



**HAL**  
open science

# On the pointwise mean squared error of a multidimensional term-by-term thresholding wavelet estimator

Christophe Chesneau, Fabien Navarro

► **To cite this version:**

Christophe Chesneau, Fabien Navarro. On the pointwise mean squared error of a multidimensional term-by-term thresholding wavelet estimator. *Communications in Statistics - Theory and Methods*, 2017, 46 (11), 10.1080/03610926.2015.1107587. hal-01188362v3

**HAL Id: hal-01188362**

**<https://hal.science/hal-01188362v3>**

Submitted on 13 Dec 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the pointwise mean squared error of a multidimensional term-by-term thresholding wavelet estimator

Christophe Chesneau<sup>1</sup>, Fabien Navarro<sup>2</sup>

<sup>1</sup>*Laboratoire de Mathématiques Nicolas Oresme,  
Université de Caen BP 5186, F 14032 Caen Cedex, France. e-mail:  
christophe.chesneau@unicaen.fr*

<sup>2</sup>*CREST-ENSAI,  
Campus de Ker-Lann,  
Rue Blaise Pascal - BP 37203, 35172 BRUZ cedex, France. e-mail:  
fabien.navarro@ensai.fr*

**Abstract:** In this paper we provide a theoretical contribution to the pointwise mean squared error of an adaptive multidimensional term-by-term thresholding wavelet estimator. A general result exhibiting fast rates of convergence under mild assumptions on the model is proved. It can be applied for a wide range of nonparametric models including possible dependent observations. We give applications of this result for the nonparametric regression function estimation problem (with random design) and the conditional density estimation problem.

**AMS 2000 subject classifications:** 62G07, 62G20.

**Keywords and phrases:** Pointwise mean squared error, Wavelet methods, Density estimation, Nonparametric regression.

## 1. Introduction

Over the last decade, the theoretical and practical aspects of thresholding wavelet methods in nonparametric estimation have been well-developed. The standard scheme is to expand the unknown function of interest  $f$  on a wavelet basis, estimate only the wavelet coefficients having a great magnitude and reconstruct these wavelet coefficients estimators at suitable levels to form an estimator  $\hat{f}$ . The most popular selections of the wavelet coefficient estimators are based on term-by-term thresholding rules introduced by Donoho and Johnstone (1994) and Donoho *et al.* (1996) and block thresholding rules introduced by Hall *et al.* (1999, 1997) and Cai (1999). The main advantage of these techniques is to provide adaptive estimators in the sense that they are relatively unaffected by discontinuities of  $f$ .

From the theoretical point of view, rates of convergence under global or local error over wide function spaces have been determined. In particular, if we focus our attention on the pointwise mean squared error:  $R(\hat{f}, f)(x_0) = E\left((\hat{f}(x_0) - f(x_0))^2\right)$ , where  $x_0$  is a fixed point in  $\mathbb{R}$ , numerous results exist for unidimensional nonparametric models with independent observations.

In the context of the nonparametric regression model, see for instance Cai and Brown (1998) for the term-by-term thresholding wavelet estimator, Cai (1999, 2002a, 2003), Picard and Tribouley (2000) and Efromovich (2005) for the block thresholding wavelet estimators, Bochkina and Sapatinas (2006, 2009) and Abramovich *et al.* (2007) for the Bayes factor wavelet estimators. See also Cai (2002b) for the block thresholding wavelet estimators related to statistical inverse problems and Chicken and Cai (2005) for similar estimators but for the density estimation problem. However, to the best of our knowledge, there is a lack of theoretical results on the adaptive wavelet estimation for the multidimensional setting including possible dependent observations, under the pointwise mean squared error. This motivates this study.

In the present paper, we consider a general nonparametric framework where a  $d$ -multidimensional function where  $d$  is a positive integer:  $f : [0, 1]^d \rightarrow \mathbb{R}$ , needs to be estimated from  $n$  observations. We propose a general form of a multidimensional term-by-term thresholding wavelet estimator  $\hat{f} : [0, 1]^d \rightarrow \mathbb{R}$ . Our main result proves that, under suitable assumptions on the wavelet coefficient estimators, the term-by-term thresholding rule, the tuning parameters and the local smoothness of  $f$ ,  $\hat{f}$  attains a fast rate of convergence under the pointwise mean squared error:  $R_d(\hat{f}, f)(\mathbf{x}_0) = E\left((\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2\right)$ , where  $x_0$  is a fixed point in  $[0, 1]^d$ . The main interest of this result is to be sharp and very flexible; it can be applied for a wide variety of nonparametric models, including those based on dependent observations. We illustrate our general result by considering two different nonparametric estimation problems: the nonparametric regression function estimation problem in the context of random design and the conditional density estimation problem. To the best of knowledge, these two applications provide new results in terms of rate of convergence of  $\hat{f}$  in such a multidimensional setting.

The rest of this paper is organized as follows. Section 2 introduces the considered multidimensional wavelet basis and the function spaces used in the study. The main wavelet estimator and its pointwise mean squared error properties are presented in Section 3. Applications are given in Section 4. The technical proofs are postponed in Section 5.

## 2. Multidimensional wavelet bases

For any positive integer  $m$ , define the  $\mathbb{L}_2([0, 1]^m)$  spaces as

$$\mathbb{L}_2([0, 1]^m) = \left\{ f : [0, 1]^m \rightarrow \mathbb{R}; \int_{[0, 1]^m} (f(\mathbf{x}))^2 d\mathbf{x} < \infty \right\}.$$

Let  $R$  and  $d$  be positive integers. In this study, we consider a  $d$ -multidimensional wavelet bases on  $[0, 1]^d$  based on the scaling and wavelet functions  $\phi$  and  $\psi$  respectively from Daubechies family  $\text{db}_{2R}$  (see Daubechies (1992)). For any

$\mathbf{x} = (x_1, \dots, x_d) \in [0, 1]^d$ , we set

$$\Phi(\mathbf{x}) = \prod_{v=1}^d \phi(x_v),$$

and

$$\Psi_u(\mathbf{x}) = \begin{cases} \psi(x_u) \prod_{\substack{v=1 \\ v \neq u}}^d \phi(x_v) & \text{for } u \in \{1, \dots, d\}, \\ \prod_{v \in A_u} \psi(x_v) \prod_{v \notin A_u} \phi(x_v) & \text{for } u \in \{d+1, \dots, 2^d - 1\}, \end{cases}$$

where  $(A_u)_{u \in \{d+1, \dots, 2^d - 1\}}$  forms the set of all non void subsets of  $\{1, \dots, d\}$  of cardinality greater or equal to 2.

For any integer  $j$  and any  $\mathbf{k} = (k_1, \dots, k_d)$ , we consider

$$\Phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2} \Phi(2^j x_1 - k_1, \dots, 2^j x_d - k_d),$$

$$\Psi_{j,\mathbf{k},u}(\mathbf{x}) = 2^{jd/2} \Psi_u(2^j x_1 - k_1, \dots, 2^j x_d - k_d), \text{ for any } u \in \{1, \dots, 2^d - 1\}.$$

Let  $\mathbf{D}_j = \{0, \dots, 2^j - 1\}^d$ . Then, with an appropriate treatment at the boundaries, there exists an integer  $\tau$  such that, for any integer  $j_* \geq \tau$ , the collection

$$\{\Phi_{\tau,\mathbf{k}}, \mathbf{k} \in \mathbf{D}_{j_*}; (\Psi_{j,\mathbf{k},u})_{u \in \{1, \dots, 2^d - 1\}}, j \in \mathbb{N} - \{0, \dots, j_* - 1\}, \mathbf{k} \in \mathbf{D}_j\}$$

forms an orthonormal basis of  $\mathbb{L}_2([0, 1]^d)$ .

For any integer  $j_*$  such that  $j_* \geq \tau$ , a function  $f \in \mathbb{L}_2([0, 1]^d)$  can be expanded into a wavelet series as

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbf{D}_{j_*}} c_{j_*,\mathbf{k}} \Phi_{j_*,\mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_*}^{\infty} \sum_{\mathbf{k} \in \mathbf{D}_j} d_{j,\mathbf{k},u} \Psi_{j,\mathbf{k},u}(\mathbf{x}), \quad \mathbf{x} \in [0, 1]^d, \quad (2.1)$$

where

$$c_{j_*,\mathbf{k}} = \int_{[0,1]^d} f(\mathbf{x}) \Phi_{j_*,\mathbf{k}}(\mathbf{x}) d\mathbf{x}, \quad d_{j,\mathbf{k},u} = \int_{[0,1]^d} f(\mathbf{x}) \Psi_{j,\mathbf{k},u}(\mathbf{x}) d\mathbf{x}. \quad (2.2)$$

All the details can be found in, e.g., Meyer (1992), Daubechies (1992), Cohen *et al.* (1993) and Mallat (2009).

Let  $M > 0$ ,  $\alpha \geq 0$  and  $\mathbf{x} \in [0, 1]^d$ . Based on the expansion (2.1) and the wavelet coefficients (2.2), we define the function spaces  $\Lambda_d^\alpha(\mathbf{x}, M)$  as

$$\Lambda_d^\alpha(\mathbf{x}, M) = \left\{ f \in \mathbb{L}_2([0, 1]^d) : \sup_{u \in \{1, \dots, 2^d - 1\}} \sup_{j \geq \tau} \sup_{\mathbf{k} \in \mathcal{K}_{u,j}(\mathbf{x})} 2^{j(d/2+\alpha)} |d_{j,\mathbf{k},u}| \leq M \right\},$$

where  $\mathcal{K}_{u,j}(\mathbf{x}) = \{\mathbf{k} \in \mathbf{D}_j; \Psi_{j,\mathbf{k},u}(\mathbf{x}) \neq 0\}$ .

This class of functions can be viewed as a multidimensional version of the one considered in Efromovich (2002). One can prove that  $\Lambda_d^\alpha(\mathbf{x}, M)$  contains  $d$ -dimensional Besov balls  $\mathbf{B}_{d,p,q}^s(M)$  with  $\alpha = s - d/p$  (see Delyon and Juditsky (1996)).

### 3. Main theorem

Let us consider a general nonparametric model where an unknown function  $f \in \mathbb{L}_2([0, 1]^d)$  needs to be estimated from  $n$  observations of a random process defined on a probability space  $(\Omega, \mathcal{A}, P)$ . Adopting the notations of the wavelet series expansion (2.1) of  $f$ , we define the term-by-term thresholding estimator  $\hat{f}$  by

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} \hat{c}_{j_0, \mathbf{k}} \Phi_{j_0, \mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \mathcal{T}(\hat{d}_{j, \mathbf{k}, u}, \kappa 2^{\omega j} \lambda_n) \Psi_{j, \mathbf{k}, u}(\mathbf{x}), \quad (3.1)$$

$\mathbf{x} \in [0, 1]^d$ , where  $\hat{c}_{j_0, \mathbf{k}}$  and  $\hat{d}_{j, \mathbf{k}, u}$  are wavelet coefficients estimators of  $c_{j_0, \mathbf{k}}$  and  $d_{j, \mathbf{k}, u}$  respectively,  $\mathcal{T} : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$  satisfies the inequality:

$$|\mathcal{T}(v, \lambda) - v| \leq C \left( \min(|v|, \lambda) + |v - u| 1_{\{|v-u| > \lambda/2\}} \right), \quad (3.2)$$

$\kappa$  is a large enough constant,  $\lambda_n$  is a threshold depending on  $n$ , and  $j_0$  and  $j_1$  are integers such that

$$\frac{1}{2} 2^{\tau d} (\ln n)^\nu < 2^{j_0(d+2\omega)} \leq 2^{\tau d} (\ln n)^\nu, \quad \frac{1}{2} \frac{1}{\lambda_n^2 (\ln n)^\varrho} \leq 2^{j_1(d+2\omega)} \leq \frac{1}{\lambda_n^2 (\ln n)^\varrho},$$

with  $\nu \geq 0$ ,  $\varrho \geq 0$  and  $\omega \geq 0$ .

Under some assumptions on  $\hat{c}_{j, \mathbf{k}}$ ,  $\hat{d}_{j, \mathbf{k}, u}$ ,  $\kappa$ ,  $\lambda_n$ ,  $\nu$ ,  $\varrho$  and  $\omega$ , Theorem 3.1 below explores the performance of  $\hat{f}$  (3.1) in terms of rate of convergence under pointwise mean squared error over  $\Lambda_d^\alpha(\mathbf{x}_0, M)$ .

**Theorem 3.1.** *Let  $\hat{f}$  be (3.1), with the associated notations. We suppose that  $\hat{c}_{j, \mathbf{k}}$ ,  $\hat{d}_{j, \mathbf{k}, u}$ ,  $\kappa$ ,  $\lambda_n$ ,  $\nu$ ,  $\varrho$  and  $\omega$  satisfy the following properties:*

(a) *there exists a constant  $C > 0$  such that, for any  $\mathbf{k} \in \mathbf{D}_{j_0}$ ,*

$$E((\hat{c}_{j_0, \mathbf{k}} - c_{j_0, \mathbf{k}})^2) \leq C 2^{2\omega j_0} \lambda_n^2,$$

(b) *there exists a constant  $C > 0$  such that, for any  $j \in \{j_0, \dots, j_1\}$ ,  $\mathbf{k} \in \mathbf{D}_j$  and  $u \in \{1, \dots, 2^d - 1\}$ ,*

$$E((\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u})^4) P\left(|\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u}| \geq \frac{\kappa}{2} 2^{\omega j} \lambda_n\right) \leq C 2^{4\omega j} \lambda_n^8,$$

(c)  $\lim_{n \rightarrow \infty} (\ln n)^{\max(\nu, \varrho)} \lambda_n^{2(1-\nu)} = 0$  for any  $\nu \in [0, 1)$ .

Moreover, we suppose that  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$  with  $M > 0$ ,  $\alpha \geq 0$  and  $\mathbf{x}_0 \in [0, 1]^d$ .

Then there exists a constant  $C > 0$  such that, for  $n$  large enough,

$$E\left((\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2\right) \leq C (\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}.$$

The proof of Theorem 3.1 is based on a suitable decomposition of the pointwise mean squared error and sharp upper bounds using (a), (b), (c) and the calibration of the parameters in  $\hat{f}$ .

**Remark 3.1** (On the application  $\mathcal{T}$ ). *Examples of applications  $\mathcal{T} : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$  satisfying (3.2) are term-by-term thresholding rules. The most popular of them are the hard thresholding rule defined by  $\mathcal{T}(v, \lambda) = v \mathbf{1}_{\{|v| \geq \lambda\}}$ , where  $\mathbf{1}$  denotes the indicator function, the soft thresholding rule defined by  $\mathcal{T}(v, \lambda) = \text{sign}(v) \max(|v| - \lambda, 0)$ , where  $\text{sign}$  denotes the sign function and the non-negative garrote thresholding rule defined by  $\mathcal{T}(v, \lambda) = v \max(1 - \lambda^2/v^2, 0)$ . We refer to (Delyon and Juditsky, 1996, Lemma 1) for the technical details.*

For a wide variety of nonparametric models one can find  $(d, \lambda_n, \nu, \varrho, \omega)$  such that (a), (b) and (c) are satisfied. The most common configuration is

$$(d, \lambda_n, \nu, \varrho, \omega) = (1, \sqrt{\ln n/n}, 0, 0, 0).$$

It is the one considered in Delyon and Juditsky (1996) for Besov type errors. Then the obtained rate of convergence becomes  $(\ln n/n)^{2\alpha/(2\alpha+d)}$ , which is the optimal one in the minimax sense up to a logarithmic term for the standard white noise model (see Cai (2003) for  $d = 1$ ).

Thanks to the parameter  $\omega$ , our Theorem can be applied to some nonparametric inverse problems. It is often related to the smoothness of an auxiliary (known or unknown) function appearing in the model. See for instance Johnstone and Silverman (1997) for sequential inverse problems and Fan and Koo (2002) for the density deconvolution estimation problem.

The presence of the parameters  $\nu$  and  $\varrho$  is justified when we deal with dependent data. For the well-known  $\alpha$ -mixing case, see for instance Chesneau (2013, 2014) and Chesneau *et al.* (2015) where they play an important role for several intermediary results.

#### 4. Applications of Theorem 3.1

The interest of Theorem 3.1 is to provide new theoretical results on the rate of convergence related to the term-by-term thresholding wavelet estimators under the pointwise mean squared error. We illustrate this aspect by considering two well-known estimation problems: the regression function estimation problem in a dependent setting and the conditional density estimation problem.

##### 4.1. Regression function estimation in a dependent setting

**Model.** Let  $d$  be a positive integer,  $(\mathbf{Z}_t)_{t \in \mathbb{Z}} = ((Y_t, \mathbf{X}_t))_{t \in \mathbb{Z}}$  be a strictly stationary bivariate random process defined on the probability space  $(\mathbb{R} \times [0, 1]^d, \mathcal{B}(\mathbb{R} \times [0, 1]^d), P)$  where

$$Y_t = f(\mathbf{X}_t) + \epsilon_t, \quad t \in \mathbb{Z},$$

$(\mathbf{X}_t)_{t \in \mathbb{Z}}$  is a stationary random process following the uniform distribution on  $[0, 1]^d$ ,  $(\epsilon_t)_{t \in \mathbb{Z}}$  is a stationary random process with  $E(\epsilon_1) = 0$ , and  $f : [0, 1]^d \rightarrow \mathbb{R}$  is an unknown regression function. Moreover, it is understood that  $\epsilon_t$  is independent of  $X_t$ , for any  $t \in \mathbb{Z}$ . We aim to estimate  $f$  from  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$  in a dependent setting; we assume that  $(\mathbf{Z}_t)_{t \in \mathbb{Z}}$  is  $\alpha$ -mixing.

Numerous applications exist for this problem in dynamic economic systems and financial time series. We refer to Härdle (1990) and the references therein. Recent results on wavelet methods for this problem can be found in, e.g., Masry (2000), Patil and Truong (2001), Chaubey *et al.* (2013) and Chesneau (2013, 2014).

The contribution of our study is to prove that, under mild assumptions on the noise and the dependence structure, one can construct an adaptive multidimensional wavelet estimator which is efficient in terms of pointwise mean squared error properties.

**Definitions.** Let  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  be a strictly stationary random process. For  $j \in \mathbb{Z}$ , define the  $\sigma$ -fields  $\mathcal{F}_{-\infty, j}^{\mathbf{U}} = \sigma(\mathbf{U}_i, i \leq j)$  and  $\mathcal{F}_{j, \infty}^{\mathbf{U}} = \sigma(\mathbf{U}_i, i \geq j)$ . For any  $m \in \mathbb{Z}$ , we define the  $m$ -th  $\alpha$ -mixing coefficient of  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  by

$$\alpha_m = \sup_{(A, B) \in \mathcal{F}_{-\infty, 0}^{\mathbf{U}} \times \mathcal{F}_{m, \infty}^{\mathbf{U}}} |P(A \cap B) - P(A)P(B)|. \quad (4.1)$$

We say that  $(\mathbf{U}_t)_{t \in \mathbb{Z}}$  is  $\alpha$ -mixing if and only if  $\lim_{m \rightarrow \infty} \alpha_m = 0$ .

Full details on the  $\alpha$ -mixing dependence can be found in, e.g., Doukhan (1994) and Carrasco and Chen (2002).

**Assumptions.** We formulate the following assumptions.

(A1) There exist two constants  $\sigma > 0$  and  $\theta > 0$  such that, for any  $t \in \mathbb{R}$ ,

$$E(e^{t\epsilon_1}) \leq \theta e^{t^2 \sigma^2 / 2}.$$

(A2) There exists a known constant  $C > 0$  such that

$$\sup_{\mathbf{x} \in [0, 1]^d} |f(\mathbf{x})| \leq C.$$

(A3) For any  $m \in \mathbb{Z}$ , let  $g_{(X_0, X_m)}$  be the density of  $(\mathbf{X}_0, \mathbf{X}_m)$ . We suppose that there exists a known constant  $C > 0$  such that

$$\sup_{m \geq 1} \sup_{(\mathbf{x}, \mathbf{x}_*) \in [0, 1]^{2d}} g_{(X_0, X_m)}(\mathbf{x}, \mathbf{x}_*) \leq C.$$

(A4) There exist two constants  $a > 0$  and  $b > 0$  such that the  $m$ -th  $\alpha$ -mixing coefficient (4.1) of  $(\mathbf{Z}_t)_{t \in \mathbb{Z}}$  satisfies

$$\alpha_m \leq a e^{-bm}.$$

**Wavelet estimator.** We consider the following estimator  $\hat{f}$  for  $f$ :

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} \hat{c}_{j_0, \mathbf{k}} \Phi_{j_0, \mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \hat{d}_{j, \mathbf{k}, u} \mathbf{1}_{\{|\hat{d}_{j, \mathbf{k}, u}| \geq \kappa \sqrt{\ln n / n}\}} \Psi_{j, \mathbf{k}, u}(\mathbf{x}), \quad (4.2)$$

where

$$\hat{c}_{j_0, \mathbf{k}} = \frac{1}{n} \sum_{i=1}^n Y_i \Phi_{j_0, \mathbf{k}}(\mathbf{X}_i), \quad \hat{d}_{j, \mathbf{k}, u} = \frac{1}{n} \sum_{i=1}^n Y_i \Psi_{j, \mathbf{k}, u}(\mathbf{X}_i), \quad (4.3)$$

$\kappa$  is a large enough constant and  $j_0$  and  $j_1$  are integers such that

$$\frac{1}{2}2^{\tau d}(\ln n)^2 < 2^{j_0 d} \leq 2^{\tau d}(\ln n)^2, \quad \frac{1}{2} \frac{n}{(\ln n)^4} \leq 2^{j_1 d} \leq \frac{n}{(\ln n)^4}.$$

This estimator is adaptive; its construction does not depend on the smoothness of  $f$  in its construction. It is a particular case of (3.1).

**Result.** The following result determines the rate of convergence of  $\hat{f}$  under the pointwise mean squared error.

**Proposition 4.1.** *Let  $\hat{f}$  be defined by (4.2). Suppose that (A1)-(A4) hold and  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$  with  $M > 0$ ,  $\alpha \geq 0$  and  $\mathbf{x}_0 \in [0, 1]^d$ . Then there exists a constant  $C > 0$  such that, for  $n$  large enough,*

$$E \left( (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2\alpha/(2\alpha+d)}.$$

This result completes (Chesneau, 2013, Theorem 4.1) where the mean integrated squared error and Besov balls are considered.

#### 4.2. Conditional density estimation

**Model.** Let  $d$  and  $n$  be a positive integers,  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be  $n$  *i.i.d.* random vectors defined on the probability space  $([0, 1]^d, \mathcal{B}([0, 1]^d), P)$ . The density function of  $\mathbf{X}_1$  is given by  $f$ . Let  $d^* \in \{1, \dots, d-1\}$  and, for any  $i \in \{1, \dots, n\}$ , let  $\mathbf{X}_i^*$  be the first  $d^*$  components of  $\mathbf{X}_i$  and  $\mathbf{X}_i^o$  be the others  $d^o = d - d^*$  components. So  $\mathbf{X}_i = (\mathbf{X}_i^*, \mathbf{X}_i^o)$ . We define the conditional density function  $g$  by

$$g(\mathbf{x}) = f_{\mathbf{X}_1^*}(\mathbf{x}^* | \mathbf{X}_1^o = \mathbf{x}^o) = \frac{f(\mathbf{x})}{f_{\mathbf{X}_1^o}(\mathbf{x}^o)}, \quad \mathbf{x} = (\mathbf{x}^*, \mathbf{x}^o) \in [0, 1]^d, \quad (4.4)$$

where  $f_{\mathbf{X}_1^*}(\mathbf{x}^* | \mathbf{X}_1^o = \mathbf{x}^o)$  denotes the density function of  $\mathbf{X}_1^*$  conditionally to the event  $\{\mathbf{X}_1^o = \mathbf{x}^o\}$  and  $f_{\mathbf{X}_1^o}$  denotes the density function of  $\mathbf{X}_1^o$ . We aim to estimate  $g$  from  $\mathbf{X}_1, \dots, \mathbf{X}_n$ .

The literature about the conditional density estimation problem is very vast. We refer the reader to Akakpo and Lacour (2011), Le Penec and Cohen (2013), Chagny (2013) and the references therein. Our contribution to the subject is to provide an adaptive wavelet estimator which is efficient in a multidimensional setting and under the pointwise mean squared error.

**Assumptions.** We formulate the following assumptions.

(B1) There exists a known constant  $C > 0$  such that

$$\sup_{\mathbf{x} \in [0, 1]^d} f(\mathbf{x}) \leq C.$$

(B2) There exists a known constant  $c > 0$  such that

$$c \leq \inf_{\mathbf{x} \in [0, 1]^d} f(\mathbf{x}).$$



**Wavelet estimators.** We consider the following ratio estimator  $\hat{g}$  for  $g$ :

$$\hat{g}(\mathbf{x}) = \frac{\hat{f}(\mathbf{x})}{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}^o)} \mathbf{1}_{\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}^o) \geq c/2\}}, \quad \mathbf{x} = (\mathbf{x}^*, \mathbf{x}^o) \in [0, 1]^d, \quad (4.5)$$

where  $c$  refers to the constant in **(B2)**,

- the estimator  $\hat{f}$  is defined by

$$\hat{f}(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} \hat{c}_{j_0, \mathbf{k}} \Phi_{j_0, \mathbf{k}}(\mathbf{x}) + \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \hat{d}_{j, \mathbf{k}, u} \mathbf{1}_{\{|\hat{d}_{j, \mathbf{k}, u}| \geq \kappa \sqrt{\ln n/n}\}} \Psi_{j, \mathbf{k}, u}(\mathbf{x}),$$

where

$$\hat{c}_{j_0, \mathbf{k}} = \frac{1}{n} \sum_{i=1}^n \Phi_{j_0, \mathbf{k}}(\mathbf{X}_i), \quad \hat{d}_{j, \mathbf{k}, u} = \frac{1}{n} \sum_{i=1}^n \Psi_{j, \mathbf{k}, u}(\mathbf{X}_i), \quad (4.6)$$

$\kappa$  is a large enough constant and  $j_0$  and  $j_1$  are integers such that

$$\frac{1}{2} 2^{\tau d} < 2^{j_0 d} \leq 2^{\tau d}, \quad \frac{1}{2} \frac{n}{\ln n} \leq 2^{j_1 d} \leq \frac{n}{\ln n},$$

- the estimator  $\hat{f}_{\mathbf{X}_1^o}$  is the analog of  $\hat{f}$  but with  $d^o$  instead of  $d$  and  $\mathbf{X}_i^o$  instead of  $\mathbf{X}_i$ .

**Remark 4.1.** *The thresholding in (4.5) is to ensure that  $\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}^o)$  is large enough, and a fortiori, justified the ratio form of  $\hat{g}$ . This idea was recently developed by Vasiliev (2014) in a general context under mean integrated errors.*

The following result investigates the rate of convergence attained by  $\hat{g}$  under the pointwise mean squared error.

**Proposition 4.2.** *Let  $g$  be (4.4),  $\hat{g}$  be defined by (4.5) and  $\mathbf{x}_0 = (\mathbf{x}_0^*, \mathbf{x}_0^o) \in [0, 1]^d$ . Suppose that **(B1)** and **(B2)** hold and  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$  with  $M > 0$  and  $\alpha \geq 0$  and  $f_{\mathbf{X}_1^o} \in \Lambda_{d^o}^\beta(\mathbf{x}_0^o, M^o)$  with  $M^o > 0$  and  $\beta \geq 0$ . Then there exists a constant  $C > 0$  such that, for  $n$  large enough,*

$$E((\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0))^2) \leq C \left( \frac{\ln n}{n} \right)^{\min(2\alpha/(2\alpha+d), 2\beta/(2\beta+d^o))}.$$

The proof of Proposition 4.2 uses a suitable decomposition of the pointwise mean squared error of  $\hat{g}$  and Theorem 3.1 applied to  $\hat{f}$  and  $\hat{f}_{\mathbf{X}_1^o}$ . This result shows the consistence of  $\hat{g}$  and the influence of the smoothness of  $f$  and  $f_{\mathbf{X}_1^o}$  in the estimation of  $g$  by  $\hat{g}$ .

## 5. Proofs

In this section,  $C$  denotes any constant that does not depend on  $j$ ,  $\mathbf{k}$  and  $n$ . Its value may change from one term to another.

Proof of Theorem 3.1. Using the triangular inequality and the inequality:  $(x + y + z)^2 \leq 3(x^2 + y^2 + z^2)$ ,  $(x, y, z) \in \mathbb{R}^3$ , we obtain

$$E \left( (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2 \right) \leq 3(Q_1 + Q_2 + Q_3), \quad (5.1)$$

where

$$Q_1 = E \left( \left( \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} |\hat{c}_{j_0, \mathbf{k}} - c_{j_0, \mathbf{k}}| |\Phi_{j_0, \mathbf{k}}(\mathbf{x}_0)| \right)^2 \right),$$

$$Q_2 = E \left( \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} |\mathcal{T}(\hat{d}_{j, \mathbf{k}, u}, \kappa 2^{\omega j} \lambda_n) - d_{j, \mathbf{k}, u}| |\Psi_{j, \mathbf{k}, u}(\mathbf{x}_0)| \right)^2 \right)$$

and

$$Q_3 = \left( \sum_{u=1}^{2^d-1} \sum_{j=j_1+1}^{\infty} \sum_{\mathbf{k} \in \mathbf{D}_j} |d_{j, \mathbf{k}, u}| |\Psi_{j, \mathbf{k}, u}(\mathbf{x}_0)| \right)^2.$$

*Bound for  $Q_1$ :* Using the Cauchy-Schwarz inequality, **(a)**, the inequality  $\sum_{\mathbf{k} \in \mathbf{D}_{j_0}} |\Phi_{j_0, \mathbf{k}}(\mathbf{x}_0)| \leq C 2^{j_0 d/2}$  (since  $\text{Card}(\{\mathbf{k} \in \mathbf{D}_{j_0}; \Phi_{j_0, \mathbf{k}}(\mathbf{x}_0) \neq 0\}) \leq C$  and  $\sup_{\mathbf{x} \in [0,1]^d} |\Phi_{j_0, \mathbf{k}}(\mathbf{x})| \leq C 2^{j_0 d/2}$ ) and **(c)**, we have

$$\begin{aligned} Q_1 &\leq \left( \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} (E((\hat{c}_{j_0, \mathbf{k}} - c_{j_0, \mathbf{k}})^2))^{1/2} |\Phi_{j_0, \mathbf{k}}(\mathbf{x}_0)| \right)^2 \\ &\leq C \lambda_n^2 2^{2\omega j_0} \left( \sum_{\mathbf{k} \in \mathbf{D}_{j_0}} |\Phi_{j_0, \mathbf{k}}(\mathbf{x}_0)| \right)^2 \leq C \lambda_n^2 2^{j_0(d+2\omega)} \leq C \lambda_n^2 (\ln n)^\nu \\ &\leq C (\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}. \end{aligned} \quad (5.2)$$

*Bound for  $Q_2$ :* By the inequality:  $(x + y)^2 \leq 2(x^2 + y^2)$ ,  $(x, y) \in \mathbb{R}^2$ , and the definition of the term-by-term thresholding (3.2), we obtain

$$Q_2 \leq C(Q_{2,1} + Q_{2,2}), \quad (5.3)$$

where

$$Q_{2,1} = \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \min(|d_{j, \mathbf{k}, u}|, \kappa 2^{\omega j} \lambda_n) |\Psi_{j, \mathbf{k}, u}(\mathbf{x}_0)| \right)^2$$

and

$$Q_{2,2} = E \left( \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} |\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u}| \mathbf{1}_{\{|\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u}| \geq \kappa 2^{\omega j} \lambda_n / 2\}} |\Psi_{j, \mathbf{k}, u}(\mathbf{x}_0)| \right)^2 \right).$$

*Bound for  $Q_{2,1}$ :* Recall that  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$  implies that, for any  $\mathbf{k} \in \mathcal{K}_{u,j}(\mathbf{x}_0)$ ,  $|d_{j,\mathbf{k},u}| \leq M2^{-j(d/2+\alpha)}$ . Let  $j_2$  be an integer satisfying

$$\frac{1}{2} \left( \frac{1}{\lambda_n^2} \right)^{1/(2\alpha+2\omega+d)} < 2^{j_2} \leq \left( \frac{1}{\lambda_n^2} \right)^{1/(2\alpha+2\omega+d)}.$$

This inequality with:  $(x+y)^2 \leq 2(x^2+y^2)$ ,  $(x, y) \in \mathbb{R}^2$ , and  $\sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \leq C2^{jd/2}$  (since  $\text{Card}(\{\mathbf{k} \in \mathbf{D}_j; \Psi_{j,\mathbf{k},u}(\mathbf{x}_0) \neq 0\}) \leq C$  and  $\sup_{\mathbf{x} \in [0,1]^d} |\Psi_{j,\mathbf{k},u}(\mathbf{x})| \leq C2^{jd/2}$ ), yield

$$\begin{aligned} Q_{2,1} &\leq 2 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} \sum_{\mathbf{k} \in \mathbf{D}_j} \min(|d_{j,\mathbf{k},u}|, \kappa 2^{\omega j} \lambda_n) |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \\ &\quad + 2 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \min(|d_{j,\mathbf{k},u}|, \kappa 2^{\omega j} \lambda_n) |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \\ &\leq 2\kappa^2 \lambda_n^2 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_2} 2^{\omega j} \sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \\ &\quad + 2M^2 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_2+1}^{j_1} 2^{-j(d/2+\alpha)} \sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \\ &\leq C \left( \lambda_n^2 \left( \sum_{j=\tau}^{j_2} 2^{j(d+2\omega)/2} \right)^2 + \left( \sum_{j=j_2+1}^{\infty} 2^{-j\alpha} \right)^2 \right) \\ &\leq C \left( \lambda_n^2 2^{j_2(d+2\omega)} + 2^{-2j_2\alpha} \right) \leq C (\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}. \end{aligned}$$

*Bound for  $Q_{2,2}$ :* By the Cauchy-Schwarz inequality we get

$$Q_{2,2} \leq \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} \sum_{\mathbf{k} \in \mathbf{D}_j} \left( E \left( (\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u})^2 1_{\{|\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u}| \geq \kappa 2^{\omega j} \lambda_n / 2\}} \right) \right)^{1/2} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2.$$

It follows from the Cauchy-Schwarz inequality and **(b)** that

$$\begin{aligned} &\left( E \left( (\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u})^2 1_{\{|\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u}| \geq \kappa 2^{\omega j} \lambda_n / 2\}} \right) \right)^{1/2} \\ &\leq \left( E \left( (\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u})^4 \right) P \left( |\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u}| > \kappa 2^{\omega j} \lambda_n / 2 \right) \right)^{1/4} \\ &\leq C 2^{\omega j} \lambda_n^2. \end{aligned}$$

Owing to the inequality  $\sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \leq C2^{jd/2}$  and **(c)**, we have

$$\begin{aligned} Q_{2,2} &\leq C\lambda_n^4 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_0}^{j_1} 2^{\omega j} \sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \leq C\lambda_n^4 \left( \sum_{j=\tau}^{j_1} 2^{j(d+2\omega)/2} \right)^2 \\ &\leq C\lambda_n^4 2^{j_1(d+2\omega)} \leq C\lambda_n^4 \frac{1}{\lambda_n^2 (\ln n)^\varrho} \leq C\lambda_n^2 \leq C(\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}. \end{aligned} \quad (5.4)$$

Putting (5.3) and (5.4) together, we obtain

$$Q_2 \leq C(\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}. \quad (5.5)$$

*Bound for  $Q_3$ :* Owing to  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$ , again  $\sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \leq C2^{jd/2}$  and **(c)**, we have

$$\begin{aligned} Q_3 &\leq M^2 \left( \sum_{u=1}^{2^d-1} \sum_{j=j_1+1}^{\infty} 2^{-j(d/2+\alpha)} \sum_{\mathbf{k} \in \mathbf{D}_j} |\Psi_{j,\mathbf{k},u}(\mathbf{x}_0)| \right)^2 \leq C \left( \sum_{j=j_1+1}^{\infty} 2^{-j\alpha} \right)^2 \\ &\leq C2^{-2j_1\alpha} \leq C(\lambda_n^2 (\ln n)^\varrho)^{2\alpha/(d+2\omega)} \leq C(\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}. \end{aligned} \quad (5.6)$$

Combining (5.1), (5.2), (5.5) and (5.6), we prove that

$$E\left((\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2\right) \leq C(\lambda_n^2)^{2\alpha/(2\alpha+2\omega+d)}.$$

The proof of Theorem 3.1 is completed.  $\square$

*Proof of Proposition 4.1.* Let us investigate the assumptions **(a)**, **(b)** and **(c)** of Theorem 3.1 with the configuration  $(d, \lambda_n, \nu, \varrho, \omega) = (d, \sqrt{\ln n/n}, 2, 4, 0)$ . Using **(A1)**-**(A4)**, it follows from (Chesneau, 2013, Proposition 5.1) that

$$E((\hat{c}_{j_0,\mathbf{k}} - c_{j_0,\mathbf{k}})^2) \leq C\frac{1}{n} \leq C\frac{\ln n}{n}, \quad E((\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u})^4) \leq Cn$$

and, for  $\kappa$  large enough,

$$P\left(|\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^5}.$$

Hence

$$E((\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u})^4) P\left(|\hat{d}_{j,\mathbf{k},u} - d_{j,\mathbf{k},u}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}}\right) \leq C\frac{1}{n^4} \leq C\left(\frac{\ln n}{n}\right)^4.$$

So **(a)** and **(b)** are satisfied. Note that **(c)** holds since

$$\lim_{n \rightarrow \infty} (\ln n)^{\max(\nu, \varrho)} \left(\frac{\ln n}{n}\right)^{(1-\nu)} = 0$$

for any  $v \in [0, 1)$ . Theorem 3.1 can be applied with  $(d, \lambda_n, \nu, \varrho, \omega) = (d, \sqrt{\ln n/n}, 2, 4, 0)$  which yields the desired result: if  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$ , there exists a constant  $C > 0$  such that, for  $n$  large enough,

$$E \left( (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2\alpha/(2\alpha+d)}.$$

This ends the proof Proposition 4.1. □

Proof of Proposition 4.2. First of all, let us investigate the rates of convergence attained by  $\hat{f}$  and  $\hat{f}_{\mathbf{X}_1^\varrho}$  under the pointwise mean squared error. Let us investigate the assumptions **(a)**, **(b)** and **(c)** of Theorem 3.1 with the configuration  $(d, \lambda_n, \nu, \varrho, \omega) = (d, \sqrt{\ln n/n}, 0, 0, 0)$ . Using **(B1)**, similar arguments to (Donoho *et al.*, 1996, Section 5.1.1.) (i.e., the Rosenthal inequality and the Bernstein inequality) give

$$E((\hat{c}_{j_0, \mathbf{k}} - c_{j_0, \mathbf{k}})^2) \leq C \frac{1}{n} \leq C \frac{\ln n}{n}, \quad E((\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u})^4) \leq C \frac{1}{n^2}$$

and, for  $\kappa$  large enough,

$$P \left( |\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) \leq C \frac{1}{n^2}.$$

So

$$E((\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u})^4) P \left( |\hat{d}_{j, \mathbf{k}, u} - d_{j, \mathbf{k}, u}| \geq \frac{\kappa}{2} \sqrt{\frac{\ln n}{n}} \right) \leq C \frac{1}{n^4} \leq C \left( \frac{\ln n}{n} \right)^4.$$

So **(a)** and **(b)** are satisfied. Note that **(c)** holds since

$$\lim_{n \rightarrow \infty} (\ln n)^{\max(\nu, \varrho)} \left( \frac{\ln n}{n} \right)^{(1-\nu)} = 0$$

for any  $v \in [0, 1)$ . Therefore, we can apply Theorem 3.1 with  $(d, \lambda_n, \nu, \varrho, \omega) = (d, \sqrt{\ln n/n}, 0, 0, 0)$  which yields the desired result: if  $f \in \Lambda_d^\alpha(\mathbf{x}_0, M)$ , there exists a constant  $C > 0$  such that, for  $n$  large enough,

$$E \left( (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2\alpha/(2\alpha+d)}. \quad (5.7)$$

On the other hand, observe that **(B1)** imply  $\sup_{\mathbf{x} \in [0, 1]^{d^\circ}} f_{\mathbf{X}_1^\varrho}(\mathbf{x}) \leq C$ . Therefore, using similar arguments, one can prove **(a)**, **(b)** and **(c)** with the simple configuration  $(d, \lambda_n, \nu, \varrho, \omega) = (d^\circ, \sqrt{\ln n/n}, 0, 0, 0)$ . Owing to Theorem 3.1, if  $f_{\mathbf{X}_1^\varrho} \in \Lambda_{d^\circ}^\beta(\mathbf{x}_0^\circ, M^\circ)$ , then there exists a constant  $C > 0$  such that, for  $n$  large enough,

$$E \left( (\hat{f}_{\mathbf{X}_1^\varrho}(\mathbf{x}_0^\circ) - f_{\mathbf{X}_1^\varrho}(\mathbf{x}_0^\circ))^2 \right) \leq C \left( \frac{\ln n}{n} \right)^{2\beta/(2\beta+d^\circ)}. \quad (5.8)$$

Let us now determine the rate of convergence attained by  $\hat{g}$  under the pointwise mean squared error. We use the following natural decomposition:

$$\begin{aligned} \hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0) &= \frac{\hat{f}(\mathbf{x}_0)}{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o)} \mathbf{1}_{\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) \geq c/2\}} - \frac{f(\mathbf{x}_0)}{f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)} \\ &= \frac{1}{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)} \times \\ &\quad \left( f_{\mathbf{X}_1^o}(\mathbf{x}_0^o) (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0)) + f(\mathbf{x}_0) (f_{\mathbf{X}_1^o}(\mathbf{x}_0^o) - \hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o)) \right) \mathbf{1}_{\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) \geq c/2\}} \\ &\quad - \frac{f(\mathbf{x}_0)}{f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)} \mathbf{1}_{\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) < c/2\}}. \end{aligned}$$

Observe that **(B2)** gives  $\inf_{\mathbf{x} \in [0,1]^{d^o}} f_{\mathbf{X}_1^o}(\mathbf{x}) \geq c$  which implies  $\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) < c/2\} \subseteq \{|\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) - f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)| > c/2\}$  and, by the Markov inequality,

$$\mathbf{1}_{\{\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) < c/2\}} \leq \frac{2}{c} |\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) - f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)|.$$

The triangular inequality, the above inequality and the boundedness assumptions on the functions yield

$$|\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0)| \leq C(|\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0)| + |\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) - f_{\mathbf{X}_1^o}(\mathbf{x}_0^o)|).$$

By the inequality:  $(x + y)^2 \leq 2(x^2 + y^2)$ ,  $(x, y) \in \mathbb{R}^2$ , (5.7) and (5.8), we obtain

$$\begin{aligned} E((\hat{g}(\mathbf{x}_0) - g(\mathbf{x}_0))^2) &\leq C \left( E \left( (\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0))^2 \right) + E \left( (\hat{f}_{\mathbf{X}_1^o}(\mathbf{x}_0^o) - f_{\mathbf{X}_1^o}(\mathbf{x}_0^o))^2 \right) \right) \\ &\leq C \left( \left( \frac{\ln n}{n} \right)^{2\alpha/(2\alpha+d)} + \left( \frac{\ln n}{n} \right)^{2\beta/(2\beta+d^o)} \right) \\ &\leq C \left( \frac{\ln n}{n} \right)^{\min(2\alpha/(2\alpha+d), 2\beta/(2\beta+d^o))}. \end{aligned}$$

This ends the proof Proposition 4.2. □

**Acknowledgments.** We thank the three referees and an Associate Editor for thorough and useful comments which have significantly improved the presentation of the paper.

## References

Akakpo, N. and Lacour, C. (2011). Inhomogeneous and anisotropic conditional density estimation from dependent data, *Electron. Journal of Statistics*, 5, 1618-1653.

- Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society Series B*, 6, 97-144.
- Abramovich, F., Angelini, C. and De Canditiis, D. (2007). Pointwise optimality of Bayesian wavelet estimators, *Ann. Inst. Statist. Math.*, 59, 425-434
- Bochkina, N. and Sapatinas, T. (2006). On pointwise optimality of Bayes factor wavelet regression estimators, *Sankhya*, 68, 513-541.
- Bochkina, N. and Sapatinas, T. (2009). Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks, *Statistica Sinica*, 19, 1389-1406.
- Cai, T.T. and Brown, L.D. (1998). Wavelet shrinkage for nonequispaced samples, *The Annals of Statistics*, 26, 1783-1799.
- Cai, T.T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, 27, 898-924.
- Cai, T.T. (2002a). On block thresholding in wavelet regression: Adaptivity, block size, and threshold level, *Statistica Sinica*, 12, 1241-1273.
- Cai, T.T. (2002b). On adaptive wavelet estimation of a derivative and other related linear inverse problems, *J. Statistical Planning and Inference*, 108, 329-349.
- Cai, T.T. (2003). Rates of convergence and adaptation over Besov spaces under pointwise risk, *Statistica Sinica*, 13, 881-902.
- Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18, 17-39.
- Chagny, G. (2013). Warped bases for conditional density estimation, *Mathematical Methods of Statistics*, 22, 4, 253-282.
- Chaubey, Y.P., Chesneau, C. and Shirazi, E. (2013). Wavelet-based estimation of regression function for dependent biased data under a given random design, *Journal of Nonparametric Statistics*, 25, 1, 53-71.
- Chesneau, C. (2013). On the adaptive wavelet estimation of a multidimensional regression function under  $\alpha$ -mixing dependence: Beyond the standard assumptions on the noise, *Commentationes Mathematicae Universitatis Carolinae*, 4, 527-556.
- Chesneau, C. (2014). A general result on the mean integrated squared error of the hard thresholding wavelet estimator under  $\alpha$ -mixing dependence, *Journal of Probability and Statistics*, Volume 2014, Article ID 403764, 12 pages.
- Chesneau, C., Fadili, J. and Maillot, B. (2015). Adaptive estimation of an additive regression function from weakly dependent data, *Journal of Multivariate Analysis*, 133, 1, 77-94.
- Chicken, E. and Cai, T.T. (2005). Block thresholding for density estimation: Local and global adaptivity, *Journal of Multivariate Analysis*, 95, 76-106.
- Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis*, 24, 1, 54-81.
- Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM.
- Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, 3, 215-228.
- Donoho, D.L. and Johnstone, I.M., (1994). Ideal spatial adaptation by wavelet

- shrinkage, *Biometrika*, **81**, 425-455.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, **24**, 508-539.
- Doukhan, P. (1994). *Mixing. Properties and Examples*. Lecture Notes in Statistics 85. Springer Verlag, New York.
- Efromovich, S. (2002). On blockwise shrinkage estimation, Technical Report, Univ. New Mexico.
- Efromovich, S. (2005). On logarithmic penalty in adaptive pointwise estimation and blockwise wavelet shrinkage, *Journal of Computational and Graphical Statistics*, **14**, 1, 20-40.
- Fan, J. and Koo, J.Y. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, **48**, 734-747.
- Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators, *Statist. Sinica*, **9**, 33-50.
- Hall, P., Penev, S., Kerkyacharian, G. and Picard, D. (1997). Numerical performance of block thresholded wavelet estimators, *Statist. Comput.*, **7**, 115-124.
- Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York, **129**, Springer Verlag.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press. *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics New York, **129**, Springer Verlag.
- Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise, *Journal of the Royal Statistical Society, Series B, Methodological*, **59**, 319-351.
- Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, **9**, 2, 283-345.
- Le Pennec, E. and Cohen, S. (2013). Partition-based conditional density estimation, *ESAIM: Probability and Statistics*, eFirst.
- Mallat, S. (2009). *A wavelet tour of signal processing*, Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- Masry, E. (2000). Wavelet-Based estimation of multivariate regression functions in besov spaces, *Journal of Nonparametric Statistics*, **12**, 2, 283-308.
- Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.
- Patil, P.N. and Truong, Y.K. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series, *Annals of the Institute of Statistical Mathematics*, **53**, 1, 159-178.
- Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation, *Ann. Stat.*, **28**, 298-335.
- Tsybakov, A. (2004). *Introduction à l'estimation nonparamétrique*, Springer Verlag, Berlin.
- Vasiliev, V. (2014). A truncated estimation method with guaranteed accuracy, *Ann. Inst. Stat. Math*, **66**, 1, 141-163.



Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.