability of wavelet estimators in handling discontinuities. They showed that discontinuities of the unknown curve have a negligible effect on the performance of nonlinear wavelet curve estimators. All of the aforementioned works are under the assumption that the errors are independent.

There are a few papers which consider the estimation of regression functions with long memory Gaussian noise. Among them, Wang (1996), Johnstone and Silverman (1997) and Johnstone (1999) examined the asymptotic properties of wavelet-based estimators of mean regression functions with long memory Gaussian noise. They showed that these estimators achieve minimax rates over wide range of Besov spaces. However, they didn't study the so called "sampled data model" (1.1) directly, instead, they studied its asymptotic model or the "sequence space model"

$$y_\lambda = \theta_\lambda + \varepsilon^\alpha \gamma_j z_\lambda, \qquad \lambda \in \Lambda,$$

which is derived from the empirical wavelet transformation of the corresponding "white-noise model"

$$Y_\epsilon(dt) = f(t)dt + \varepsilon\, W(dt), \qquad t \in [0,1]. \tag{1.2}$$

For details regarding the above ε, γ_j and Λ, see Johnstone and Silverman (1997, p.339). Then they argued that the results derived from the sequence space model can be carried over to the sampled data model, based on the asymptotic equivalence between models (1.1) and (1.2) when the underlying curves are smooth enough. Based on the above assumption that the optimal minimax convergence rates obtained in the wavelet domain or sequence space model are equivalent to those derived from the original time domain or sampled data model, they proved that the unbiased risk based thresholding estimators attain optimal minimax convergence rates over a broad range of Besov classes. However, this implication may not be true, when the underlying curve f is not sufficiently smooth. In particular, when f belongs to a class of functions that may not even be continuous, the results for the sequence space model may be questionable for the sampled data model. In this paper, we consider the sampled data model (1.1) directly with g belonging to a function space $\mathcal{H}$, which includes many discontinuous functions.

3

In the next section, we introduce the regression model, function space $\mathcal{H}$, elements of wavelet transform and the block thresholding estimator of the mean regression function. The main results are described in Section 3, while their proofs appear in Section 4.

2   Notations and Estimators

We consider the nonparametric regression model (1.1) with long memory Gaussian errors. Formally, we assume that $\{\varepsilon_n,\ n \geq 1\}$ is a sequence of long-range dependent or long memory stationary Gaussian random variables with mean 0 and variance σ^2. By long-range dependence, we mean there exist two constants $C_0 > 0$ and $\alpha \in (0, 1]$ such that

$$r(j) = E\big(\varepsilon_1\varepsilon_{1+j}\big) \sim C_0|j|^{-\alpha}, \tag{2.1}$$

where $a_j \sim b_j$ means that $a_j/b_j \to 1$ when $j \to \infty$.

As in Hall, et al. (1999), we will assume that the unknown mean regression function g belongs to a large function class $\mathcal{H}$, which is defined as follows:

Definition 2.1 *For given integers $v, N \geq 1$ and positive constants $C_1, C_2, C_3, 0 \leq s_1 < s_2 < N$ and $0 \leq \gamma < (2s_1 + 1)/(2s_2 + 1)$, let $\mathcal{H} = \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$ denote the class of functions g such that for any $i \geq 0$ there exists a set of integers S_i for which the following conditions hold: $\mathrm{card}(S_i) \leq C_3 2^{i\gamma}$ and*

   1. for each $j \in S_i$ there exist constants $a_0 = g(j/2^i), a_1, \ldots, a_{N-1}$ such that

$$\left| g(x) - \sum_{l=0}^{N-1} a_l(x - 2^{-i}j)^l \right| \leq C_1 2^{-is_1} \quad \text{for all}\ \ x \in [j/2^i, (j+v)/2^i];$$

   and

   2. for each $j \notin S_i$ there exist constants $a_0 = g(j/2^i), a_1, \ldots, a_{N-1}$ such that

$$\left| g(x) - \sum_{l=0}^{N-1} a_l(x - 2^{-i}j)^l \right| \leq C_2 2^{-is_2} \quad \text{for all } x \in [j/2^i, (j+v)/2^i].$$

Note that when $C_3 = 0$, the constant C_1 plays no role in the above definition and the function class $\mathcal{H}(s_1, s_2, \gamma, C_1, C_2, 0, N, v)$ contains the Besov class $B^{s_2}_{\infty\infty}(C_2)$ as a subset for all $s_1 < s_2$, $\gamma > 0$ and with $C_1 > 0$ depending on choice of the other constants. Furthermore, as pointed out in Hall, et al. (1999), a function $g \in \mathcal{H}$ can be regarded as the superposition of a smooth function g_2 from the Besov space $B^{s_2}_{\infty\infty}$ with a function g_1 which may have irregularities of different types – such as jump discontinuities and high frequency oscillations.

However, the irregularities of g_1 are controlled by the constants C_3 and γ so that they do not overwhelm the fundamental structure of g. We refer to Hall, et al. (1998, 1999) and Cai (2002) for more discussions about the function classes $\mathcal{H}$.

Next we introduce some facts about wavelets that will be used in the sequel. Let $\phi(x)$ and $\psi(x)$ be father and mother wavelets, having the following properties: ϕ and ψ are bounded and compactly supported, and $\int \phi = 1$. Let

$$\phi_{i_0 j}(x) = 2^{i_0/2}\phi(2^{i_0}x - j), \quad \psi_{ij}(x) = 2^{i/2}\psi(2^i x - j), \quad x \in \mathbb{R}, \ i_0, \ i, j \in \mathbb{Z}.$$

Then the collection $\{\phi_{i_0 j}, \psi_{ij}, i \geq i_0, j \in \mathbb{Z}\}$ is an orthonormal basis (ONB) of $L^2(\mathcal{R})$. Furthermore, let V_{i_0} and W_i be linear subspaces of $L^2(\mathcal{R})$ with the ONB $\phi_{i_0 j}, \ j \in \mathbb{Z}$ and $\psi_{ij}, \ j \in \mathbb{Z}$, respectively, we have the following decomposition

$$L^2(\mathcal{R}) = V_{i_0} \oplus W_{i_0} \oplus W_{i_0+1} \oplus W_{i_0+2} \oplus \cdots .$$

Therefore, for all $f \in L^2(\mathcal{R})$,

$$f(x) = \sum_{j \in \mathbb{Z}} \alpha_{i_0 j}\phi_{i_0 j}(x) + \sum_{i \geq i_0}\sum_{j \in \mathbb{Z}} \beta_{ij}\psi_{ij}(x),$$

where

$$\alpha_{i_0 j} = \int f(x)\phi_{i_0 j}(x)\,dx, \quad \beta_{ij} = \int f(x)\psi_{ij}(x)\,dx.$$

The orthogonality properties of ϕ and ψ imply:

$$\int \phi_{i_0 j_1}\phi_{i_0 j_2} = \delta_{j_1 j_2}, \quad \int \psi_{i_1 j_1}\psi_{i_2 j_2} = \delta_{i_1 i_2}\delta_{j_1 j_2}, \quad \int \phi_{i_0 j_1}\psi_{ij_2} = 0, \ \forall i_0 \leq i, \qquad (2.2)$$

where δ_{ij} denotes the Kronecker delta, i.e., $\delta_{ij} = 1$, if $i = j$; and $\delta_{ij} = 0$, otherwise.

In our regression model, the mean function g is supported on the unit interval $[0, 1]$. Therefore, we confine our attention to the wavelet basis of $L^2[0, 1]$ given by Cohen, et al. (1993). That is, we assume that the collection of $\{\phi_{i_0 j}, j = 0, 1, \ldots, 2^{i_0-1}; \ \psi_{ij}, i \geq i_0 \geq 0, j = 0, 1, \ldots, 2^i - 1\}$ forms an orthonormal basis of $L^2[0, 1]$. Since, in this paper, we require varnishing moments up to $N - 1$ for both ϕ and ψ ($\int x^k\phi(x)dx = 0, k = 1, 2, \ldots, N - 1; \int x^k\psi(x)dx = 0, k = 0, 1, \ldots, N - 1$), the so-called Coiflets will be used here. Moreover, we will assume the wavelets ϕ and ψ are continuous in $[0, 1]$. Since the supports of our wavelets are contained in the interval $[0, 1]$, we confine the function spaces $\mathcal{H}$ with $v = 1$. For more on these compactly supported wavelets with high order varnish moments and continuous derivative, see Daubechies (1992).

Hence, the corresponding wavelet expansion of $g(x)$ is

$$g(x) = \sum_{j=0}^{2^{i_0}-1} \alpha_{i_0 j}\phi_{i_0 j}(x) + \sum_{i \geq i_0}\sum_{j=0}^{2^i-1} \beta_{ij}\psi_{ij}(x), \qquad (2.3)$$

5

where

$$\alpha_{i_0 j} = \int_0^1 g(x)\phi_{i_0 j}(x)\, dx \quad \text{and} \quad \beta_{ij} = \int_0^1 g(x)\psi_{ij}(x)\, dx.$$

Before we provide the wavelet estimator of the mean regression function, we assume the sample size $n = 2^{i_1}$ for the convenience of exposition. Also in the statement below, the notation $2^{i(n)} \simeq h(n)$ means that $i(n)$ is chosen to satisfy the inequalities $2^{i(n)} \le h(n) < 2^{i(n)+1}$. For the sake of simplicity, we always omit the dependence on n for i. The idea of defining a block thresholding estimator is to threshold empirical wavelet coefficients in groups rather than individually (Hall, et al. 1998, 1999). At each resolution level i, the integers $\{0, 1, \ldots, 2^i - 1\}$ are divided among consecutive, nonoverlapping blocks of length l, say $\Gamma_{ik} = \{j : (k-1)l + 1 \le j \le kl\}$, $-\infty < k < \infty$.

Since the convergence rates of the estimators are different for long memory parameter $\alpha \in (0,1)$ and $\alpha = 1$, we treat these two cases separately (see Theorem 3.1 in next section).

Define

$$\hat{G}_{i_1}(x) = n^{-1/2} \sum_{m=1}^{n} Y_m \phi_{i_1 m}(x).$$

Let the coefficients $\hat{\alpha}_{i_0 j}$ and $\hat{\beta}_{ij}$ be given by

$$\text{Proj}_{V_{i_0}}(\hat{G}_{i_1}) = \sum_{j=0}^{2^{i_0}-1} \hat{\alpha}_{i_0 j}\phi_{i_0 j} \quad \text{and} \quad \text{Proj}_{W_i}(\hat{G}_{i_1}) = \sum_{j=0}^{2^i-1} \hat{\beta}_{ij}\psi_{ij},$$

and put $\hat{B}_{ik} = l^{-1}\sum_{(ik)}\hat{\beta}_{ij}^2$, where $\sum_{(ik)}$ denotes summation over $j \in \Gamma_{ik}$ and l denotes the block length.

When $\alpha \in (0,1)$, our wavelet estimator of g is defined as

$$\hat{g}(x) = \sum_{j=0}^{2^{i_0}-1} \hat{\alpha}_{i_0 j}\phi_{i_0 j}(x) + \sum_{i=i_0}^{i_1-1} \sum_{k=-\infty}^{\infty} \Big(\sum_{(ik)}\hat{\beta}_{ij}\psi_{ij}(x)\Big) I\big(\hat{B}_{ik} > \delta_i\big), \quad (2.4)$$

where the smoothing parameter i_0 satisfies $2^{i_0} \simeq n^{\alpha/(2N+\alpha)}$, the block-length $l = (\log n)^{\theta}$ with $\theta > 1/\alpha$ and δ_i $(i_0 \le i < i_1)$ are the level-dependent thresholding satisfying $\delta_i = 48\tau_i^2$ with $\tau_i^2 = C_4 n^{-\alpha} 2^{-i(1-\alpha)}$, where $C_4 > 0$ is the constant defined by

$$C_4 = C_0 \int_0^1 \int_0^1 |x-y|^{-\alpha}\, \psi(x)\, \psi(y)\, dxdy. \quad (2.5)$$

Similarly for $\alpha = 1$, our wavelet estimator of g is defined as

$$\hat{g}_1(x) = \sum_{j=0}^{2^{i_0}-1} \hat{\alpha}_{i_0 j}\phi_{i_0 j}(x) + \sum_{i=i_0}^{i_1-1} \sum_{k=-\infty}^{\infty} \Big(\sum_{(ik)}\hat{\beta}_{ij}\psi_{ij}(x)\Big) I\big(\hat{B}_{ik} > \delta_i\big), \quad (2.6)$$

where the smoothing parameter i_0 is chosen to satisfy $2^{i_0} \simeq n^{1/(2N+1)}$, the block-length $l = (\log n)^{\theta}$ with $\theta > 1$ and δ_i $(i_0 \leq i < i_1)$ are the level-dependent thresholding satisfying $\delta_i = 48\xi_i^2$ with $\xi_i^2 = 2C_0 n^{-1} \log(n2^{-i}e)$. Note that the only difference for $\alpha = 1$ is in the threshold δ_i.

In practice, the empirical coefficients $\hat{\alpha}_{i_0j}$ and $\hat{\beta}_{ij}$ in (2.4) and (2.6) can be computed from the observations Y_m by "subband filtering schemes". It is worthwhile to note that, for each fixed $i_0 \leq i < i_1$ in (2.4) and (2.6), the sum in k only contains at most $2^i l^{-1}$ non-zero terms.

Throughout this paper, we use C to denote positive and finite constants whose value don't depend on sample size n and may change from line to line. Specific constants are denoted by C_0, C_1, C_2 and so on.

3   Main results and discussions

The following theorem shows that the wavelet-based estimator, based on block thresholding of the wavelet coefficients, attains exactly the optimal convergence rate over a large range of function classes.

Theorem 3.1 *Let the wavelets ϕ and ψ, and the estimators $\hat{g}$ and $\hat{g}_1$ be given as in Section 2. Then there exists a constant $C_5 = C(s_1, s_2, \gamma, C_1, C_2, C_3, N, v) > 0$ such that the following hold:*

(i). When $\alpha \in (0, 1)$,

$$\sup_{g \in \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)} E \int (\hat{g} - g)^2 \leq C_5 \, n^{-2s_2\alpha/(2s_2+\alpha)}.$$

(ii). When $\alpha = 1$,

$$\sup_{g \in \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)} E \int (\hat{g}_1 - g)^2 \leq C_5 \left(\frac{\log n}{n}\right)^{2s_2/(2s_2+1)}.$$

Remark 3.1 Hall and Hart (1990, Theorem 2.2 and Remarks on p.343–345) showed that, for the regression model (1.1) with Gaussian autoregression errors $\{\varepsilon_m\}$, the minimax convergence rate over the Hölder class $H^{s_2}(C_2)$ (where $s_2 \geq 2$ is an integer) is $n^{-2s_2\alpha/(2s_2+\alpha)}$ when $\alpha \in (0, 1)$ and $(n^{-1}\log n)^{2s_2/(2s_2+1)}$ when $\alpha = 1$, respectively. Since $\mathcal{H}$ contains $H^{s_2}(C_2)$ as a subset, our block thresholded estimator attains optimal convergence rates over a larger range of function classes.

Remark 3.2 Cai (2002) investigated the asymptotic and numerical properties of a class of block thresholding estimators for wavelet regression, when the errors $\{\varepsilon_m\}$ are independent homoscedastic Gaussian noise. Our results can easily be extended to heteroscedastic errors case. Cai (2002) also showed that the block length $l = (\log n)^s$, $s = 1$ attains both global and local optimal convergence rates for i.i.d errors, i.e., in order to attain optimal convergence rates, the block length size $l = (\log n)^s$ shouldn't be too large ($s \leq 1$) for local estimation and shouldn't be too small ($s \geq 1$) for global estimation. Hall, et al. (1999) consider block length $l = (\log n)^2$ for i.i.d. errors and suggest block length $l = n^c$, for some $c > 0$ for long range dependent errors. We find that, in order to attain the minimax convergence rates in the long memory case, we can choose the block lengths $l = (\log n)^\theta$ with $\theta > 1/\alpha$ for $\alpha \in (0, 1]$. It is interesting to note that our block lengths l depend on the long memory parameter α, but they are much smaller than $l = n^c$ suggested by Hall, et al. (1999). From Cai (2002), one would expect that our estimator could achieve simultaneously the global and local adaptivity.

Remark 3.3 This is a technical remark about the choice of block length related to Remark 3.2. From Lemma 4.2 and the proof of Theorem 3.1, it can be seen that, for any $\alpha \in (0, 1]$, one can choose the block length $l = C (\log n)^{1/\alpha} (\log \log n)$, where $C > 0$ is a large constant and the same conclusions of Theorem 3.1 hold. We have chosen $l = (\log n)^\theta$ so that it doesn't involve an unknown constant C in the block length.

Remark 3.4 For long-range dependent errors, one generally assumes $r(j) = E(\varepsilon_1 \varepsilon_{1+j}) = L(j) \cdot |j|^{-\alpha}$, where $L(x)$ is a slowly varying function at ∞, i.e., for all $a > 0$,

$$\lim_{x \to \infty} L(ax)/L(x) = 1.$$

See Bingham, et al. (1987) for more information on slowly varying functions. In this paper, we have taken the slowly varying function L to be the constant C_0 in (2.1) to simplify the presentations of our results (The threshold δ_i can be expressed in terms of C_0 explicitly and the proof later on is much clearer than that involving $L(x)$). With some modifications, one can show that results similar to Theorems 3.1 hold under the more general slowly varying functions L assumption.

Remark 3.5 Because of the long-range dependence, our thresholds δ_i depend not only on the level but also on the unknown long memory parameter α. So are the smoothing parameter i_0 and the block length l. Wang (1996, p.480) and Johnstone and Silverman (1997, p.340) have provided simple methods to estimate the long memory parameter α. Delbeke and Van Assche (1998) and Abry and Veitch (1998) have also provided wavelet-based, statistically and computationally efficient estimators of α based on the wavelet coefficients and have shown

that these estimators are unbiased, consistent and have asymptotically a normal distribution. Thus, in practice, we assume that the parameter α has been estimated and treat it as known. Another remark is on the threshold $\delta_i = 48\tau_i^2$ for $\alpha \in (0,1)$ or $\delta_i = 48\xi_i^2$ for $\alpha = 1$, where $\tau_i^2 = \sigma_i^2 = \text{Var}(\hat{\beta}_{ij}) = C_4 n^{-\alpha} 2^{-i(1-\alpha)}$ or $\xi_i^2 = \sigma_i^2 = 2C_0 n^{-1}\log(n2^{-i}e)$ (for details, see Section 4). This noise variance σ_i^2 at each level i can be estimated from the data by using, for example, the robust median absolute deviation estimator $\hat{\sigma}_i = MAD\{\hat{\beta}_{ij}, j = 0, 1, \cdots, 2^i - 1\}/0.6745$. Hence, we can also treat σ_i^2 as known (see Johnstone and Silverman (1997) for more details).

4 Proof of Theorem 3.1

The method of proving Theorem 3.1 is similar to that of Theorem 4.1 of Hall, et al. (1999). The difference is that we consider the errors $\{\varepsilon_m, m \geq 1\}$ to be a stationary Gaussian process with long memory, instead of i.i.d. random variables in their paper. Hence, several nontrivial technical difficulties have to be overcome.

We will break the proof of Theorem 3.1 into several parts. Observing that the orthogonality (2.2) implies

$$E\|\hat{g} - g\|_2^2 = T_1 + T_2 + T_3 + T_4, \tag{4.1}$$

where

$$T_1 = \sum_{i=i_1}^{\infty} \sum_{j=0}^{2^i-1} \beta_{ij}^2,$$

$$T_2 = \sum_{j=0}^{2^{i_0}-1} E(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j})^2 = E\|\text{Proj}_{V_{i_0}}(\hat{G}_{i_1} - g)\|_2^2,$$

$$T_3 = \sum_{i=i_0}^{i_1-1} \sum_{k=-\infty}^{\infty} E\left\{ I(\hat{B}_{ik} > \delta_i) \sum_{(ik)} (\hat{\beta}_{ij} - \beta_{ij})^2 \right\},$$

$$T_4 = \sum_{i=i_0}^{i_1-1} \sum_{k=-\infty}^{\infty} P\left(\hat{B}_{ik} \leq \delta_i\right) \sum_{(ik)} \beta_{ij}^2.$$

The reminder of the proof consists of bounding $T_1, \ldots, T_4$. For this purpose, we need some preparatory results.

The first lemma, which characterizes some properties of the wavelet coefficients of $g \in \mathcal{H}$, is due to Hall, et al. (1999, Proposition 3.2).

Lemma 4.1 *For every function $g \in \mathcal{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$ and our selected Coiflets,*

the wavelet coefficients of g, denoted by α_{ij} and β_{ij}, have the following properties:

$$
\begin{aligned}
|\beta_{ij}| &\leq ||\psi||_1 C_1 2^{-i(s_1+1/2)} & \text{if} \quad j \in S_i, \\
|\beta_{ij}| &\leq ||\psi||_1 C_2 2^{-i(s_2+1/2)} & \text{if} \quad j \notin S_i, \\
|\alpha_{ij} - 2^{-i/2} g(j/2^i)| &\leq ||\phi||_1 C_1 2^{-i(s_1+1/2)} & \text{if} \quad j \in S_i, \\
|\alpha_{ij} - 2^{-i/2} g(j/2^i)| &\leq ||\phi||_1 C_2 2^{-i(s_2+1/2)} & \text{if} \quad j \notin S_i.
\end{aligned}
\tag{4.2}
$$

As in Hall, et al. (1999, p.42), there exist real numbers $r_{i_1 m}$ $(m = 1, \ldots, n)$, which are small when n is large, such that

$$
\begin{aligned}
\alpha_{i_1 m} &= \int g(x)\phi_{i_1 m}(x)dx \\
&= n^{-1/2} \int g\left(\frac{m}{n} + \frac{y}{n}\right)\phi(y)dy \\
&=: n^{-1/2} g\left(\frac{m}{n}\right) - r_{i_1 m}.
\end{aligned}
\tag{4.3}
$$

Thus, we can write (2.4) as

$$
\hat{G}_{i_1}(x) = \sum_{m=1}^{n} (\alpha_{i_1 m} + r_{i_1 m})\phi_{i_1 m}(x) + n^{-1/2} \sum_{m=1}^{n} \varepsilon_m \phi_{i_1 m}(x).
$$

Consequently, we may write for every integer $0 \leq i < i_1$,

$$
\text{Proj}_{W_i}(\hat{G}_{i_1}) = \sum_{j=0}^{2^i - 1} (\beta_{ij} + u_{ij} + U_{ij})\psi_{ij}(x),
$$

$$
\text{Proj}_{V_{i_0}}(\hat{G}_{i_1}) = \sum_{j=0}^{2^i - 1} (\alpha_{i_0 j} + v_{i_0 j} + V_{i_0 j})\phi_{i_0 j}(x).
$$

Since $\text{Proj}_{V_{i_1}}(g) = \sum_j \alpha_{i_1 j}\phi_{i_1 j}(x)$, we have, for $0 \leq i < i_1$, $\text{Proj}_{W_i}(g) = \text{Proj}_{W_i}(\text{Proj}_{V_{i_1}}(g))$. Now from $\text{Proj}_{W_i}(g) = \sum_j \beta_{ij}\psi_{ij}(x)$, we have

$$
\beta_{ij} = \sum_{m=1}^{n} \alpha_{i_1 m}\langle \phi_{i_1 m}, \ \psi_{ij}\rangle.
$$

There are several results related to u_{ij}, $v_{i_0 j}$, U_{ij} and $V_{i_0 j}$ which will be used in the sequel. The real numbers u_{ij} and $v_{i_0 j}$ can be expressed as

$$
u_{ij} = \sum_{m=1}^{n} r_{i_1 m}\langle \phi_{i_1 m}, \ \psi_{ij}\rangle, \qquad v_{i_0 j} = \sum_{m=1}^{n} r_{i_1 m}\langle \phi_{i_1 m}, \ \phi_{i_0 j}\rangle,
\tag{4.4}
$$

and

$$
U_{ij} = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} \varepsilon_m \langle \phi_{i_1 m}, \ \psi_{ij}\rangle, \qquad V_{i_0 j} = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} \varepsilon_m \langle \phi_{i_1 m}, \ \phi_{i_0 j}\rangle.
\tag{4.5}
$$

In the above, $\langle f, g \rangle = \int fg$ is the inner product in $L^2([0,1])$. It follows from Parseval's identity that

$$\sum_{i=i_0}^{i_1-1} \sum_{j=0}^{2^i-1} u_{ij}^2 + \sum_{j=0}^{2^{i_0}-1} v_{i_0j}^2 = \sum_{m=1}^{n} r_{i_1m}^2.$$

Thus, from (4.2), (4.3) and our choice of γ, we have

$$\sum_{m=1}^{n} r_{i_1m}^2 \leq C_1 C_3 n^{-(2s_1+1-\gamma)} + C_2 n^{-2s_2} \leq C n^{-2s_2/(2s_2+1)}. \tag{4.6}$$

Because our wavelets have compact support, there are at most 2^{i_1-i} non-zero terms of $\langle \phi_{i_1m}, \psi_{ij} \rangle, m = 1, 2, \ldots, n$. Moreover,

$$|\langle \phi_{i_1l}, \psi_{ij} \rangle| \leq 2^{i/2-i_1/2} ||\psi||_\infty ||\phi||_1 \quad \text{and} \quad |r_{i_1l}| \leq (C_1 \vee C_2) 2^{-i_1(s_1+1/2)}.$$

Hence we have, for $\forall i \geq i_0$,

$$|u_{ij}| \leq C 2^{i_1-i} 2^{-i_1(s_1+1/2)} 2^{i/2-i_1/2} \leq C 2^{-i(s_1+1/2)}. \tag{4.7}$$

Now let us calculate the variance of U_{ij}. Since $EU_{ij} = 0$, we have

$$\begin{aligned}
\mathrm{Var}(U_{ij}^2) &= \frac{1}{n} \sum_{m=1}^{n} E(\varepsilon_m^2) \langle \phi_{i_1m}, \psi_{ij} \rangle^2 + \frac{1}{n} \sum_{m=1}^{n} \sum_{k \neq m} E(\varepsilon_m \varepsilon_k) \langle \phi_{i_1m}, \psi_{ij} \rangle \langle \phi_{i_1k}, \psi_{ij} \rangle \\
&= \frac{\sigma^2}{n} + \frac{1}{n} \sum_{m=1}^{n} \sum_{k \neq m} r(m-k) \langle \phi_{i_1m}, \psi_{ij} \rangle \langle \phi_{i_1k}, \psi_{ij} \rangle \\
&=: \frac{\sigma^2}{n} + I_1.
\end{aligned} \tag{4.8}$$

Recall that $n = 2^{i_1}$ and by a change of variables, we may write

$$\begin{aligned}
I_1 &= 2^i \sum_{m=1}^{n} \sum_{k \neq m} r(m-k) \iint \phi(2^{i_1}x - m)\phi(2^{i_1}y - k)\psi(2^i x - j)\psi(2^i y - j) \, dxdy \\
&= 2^i \iint \phi(u)\phi(v) \left\{ \sum_{m=1}^{n} \sum_{k \neq m} r(m-k)\psi\left(2^i \frac{u+m}{n} - j\right)\psi\left(2^i \frac{v+k}{n} - j\right) \frac{1}{n^2} \right\} dudv.
\end{aligned} \tag{4.9}$$

We first consider the case when $\alpha \in (0,1)$. It follows from (2.1) that as $n \to \infty$,

$$\begin{aligned}
&\sum_{m=1}^{n} \sum_{k \neq m} r(m-k) \, \psi\left(2^i \frac{u+m}{n} - j\right)\psi\left(2^i \frac{v+k}{n} - j\right) \frac{1}{n^2} \\
&\sim C_0 \, n^{-\alpha} \sum_{m=1}^{n} \sum_{k \neq m} \left|\frac{m}{n} - \frac{k}{n}\right|^{-\alpha} \psi\left(2^i \frac{u+m}{n} - j\right)\psi\left(2^i \frac{v+k}{n} - j\right) \frac{1}{n^2} \\
&\sim C_0 \, n^{-\alpha} \int_0^1 \int_0^1 |x-y|^{-\alpha} \, \psi(2^i x - j)\psi(2^i y - j) \, dxdy \\
&= C_0 \, n^{-\alpha} 2^{(\alpha-2)i} \int_0^1 \int_0^1 |x-y|^{-\alpha} \, \psi(x)\psi(y) \, dxdy.
\end{aligned} \tag{4.10}$$

11

uniformly for all u, v in the support of ϕ. Combining (4.9) and (4.10), we have

$$I_1 \sim C_4 n^{-\alpha} 2^{-i(1-\alpha)} \quad \text{as } n \to \infty, \tag{4.11}$$

where $C_4 > 0$ is the constant defined by (2.5). Thus, it follows from (4.8), (4.11) and the fact that $0 < \alpha < 1$ that for any fixed integer i,

$$\text{Var}(U_{ij}) \sim C_4 \, n^{-\alpha} \, 2^{-i(1-\alpha)} \quad \text{as } n \to \infty. \tag{4.12}$$

Similarly, we have $EV_{i_0 j} = 0$ and for any fixed integer i,

$$\text{Var}(V_{i_0 j}) \sim C_6 \, n^{-\alpha} \, 2^{-i_0(1-\alpha)} \quad \text{as } n \to \infty, \tag{4.13}$$

where $C_6 > 0$ is the constant given by

$$C_6 = C_0 \int_0^1 \int_0^1 |x - y|^{-\alpha} \, \phi(x) \, \phi(y) \, dx dy.$$

Next we consider the case when $\alpha = 1$. Similar to (4.9), we have for each $c > 0$,

$$I_1 \sim 2^i \sum_{n=1}^{n} \sum_{|m-k|>c} r(m-k) \iint \phi(2^{i_1}x - m)\phi(2^{i_1}y - k)\psi(2^i x - j)\psi(2^i y - j) \, dx dy$$

$$\sim 2^i \iint \phi(u)\phi(v) \left\{ \sum_{m=1}^{n} \sum_{|m-k|>c} r(m-k)\psi\left(2^i \frac{u+m}{n} - j\right)\psi\left(2^i \frac{v+k}{n} - j\right) \frac{1}{n^2} \right\} du dv. \tag{4.14}$$

Apply similar argument as in (4.10) (see also Hall and Hart (1990, p.350), we have

$$\sum_{m=1}^{n} \sum_{|m-k|>c} r(m-k) \, \psi\left(2^i \frac{u+m}{n} - j\right)\psi\left(2^i \frac{v+k}{n} - j\right) \frac{1}{n^2}$$

$$\sim C_0 \, n^{-1} \iint_{|x-y|>c/n} |x - y|^{-1} \, \psi(2^i x + 2^i u/n)\psi(2^i y + 2^i v/n) \, dx dy$$

$$\sim C_0 \, n^{-1} 2^{-i} \iint_{|p-q|>c2^i/n} |p - q|^{-1} \, \psi(p)\psi(q) \, dp dq \tag{4.15}$$

$$= C_0 \, n^{-1} 2^{-i} \int \psi(q) \int_{|y|>c2^i/n} |y|^{-1} \psi(y + q) dy dq$$

$$\sim C_0 \, n^{-1} 2^{-i} \, 2 \log(n2^{-i}e) \int \psi^2(q) dq$$

$$= 2C_0 \, n^{-1} 2^{-i} \, \log(n2^{-i}e).$$

Therefore, when $\alpha = 1$, from (4.8), (4.14) and (4.15), we have

$$\text{Var}(U_{ij}) \sim 2C_0 \, n^{-1} \log(n2^{-i}e). \tag{4.16}$$

12

Similarly, we have

$$EV_{i_0j} = 0, \qquad \mathrm{Var}(V_{i_0j}) \sim 2C_0\, n^{-1} \log(n2^{-i_0}e). \qquad (4.17)$$

Now we are in the position to bound the four terms $T_1, \ldots, T_4$ in (4.1), respectively.
Bound for T_1: Since $g \in \mathcal{H}$, we use Lemma 4.1 to derive

$$\begin{aligned}
T_1 &= \sum_{i=i_1}^{\infty} \Big(\sum_{j \in S_i} + \sum_{j \notin S_i} \Big) \beta_{ij}^2 \\
&\le C \sum_{i=i_1}^{\infty} 2^{i\gamma} 2^{-i(2s_1+1)} + C \sum_{i=i_1}^{\infty} 2^i 2^{-i(2s_2+1)} \\
&\le Cn^{-(2s_1+1-\gamma)} + Cn^{-2s_2} \\
&\le Cn^{-2s_2/(2s_2+1)}.
\end{aligned}$$

Therefore, we have $T_1 \le Cn^{-2s_2\alpha/(2s_2+\alpha)}$ when $\alpha \in (0,1)$, and $T_1 \le C\big(n^{-1}\log n\big)^{2s_2/(2s_2+1)}$ when $\alpha = 1$.
Bound for T_2. From the definition of $\hat\alpha_{i_0j}$, (4.6) and (4.13), we have for $\alpha \in (0,1)$

$$\begin{aligned}
T_2 &= \sum_{j=0}^{2^{i_0}-1} v_{i_0j}^2 + \sum_{j=0}^{2^{i_0}-1} EV_{i_0j}^2 \\
&\le Cn^{-2s_2/(2s_2+1)} + C2^{i_0} n^{-\alpha} 2^{-i_0(1-\alpha)} \\
&\le Cn^{-2s_2\alpha/(2s_2+\alpha)},
\end{aligned}$$

where the last inequality follows from our choice of i_0 in (2.4).

When $\alpha = 1$, from (4.6) and (4.17), we have

$$\begin{aligned}
T_2 &\le Cn^{-2s_2/(2s_2+1)} + 2^{i_0} 2C_0\, n^{-1} \log(n2^{-i_0}e) \\
&\le C\Big(\frac{\log n}{n}\Big)^{2s_2/(2s_2+1)},
\end{aligned}$$

where the last inequality follows from our choice of i_0 also and $N > s_2$.
Bound for T_3. In order to derive an upper bound for T_3, we will make use of the following lemma, whose proof will be deferred to the end of our paper.

Lemma 4.2 _Let U_{ij} be the Gaussian random variables defined as in (4.5) and let $\tau_i^2 = C_4 n^{-\alpha} 2^{-i(1-\alpha)}$ and $\xi_i^2 = 2C_0 n^{-1} \log(n2^{-i}e)$._

(i). If $\alpha \in (0,1)$, then for all integers i, k, and for all real numbers $\lambda \ge 4l\,\tau_i^2$,

$$P\Big\{ \sum_{(ik)} U_{ij}^2 \ge \lambda \Big\} \le \exp\Big(-\frac{\lambda}{C_7\, l^{1-\alpha}\, \tau_i^2} \Big), \qquad (4.18)$$

where $C_7 > 0$ is an absolute constant.

13

(ii). If $\alpha = 1$, then for all real numbers $\lambda \geq 4l\,\xi_i^2$,

$$P\Big\{\sum_{(ik)} U_{ij}^2 \geq \lambda\Big\} \leq \exp\Big(-\frac{\lambda}{C_8\,\xi_i^2\,\log l}\Big), \qquad (4.19)$$

where $C_8 > 0$ is an absolute constant.

In order to bound T_3, we write it as

$$
\begin{aligned}
T_3 &= \sum_{i=i_0}^{i_1-1}\sum_{k=-\infty}^{\infty} E\Big\{I(\hat{B}_{ik} > \delta_i)\sum_{(ik)}(u_{ij}+U_{ij})^2\Big\}\\
&\leq 2\sum_{i=i_0}^{i_1-1}\sum_{k} E\Big\{I(\hat{B}_{ik} > \delta_i)\sum_{(ik)}U_{ij}^2\Big\} + 2\sum_{i=i_0}^{i_1-1}\sum_{k} E\Big\{I(\hat{B}_{ik} > \delta_i)\sum_{(ik)}u_{ij}^2\Big\}\\
&=: 2T_3' + 2T_3''.
\end{aligned}
\qquad (4.20)
$$

It follows from (4.6) that

$$T_3'' \leq \sum_{i=i_0}^{i_1-1}\sum_{k}\sum_{(ik)} u_{ij}^2 \leq \sum_{i=i_0}^{i_1-1}\sum_{j} u_{ij}^2 \leq C n^{-2s_2/(2s_2+1)}.$$

Thus, we only need to bound T_3'. For this purpose, let

$\quad A_i = \{\text{blocks at level } i \text{ contains at least one coefficient } \beta_{ij} \text{ with indices in } S_i\};$

$\quad A_i' = \{\text{blocks at level } i \text{ contains no coefficient } \beta_{ij} \text{ with indices in } S_i\}.$

As in Hall, et al. (1999, p.44), we may split T_3' into several parts:

$$
\begin{aligned}
T_3' &= \sum_{i=i_0}^{i_s}\sum_{k} E\Big\{I(\hat{B}_{ik} > \delta_i)\sum_{(ik)}U_{ij}^2\Big\}\\
&\quad + \sum_{i=i_s+1}^{i_1-1}\sum_{k\in A_i} E\Big\{I(\hat{B}_{ik} > \delta_i)I(B_{ik} > \delta_i/2)\sum_{(ik)}U_{ij}^2\Big\}\\
&\quad + \sum_{i=i_s+1}^{i_1-1}\sum_{k\in A_i'} E\Big\{I(\hat{B}_{ik} > \delta_i)I(B_{ik} > \delta_i/2)\sum_{(ik)}U_{ij}^2\Big\}\\
&\quad + \sum_{i=i_s+1}^{i_1-1}\sum_{k} E\Big\{I(\hat{B}_{ik} > \delta_i)I(B_{ik} \leq \delta_i/2)\sum_{(ik)}U_{ij}^2\Big\}\\
&=: T_{31} + T_{32} + T_{33} + T_{34}.
\end{aligned}
\qquad (4.21)
$$

Here $B_{ik} = l^{-1}\sum_{(ik)}(\beta_{ij}+u_{ij})^2$ and $l = (\log n)^\theta$ is the block length.

14

We will distinguish two cases: $\alpha \in (0,1)$ and $\alpha = 1$. In the first case, let i_s be an integer satisfying $2^{i_s} \simeq n^{\alpha/(2s_2+\alpha)}$. By (4.12), we see that the first term in (4.21) satisfies

$$
\begin{aligned}
T_{31} &\leq \sum_{i=i_0}^{i_s} \sum_k \sum_{(ik)} E(U_{ij}^2) \leq C \sum_{i=i_0}^{i_s} \sum_{j=0}^{2^i-1} n^{-\alpha} 2^{-i(1-\alpha)} \\
&= C \sum_{i=i_0}^{i_s} n^{-\alpha} 2^{i\alpha} = C\, n^{-\alpha} 2^{i_s \alpha} \\
&\leq C n^{-2s_2\alpha/(2s_2+\alpha)},
\end{aligned}
\tag{4.22}
$$

where the last inequality follows from our choice of i_s.

In order to estimate T_{32}, we recall that $\delta_i = 48\tau_i^2$ in (2.4), where $\tau_i^2 = C_4 n^{-\alpha} 2^{-i(1-\alpha)}$. From (4.2), (4.7) and (4.12), we may write, for all $t > 0$,

$$
\begin{aligned}
T_{32} &\leq \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} I(B_{ik} > \delta_i/2) \sum_{(ik)} E(U_{ij}^2) \\
&\leq C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} \delta_i^{-t} B_{ik}^t \sum_{(ik)} E(U_{ij}^2) \\
&\leq Cl \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} n^{-\alpha(1-t)} 2^{-i(1-\alpha)(1-t)} 2^{-i(2s_1+1)t} \\
&\leq Cl \sum_{i=i_s+1}^{i_1-1} 2^{i\gamma} n^{-\alpha(1-t)} 2^{-i(1-\alpha)(1-t)} 2^{-i(2s_1+1)t} \\
&\leq C n^{-2s_2\alpha/(2s_2+\alpha)},
\end{aligned}
\tag{4.23}
$$

the last inequality follows from our choice of t such that $0 < t < (1-\gamma)/[2(s_2 - s_1)]$.

As to T_{33}, from (4.12) and the definition of δ_i and B_{ik}, we have

$$
\begin{aligned}
T_{33} &\leq \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i'} I(B_{ik} > \delta_i/2) \sum_{(ik)} EU_{ij}^2 \leq C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i'} \delta_i^{-1} B_{ik}\, l\delta_i \\
&= C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i'} \left(\sum_{(ik)} (\beta_{ij} + u_{ij})^2 \right) \\
&\leq C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i'} \sum_{(ik)} \beta_{ij}^2 + C \sum_{i=i_s+1}^{i_1-1} \sum_j u_{ij}^2 \\
&\leq C \sum_{i=i_s+1}^{i_1-1} 2^i 2^{-i(2s_2+1)} + C \sum_{i=i_s+1}^{i_1-1} \sum_j u_{ij}^2 \\
&\leq C n^{-2s_2\alpha/(2s_2+\alpha)},
\end{aligned}
\tag{4.24}
$$

15

the last inequality follows from the choice of i_s, (4.6) and the fact that $0 < \alpha < 1$.

Therefore, in order to bound T_3, it remains to bound the last term T_{34}. Applying the same argument as Lemma 5.1 in Hall, et al. (1999, p.46), we have

$$\left\{ \hat{B}_{ik} > \delta_i \right\} \cap \left\{ B_{ik} \leq \delta_i/2 \right\} \subset \left\{ \sum_{(ik)} U_{ij}^2 \geq \frac{l}{12}\delta_i \right\} = \left\{ \sum_{(ik)} U_{ij}^2 \geq 4l\tau_i^2 \right\}. \tag{4.25}$$

Hence, it follows from (4.21) and (4.25) that

$$\begin{aligned} T_{34} &\leq \sum_{i=i_s+1}^{i_1-1} \sum_k E\left\{ \sum_{(ik)} U_{ij}^2 \, I\left(\sum_{(ik)} U_{ij}^2 \geq 4l\tau_i^2 \right) \right\} \\ &= \sum_{i=i_s+1}^{i_1-1} \sum_k \int_{4l\tau_i^2}^{\infty} P\left\{ \sum_{(ik)} U_{ij}^2 \geq \lambda \right\} d\lambda. \end{aligned} \tag{4.26}$$

By (i) of Lemma 4.2, we have

$$\begin{aligned} \int_{4l\tau_i^2}^{\infty} P\left\{ \sum_{(ik)} U_{ij}^2 \geq \lambda \right\} d\lambda &\leq \int_{4l\tau_i^2}^{\infty} \exp\left(-\frac{\lambda}{C_7 \, l^{1-\alpha}\tau_i^2} \right) d\lambda \\ &= C_7 \, l^{1-\alpha}\tau_i^2 \, \exp\left(-\frac{4\, l^\alpha}{C_7} \right). \end{aligned} \tag{4.27}$$

Combining (4.26) and (4.27), we get

$$\begin{aligned} T_{34} &\leq C\, l^{1-\alpha} \sum_{i=i_s+1}^{i_1-1} 2^i \, \tau_i^2 \, \exp\left(-\frac{4\, l^\alpha}{C_7} \right) \\ &\leq C\, (\log n)^{\theta(1-\alpha)} \exp\left(-\frac{4\, (\log n)^{\alpha\theta}}{C_7} \right). \end{aligned} \tag{4.28}$$

Since $\alpha\theta > 1$, we see that for all constant $\eta > 0$,

$$T_{34} \leq n^{-\eta} \tag{4.29}$$

for all n large enough. Combining the inequalities (4.22), (4.23), (4.24) and (4.29) together, we have proved that, for $\alpha \in (0, 1)$, $T_3 \leq C\, n^{-2s_2\alpha/(2s_2+\alpha)}$.

Now we consider the case of $\alpha = 1$ and show that

$$T_3 \leq C \left(\frac{\log n}{n} \right)^{2s_2/(2s_2+1)}. \tag{4.30}$$

The proof of (4.30) follows the same line as that of T_3 for $\alpha \in (0, 1)$. Hence we only present the modifications needed.

When $\alpha = 1$, let i_s be an integer satisfying $2^{i_s} \simeq (n^{-1} \log n)^{-1/(2s_2+1)}$, which is different from the previous one. From (4.16) and (4.22), we have

$$
\begin{aligned}
T_{31} &\le C \sum_{i=i_0}^{i_s} \sum_{j=0}^{2^i-1} 2C_0 n^{-1} \log(n2^{-i}e) \\
&\le Cn^{-1} \log n \sum_{i=i_0}^{i_s} 2^i \\
&= Cn^{-1} \log n \, 2^{i_s} \\
&\le C \Big(\frac{\log n}{n} \Big)^{2s_2/(2s_2+1)},
\end{aligned}
\tag{4.31}
$$

where the last inequality follows from our choice of i_s.

In view of (4.23), when $\alpha = 1$, we have for all $t > 1/(2s_2 + 1)$,

$$
\begin{aligned}
T_{32} &\le C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} \delta_i^{-t} B_{ik}^t \sum_{(ik)} E(U_{ij}^2) \\
&\le Cl (\log n2^{-i_s})^{1-t} n^{-(1-t)} \sum_{i=i_s+1}^{i_1-1} 2^{-i((2s_1+1)t-\gamma)} \\
&\le C \Big(\frac{\log n}{n} \Big)^{2s_2/(2s_2+1)},
\end{aligned}
\tag{4.32}
$$

the last inequality follows from our choice of t such that $1/(2s_2+1) < t < (1-\gamma)/[2(s_2-s_1)]$.

In view of (4.24), when $\alpha = 1$, we have

$$
\begin{aligned}
T_{33} &\le C \sum_{i=i_s+1}^{i_1-1} 2^i 2^{-i(2s_2+1)} + C \sum_{i=i_s+1}^{i_1-1} \sum_j u_{ij}^2 \\
&\le C \sum_{i=i_s+1}^{i_1-1} 2^{-2is_2} + Cn^{-2s_2/(2s_2+1)} \\
&\le C \Big(\frac{\log n}{n} \Big)^{2s_2/(2s_2+1)}.
\end{aligned}
\tag{4.33}
$$

In order to estimate T_{34}, note that the inequality (4.26) with τ_i^2 replaced by ξ_i^2 holds. Applying Part (ii) of Lemma 4.2, we have

$$
\begin{aligned}
\int_{4l\xi_i^2}^{\infty} P\Big\{ \sum_{(ik)} U_{ij}^2 \ge \lambda \Big\} d\lambda &\le \int_{4l\xi_i^2}^{\infty} \exp\Big(-\frac{\lambda}{C_8 \, \xi_i^2 \log l} \Big) d\lambda \\
&= C \, \xi_i^2 \, (\log l) \exp\Big(-\frac{4\, l}{C_8 \log l} \Big).
\end{aligned}
\tag{4.34}
$$

Therefore, we bound term T_{34} with

$$
T_{34} \le C \log l \sum_{i=i_s+1}^{i_1-1} 2^i \, \tau_i^2 \exp\Big(-\frac{4l}{C_8 \log l} \Big).
\tag{4.35}
$$

17

Since we have chosen the block-length $l = (\log n)^\theta$ with $\theta > 1$, we have for any constant $\eta > 0$, we obtain

$$T_{34} \le n^{-\eta}. \tag{4.36}$$

Then (4.30) follows from (4.31), (4.32), (4.33) and (4.36). Hence we have bounded the term T_3 as desired.

Bound for T_4: As to the last term in (4.1), we may write

$$
\begin{aligned}
T_4 \le & \sum_{i=i_0}^{i_1-1} \sum_{k \in A_i} P\left(\hat{B}_{ik} \le \delta_i \quad \text{and} \quad B_{ik} \ge 2\delta_i\right) \sum_{(ik)} \beta_{ij}^2 \\
& + \sum_{i=i_0}^{i_s} \sum_{k \in A_i'} P\left(\hat{B}_{ik} \le \delta_i \quad \text{and} \quad B_{ik} \ge 2\delta_i\right) \sum_{(ik)} \beta_{ij}^2 \\
& + \sum_{i=i_0}^{i_s} \sum_{k} P\left(\hat{B}_{ik} \le \delta_i \quad \text{and} \quad B_{ik} < 2\delta_i\right) \sum_{(ik)} \beta_{ij}^2 \\
& + \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} P\left(\hat{B}_{ik} \le \delta_i \quad \text{and} \quad B_{ik} < 2\delta_i\right) \sum_{(ik)} \beta_{ij}^2 \\
& + \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i'} P\left(\hat{B}_{ik} \le \delta_i\right) \sum_{(ik)} \beta_{ij}^2 \\
=: & \ T_{41} + T_{42} + T_{43} + T_{44} + T_{45}.
\end{aligned}
$$

Again, we first consider $\alpha \in (0,1)$ case. The proofs of T_{41} and T_{42} are similar, which involve the large deviation result. Here we only provide the proof of T_{41}. Since

$$\left\{\hat{B}_{ik} \le \delta_i\right\} \cap \left\{B_{ik} \ge 2\delta_i\right\} \subset \left\{\sum_{(ik)} U_{ij}^2 \ge \frac{l}{6}\delta_i\right\} \subset \left\{\sum_{(ik)} U_{ij}^2 \ge 4l\tau_i^2\right\},$$

thus, from Lemma 4.2, we have for any $t > 0$

$$T_{41} \le C \sum_{i=i_0}^{i_1-1} n^{-t} 2^{i\gamma} 2^{-i(2s_1+1)} \le C n^{-2s_2\alpha/(2s_2+\alpha)}.$$

As to T_{43}, since $\frac{1}{2} \sum_{(ik)} \beta_{ij}^2 \le \sum_{(ik)}(\beta_{ij}+u_{ij})^2 + \sum_{(ik)} u_{ij}^2$, and recall B_{ik}, we have

$$T_{43} \le C \sum_{i=i_0}^{i_s} 2^i \delta_i + C \sum_{i=i_0}^{i_s} \sum_j u_{ij}^2 \le C n^{-2s_2\alpha/(2s_2+\alpha)},$$

the last inequality follows from our choice i_s and (4.6).

18

Similar to T_{43}, for any $t > 0$, we have

$$T_{44} \leq 2 \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} I\left(B_{ik} < 2\delta_i\right) \sum_{(ik)} (\beta_{ij} + u_{ij})^2 + 2 \sum_{i=i_s+1}^{i_1-1} \sum_j u_{ij}^2$$

$$\leq C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} \delta_i^t (B_{ik})^{-t} \sum_{(ik)} (\beta_{ij} + u_{ij})^2 + Cn^{-2s_2/(2s_2+1)}$$

$$\leq C \sum_{i=i_s+1}^{i_1-1} \sum_{k \in A_i} n^{-\alpha t} 2^{-i(1-\alpha)t} l^t \left[\sum_{(ik)} (\beta_{ij} + u_{ij})^2\right]^{1-t} + Cn^{-2s_2/(2s_2+1)}$$

$$\leq Cln^{-\alpha t} \sum_{i=i_s+1}^{i_1-1} 2^{i\gamma} 2^{-i(1-\alpha)t} 2^{-i(2s_1+1)(1-t)} + Cn^{-2s_2/(2s_2+1)}$$

$$\leq Cn^{-2s_2\alpha/(2s_2+\alpha)},$$

The last inequality follows from the choice of t such that $1 > t > 1 - (1-\gamma)/[2(s_2 - s_1)]$.

For the last term, we have $T_{45} \leq C \sum_{i=i_s+1}^{i_1-1} 2^i 2^{-i(2s_2+1)} \leq Cn^{-2s_2/(2s_2+1)}$. Together with the other four terms, it bounds the T_4, hence we have completed the proof of the theorem for $\alpha \in (0,1)$.

When $\alpha = 1$, the proof is similar to and simpler than that of $\alpha \in (0,1)$. Simply replace δ_i with $Cn^{-1}\log(n2^{-i}e)$, we have

$$T_{41} \leq C \sum_{i=i_0}^{i_1-1} n^{-t} 2^{i\gamma} 2^{-i(2s_1+1)} = Cn^{-t}(2^{i_0})^{-(2s_1+1-\gamma)} \leq C\left(\frac{\log n}{n}\right)^{2s_2/(2s_2+1)}.$$

$$T_{43} \leq C \sum_{i=i_0}^{i_s} 2^i \delta_i + C \sum_{i=i_0}^{i_s} \sum_j u_{ij}^2$$

$$\leq C \sum_{i=i_0}^{i_s} 2^i n^{-1} \log(n2^i e) + Cn^{-2s_2/(2s_2+1)} \quad \leq C\left(\frac{\log n}{n}\right)^{2s_2/(2s_2+1)}.$$

$$T_{44} \leq C \sum_{i=i_s+1}^{i_1-1} 2^{i\gamma} n^{-t} \left(\log(n2^{-i}e)\right)^t l^t \left[\sum_{(ik)} (\beta_{ij} + u_{ij})^2\right]^{1-t} + Cn^{-2s_2/(2s_2+1)}$$

$$\leq Cln^{-t} \left(\log(n2^{-i_s}e)\right)^t \sum_{i=i_s+1}^{i_1-1} 2^{-i[(2s_1+1)(1-t)-\gamma]} + Cn^{-2s_2/(2s_2+1)}$$

$$\leq C\left(\frac{\log n}{n}\right)^{2s_2/(2s_2+1)},$$

where the last inequality follows from the choice of t such that $1 - (1-\gamma)/[2(s_2 - s_1)] < t < 1 - \gamma/(2s_1 + 1)$. The term $T_{45} \leq C\left(n^{-1}\log n\right)^{2s_2/(2s_2+1)}$ is obvious. Combining all the above terms, we have proved the theorem for the $\alpha = 1$ case.

Finally, we prove Lemma 4.2.

Proof of Lemma 4.2 Let $\mathcal{A} = \{a = (a_1, \ldots, a_l) \in \mathbb{R}^l : \sum_{j=1}^l a_j^2 = 1\}$ be the unit sphere in $\mathbb{R}^l$. Note that for all integers i and k, we have

$$\Big(\sum_{(ik)} U_{ij}^2 \Big)^{1/2} = \sup_{a \in \mathcal{A}} \sum_{j=1}^l a_j U_{ij}. \tag{4.37}$$

Hence, in order to prove (4.18) for $0 < \alpha < 1$, it is sufficient to show that for all $u > 2\sqrt{l}\,\tau_i$,

$$P\Big\{ \sup_{a \in \mathcal{A}} \sum_{j=1}^l a_j U_{ij} \geq u \Big\} \leq \exp\Big(-\frac{u^2}{C_7\, l^{1-\alpha}\tau_i^2} \Big) \tag{4.38}$$

for some constant $C_7 > 0$. To this end, we appeal to Borell's inequality [see, e.g., Adler (1990)] about tail probability of the supremum of a general Gaussian process.

Consider the centered Gaussian process $\{Z(a), a \in \mathcal{A}\}$ defined by

$$Z(a) = \sum_{j=1}^l a_j U_{ij}. \tag{4.39}$$

Firstly, by the Cauchy-Schwarz inequality, Jensen's inequality and (4.12), we have

$$\begin{aligned}
E\Big(\sup_{a \in \mathcal{A}} Z(a) \Big) &\leq E\Big\{ \Big(\sum_{j=1}^l U_{ij}^2 \Big)^{1/2} \Big\} \\
&\leq \Big\{ \sum_{j=1}^l E(U_{ij}^2) \Big\}^{1/2} \\
&\leq \sqrt{l}\,\tau_i.
\end{aligned} \tag{4.40}$$

Secondly, for every $a \in \mathcal{A}$, by (4.39) and (4.5) we have

$$\begin{aligned}
E\big(Z(a)^2\big) &= E\Big(\frac{1}{\sqrt{n}} \sum_{m=1}^n \varepsilon_m \langle \phi_{i_1 m}, \sum_{j=1}^l a_j \psi_{ij} \rangle \Big)^2 \\
&= \frac{\sigma^2}{n} \sum_{m=1}^n \langle \phi_{i_1,m}, \sum_{j=1}^l a_j \psi_{ij} \rangle^2 \\
&\quad + \frac{1}{n} \sum_{m=1}^n \sum_{m' \neq m} r(m - m') \langle \phi_{i_1,m}, \sum_{j=1}^l a_j \psi_{ij} \rangle \langle \phi_{i_1,m'}, \sum_{j=1}^l a_j \psi_{ij} \rangle \\
&=: J_1 + J_2.
\end{aligned} \tag{4.41}$$

By the orthogonality of $\phi_{i_1,m}$ and ψ_{ij}, we have

$$\begin{aligned}
J_1 &\leq \frac{\sigma^2}{n} \Big\| \sum_{j=1}^l a_j\,\psi_{ij} \Big\|_2^2 \\
&= \frac{\sigma^2}{n} \sum_{j=1}^l a_j^2 = \frac{\sigma^2}{n}.
\end{aligned} \tag{4.42}$$

20

In the above, $\| \cdot \|_2$ denotes the usual norm in $L^2(\mathbb{R})$.

In order to bound J_2, we note that

$$J_2 = \frac{1}{n} \sum_{j=1}^{l} \sum_{j'=1}^{l} a_j \, a_{j'} \left\{ \sum_{m=1}^{n} \sum_{m' \neq m} r(m - m') \, \langle \phi_{i_1, m}, \, \psi_{ij} \rangle \, \langle \phi_{i_1, m'}, \, \psi_{ij'} \rangle \right\}$$

$$\leq \int \int \phi(u) \phi(v) \sum_{j=1}^{l} \sum_{j'=1}^{l} a_j \, a_{j'} \tag{4.43}$$

$$\cdot \left\{ \frac{2^i}{n^2} \sum_{m=1}^{n} \sum_{m' \neq m} r(|m - m'|) \, \psi \big(2^i \, \frac{u + m}{n} - j \big) \, \psi \big(2^i \, \frac{v + m'}{n} - j' \big) \right\} du dv.$$

Using the same argument in (4.9)–(4.11), we derive that for all integers $j, j' \geq 1$ and all $u, v \in [0, 1]$,

$$\frac{2^i}{n^2} \sum_{m=1}^{n} \sum_{m' \neq m} r(|m - m'|) \, \psi \big(2^i \, \frac{u + m}{n} - j \big) \, \psi \big(2^i \, \frac{v + m'}{n} - j' \big)$$

$$\leq C \, n^{-\alpha} \, 2^{-i(1-\alpha)} \int \int \frac{\psi(x)\psi(y)}{|x - y + j - j'|^\alpha} \, dx dy \tag{4.44}$$

$$\leq \frac{C \, \tau_i^2}{(1 + |j - j'|)^\alpha}.$$

In deriving the first inequality, we have used the fact that ψ is supported on $[0, 1]$.

In the case of $\alpha \in (0, 1)$, we have for all $a \in \mathcal{A}$,

$$\sum_{j=1}^{l} \sum_{j'=1}^{l} a_j \, a_{j'} \frac{1}{(1 + |j - j'|)^\alpha} = \sum_{j=1}^{l} a_j^2 + 2 \sum_{j=1}^{l} \sum_{j'=1}^{j-1} \frac{a_j \, a_{j'}}{(1 + |j - j'|)^\alpha}$$

$$\leq 1 + \sum_{j=1}^{l} \sum_{k=1}^{j-1} \frac{a_j^2 + a_{j-k}^2}{(1 + k)^\alpha}$$

$$= 1 + \sum_{j=1}^{l} a_j^2 \sum_{k=1}^{j-1} \frac{1}{(1 + k)^\alpha} + \sum_{j=1}^{l} \sum_{k=1}^{j-1} \frac{a_{j-k}^2}{(1 + k)^\alpha} \tag{4.45}$$

$$\leq 1 + \sum_{k=1}^{l-1} \frac{1}{(1 + k)^\alpha} + \sum_{k=1}^{l-1} \frac{1}{(1 + k)^\alpha} \sum_{j=k+1}^{l} a_{j-k}^2$$

$$\leq C \, l^{1-\alpha}.$$

Combining (4.43), (4.44) and (4.45), we obtain

$$J_2 \leq C_9 \, l^{1-\alpha} \, \tau_i^2. \tag{4.46}$$

Since $0 < \alpha < 1$, (4.42) and (4.46) together yield

$$D^2 := \sup_{a \in \mathcal{A}} E\big(Z(a)^2\big) \leq C_{10} \, l^{1-\alpha} \, \tau_i^2. \tag{4.47}$$

Denote $\tilde{m} := E(\sup_{a \in \mathcal{A}} Z(a))$. It follows from Borell's inequality [see, Adler (1990)], (4.40) and (4.47) that for all $u \geq 2\tilde{m}$ [by (4.40), this holds whenever $u \geq 2\sqrt{l}\,\tau_i$],

$$
\begin{aligned}
P\left\{ \sup_{a \in \mathcal{A}} \sum_{j=1}^{l} a_j U_{ij} \geq u \right\} &\leq \exp\left(-\frac{(u-\tilde{m})^2}{2D^2} \right) \\
&\leq \exp\left(-\frac{u^2}{8\,C_{10}\,l^{1-\alpha}\,\tau_i^2} \right).
\end{aligned}
\tag{4.48}
$$

This proves (4.38) with $C_7 = 8\,C_{10}$.

Now we consider the case $\alpha = 1$. The proof of (4.19) is similar to the above, but we need to modify the estimation of J_2. When $j = j'$, similar to (4.14) and (4.15), we use the fact that ψ is bounded and has its support in $[0,1]$ to derive that for all $u, v \in [0,1]$,

$$
\begin{aligned}
&\frac{2^i}{n^2} \sum_{m=1}^{n} \sum_{m' \neq m} r(|m-m'|)\,\psi\big(2^i\,\frac{u+m}{n} - j\big)\,\psi\big(2^i\,\frac{v+m'}{n} - j\big) \\
&\leq C\,n^{-1}\,\log(n\,2^{-i}\,e).
\end{aligned}
\tag{4.49}
$$

When $j \neq j'$, we apply the same method as above and the elementary inequalities $\sum_{k=a}^{b} k^{-1} \leq C\log(b/a)$ and $\log(1+x) \leq x$ $(\forall\, x > 0)$ to derive the following better bound:

$$
\begin{aligned}
&\frac{2^i}{n^2} \sum_{m=1}^{n} \sum_{m' \neq m} r(|m-m'|)\,\psi\big(2^i\,\frac{u+m}{n} - j\big)\,\psi\big(2^i\,\frac{v+m'}{n} - j'\big) \\
&\leq C\,n^{-1}\,\frac{1}{|j-j'|}.
\end{aligned}
\tag{4.50}
$$

It follows from (4.43), (4.49), (4.50) and the same argument as in deriving (4.45) that

$$
\begin{aligned}
J_2 &\leq C\left\{ \sum_{j=1}^{l} a_j^2\,n^{-1}\,\log(n\,2^{-i}\,e) + \sum_{j=1}^{l} \sum_{j' \neq j} \frac{a_j a_{j'}}{|j-j'|}\,n^{-1} \right\} \\
&\leq C\left(n^{-1}\,\log(n\,2^{-i}\,e) + n^{-1}\,\log l \right) \\
&\leq C\,\xi_i^2\,\log l.
\end{aligned}
\tag{4.51}
$$

Hence, we have $D^2 \leq C\,\xi_i^2\,\log l$ and (4.19) follows from Borell's inequality. The proof of Lemma 4.2 is finished.

Acknowledgements

REFERENCES

Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE. Trans. on Inform. Theory*, 44, 2-15.

Adler, R. J. (1990), *An introduction to continuity, extrema, and related topics for general Gaussian processes.* Institute of Mathematical Statistics, Hayward, CA.

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

Bingham, N., Goldie, C. and Teugels, J. (1987). *Regular Variation*, Cambridge University Press, Cambridge.

Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size, and threshold level. *Statistica Sinica* **12**, 1241-1273.

Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harm. Anal.* **1**, 54–82.

Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.* **23**, 1000-1014.

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

Delbeke, L and Van Assche, Walter (1998) A wavelet based estimator for the parameter of self-similarity of fractional Brownian motion. In: *3rd International Conference on Approximation and Optimization in the Caribbean(Puebla, 1995)*, pp.65–76, Aportaciones Mat. Comun., **24**, Soc. Mat. Mexicana, México.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinking. *J. Am. Statist. Assoc.* **90**, 1200–1224.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26**, 879–921.

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

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539.

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

Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet method. *Ann. Statist.* **26**, 922-942.

Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statistica Sinica* **9**, 33-50.

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

Hall, P. and Patil, P. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet wavelet methods. *J. Roy. Statist. Soc. Ser. B.* **58**, 361–377.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation and Statistical Applications.* Lecture Notes in Statistics **129**, Springer, New York.

Johnstone, I.M. (1999). Wavelet threshold estimators for correlated data and inverse problems: Adaptivity results. *Statistica Sinica* **9**, 51–83.

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.

Robinson, P.M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25**, 2054–2083.

Robinson, P.M. and Hidalgo, F.J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25**, 77–104.

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466-484.