Journal of Sciences, Islamic Republic of Iran 17(2): 153-159 (2006) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

On the Minimax Optimality of Block Thresholded Wavelets Estimators for *ρ*-Mixing Process

H. Doosti* and H.A. Niroumand

Department of Statistics, School of Mathematical Sciences, Ferdowsi University, P.O. Box 1159-91775, Mashhad, Islamic Republic of Iran

Abstract

We propose a wavelet based regression function estimator for the estimation of the regression function for a sequence of *ρ*-missing random variables with a common one-dimensional probability density function. Some asymptotic properties of the proposed estimator based on block thresholding are investigated. It is found that the estimators achieve optimal minimax convergence rates over large classes of functions that involve many irregularities of a wide variety of types, including chirp and Doppler functions and jump discontinuities.

Keywords: Block thresholded; Non-linear wavelet-based estimator; Rates of convergence; Minimax estimation splines recieved

1. Introduction

Let $(\Omega, \mathbf{F}, P)$ be a probability space. The random variables we deal with are all defined on $(\Omega, \mathbf{F}, P)$. Let $\mathbf{N}_k^m$ denote the σ-algebra generated by the events

$$\{X_k \in A_k, ..., X_m \in A_m\}.$$

A sequence of random variables $\{X_n, n \geq 1\}$ is said to be ρ-mixing if

$$\sup_m \sup_{X \in L_2(\mathbf{N}_1^m), Y \in L_2(\mathbf{N}_{m+s}^\infty)} |corr(X, Y)| = \rho(s) \to 0, \quad as \quad s \to \infty.$$

The problem of interest is the estimation of nonparametric regression function

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

where $x_m = \frac{m}{n} \in [0,1]$ and the variables ε_m are ρ-dependent random variables with a common one-dimensional normal density function with zero mean and variance σ^2 and g belongs to a large function class $\mathbf{H}$ (definition will be given in the next section).

Hall *et al.* [8] considered model (1.1) when $\varepsilon_1, ..., \varepsilon_n$ are independent, identically distributed (i.i.d.) normal random variables with mean 0 and variance σ^2. They introduced a local block thresholding estimator which thresholds empirical wavelet coefficients in groups rather than individually and showed that the estimators achieve optimal minimax convergence rates over a large class of functions $\mathbf{H}$ 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 [1] considered the asymptotic and numerical properties of a class of block thresholding estimators for model (1.1)

*E-mail: doosti@math.um.ac.ir

with i.i.d. Gaussian errors. He investigated the block size and the thresholding constant such that the corresponding block thresholding estimators obtain optimal convergence rates for both global and local estimation over a large classes of functions as in [8]. Doosti and Niroumand [7] considered a stochastic regression model with pairwise negative quadrant dependent noise.

Wavelet methods in nonparametric curve estimation have become a well-known technique. For a systematic discussion of wavelets and their applications in statistics, see the recent monograph by Hardle *et al.* [11]. The major advantage of wavelet method is ability to adapt to the degree of smoothness of the underlying unknown curve. These wavelet estimators typically achieve the optimal convergence rates over exceptionally large function spaces. For reference, see [4-6]. Hall and Patil [9,10] also have 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 performance of nonlinear wavelet curve estimators.

This paper first establishes some necessary basic mathematical background and terminology relating to wavelets in Section 2. The main results are described in Section 3.

2. Preliminaries

2.1. Wavelet Estimators

For any function $f \in \mathbf{L}^2(\mathbf{R})$, we can write a formal expansion (see [3]):

$$f = \sum_{k\in Z}\alpha_{i_0,k}\phi_{i_0,k} + \sum_{j\ge i_0}\sum_{k\in Z}\beta_{j,k}\psi_{j,k},$$

where the functions

$$\phi_{i_0,k}(x) = 2^{i_0/2}\phi(2^{i_0}x - k)$$

and

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^{j}x - k),$$

constitute an (inhomogeneous) orthonormal basis of $\mathbf{L}^2(\mathbf{R})$. Here $\phi(x)$ and $\psi(x)$ are the scale function and the orthogonal wavelet, respectively. $\phi(x)$ and $\psi(x)$ are bounded and compactly supported and $\int\phi = 1$. Wavelet coefficients are given by the integrals

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

The orthogonality properties of ϕ and ψ imply:

$$\int\phi_{i_0,j_1}\phi_{i_0,j_2} = \delta_{j_1j_2},$$
$$\int\psi_{i_1,j_1}\psi_{i_2,j_2} = \delta_{i_1i_2}\delta_{j_1j_2}, \tag{2.1}$$
$$\int\phi_{i_0,j_1}\psi_{i,j_2} = 0, \quad \forall i_0 \le i,$$

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 a fixed unit interval [0,1]. Therefore, we confine our attention to the wavelet basis of [0,1] intervals given by [2], that is, the collection of $\{\phi_{i_0,j}$, $j = 0,1,\dots,2^{i_0-1}$; $\psi_{i,j}$, $i \ge i_0 \ge 0$, $j = 0,1,\dots, 2^j - 1\}$ forms an orthonormal basis of $\mathbf{L}^2[0,1]$. Since, in this paper, we require vanishing moments up to N-1 for both ϕ and ψ ($\int x^k\phi(x)dx = 0$, $k = 1,2,\dots, N-1$; $\int x^k\psi(x)dx = 0$, $k = 1,2,\dots,N-1$), the so-called Coiflets will be used here. Hence, the corresponding wavelet expansion of $g(x)$, is

$$g(x) = \sum_{j\in Z}\alpha_{i_0,j}\phi_{i_0,j} + \sum_{i\ge i_0}\sum_{j\in Z}\beta_{i,j}\psi_{i,j}(x), \tag{2.2}$$

where

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

An empirical wavelet expansion based on term-by-term thresholding is given by

$$\bar{g}(x) = \sum_{j\in Z}\bar{\alpha}_{i_0,j}\phi_{i_0,j} + \sum_{i\ge i_0}\sum_{j\in Z}\bar{\beta}_{i,j}\psi_{i,j}(x)I(\bar{\beta}_{ij}^2 > cn^{-1}\log n), \tag{2.3}$$

where $\bar{\alpha}_{i,j} = n^{-1}\sum_m Y_m\phi_{i,j}(x_m)$, $\bar{\beta}_{i,j} = n^{-1}\sum_m Y_m\psi_{i,j}(x_m)$, c is an appropriate threshold constant, and $i_1 > i_0$ is a truncating point. Note that here; a thresholding decision is made about each term in ψ_{ij}.

In block thresholding, the integers j are divided among consecutive, nonoverlapping blocks of length l_i, say $\mathbf{B}_{ik} = \{j : (k-1)l_i + v + 1 \le j \le kl_i + v\}$, $-\infty < k < \infty$, where v is an arbitrary integer. (It simplifies notation a little if we take $v = 0$ which we shall do.) In this approach, all terms involving the functions ψ_{ij} for

$j \in \mathbf{B}_{ik}$ are included in or excluded from the empirical wavelet transform. This leads to the estimator,

$$\tilde{g}(x) = \sum_{j \in Z} \overline{\alpha}_{i_0,j}\phi_{i_0,j} + \sum_{i=i_0}^{i_1-1}\sum_{k \in Z}(\sum_{(ik)}\overline{\beta}_{i,j}\psi_{i,j}(x))I(\hat{B}_{ij} > cn^{-1}), \tag{2.4}$$

where $\sum_{(ik)}$ denotes summation over $j \in \mathbf{B}_{ik}$, and $\hat{B}_{ik}$ is an estimator of the "average" value of β_{ij}^2 for $j \in \mathbf{B}_{ik}$. Let V_i and W_i be the spaces spanned by $\{\phi_{ij},\ j \in Z\}$ and $\{\psi_{ij},\ j \in Z\}$, respectively, and let $\mathrm{Pr}oj_{V_i}(.)$ and $\mathrm{Pr}oj_{W_i}(.)$ be the projection operators on these spaces. If $i < i_1$ and $f \in V_{i_1}$ then the coefficients of $\mathrm{Pr}oj_{V_i}(f)$ and $\mathrm{Pr}oj_{W_i}(f)$ may be computed from the values of $\int f\phi_{i_1,j}$, $j \in Z$, using "subband filtering schemes" discussed by [3], chapter 5. Define

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

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

$$\mathrm{Pr}oj_{W_i}(\hat{G}_{i_1}) = \sum_{j \in Z}\hat{\beta}_{i,j}\psi_{i,j}$$

and

$$\mathrm{Pr}oj_{V_{i_0}}(\hat{G}_{i_1}) = \sum_{j \in Z}\hat{\alpha}_{i_0,j}\phi_{i_0,j},$$

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

$$\hat{g} = \sum_{j \in Z}\hat{\alpha}_{i_0,j}\phi_{i_0,j} + \sum_{i=i_0}^{i_1-1}\sum_{k \in Z}(\sum_{(ik)}\hat{\beta}_{i,j}\psi_{i,j}(x))I(\hat{B}_{ij} > cn^{-1}). \tag{2.5}$$

Choice of i_0, i_1, l_i and c will be discussed in next section.

2.2. The Class of Functions, $\mathbf{H}$

Given $0 < s_1 < s_2 < N$ and $\gamma, C_1, C_2, C_3, v \geq 0$, we shall define a class of functions $\mathbf{H} = \mathbf{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$.

Definition 2.1. *For given constants* $0 < s_1 < s_2 < N$, *let* $\mathbf{H} = \mathbf{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$ *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 is true:* $card(S_i) \leq C_3 2^{i\gamma}$ *and*

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

$$|g(x) - \sum_{l=0}^{N-1} a_l (x - 2^{-i}j)^l| \leq C_1 2^{-is_1} \textit{ for all}$$

$x \in [j/2^i, (j+v)/2^i]$;

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

$$|g(x) - \sum_{l=0}^{N-1} a_l (x - 2^{-i}j)^l| \leq C_2 2^{-is_2} \textit{ for all}$$

$x \in [j/2^i, (j+v)/2^i]$.

The function class $\mathbf{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$ contains the Besov class $B_{\infty\infty}^{s_2}(C_2)$ as a subset for all $s_1 < s_2, \gamma > 0$ and with $C_1 > 0$ depending on the choice of the other constants. Furthermore, as pointed out in [8], a function $g \in \mathbf{H}$ can be regarded as the superposition of a smooth function g_2 from the Besov space $B_{\infty\infty}^{s_2}$ 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 [8] and [1] for more discussions about the function classes $\mathbf{H}$.

Since our wavelets' support is contained in the interval [0,1], we confine attention to the function space $\mathbf{H}$ with $v = 1$.

The following lemma which characterizes some properties of the wavelet coefficients of $g \in \mathbf{H}$, is due to [8], Proposition 3.1.

Lemma 2.1. *For every function* $g \in \mathbf{H}(s_1, s_2, \gamma, C_1, C_2, C_3, N, v)$ *and our selected Coiflets, the wavelet coefficients of* g*, denoted with* $\alpha_{i,j}$ *and* $\beta_{i,j}$ *have following properties:*

$$|\beta_{i,j}| \leq |\psi|_1 C_1 2^{-i(s_1+1/2)} \textit{ ifj } \in S_i,$$

$$|\beta_{i,j}| \leq |\psi|_1 C_2 2^{-i(s_2+1/2)} \textit{ ifj } \notin S_i,$$

$$|\alpha_{i,j} - 2^{-i/2} g(j/2^i)| \leq |\phi|_1 C_1 2^{-i(s_1+1/2)} \textit{ ifj } \in S_i,$$

$$|\alpha_{i,j} - 2^{-i/2} g(j/2^i)| \leq |\phi|_1 C_2 2^{-i(s_2+1/2)} \textit{ ifj } \notin S_i.$$

3. Main Results

Our main theorem provides an upper bound to convergence rates uniformly over functions in $\mathbf{H}$. Since the bound is of the same size as the minimax lower bound, then it is optimal.

Let ϕ be a Coiflet, and ψ the associated wavelet, with Daubechies number N and support contained in the interval $[0,1]$. Define the indices i_0 and i_1 in terms of N by $2^{i_0-1} \le n^{1/(2N+1)} \le 2^{i_0}$ and $2^{i_1-1} \le n \le 2^{i_1}$. Assume that the errors ε_m in the model at (1.1) form the ρ-mixing sequence of random variables which

$$\sum_k \rho(k) < \infty$$

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

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

(Recall that c is the threshold constant in the formula for $\hat{g}$.) We call these conditions (C). Hall *et al.* [8] considered model (1.1) and provided the following theorem when $\varepsilon_1,...,\varepsilon_n$ were independent, identically distributed (i.i.d.) normal random variables with mean 0 and variance σ^2. Here we extend their results when variables ε_m form a ρ-dependent processes.

Theorem 3.1. *If conditions (C) hold, and if the estimator* $\hat{g}$ *is as defined at (2.5), then for each* $C_1, C_2 > 0$ *there exist a constant* $K = K(s_1, s_2, \gamma, C_1, C_2, V, N, v) > 0$ *such that*

$$\sup_{g \in \mathbf{H}(s_1,s_2,\gamma,C_1,C_2,V,N,v)} \int \mathbf{E}(\hat{g} - g)^2 \le n^{\frac{-2s_2}{2s_2+1}}(K + o(1)).$$

Proof. The proof of this Theorem is similar to that of Theorem 4.1 of [8]. The difference is that we consider the errors $\{\varepsilon_m,\ m \ge 1\}$ to be a ρ-mixing process, instead of i.i.d. random variables in their paper. Hence, several technical difficulties have to be overcome.

We will break the proof of Theorem 3.1 into several parts.

Part (a). Properties of the projection operator. As in [8] page 42, there exists small number $r_{i_1 m}$, such that

$$\alpha_{i_1 m} = \int g(x)\phi_{i_1 m}(x)dx$$

$$= n^{1/2}\int g\left(\frac{m+y}{n}\right)\phi(y)dy \qquad (3.1)$$

$$=: n^{1/2} g\left(\frac{m}{n}\right) - r_{i_1 m}.$$

Thus, we have

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

In similar way, we may write for every integer $0 \le i < i_1$,

$$\mathrm{Proj}_{W_i}(\hat{G}_{i_1}) = \sum_{j \in Z}(\beta_{ij} + u_{ij} + U_{ij})\psi_{ij}(x),$$

$$\mathrm{Proj}_{V_{i_0}}(\hat{G}_{i_1}) = \sum_{j \in Z}(\alpha_{i_0 j} + v_{i_0 j} + V_{i_0 j})\phi_{i_0 j}(x),$$

where u_{ij} and $v_{i_0 j}$ are real numbers,

$$U_{ij} = \frac{1}{\sqrt{n}}\sum_{m=1}^{n}\varepsilon_m <\phi_{i_1 m}, \psi_{ij}>,$$
$$V_{i_0 j} = \frac{1}{\sqrt{n}}\sum_{m=1}^{n}\varepsilon_m <\phi_{i_1 m}, \phi_{i_0 j}>. \qquad (3.2)$$

In the above, $<f, g> = \int fg$ is the inner product in $L^2([0,1])$. In this notation, we may write

$$u_{ij} = \sum_{m=1}^{n} r_{i_1 m} <\phi_{i_1 m}, \psi_{ij}>.$$

By Parseval's identity,

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

Hall *et al.* [8, p.43] showed that

$$\sum_m r_{i_1 m}^2 \le C n^{\frac{-2s_2}{2s_2+1}}, \qquad (3.3)$$

and

$$|u_{ij}| \le C\, 2^{-i(s_1+1/2)}. \quad (3.4)$$

Because of the compact support of our wavelets, there are at most 2^{i_1-i} none zero terms of $<\phi_{i_1 l}, \psi_{ij}>$, $l = 1,2,...,n$, and also $<\phi_{i_1 l}, \psi_{ij}>$, $l = 1,2,...,n$.

At last, let's calculate the variance of U_{ij} and $V_{i_0 j}$.

$$\mathbf{E}U_{ij}^2 = \frac{1}{n}\sum_{m=1}^{n}\mathbf{E}(\varepsilon_m^2) < \phi_{i_1 m}, \psi_{ij} >^2$$

$$+\frac{2}{n}(\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}\mathbf{E}\varepsilon_k\varepsilon_l < \phi_{i_1 k}, \psi_{ij} > . < \phi_{i_1 l}, \psi_{ij} >)$$

$$= \frac{\sigma^2}{n} + \frac{2\sigma^2}{n}\sum_{k=1}^{n-1}\rho(k) \tag{3.5}$$

$$[\sum_{r=k}^{n-k} < \phi_{i_1 r}, \psi_{ij} > . < \phi_{i_1 (r+k)}, \psi_{ij} >]$$

$$\le \frac{\sigma^2}{n} + \frac{2\sigma^2}{n}\sum_{k=1}^{n-1}\rho(k)$$

$$= O(\frac{1}{n}).$$

Similarly, we have

$$\mathbf{E}V_{i_0 j}^2 = O(\frac{1}{n}) \tag{3.6}$$

Therefore, U_{ij} and $V_{i_0 j}$ are both normally distributed with zero means with variance σ^2/n.

Part (b). Decomposition of the quadratic risk. Observing that the orthogonality (2.1) implies that

$$\mathbf{E}\,|\,\hat{g} - g\,|_2^2 = T_1 + T_2 + T_3 + T_4, \tag{3.7}$$

where

$$T_1 = \sum_{i=i_1}^{\infty}\sum_{j\in Z}\beta_{ij}^2,$$

$$T_2 = \sum_{j\in Z}\mathbf{E}(\hat{\alpha}_{i_0 j} - \alpha_{i_0 j})^2 = \mathbf{E}\,?\,\mathrm{Pr}oj_{V_{i_0}}(\hat{G}_{i_1} - g)\,?_2^2,$$

$$T_3 = \sum_{i=i_0}^{i_1-1}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}(\hat{\beta}_{ij} - \beta_{ij})^2\}$$

$$= \sum_{i=i_0}^{i_1-1}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}(u_{ij} - U_{ij})^2\},$$

$$T_4 = \sum_{i=i_0}^{i_1-1}\sum_{k\in Z}P(\hat{B}_{ik} \le n^{-1}c)\sum_{(ik)}\beta_{ij}^2.$$

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

Bound for T_1: By Considering Equation (5.5) of [8]

$$T_1 = O(n^{\frac{-2s_2}{2s_2+1}}). \tag{3.8}$$

Bound for T_2: From the definition of $\hat{\alpha}_{i_0 j}$, (3.3) and (3.6), we have

$$T_3 = \sum_{j=0}^{2^{i_0}-1}v_{i_0 j}^2 + \sum_{j=0}^{2^{i_0}-1}\mathbf{E}V_{i_0 j}^2$$

$$\le Cn^{\frac{-2s_2}{2s_2+1}} + n^{\frac{-2N}{2N+1}}\sigma^2 \tag{3.9}$$

$$= O(n^{\frac{-2s_2}{2s_2+1}}).$$

Bound for T_3:

$$T_3 = \sum_{i=i_0}^{i_1-1}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}(u_{ij} - U_{ij})^2\}$$

$$\le 2\sum_{i=i_0}^{i_1-1}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}U_{ij}^2\} \tag{3.10}$$

$$+ 2\sum_{i=i_0}^{i_1-1}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}u_{ij}^2\}$$

$$=: 2T_3' + 2T_3''.$$

It follows from (3.3) that

$$T_3'' \le \sum_{i=i_0}^{i_1-1}\sum_{k}\sum_{(ik)}u_{ij}^2 \le \sum_{i=i_0}^{i_1-1}\sum_{j}u_{ij}^2 \le Cn^{\frac{-2s_2}{2s_2+1}}. \tag{3.11}$$

Thus, we only need to bind T_3'. Let $i-1$ denote the integer part of the base-2 logarithm of $n^{1/(2s_2+1)}$; thus, 2^{-i} is of the optimal order for a bandwidth in kernel estimation of a function of known smoothness s_2. Put $B_{ik} = l^{-1}\sum_{(ik)}(\beta_{ij} + u_{ij})^2$, where $l = l_i$ denotes block length. As in [8] page 44, we may split T_3' into several parts:

$$T_3' = T_{31} + T_{32} + T_{33} + T_{34}, \tag{3.12}$$

where

$$T_{31} = \sum_{i=i_0}^{i}\sum_{k\in Z}\mathbf{E}\{I(\hat{B}_{ik} > n^{-1}c)\sum_{(ik)}U_{ij}^2\},$$

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

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

$$T_{34} = \sum_{i=i+1}^{i_1-1}\sum_{k\in Z}\mathbf{E}[I(\hat{B}_{ik} > n^{-1}c)I(B_{ik} \le (2n)^{-1}c)\sum_{(ik)}U_{ij}^2].$$

From (3.5), we have

$$T_{31} \le \sum_{i=i_0}^{i'} \sum_k \sum_{(ik)} \mathbf{E}(U_{ij}^2)$$

$$\le Cn^{-1} \sum_{i=i_0}^{i'} \sum_{j=0}^{2^i-1} \tag{3.13}$$

$$= O(n^{\frac{-2s_2}{2s_2+1}}).$$

By considering Equations (5.10) and (5.11) of [8] page 45 we have

$$T_{32} = O(n^{\frac{-2s_2}{2s_2+1}}), \tag{3.14}$$

$$T_{33} = O(n^{\frac{-2s_2}{2s_2+1}}). \tag{3.15}$$

By Lemma 3.1 below, we can show that for all $\lambda > 0$,

$$T_{34} = O(n^{-\lambda}). \tag{3.16}$$

Lemma 3.1. *Let the integer* n *be large enough. Then for all integers* i, k, *and for all* $t > 0$,

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

Proof. Let $\mathbf{A} = \{a = (a_1,...,a_l) \in \mathbf{R}^l \ : \ \sum_{j=1}^l a_j^2 = 1\}$. Note that for all integer i and k, we have

$$(\sum_{(ij)} U_{ij}^2)^1 / 2 = \sup_{a\in\mathbf{A}} \sum_{j=1}^l a_j U_{ij}. \tag{3.17}$$

Consider the centered Gaussian process $\{Z(a), a \in \mathbf{A}\}$ defined by $Z(a) = \sum_{j=1}^l a_j U_{ij}$. Firstly, by the Cauchy-Schwarz inequality, Jenson's inequality and (3.5), we have

$$\mathbf{E}(\sup_{a\in\mathbf{A}} Z(a)) \le \mathbf{E}\{(\sum_{j=1}^l U_{ij}^2)^{1/2} \le (l\sigma^2/n)^{1/2}. \tag{3.18}$$

Secondly, for every $a \in \mathbf{A}$, we have

$$\mathbf{E}(Z(a)^2) = \sum_{j=1}^l a_j^2 \mathbf{E}(U_{ij}^2) + 2\sum_{j=1}^{l-1}\sum_{k=j+1}^{l} a_j a_k \mathbf{E}(U_{ij}U_{ik}) = J_1 + J_2. \tag{3.19}$$

By an argument similar to that for proving (3.5), we can show the following inequalities:

$$J_1 \le \frac{1}{n}\sum_{m=1}^n <\phi_{i_1 m}, \sum_{j=1}^l a_j \psi_{ij} >^2 \le \frac{1}{n} ?\sum_{j=1}^l a_j\psi_{ij} ?_2^2 \le \frac{1}{n}\sum_{j=1}^l a_j^2 = \frac{1}{n}, \tag{3.20}$$

$$J_2 \le \frac{2}{n}\sum_{j=1}^l \sum_{j\ne j'} a_j a_{j'} \{\sum_{m_1=1}^{n-1}\sum_{m_2=m_1+1}^{n} \rho(m_2-m_1) <\phi_{i_1 m_1}\psi_{ij}><\phi_{i_1 m_2}\psi_{ij'}\}$$
$$\le \frac{2}{n}\sum_{j=1}^l a_j^2 \{\sum_{m_1=1}^{n-1}\sum_{m_2=m_1+1}^{n} \rho(m_2-m_1) <\phi_{i_1 m_1}\psi_{ij}><\phi_{i_1 m_2}\psi_{ij}\} \tag{3.21}$$
$$= O(\frac{1}{n}).$$

In deriving the first inequality, we have used the fact that ϕ and ψ are supported on $[0,1]$. Again (3.20) and (3.21) together yield

$$D^2 := \sup_{a\in\mathbf{A}} \mathbf{E}(Z(a)^2) = O(\frac{1}{n}) \tag{3.22}$$

Now we denote $\tilde{m} := \mathbf{E}(\sup_{a\in\mathbf{A}} Z(a))$. It follows from Borel's inequality, (3.18) and (3.22) that for all $u > 2\tilde{m}$,

$$P\{\sup_{a\in\mathbf{A}} \sum_{j=1}^l a_j U_{ij} \ge u\} \le \exp\{-\frac{(u-\tilde{m})}{2D^2}\} \le \exp\{-\frac{nu^2}{2\sigma^2}\}. \tag{3.23}$$

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

<u>Bound for</u> T_4: As to the last term, we may write

$$T_4 \le \sum_{i=i_0}^{i_1-1}\sum_{k\in S_i} P(\hat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} \ge 2n^{-1}c)\sum_{(ik)}\beta_{ij}^2$$
$$+\sum_{i=i_0}^{i'}\sum_{k\notin S_i} P(\hat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} \ge 2n^{-1}c)\sum_{(ik)}\beta_{ij}^2$$
$$+\sum_{i=i_0}^{i'}\sum_{k\in Z} P(\hat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} < 2n^{-1}c)\sum_{(ik)}\beta_{ij}^2$$
$$+\sum_{i=i'+1}^{i_1-1}\sum_{k\in S_i} P(\hat{B}_{ik} \le n^{-1}c \text{ and } B_{jk} < 2n^{-1}c)\sum_{(ik)}\beta_{ij}^2$$
$$+\sum_{i=i'+1}^{i_1-1}\sum_{k\notin S_i} P(\hat{B}_{ik} \le n^{-1}c\sum_{(ik)}\beta_{ij}^2$$
$$= T_{41} + T_{42} + T_{43} + T_{44} + T_{45}.$$

As in [8] page 47 we could show

$$T_{41} = o(n^{\frac{-2s_2}{2s_2+1}}), \tag{3.24}$$

$$T_{42} = o(n^{\frac{-2s_2}{2s_2+1}}), \tag{3.25}$$

$$T_{43} = O(n^{\frac{-2s_2}{2s_2+1}}), \tag{3.26}$$

$$T_{44} = o(n^{\frac{-2s_2}{2s_2+1}}), \tag{3.27}$$

$$T_{45} = O(n^{\frac{-2s_2}{2s_2+1}}). \tag{3.28}$$

This in conjunction with (3.8)-(3.16), gives Theorem 3.1.

Acknowledgements

The authors are grateful to the editor and anonymous referees for contributing to the improvement of this paper through many useful suggestions and detail comments.

References

1. Cai T. On block thresholding in wavelet regression: Adaptivity, block size and threshold. *Statistica Sinica*, **12**: 1241-1273 (2002).
2. Cohen A., Daubechies I. and Vail P. Wavelets on the interval and fast wavelets transforms. *Appl. Comput. Harm. Anal.*, **1**: 54-82 (1993).
3. Daubechies I. Ten Lectures on Wavelets, CBMS-NSF regional conferences series in applied mathematics. SIAM, Philadelphia (1992).
4. Donoho D.L., Johnstone I.M., Kerkyacharian G., and Picard D. Wavelet shrinkage: asymptopia (with discussion). *Journal of Royal Statistical Society*, Ser. B, **57**(2): 301-370 (1995).
5. Donoho D.L., Johnstone I.M., Kerkyacharian G., and Picard D. Density estimation by wavelet thresholding. *The Annals of Statistics*, **2**: 508-539 (1996).
6. Donoho D.L. and Johnstone I. M. Minimax estimation via wavelet shrinkage. *Ibid.*, **26**(2): 879-921 (1998).
7. Doosti H. and Niroumand H.A. Wavelets for Non-parametric Stochastic Regression With pairwise negative quadrant dependent random variables. *J. Sci. Islam. Republ. Iran*, Preprint (2005).
8. Hall P., Kerkyacharian G., and Picard D. On the minimax optimality of block thresholded wavelet estimators. *Statistica Sinica*, **9**: 33-49 (1999).
9. Hall P. and Patil P. Formulate for mean integrated squared error of non-linear wavelet-based density estimators. *The Annals of Statistics*, **23**: 905-928 (1995).
10. Hall P. and Patil P. On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *Journal of Royal Statistical Society*, Ser. B, **58**(2): 361-377 (1996).
11. Hardle W., Kerkyacharian G., Picard D., and Tsybabov A. *Wavelets: Approximation and Statistical Applications*. Springer-Verlag, NewYork (1998).