

# Estimation of Regression Function for Nonequispaced Samples Based on Warped Wavelets

**Nargess Hosseinioun**

Department of Statistics, Payame Noor University, 19395-4697, Tehran, Iran  
Email: Narges\_8h@yahoo.com

Received 21 December 2015; accepted 20 February 2016; published 23 February 2016

Copyright © 2016 by author and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

We consider the problem of estimating an unknown density and its derivatives in a regression setting with random design. Instead of expanding the function on a regular wavelet basis, we expand it on the basis  $\phi(G(x))$ , a warped wavelet basis. We investigate the properties of this new basis and evaluate its asymptotic performance by determining an upper bound of the mean integrated squared error under different dependence structures. We prove that it attains a sharp rate of convergence for a wide class of unknown regression functions.

## Keywords

Dependent Sequence, Nonparametric Regression, Random Design, Warped Wavelet Basis

---

## 1. Introduction

In nonparametric regression, it is often of interest to estimate some functionals of a regression function, such as its derivatives. For example, in the study of growth curves, the first (speed) and second (spurt) derivatives of the height as a function of age are important parameters for study (Muller [1]). Other needs for derivative estimation often arise in nonparametric regressions themselves. For example, in constructing interval estimates for a regression function and kernel bandwidth selection (Ruppert and Wand [2]), estimators of higher order derivatives are employed in estimating the leading bias terms. Suppose  $n$  independent variables  $Y_1, \dots, Y_n$  with

$$Y_i = f(X_i) + \varepsilon_i,$$

where  $X_i$  and  $\varepsilon_i$  are independent random variables,  $\varepsilon_i$  assumed to have normal distribution with mean zero

and variance  $\sigma^2 = 1$  for simplicity. The  $X_i$ 's have a density  $g$  which may be known or unknown, but assumed to be compactly supported on the interval  $I = [a, b]$ , as well as  $f$ . We aim to estimate  $f^{(d)}$ , that is, the  $d$ th derivative of  $f$ , for any integer  $d$ .

Considerable research has been devoted to the subject of estimation, mainly the kernel methods, see, *e.g.*, [3]-[8], the smoothing splines, and local polynomial methods, see, *e.g.*, [9]-[11]. One may also be interested in more traditional approaches to nonparametric regression, mainly fixed-bandwidth kernel methods, orthogonal series methods and linear spline smoothers. These methods are not adaptive. The estimators based on these methods may achieve substantially slower rate of convergence if the smoothness of the underlying regression functions is misspecified. The recent development of wavelet bases based on multiresolution analyses suggests new techniques for nonparametric function estimation. Wavelet analysis plays important roles in both pure and applied mathematics such as signal processing, image compressing, and numerical solutions. The application of wavelet theory to the field of statistical function estimation is pioneered by Donoho and Johnstone. In a series of important papers (see, *e.g.*, [12]-[15]), Donoho and Johnstone and coauthors present a coherent set of procedures that are spatially adaptive and near optimal over a range of function spaces of inhomogeneous smoothness. They enjoy excellent mean squared error properties when are used to estimate functions that are only piecewise smooth and have near optimal convergence rates over large function classes.

Recently a quite different algorithm is developed by Kerkyacharian and Picard [16]. The procedure stays very close to the equispaced Donoho and Johnstone's Visushrink procedure, and thus is very simple in its form and in its implementation. Simply, the projection is done on an unusual non-orthonormal basis, called *warped* wavelet basis. Assuming that  $g$  is known but with no boundedness assumptions on it, two new estimators have been introduced based on a warped wavelet basis. The features of this basis consist of a standard wavelet basis and of the definition of  $G$  related to the model. Of course, the properties of this basis truly depend on the warping factor  $G$ . Such a technique has been already used with success in the framework of nonparametric regression with random design by Kerkyacharian and Picard [16]. Recent works on warped wavelet basis in nonparametric statistics can be found in [17]-[20]. To the best of our knowledge, only Cai [21] and Petsa and Sapatinas [22] have proposed wavelet estimators for  $f^{(d)}$ , but defined with a deterministic equidistant design; that is,

$$X_i = \frac{i}{n}.$$

The consideration of a random design with *warped* wavelet complicates significantly the problem and no wavelet estimators for derivative of regression function exist in this case. This motivates us to study the case under different dependence structures: the strong mixing case and the  $\rho$ -mixing case. Asymptotic mean integrated squared error properties for derivatives of regression function has been explored. In each case, we prove that warped wavelet estimator attains a fast rate of convergence. Another important advantage of the *warped* basis estimators is that they are near optimal in the minimax sense over a large class of function spaces for a wide variety of design densities, not necessarily bounded above and below as generally required by other wavelet estimators. Basically, the condition on the design refers to the Muckenhoupt weights theory introduced in Muckenhoupt [23].

The rest of the paper is organized as follows. Section 2 describes the warped wavelet basis and nonquispaced procedure. Optimality of the estimators will be presented in Section 3, while Section 4 contains proofs of the main results.

## 2. Assumptions

We aim to estimate derivative of regression function when  $f \in L^2([0;1])$  via  $n$  random variables (or vectors)  $Y_1 \cdots Y_n$  from a strictly stationary stochastic process  $(Y_i)_{i \in \mathbb{Z}}$  defined on a probability space  $(\Omega, A, P)$ .

**Condition 1.** We define the  $m$ -th strong mixing coefficient of  $(Y_i)_{i \in \mathbb{Z}}$  by

$$\alpha_m = \sup_{(A,B) \in \sigma_{-\infty,0} \times \sigma_{m,\infty}} |P(AB) - P(A)P(B)|.$$

We define  $\sigma_{-\infty,0}$  as the  $\sigma$ -algebra generated by the random variables (or vectors)  $\cdots, Y_{-1}, Y_0$  and  $\sigma_{m,\infty}$  as the  $\sigma$ -algebra generated by the random variables (or vectors)  $Y_m, Y_{m+1}, \cdots$ . We say that  $(Y_i)_{i \in \mathbb{Z}}$  is strong mixing if and only if  $\lim_{m \rightarrow \infty} \alpha_m = 0$ . Furthermore, there exist two constants  $\gamma, c > 0$  such that, for any integer  $m \geq 1$ ,

$$\alpha_m \leq \gamma \exp(-cm).$$

Applications on strong mixing can be found in [24]-[26]. Among various mixing conditions used in the literature,  $\alpha$ -mixing has many practical applications. Many stochastic processes and time series are known to be  $\alpha$ -mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with exponential mixing coefficients. The  $\alpha$ -mixing dependence is reasonably weak; it is satisfied by a wide variety of models including Markov chains, GARCH-type models and discretely observed discussions.

**Condition 2.** Let  $(Y_i)_{i \in \mathbb{Z}}$  be a strictly stationary random sequence. For any  $m \in \mathbb{Z}$ , we define the  $m$ -th maximal correlation coefficient of  $(Y_i)_{i \in \mathbb{Z}}$  by  $\rho$ -mixing:

$$\rho_{(m)} = \sup_{(U,V) \in L^2(\sigma_{-\infty,0}) \times L^2(\sigma_{m,\infty})} |\text{corr}(U,V)|.$$

### 3. Warped Basis and Estimation Framework

Let  $N$  be a positive integer. We consider an orthonormal wavelet basis generated by dilations and translations of a father Daubechies-type wavelet and a mother Daubechies type wavelet of the family db2N (see [27]). Further details on wavelet theory can be found in Daubechies [27] and Meyer [28]. In particular, mention that  $\phi$  and  $\psi$  have compact supports. For any  $j \geq 0$ , we set  $\Lambda_j = \{0, \dots, 2^j - 1\}$  and for  $k \in \Lambda_j$ , we define  $\phi$  and  $\psi$  as father and mother wavelet:

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

With appropriated treatments at the boundaries, there exists an integer  $\tau$  such that, for any integer  $l \geq \tau$ ,

$$\Gamma = \{\phi_{l,k}, k \in \Lambda_l; \psi_{j,k}, j \in N - \{0, \dots, L-1\}, K \in \Lambda_j\},$$

forms an orthonormal basis of  $L^2[0,1]$ . For any integer  $l \geq \tau$  and  $f \in L^2[0,1]$ , we have the following wavelet expansion:

$$f(x) = \sum_{k \in \Lambda_l} \alpha_{l,k} \phi_{l,k}(x) + \sum_{j=l}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x), x \in [0,1],$$

where  $\alpha_{l,k} = \int f \phi_{l,k}$  and  $\beta_{j,k} = \int f \psi_{j,k}$ . Furthermore we consider the following wavelet sequential definition of the Besov balls. We say  $f \in B_{p,r}^s(d)$ , with  $s > 0, p, r \geq 1$ , and  $M > 0$  if there exists a constant  $C > 0$ , such that

$$2^{\tau(1/2-1/p)} \left( \sum_{k \in \Lambda_{\tau}} |\alpha_{\tau,k}|^p \right)^{1/p} + \left( \sum_{j=\tau}^{\infty} \left( 2^{j(s+1/2-1/p)} \left( \sum_{k \in \Lambda_j} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq C,$$

with the usual modifications if  $p=1$  or  $r=1$ . Note that, for particular choices of  $s, p$  and  $r, B_{p,r}^s(M)$  contains the classical Holder and Sobolev balls. See, e.g., Meyer [28] and Hardle *et al.* [29]. Now we consider the wavelet basis  $\Gamma$  with  $N > 5m$  and  $\phi$  and  $\psi$  have  $d$  derivatives, then the generalized expansion of derivative of  $f$  is

$$f^{(d)}(x) = \sum_{k \in \Lambda_l} \alpha_{l,k}^{(d)} \phi_{l,k}^{(d)}(x) + \sum_{j=l}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}^{(d)} \psi_{j,k}^{(d)}(x), \quad (3.1)$$

where the coefficients are

$$\alpha_{l,k}^{(d)}(x) = \int f^{(d)}(x) \phi_{l,k}(x) dx, \quad (3.2)$$

and

$$\beta_{j,k}^{(d)}(x) = \int f^{(d)}(x) \psi_{j,k}(x) dx. \quad (3.3)$$

We define the linear wavelet estimator  $\hat{f}^{(d)}(x)$  by

$$\hat{f}^{(d)}(x) = \sum_{k \in \Lambda_l} \hat{\alpha}_{l,k}^{(d)} \phi_{l,k}(x), \quad (3.4)$$

where

$$\hat{\alpha}_{l,k}^{(d)} = \frac{(-1)^m}{n} \sum_{k \in \Lambda_{j_0}} \phi_{l,k}^{(d)}(x) Y_i, \quad (3.5)$$

where  $j_0$  is an integer a posteriori. For more on estimating of derivatives of density function see [30] and [31]. Kerkyacharian and Picard [16] propose a construction where the unknown function is expanded on a warped basis instead of a regular wavelet basis. Proceeding in such a way, the estimates of the coefficients become more natural. Let us briefly describe the construction of this procedure. Suppose

$$G(x) = \int_a^x g(u) du,$$

is a known function, continuous and strictly monotone from  $[a, b]$  to  $[0, 1]$ , then

$$\hat{\alpha}_{l,k}^{(d)} = \frac{(-1)^m}{n} \sum_{k \in \Lambda_{j_0}} \phi_{l,k}^{(m)}(G(x)) Y_i. \quad (3.6)$$

It is clear that the above estimator is unbiased and we perform the following warped estimator:

$$\hat{f}^d(x) = \sum_{k \in \Lambda_{j_0}} \hat{\alpha}_{l,k}^{(d)} \phi_{l,k}(G(x)). \quad (3.7)$$

In the case where  $g$  is unknown, we replace  $G$  wherever it appears in the construction by the empirical distribution of the  $X_i$ 's:

$$\hat{G}_n = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x_i\}$$

Let us define the new empirical wavelet coefficients:

$$\tilde{\alpha}_{l,k}^{(d)} = \frac{(-1)^m}{n} \sum_{k \in \Lambda_{j_0}} \phi_{l,k}^{(d)}(\hat{G}(x)) Y_i. \quad (3.8)$$

Consequently we have the estimator:

$$\tilde{f}^{(d)}(x) = \sum_{k \in \Lambda_l} \tilde{\alpha}_{l,k}^{(d)} \phi_{l,k}(\hat{G}(x)). \quad (3.9)$$

This approach was initially introduced by Rao [32] for the estimation of the derivatives of a density. Note that, for  $m = 0$  the standard case, this estimator has been considered and studied in Kerkyacharian and Picard [16].

#### 4. Optimality Results

The main results of the paper are upper bounds for the mean integrated square error of the wavelet estimator  $\hat{f}^{(d)}(x)$ , which is defined as usual by

$$MISE(f^{(d)}) = E \int_R (\hat{f}^{(d)}(x) - f^{(d)}(x))^2 dx.$$

Moreover,  $C$  denotes any constant that does not depend on  $l, k$  and  $n$ .

**Proposition 4.1.** Suppose that  $(X_i)_{i \in Z}$  are independent. For any integer  $l \geq \tau$  and  $k \in \Lambda_j$ ,  $\hat{\alpha}_{l,k}^{(d)}$  is unbiased estimator of  $\alpha_{l,k}^{(d)}$  and there exists a constant  $C > 0$  such that

$$E \left( \hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)} \right)^2 \leq C 2^{2lm}.$$

**Proof of Proposition 4.1.** We have

$$\begin{aligned}
 E\left(\hat{\alpha}_{l,k}^{(d)}\right) &= \frac{(-1)^m}{n} \sum_{k \in \Lambda_{j_0}} E\left(\phi_{l,k}^{(d)}(G(X))(f(X_i) + \varepsilon_i)\right) \\
 &= (-1)^{(d)} E\left(\phi^{(d)}\right)(G(X) f(X)) \\
 &= (-1)^{(d)} \int \phi^{(d)}(G(x)) f(x) g(x) dx \\
 &= (-1)^{(d)} \int \phi^{(d)}(u) f(G^{-1}u) du = \alpha_{l,k}^{(d)}.
 \end{aligned}$$

So  $\hat{\alpha}_{l,k}^{(d)}$  is unbiased estimator of  $\alpha_{l,k}^{(d)}$ . Therefore

$$\begin{aligned}
 \text{Var}\left(\hat{\alpha}_{l,k}^{(d)}\right) &\leq \frac{1}{n} E\left(\phi_{l,k}^{(d)}(G(X))Y_i\right)^2 \leq \frac{1}{n} E\left(\phi_{l,k}^{(d)}(G(X))(f(X_i) + \varepsilon_i)\right)^2 \\
 &\leq \frac{1}{n} \left( E\left(\phi_{l,k}^{(d)}(G(X))f(X_i)\right)^2 + E\left(\phi_{l,k}^{(d)}(G(X))(\varepsilon_i)\right)^2 \right) \\
 &= T_1 + T_2,
 \end{aligned} \tag{4.1}$$

where

$$T_1 = \frac{1}{n} E\left(\phi_{l,k}^{(d)}(G(x))(f(X_i))\right)^2,$$

and

$$T_2 = \frac{1}{n} E\left(\phi_{l,k}^{(d)}(G(x))(\varepsilon_i)\right)^2.$$

For upper bound of  $T_1$ , we have

$$E\left(\phi_{l,k}^{(d)}(G(X))(\varepsilon_i)\right)^2 \leq \|f\|_\infty^2 \int \phi_{l,k}^{(d)2}(G(x)) g(X) dx = \|f\|_\infty^2 \int \phi_{l,k}^{(d)2}(u) du.$$

Using the same technic as [19] and change of variables  $y = 2^l x - k$ , we obtain

$$= 2^{2rl} \|f\|_\infty^2 \int \phi^{(d)2}(u) du \leq C 2^{2lm}. \tag{4.2}$$

Considering almost the same integral as in  $T_1$ , and the fact  $E(\varepsilon_i)^2 = 1$ , we have

$$T_2 = \frac{1}{n} E\left(\phi_{l,k}^{(d)}(G(X))\right)^2 E(\varepsilon_i)^2 \leq C 2^{2lm} \tag{4.3}$$

It follows from (4.2), (4.3) and (4.1), that

$$\text{Var}\left(\hat{\alpha}_{l,k}^{(d)}\right) \leq C \frac{2^{2lm}}{n}$$

**Proposition 4.2.** Suppose that the assumptions of Condition 1 hold. Let  $\hat{f}^{(d)}$  be (3.4). Then there exists  $C > 0$  such that

$$E\left(\hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)}\right)^2 \leq C 2^{2lm}.$$

**Proof of Proposition 4.2.** Observe that

$$\text{Var}\left(\hat{\alpha}_{l,k}^{(d)}\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left(\phi_{l,k}^{(d)}(G(X_i))Y_i, \phi_{l,k}^{(d)}(G(X_j))Y_j\right) \leq S_1 + S_2, \tag{4.4}$$

where

$$S_1 = \frac{1}{n} \text{Var}\left(\hat{\alpha}_{l,k}^{(d)}\right),$$

and

$$S_2 = \frac{2}{n^2} \left| \sum_{i=2}^n \sum_{j=1}^{i-1} \text{Cov}(\phi_{l,k}^{(d)}(G(X_i))Y_i, \phi_{l,k}^{(d)}(G(X_j))Y_j) \right|.$$

It follows from the fact that  $2^l < n$ ,

$$\begin{aligned} S_2 &= \frac{2}{n^2} \left| \sum_{i=1}^n (n-i) \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right| \\ &\leq \frac{2}{n} \sum_{i=1}^n \left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right| = A + B, \end{aligned} \quad (4.5)$$

where

$$\begin{aligned} A &= \frac{2}{n} \sum_{i=1}^{2^l-1} \left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right|, \\ B &= \frac{2}{n} \sum_{i=2^l}^n \left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right|. \end{aligned}$$

Using Proposition 6.1 in [33], and the fact that  $\phi_{l,k}^{(d)}(x) = 2^{l/2} 2^{dl} \phi^{(d)}(2^l x - k)$ , we have

$$\begin{aligned} &\left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right| \\ &\leq C \left( \int \phi_{l,k}^{(d)}(G(x)) y_i g(x) dx \right)^2 \leq C \left( \int \phi_{l,k}^{(d)}(u) y du \right)^2 \leq c 2^{-j} 2^{2ld} \end{aligned}$$

Therefore

$$A \leq \frac{C}{n} 2^{2ld} \quad (4.6)$$

Applying the Davydov inequality for strongly mixing processes (see [34]), for any  $q \in (0,1)$ , we have

$$\begin{aligned} &\left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right| \leq C \alpha_m^q \left( E \left| \phi_{l,k}^{(d)}(G(x))y \right|^{2/(1-q)} \right)^{1-q} \\ &\leq C \alpha_m^q \left( \sup_{x \in [0,1]} \left| \phi_{l,k}^{(d)}(G(x))y \right| \right)^{2q} E \left( \left| \phi_{l,k}^{(d)}(G(X))Y \right|^2 \right)^{1/q}. \end{aligned} \quad (4.7)$$

Now we have  $\sup_{x \in [0,1]} \left| \phi_{l,k}^{(d)}(G(x))y \right| \leq 2^{-l/2} 2^{dl}$  and by (4.3),

$$E \left( \left| \phi_{l,k}^{(d)}(G(X))Y \right|^2 \right)^{1-q} \leq C 2^{2dl(1-q)} \alpha_m^q. \quad (4.8)$$

Hence by applying (4.7) and (4.8), we get

$$B \leq C n^{2ld} \sum_{m=2^l}^n m^{-q} \alpha_m^q. \quad (4.9)$$

It follows from (4.5), (4.6) and (4.9) that

$$S_2 \leq C n^{2ld}. \quad (4.10)$$

Now (4.10) with **proposition 4.1** completes the proof.

**Proposition 4.3.** Suppose that the assumptions of **Condition 2** hold. Let  $\hat{f}^{(d)}$  be (3.4). Then there exists a constant  $C > 0$  such that

$$E \left( \hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)} \right)^2 \leq C 2^{2lm}.$$

**Proof of Proposition 4.3.** Having the same technique as in **Proposition 4.2**, we have

$$A = \frac{2}{n} \sum_{i=1}^{2^l-1} \left| \text{Cov}(\phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i) \right| \leq \frac{C}{n} 2^{2ld}. \quad (4.11)$$

Applying the covariance inequality for  $\rho$ -mixing processes (see Doukahn [25]), *i.e.*,

$$\begin{aligned} & \left| \text{Cov} \left( \phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i \right) \right| \\ & \leq 2\rho_i \left( E \left( \phi_{l,k}^{(d)}(G(X_0))Y_0 \right)^2 \right)^{1/2} \left( E \left( \phi_{l,k}^{(d)}(G(X_i))Y_i \right)^2 \right)^{1/2}. \end{aligned}$$

We obtain from (4.2),

$$\left| \text{Cov} \left( \phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i \right) \right| \leq C\rho_i 2^{2lm}. \quad (4.12)$$

Hence by

$$B = \frac{2}{n} \sum_{i=2}^n \left| \text{Cov} \left( \phi_{l,k}^{(d)}(G(X_0))Y_0, \phi_{l,k}^{(d)}(G(X_i))Y_i \right) \right| \leq C \frac{2^{2lm}}{n}. \quad (4.13)$$

So **Proposition 4.3** is complete from (4.12) and (4.13).

Now based on the above Proposition, we have the following main result:

**Theorem 4.1.** Suppose that the assumptions of Section 2 hold. Let  $\hat{f}^{(d)}$  be (3.4). Suppose that  $f \in B_{p,r}^s(d)$ ,  $s > 0$ ,  $p, r \geq 1$  and  $2^l = n^{1/(2s_0+2d+1)}$ . Then there exists a constant  $C > 0$  such that

$$\text{MISE} \left( f^{(d)} \right) \leq C n^{-2s_0/(2s_0+2d+1)},$$

where  $s_0 = s + \min(1/2 - 1/p, 0)$ . The rate of convergence corresponds to the one obtained in the derivatives density estimation framework. See, for example, Rao [32] and Chaubey *et al.* [30] [31].

**Proof of Theorem 4.1.** Since we set  $\Lambda$  an arthonormal set in  $L[0,1]$ , we have

$$E \left( \int \hat{f}^{(d)}(x) - (f^{(d)}(x))^2 dx \right) = \sum_{k \in \Lambda_l} \left( \hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)} \right)^2 + \sum_{j=1}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}^{(d)2}.$$

As we define  $f \in B_{p,r}^s(d)$ , there exists a constant  $C > 0$ , such that

$$\sum_{j=1}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}^{(d)2} \leq C 2^{-2ls}. \quad (4.14)$$

First consider the i.i.d case. Using (4.2) and (4.3) and the fact that  $\text{Card}(\Gamma) = 2^l$ , one can easily have

$$\sum_{k \in \Lambda_l} \left( \hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)} \right)^2 \leq C 2^l \frac{2^{2ld}}{n} \leq C n^{-2s_0/(2s_0+2d+1)}. \quad (4.15)$$

Second, suppose the assumptions of Section 2 hold. Using **Proposition 4.2** with  $2^l = n^{1/(2s_0+2d+1)}$ , we have

$$\sum_{k \in \Lambda_l} \left( \hat{\alpha}_{l,k}^{(d)} - \alpha_{l,k}^{(d)} \right)^2 \leq C 2^l \frac{2^{2ld}}{n} \leq C n^{-2s_0/(2s_0+2d+1)}. \quad (4.16)$$

**Remark 4.1.** **Theorem 4.1** shows that, under mild assumptions on the dependence of observations,  $\hat{f}^{(d)}$  attains a rate of convergence close to the one for the i.i.d. case *i.e.*,  $n^{-2s_0/(2s_0+2d+1)}$ .

## 5. Conclusion

In this paper, we proposed a wavelet-based estimator for derivatives of regression function in the random design. The proposed estimator was formulated according to the warped basis which was simple and easy for applications. The results successfully revealed that without imposing too restrictive assumptions on the model, the wavelet-based estimator attained a sharp rate of convergence under strong mixing and  $\rho$ -mixing structures.

## Acknowledgements

The author would like to express her gratitude to the referee and chief editor for their valuable suggestions which have improved the earlier version of the paper.

## References

- [1] Muller, H.G. (1988) Nonparametric Analysis of Longitudinal Data. Springer, Berlin.
- [2] Ruppert, D. and Wand, M.P. (1994) Multivariate Locally Weighted Least Squares Regression. *Annals of Statistics*, **22**, 1346-1370.
- [3] Gasser, T. and Müller, H. (1984) Estimating Regression Functions and Their Derivatives by the Kernel Method. *Scandinavian Journal of Statistics, Theory and Applications*, **11**, 171-185.
- [4] Hardle, W. and Gasser, T. (1985) On Robust Kernel Estimation of Derivatives of Regression Functions. *Scandinavian Journal of Statistics*, **12**, 233-240.
- [5] Jarrow, R., Ruppert, D. and Yu, Y. (2004) Estimating the Interest Rate Term Structure of Corporate Debt with a Semi-parametric Penalized Spline Model. *Journal of the American Statistical Association*, **99**, 57-66.
- [6] Mack, Y.P. and Muller, H.-G. (1989) Derivative Estimation in Nonparametric Regression with Random Predictor Variable. *Sankhya: The Indian Journal of Statistics, Series A*, **51**, 59-72.
- [7] Schumaker, L.L. (1981) Spline Functions: Basic Theory. Wiley, New York.
- [8] Wanda, M.P. and Jones, M.C. (1995) Kernel Smoothing. Chapman and Hall, London.
- [9] Stone, C. (1985) Additive Regression and Other Nonparametric Models. *The Annals of Statistics*, **13**, 689-705.
- [10] Wahba, G. and Wang, Y.H. (1990) When Is the Optimal Regularization Parameter Insensitive to the Choice of the  $l_0$  Function? *Communications in Statistics: Theory and Methods*, **19**, 1685-1700.
- [11] Zhou, S. and Wolfe, D.A. (2000) On Derivative Estimation in Spline Regression. *Statistica Sinica*, **10**, 93-108.
- [12] Donoho, D.L. and Johnstone, I.M. (1994) Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika*, **81**, 425-455. <http://dx.doi.org/10.1093/biomet/81.3.425>
- [13] Donoho, D.L. and Johnstone, I.M. (1995) Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association*, **90**, 1200-1224. <http://dx.doi.org/10.1080/01621459.1995.10476626>
- [14] Donoho, D.L. and Johnstone, I.M. (1996) Neo-Classic Minimax Problems, Thresholding, and Adaptation. *Bernoulli*, **2**, 39-62. <http://dx.doi.org/10.2307/3318568>
- [15] Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1995) Wavelet Shrinkage: Asymptopia. *Journal of the Royal Statistical Society, Series B*, **57**, 301-369.
- [16] Kerkycharian, G. and Picard, D. (2004) Regression in Random Design and Warped Wavelets. *Bernoulli*, **10**, 1053-1105. <http://dx.doi.org/10.3150/bj/1106314850>
- [17] Brutti, P. (2008) Warped Wavelets and Vertical Thresholding. Preprint. arXiv:0801.3319v1.
- [18] Chagny, G. (2012) Warped Bases for Conditional Density Estimation. Preprint MAP5.
- [19] Chesneau, C. and Willer, T. (2015) Estimation of a Cumulative Distribution Function under Interval Censoring "Case 1" via Warped Wavelets. *Communication in Statistics, Theory and Methods*, **44**, 3680-3702.
- [20] Pham Ngoc, T.M. (2009) Regression in Random Design and Bayesian Warped Wavelets Estimators. *Electronic Journal of Statistics*, **3**, 1084-1112.
- [21] Cai, T. (2002) On Adaptive Wavelet Estimation of a Derivative and Other Related Linear Inverse Problems. *Journal of Statistical Planning and Inference*, **108**, 329-349.
- [22] Petsa, A. and Sapatinas, T. (2011) On the Estimation of the Function and Its Derivatives in Nonparametric Regression: A Bayesian Estimation Approach. *Sankhyā: The Indian Journal of Statistics, Series A*, **73**, 231-244.
- [23] Muckenhoupt, B. (1972) Weighted Norm Inequalities for the Hardy Maximal Function. *Transactions of the American Mathematical Society*, **165**, 207-226.
- [24] Carrasco, M. and Chen, X. (2002) Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory*, **18**, 17-39. <http://dx.doi.org/10.1017/S0266466602181023>
- [25] Doukhan, P. (1994) Mixing Properties and Examples. Lecture Notes in Statistics, Vol. 85, Springer-Verlag, New York.
- [26] Withers, C.S. (1981) Conditions for Linear Processes to Be Strong Mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **57**, 477-480. <http://dx.doi.org/10.1007/BF01025869>
- [27] Daubechies, I. (1992) Ten Lectures on Wavelets, SIAM.
- [28] Meyer, Y. (1992) Wavelets and Operators. Cambridge University Press, Cambridge.
- [29] Hardle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998) Wavelet, Approximation and Statistical Applications. Lecture Notes in Statistics, Vol. 129, Springer-Verlag, New York.
- [30] Chaubey, Y.P., Doosti, H. and PrakasaRao, B.L.S. (2006) Wavelet Based Estimation of the Derivatives of a Density with Associated Variables. *International Journal of Pure and Applied Mathematics*, **27**, 97-106.



- 
- [31] Chaubey, Y.P., Chesneau, C. and Doosti, H. (2011) On Linear Wavelet Density Estimation: Some Recent Developments. *Journal of the Indian Society of Agricultural Statistics*, **65**, 169-179.
  - [32] Rao, B.L.S.P. (1996) Nonparametric Estimation of the Derivatives of a Density by the Method of Wavelets. *Bulletin of Informatics and Cybernetics*, **28**, 91-100.
  - [33] Chesneau, C. and Willer, T. (2007) Numerical Performances of a Warped Wavelet Estimation Procedure for Regression in Random Design. Preprint, HAL:hal-00133831.
  - [34] Davydov, Y. (1970) The Invariance Principle for Stationary Processes. *Theory of Probability and Its Applications*, **15**, 498-509.