Adaptive wavelet estimation of a function in an indirect regression model

Christophe Chesneau

Noname manuscript No.
(will be inserted by the editor)

Adaptive wavelet estimation of a function in an indirect regression model

Christophe Chesneau

Received:

Abstract We consider a nonparametric regression model where m noise-perturbed functions $f_1, \ldots, f_m$ are randomly observed. For a fixed $\nu \in \{1, \ldots, m\}$, we want to estimate f_ν from the observations. To reach this goal, we develop an adaptive wavelet estimator based on a hard thresholding rule. Adopting the minimax approach under the mean integrated squared error over Besov balls, we prove that it attains a sharp rate of convergence.

Keywords indirect nonparametric regression · minimax estimation · Besov balls · wavelets · hard thresholding.

2000 Mathematics Subject Classification: 62G07, 62G20.

1 Motivations

An indirect nonparametric regression model is considered: we observe n independent pairs of random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ where, for any $i \in \{1, \ldots, n\}$,

$$Y_i = f_{V_i}(X_i) + \xi_i, \tag{1}$$

$V_1, \ldots, V_n$ are n unobserved independent discrete random variables such that, for any $i \in \{1, \ldots, n\}$, the set of possible values of V_i is

$$V_i(\Omega) = \{1, \ldots, m\}, \qquad m \in \mathbb{N}^*,$$

for any $d \in \{1, \ldots, m\}$, $f_d : [0, 1] \to \mathbb{R}$ is an unknown function, $X_1, \ldots, X_n$ are n i.i.d. random variables having the uniform distribution on $[0, 1]$ and $\xi_1, \ldots, \xi_n$ are n i.i.d. unobserved random variables such that

$$\mathbb{E}(\xi_1) = 0, \qquad \mathbb{E}(\xi_1^2) < \infty.$$

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

(The distribution of ξ_1 can be unknown). We suppose that $V_1, \ldots, V_n, X_1, \ldots, X_n$, $\xi_1, \ldots, \xi_n$ are independent. For a fixed $\nu \in \{1, \ldots, m\}$, we want to estimate f_ν from $(X_1, Y_1), \ldots, (X_n, Y_n)$. An application of this estimation problem is the following: m noise-perturbed signals $f_1, \ldots, f_m$ are randomly observed and only f_ν is of interest.

To estimate f_ν, various methods can be investigated (Kernel, Spline, ...) (see e.g. Prakasa Rao (1983, 1999) and Tsybakov (2004)). In this study, we focus our attention on the wavelet methods. They are attractive for nonparametric function estimation because of their spatial adaptivity, computational efficiency and asymptotic optimality properties. They can achieve near optimal convergence rates over a wide range of function classes (Besov balls, ...) and enjoy excellent mean integrated squared error (MISE) properties when used to estimate spatially inhomogeneous function. Details on the basics on wavelet methods in function estimation can be found in Antoniadis (1997) and Härdle et al (1998).

When (1) is considered with $V_1 = \ldots = V_n = 1$, it becomes the classical nonparametric regression model. In this case, to estimate $f_1 = f$, numerous wavelet methods have been developed. See e.g. Donoho and Johnstone (1994, 1995, 1998), Donoho et al (1995), Delyon and Juditsky (1996), Antoniadis et al. (1999), Cai and Brown (1999), Zhang and Zheng (1999), Cai (1999, 2002), Pensky and Vidakovic (2001), Chicken (2003), Kerkyacharian and Picard (2004), Chesneau (2007) and Pham Ngoc (2009). However, to the best of our knowledge, there is no adaptive wavelet estimator for f_ν in the general case.

In this paper, we develop an adaptive wavelet estimator for f_ν using the hard thresholding rule. It has the originality to combine an "observations thresholding technique" introduced by Delyon and Juditsky (1996) with some technical tools taking into account the distributions of $V_1, \ldots, V_n$. We evaluate its performance via the minimax approach under the MISE over Besov balls $B^s_{p,r}(M)$ (to be defined in Section 3). Under mild assumptions on the distributions of $V_1, \ldots, V_n$, we prove that our estimator attains the rate of convergence

$$v_n = \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)},$$

where z_n depends on the distributions of $V_1, \ldots, V_n$ (see (5)). This rate is "near optimal" in the sense that it is the one attained by the best nonadaptive linear wavelet estimator (the one which minimizes the MISE) up to a logarithmic term.

The paper is organized as follows. Assumptions on the model and some notations are introduced in Section 2. Section 3 briefly describes the wavelet basis on $[0, 1]$ and the Besov balls. The estimators are presented in Section 4. The results are set in Section 5. Section 6 is devoted to the proofs.

2 Assumptions

Additional assumptions on the model (1) are presented below.

Assumption on $(f_d)_{d\in\{1,\ldots,m\}}$**.** We suppose that there exists a known constant $C_* > 0$ such that

$$\sup_{d\in\{1,\ldots,m\}} \sup_{x\in[0,1]} |f_d(x)| \leq C_*. \tag{2}$$

Assumptions on $V_1, \ldots, V_n$**.** Recall that $V_1, \ldots, V_n$ are unobserved and, for any $i \in \{1, \ldots, n\}$, we know

$$w_d(i) = \mathbb{P}(V_i = d), \qquad d \in \{1, \ldots, m\}.$$

We suppose that the matrix

$$\Gamma_n = \left(\frac{1}{n}\sum_{i=1}^{n} w_k(i)w_\ell(i)\right)_{(k,\ell)\in\{1,\ldots,m\}^2}$$

is nonsingular i.e. $\det(\Gamma_n) > 0$. For the considered ν (the one which refers to the estimation of f_ν) and any $i \in \{1, \ldots, n\}$, we set

$$a_\nu(i) = \frac{1}{\det(\Gamma_n)} \sum_{k=1}^{m} (-1)^{k+\nu} \gamma_{\nu,k}^n w_k(i), \tag{3}$$

where $\gamma_{\nu,k}^n$ denotes the determinant of the minor (ν, k) of the matrix Γ_n. Then, for any $d \in \{1, \ldots, m\}$,

$$\frac{1}{n}\sum_{i=1}^{n} a_\nu(i)w_d(i) = \begin{cases} 1 & \text{if } d = \nu, \\ \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

and

$$(a_\nu(1), \ldots, a_\nu(n)) = \underset{(b_1,\ldots,b_n)\in\mathbb{R}^n}{\text{argmin}} \frac{1}{n}\sum_{i=1}^{n} b_i^2.$$

Technical details can be found in Maiboroda (1996).
We set

$$z_n = \frac{1}{n}\sum_{i=1}^{n} a_\nu^2(i) \tag{5}$$

and we suppose that $z_n < n/e$.

In nonparametric statistics, the sequence $(a_\nu(i))_{i\in\{1,\ldots,n\}}$ has ever been used in some mixture density estimation problems. See e.g. Maiboroda (1996), Pokhyl'ko (2005) and Prakasa Rao (2010).

3 Wavelets and Besov balls

Wavelet basis. Let $N \in \mathbb{N}^*$, ϕ be a father wavelet of a multiresolution analysis on $\mathbb{R}$ and ψ be the associated mother wavelet. Assume that

- $\text{supp}(\phi) = \text{supp}(\psi) = [1-N, N]$,
- $\int_{1-N}^{N} \phi(x)dx = 1$,
- for any $v \in \{0, \ldots, N-1\}$, $\int_{1-N}^{N} x^v \psi(x)dx = 0$.

For instance, the Daubechies wavelets satisfy these assumptions. Set

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

Then there exists an integer τ satisfying $2^\tau \geq 2N$ such that the collection

$$\mathcal{B} = \{\phi_{\tau,k}(.),\ k \in \{0, \ldots, 2^\tau - 1\};\ \psi_{j,k}(.);\ j \in \mathbb{N} - \{0, \ldots, \tau - 1\},\ k \in \{0, \ldots, 2^j - 1\}\},$$

(with an appropriate treatments at the boundaries) is an orthonormal basis of $\mathbb{L}^2([0,1])$, the set of square-integrable functions on $[0,1]$. We refer to Cohen (1993).

For any integer $\ell \geq \tau$, any $h \in \mathbb{L}^2([0,1])$ can be expanded on $\mathcal{B}$ as

$$h(x) = \sum_{k=0}^{2^\ell - 1} \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k=0}^{2^j - 1} \beta_{j,k}\psi_{j,k}(x),$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of h defined by

$$\alpha_{j,k} = \int_0^1 h(x)\phi_{j,k}(x)dx, \qquad \beta_{j,k} = \int_0^1 h(x)\psi_{j,k}(x)dx. \tag{6}$$

Besov balls. Let $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. A function h belongs to $B^s_{p,r}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that the associated wavelet coefficients (6) satisfy

$$\left(\sum_{j=\tau-1}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j - 1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*.$$

(We set $\beta_{\tau-1,k} = \alpha_{\tau,k}$). In this expression, s is a smoothness parameter and p and r are norm parameters. For a particular choice of s, p and r, $B^s_{p,r}(M)$ contain the Hölder and Sobolev balls. See Meyer (1990).

4 Estimators

Wavelet coefficient estimators. The first step to estimate f_ν consists in expanding f_ν on $\mathcal{B}$ and estimating its unknown wavelet coefficients.

For any integer $j \geq \tau$ and any $k \in \{0, \ldots, 2^j - 1\}$,

– we estimate $\alpha_{j,k} = \int_0^1 f_\nu(x)\phi_{j,k}(x)dx$ by

$$\widehat{\alpha}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} a_\nu(i)Y_i\phi_{j,k}(X_i), \tag{7}$$

– we estimate $\beta_{j,k} = \int_0^1 f_\nu(x)\psi_{j,k}(x)dx$ by

$$\widehat{\beta}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Z_i \mathbf{1}_{\{|Z_i|\leq\gamma_n\}}, \tag{8}$$

where, for any $i \in \{1,\dots,n\}$,

$$Z_i = a_\nu(i)Y_i\psi_{j,k}(X_i),$$

$a_\nu(i)$ is defined by (3), for any random event $\mathcal{A}$, $\mathbf{1}_{\mathcal{A}}$ is the indicator function on $\mathcal{A}$, the threshold γ_n is defined by

$$\gamma_n = \theta\sqrt{\frac{nz_n}{\ln(n/z_n)}}, \tag{9}$$

z_n is defined by (5), $\theta = \sqrt{C_*^2 + \mathbb{E}(\xi_1^2)}$ and C_* is the one in (2).

Remark 1. Mention that $\widehat{\alpha}_{j,k}$ is an unbiased estimator of $\alpha_{j,k}$, whereas $\widehat{\beta}_{j,k}$ is not an unbiased estimator of $\beta_{j,k}$. However $(1/n)\sum_{i=1}^{n} Z_i$ is an unbiased estimator of $\beta_{j,k}$. The proofs are given in (14) and (19).

Remark 2. The "observations thresholding technique" used in (8) has been firstly introduced by Delyon and Juditsky (1996) for (1) in the classical case (i.e. $V_1 = \ldots = V_n = 1$). In our general setting, this allows us to provide a good estimator of $\beta_{j,k}$ under mild assumptions on

– $(a_\nu(i))_{i\in\{1,\dots,n\}}$ and a fortiori the distributions of $V_1,\dots,V_n$ (only $z_n < n/e$ is required),
– $\xi_1,\dots,\xi_n$ (only finite moments of order 2 are required).

Linear estimator. Assuming that $f_\nu \in B^s_{p,r}(M)$ with $p \geq 2$, we define the linear estimator $\widehat{f}^L$ by

$$\widehat{f}^L(x) = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(x), \tag{10}$$

where $\widehat{\alpha}_{j,k}$ is defined by (7) and j_0 is the integer satisfying

$$\frac{1}{2}\left(\frac{n}{z_n}\right)^{1/(2s+1)} < 2^{j_0} \leq \left(\frac{n}{z_n}\right)^{1/(2s+1)}.$$

The definition of j_0 is chosen to minimize the MISE of $\widehat{f}^L$. Note that it is not adaptive since it depends on s, the smoothness parameter of f_ν.

Hard thresholding estimator. We define the hard thresholding estimator $\widehat{f}^H$ by

$$\widehat{f}^H(x) = \sum_{k=0}^{2^\tau-1} \widehat{\alpha}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \widehat{\beta}_{j,k}\mathbf{1}_{\left\{|\widehat{\beta}_{j,k}|\geq \kappa\lambda_n\right\}}\psi_{j,k}(x), \tag{11}$$

where $\widehat{\alpha}_{j,k}$ is defined by (7), $\widehat{\beta}_{j,k}$ by (8), j_1 is the integer satisfying

$$\frac{n}{2z_n} < 2^{j_1} \leq \frac{n}{z_n},$$

$\kappa \geq 8/3 + 2 + 2\sqrt{16/9+4}$ and λ_n is the threshold

$$\lambda_n = \theta\sqrt{\frac{z_n \ln(n/z_n)}{n}}. \tag{12}$$

Further details on the hard thresholding wavelet estimator for the standard nonparametric regression model can be found in Donoho and Johnstone (1994, 1995, 1998) and Delyon and Juditsky (1996).
Note that the choice of γ_n in (9) depends on λ_n in (12): we have $\lambda_n = \theta^2 z_n/\gamma_n$. The definitions of γ_n and λ_n are based on theoretical considerations.

5 Results

Theorem 1 *Consider* (1) *under the assumptions of Section* 2. *Suppose that* $f_\nu \in B^s_{p,r}(M)$ *with* $s > 0$, $p \geq 2$ *and* $r \geq 1$. *Let* $\widehat{f}^L$ *be* (10). *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^L(x) - f_\nu(x)\right)^2 dx\right) \leq C\left(\frac{z_n}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 1 uses moment inequalities on (7) and (8), and a suitable decomposition of the MISE.

Due to our weak assumptions on $V_1, \ldots, V_n, \xi_1, \ldots, \xi_n$, the optimal lower bound of (1) seems difficult to determine (see Tsybakov (2004)). However, since $\widehat{f}^L$ is constructed to be the linear estimator which optimizes the MISE, our benchmark will be the rate of convergence $v_n = (z_n/n)^{2s/(2s+1)}$.

Remark that, in the case $V_1 = \ldots = V_n = 1$ and $\xi_1 \sim \mathcal{N}(0,1)$, we have $z_n = 1$ and v_n $(= n^{-2s/(2s+1)})$ is the optimal rate of convergence (see Tsybakov (2004)).

Theorem 2 *Consider* (1) *under the assumptions of Section* 2. *Let* $\widehat{f}^H$ *be* (11). *Suppose that* $f_\nu \in B^s_{p,r}(M)$ *with* $r \geq 1$, $\{p \geq 2$ *and* $s > 0\}$ *or* $\{p \in [1,2)$ *and* $s > 1/p\}$. *Then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^H(x) - f_\nu(x)\right)^2 dx\right) \leq C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 2 is based on several probability results (moment inequalities, concentration inequality,...) and a suitable decomposition of the MISE.

Theorem 2 proves that $\widehat{f}^H$ attains $v_n = (z_n/n)^{2s/(2s+1)}$ up to the logarithmic term $(\ln(n/z_n))^{2s/(2s+1)}$.

Naturally, when $V_1 = \ldots = V_n = 1$ and $\xi_1 \sim \mathcal{N}(0,1)$, $\widehat{f}^H$ attains the same rate of convergence than the standard hard thresholding estimator adapted to the classical nonparametric regression model (see Donoho and Johnstone (1994, 1995, 1998)). And this one is optimal in the minimax sense up to a logarithmic term.

Conclusions and perspectives. We construct an adaptive wavelet estimator to estimate the function f_ν from the sophisticated regression model (1). Under mild assumptions, we prove that it attains a sharp rate of convergence for a wide class of functions.

Possible perspectives are to

- investigate the estimation of f in (1) when X_1 has a more complex distribution than the random uniform one. In this case, the warped wavelet basis introduced in the nonparametric regression estimation by Kerkyacharian and Picard (2004) seems to be an adapted powerful tool.
- consider the case where the distributions of $V_1, \ldots, V_n$ are unknown.
- potentially improve the estimation of f_ν (and remove the extra logarithmic term). The thresholding rule named BlockJS developed in wavelet estimation by Cai (1999, 2002) seems to be a good alternative.

All these aspects need further investigations that we leave for a future work.

6 Proofs

In this section, we consider (1) under the assumptions of Section 2. Moreover, C represents a positive constant which may differ from one term to another.

6.1 Auxiliary results

Proposition 1 *For any integer $j \geq \tau$ and any $k \in \{0, \ldots, 2^j - 1\}$, let $\alpha_{j,k}$ be the wavelet coefficient* (6) *of f_ν and $\widehat{\alpha}_{j,k}$ be* (7)*. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left((\widehat{\alpha}_{j,k} - \alpha_{j,k})^2\right) \leq C\frac{z_n}{n}.$$

Proof of Proposition 1. First of all, we prove that $\widehat{\alpha}_{j,k}$ is an unbiased estimator of $\alpha_{j,k}$. For any $i \in \{1, \ldots, n\}$, set

$$W_i = a_\nu(i) Y_i \phi_{j,k}(X_i).$$

Since X_i, V_i and ξ_i are independent, and $\mathbb{E}(\xi_i) = 0$, we have

$$\begin{aligned}
\mathbb{E}(W_i) &= \mathbb{E}(a_\nu(i) Y_i \phi_{j,k}(X_i)) = \mathbb{E}(a_\nu(i)(f_{V_i}(X_i) + \xi_i)\phi_{j,k}(X_i)) \\
&= a_\nu(i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}(X_i)) + a_\nu(i)\mathbb{E}(\xi_i)\mathbb{E}(\phi_{j,k}(X_i)) \\
&= a_\nu(i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}(X_i)) \\
&= a_\nu(i)\sum_{d=1}^{m} w_d(i)\int_0^1 f_d(x)\phi_{j,k}(x)dx. \qquad (13)
\end{aligned}$$

It follows from (13) and (4) that

$$\begin{aligned}
\mathbb{E}(\widehat{\alpha}_{j,k}) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(W_i) = \frac{1}{n}\sum_{i=1}^{n}\left(a_\nu(i)\sum_{d=1}^{m} w_d(i)\int_0^1 f_d(x)\phi_{j,k}(x)dx\right) \\
&= \sum_{d=1}^{m}\int_0^1 f_d(x)\phi_{j,k}(x)dx\left(\frac{1}{n}\sum_{i=1}^{n} a_\nu(i)w_d(i)\right) \\
&= \int_0^1 f_\nu(x)\phi_{j,k}(x)dx = \alpha_{j,k}. \qquad (14)
\end{aligned}$$

So $\widehat{\alpha}_{j,k}$ is an unbiased estimator of $\alpha_{j,k}$. Therefore

$$\begin{aligned}
\mathbb{E}\left((\widehat{\alpha}_{j,k} - \alpha_{j,k})^2\right) &= \mathbb{V}\left(\widehat{\alpha}_{j,k}\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} W_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(W_i) \\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left(W_i^2\right). \qquad (15)
\end{aligned}$$

For any $i \in \{1, \ldots, n\}$, we have

$$\mathbb{E}\left(W_i^2\right) = \mathbb{E}(a_\nu^2(i) Y_i^2 \phi_{j,k}^2(X_i)) = a_\nu^2(i)\mathbb{E}\left((f_{V_i}(X_i) + \xi_i)^2\phi_{j,k}^2(X_i)\right). \qquad (16)$$

Since X_i, V_i and ξ_i are independent, $\mathbb{E}\left(\phi_{j,k}^2(X_i)\right) = \int_0^1 \phi_{j,k}^2(x)dx = 1$ and, by (2), $\sup_{d\in\{1,\ldots,m\}}\sup_{x\in[0,1]}|f_d(x)| \leq C_*$, we have

$$\begin{aligned}
&\mathbb{E}\left((f_{V_i}(X_i) + \xi_i)^2\phi_{j,k}^2(X_i)\right) \\
&= \mathbb{E}\left(f_{V_i}^2(X_i)\phi_{j,k}^2(X_i)\right) + 2\mathbb{E}(\xi_i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}^2(X_i)) + \mathbb{E}\left(\xi_i^2\right)\mathbb{E}\left(\phi_{j,k}^2(X_i)\right) \\
&= \mathbb{E}\left(f_{V_i}^2(X_i)\phi_{j,k}^2(X_i)\right) + \mathbb{E}\left(\xi_1^2\right) \leq C_*^2\mathbb{E}\left(\phi_{j,k}^2(X_i)\right) + \mathbb{E}\left(\xi_1^2\right) \\
&= C_*^2 + \mathbb{E}\left(\xi_1^2\right) = \theta^2. \qquad (17)
\end{aligned}$$

Putting (16) and (17) together, we obtain

$$\mathbb{E}\left(W_i^2\right) \leq \theta^2 a_\nu^2(i). \qquad (18)$$

It follows from (15) and (18) that

$$\mathbb{E}\left((\widehat{\alpha}_{j,k} - \alpha_{j,k})^2\right) \leq \frac{1}{n}\left(\theta^2\frac{1}{n}\sum_{i=1}^{n} a_\nu^2(i)\right) = C\frac{z_n}{n}.$$

□

Proposition 2 *For any integer $j \geq \tau$ and any $k \in \{0, \dots, 2^j - 1\}$, let $\beta_{j,k}$ be the wavelet coefficient* (6) *of f_ν and $\widehat{\beta}_{j,k}$ be* (8)*. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right) \leq C\frac{(z_n \ln(n/z_n))^2}{n^2}.$$

Proof of Proposition 2. Taking ψ instead of ϕ in (14), we obtain

$$\begin{aligned}\beta_{j,k} &= \int_0^1 f_\nu(x)\psi_{j,k}(x)dx = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(Z_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Z_i\mathbf{1}_{\{|Z_i|\leq\gamma_n\}}) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Z_i\mathbf{1}_{\{|Z_i|>\gamma_n\}}). \qquad (19)\end{aligned}$$

Therefore, by the elementary inequality $(x+y)^4 \leq 8(x^4+y^4)$, $(x,y) \in \mathbb{R}^2$, we have

$$\begin{aligned}&\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right) \\ &= \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(Z_i\mathbf{1}_{\{|Z_i|\leq\gamma_n\}} - \mathbb{E}\left(Z_i\mathbf{1}_{\{|Z_i|\leq\gamma_n\}}\right)\right) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(Z_i\mathbf{1}_{\{|Z_i|>\gamma_n\}}\right)\right)^4\right) \\ &\leq 8(A+B), \qquad (20)\end{aligned}$$

where

$$A = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(Z_i\mathbf{1}_{\{|Z_i|\leq\gamma_n\}} - \mathbb{E}(Z_i\mathbf{1}_{\{|Z_i|\leq\gamma_n\}})\right)\right)^4\right)$$

and

$$B = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(|Z_i|\mathbf{1}_{\{|Z_i|>\gamma_n\}})\right)^4.$$

Let us bound A and B, in turn.

Upper bound for A. Let us present the Rosenthal inequality (see Rosenthal (1970)).

Lemma 1 (Rosenthal's inequality) *Let $n \in \mathbb{N}^*$, $p \geq 2$ and $(U_i)_{i\in\{1,\dots,n\}}$ be n zero mean independent random variables such that $\sup_{i\in\{1,\dots,n\}}\mathbb{E}(|U_i|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n}U_i\right|^p\right) \leq C\max\left(\sum_{i=1}^{n}\mathbb{E}\left(|U_i|^p\right), \left(\sum_{i=1}^{n}\mathbb{E}\left(U_i^2\right)\right)^{p/2}\right).$$

Applying the Rosenthal inequality with $p = 4$ and, for any $i \in \{1, \ldots, n\}$,

$$U_i = Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} - \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}}\right),$$

we obtain

$$A = \frac{1}{n^4}\mathbb{E}\left(\left(\sum_{i=1}^n U_i\right)^4\right) \le C\frac{1}{n^4}\max\left(\sum_{i=1}^n \mathbb{E}\left(U_i^4\right), \left(\sum_{i=1}^n \mathbb{E}\left(U_i^2\right)\right)^2\right).$$

Using (18) (with ψ instead of ϕ), we have, for any $a \in \{2, 4\}$ and any $i \in \{1, \ldots, n\}$,

$$\mathbb{E}\left(U_i^a\right) \le 2^a \mathbb{E}\left(Z_i^a \mathbf{1}_{\{|Z_i| \le \gamma_n\}}\right) \le 2^a \gamma_n^{a-2}\mathbb{E}\left(Z_i^2\right) \le 2^a \gamma_n^{a-2}\theta^2 a_\nu^2(i).$$

Hence, using $z_n < n/e$,

$$\begin{aligned} A &\le C\frac{1}{n^4}\max\left(\gamma_n^2 \sum_{i=1}^n a_\nu^2(i), \left(\sum_{i=1}^n a_\nu^2(i)\right)^2\right) \\ &= C\frac{1}{n^4}\max\left(\frac{n^2}{\ln(n/z_n)}z_n^2, n^2 z_n^2\right) = C\frac{z_n^2}{n^2}. \end{aligned} \tag{21}$$

Upper bound for B. Using again (18) (with ψ instead of ϕ), for any $i \in \{1, \ldots, n\}$, we obtain

$$\mathbb{E}\left(|Z_i|\mathbf{1}_{\{|Z_i| > \gamma_n\}}\right) \le \frac{\mathbb{E}\left(Z_i^2\right)}{\gamma_n} \le \frac{1}{\theta}\sqrt{\frac{\ln(n/z_n)}{nz_n}}\theta^2 a_\nu^2(i) = \theta\sqrt{\frac{\ln(n/z_n)}{nz_n}}a_\nu^2(i).$$

Therefore

$$\begin{aligned} B &= \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}(|Z_i|\mathbf{1}_{\{|Z_i| > \gamma_n\}})\right)^4 \le \theta^4 \frac{(\ln(n/z_n))^2}{n^2 z_n^2}\left(\frac{1}{n}\sum_{i=1}^n a_\nu^2(i)\right)^4 \\ &= \theta^4 \frac{(\ln(n/z_n))^2}{n^2 z_n^2} z_n^4 = \theta^4 \frac{(z_n \ln(n/z_n))^2}{n^2}. \end{aligned} \tag{22}$$

Combining (20), (21) and (22) and using $z_n < n/e$, we have

$$\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right) \le C\left(\frac{1}{n^2}z_n^2 + \frac{(z_n \ln(n/z_n))^2}{n^2}\right) \le C\frac{(z_n \ln(n/z_n))^2}{n^2}.$$

□

Proposition 3 *For any integer $j \ge \tau$ and any $k \in \{0, \ldots, 2^j - 1\}$, let $\beta_{j,k}$ be the wavelet coefficient* (6) *of f_ν and $\widehat{\beta}_{j,k}$ be* (8)*. Then, for any $\kappa \ge 8/3 + 2 + 2\sqrt{16/9 + 4}$,*

$$\mathbb{P}\left(|\widehat{\beta}_{j,k} - \beta_{j,k}| \ge \kappa\lambda_n/2\right) \le 2\left(\frac{z_n}{n}\right)^2.$$

Proof of Proposition 3. By (19) we have

$$
\begin{aligned}
&|\widehat{\beta}_{j,k} - \beta_{j,k}| \\
&= \left| \frac{1}{n} \sum_{i=1}^{n} \left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} - \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \right) - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| > \gamma_n\}} \right) \right| \\
&\le \left| \frac{1}{n} \sum_{i=1}^{n} \left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} - \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \right) \right| + \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}} \right).
\end{aligned}
$$

Using (18) (with ψ instead of ϕ), we obtain

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(|Z_i| \mathbf{1}_{\{|Z_i| > \gamma_n\}} \right) &\le \frac{1}{\gamma_n} \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left(Z_i^2 \right) \right) \le \frac{1}{\gamma_n} \left(\theta^2 \frac{1}{n} \sum_{i=1}^{n} a_\nu^2(i) \right) \\
&= \frac{1}{\gamma_n} \theta^2 z_n = \frac{1}{\theta} \sqrt{\frac{\ln(n/z_n)}{n z_n}} \theta^2 z_n \\
&= \theta \sqrt{\frac{z_n \ln(n/z_n)}{n}} = \lambda_n.
\end{aligned}
$$

Hence

$$
\begin{aligned}
S &= \mathbb{P}\left(|\widehat{\beta}_{j,k} - \beta_{j,k}| \ge \kappa \lambda_n / 2 \right) \\
&\le \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} - \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \right) \right| \ge (\kappa/2 - 1) \lambda_n \right).
\end{aligned}
$$

Now we need the Bernstein inequality presented in the lemma below (see Petrov (1995)).

Lemma 2 (Bernstein's inequality) *Let $n \in \mathbb{N}^*$ and $(U_i)_{i \in \{1,\dots,n\}}$ be n zero mean independent random variables such that there exists a constant $M > 0$ satisfying $\sup_{i \in \{1,\dots,n\}} |U_i| \le M < \infty$. Then, for any $\lambda > 0$,*

$$
\mathbb{P}\left(\left| \sum_{i=1}^{n} U_i \right| \ge \lambda \right) \le 2 \exp\left(- \frac{\lambda^2}{2 \left(\sum_{i=1}^{n} \mathbb{E}\left(U_i^2 \right) + \lambda M / 3 \right)} \right).
$$

Let us set, for any $i \in \{1, \dots, n\}$,

$$
U_i = Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} - \mathbb{E}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right).
$$

For any $i \in \{1, \dots, n\}$, we have $\mathbb{E}(U_i) = 0$,

$$
|U_i| \le |Z_i| \, \mathbf{1}_{\{|Z_i| \le \gamma_n\}} + \mathbb{E}\left(|Z_i| \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \le 2\gamma_n
$$

and, using again (18) (with ψ instead of ϕ),

$$
\mathbb{E}\left(U_i^2 \right) = \mathbb{V}\left(Z_i \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \le \mathbb{E}\left(Z_i^2 \mathbf{1}_{\{|Z_i| \le \gamma_n\}} \right) \le \mathbb{E}\left(Z_i^2 \right) \le \theta^2 a_\nu^2(i).
$$

So
$$\sum_{i=1}^{n} \mathbb{E}\left(U_i^2\right) \le \theta^2 \sum_{i=1}^{n} a_\nu^2(i) = \theta^2 n z_n.$$
It follows from the Bernstein inequality that
$$S \le 2\exp\left(-\frac{n^2(\kappa/2-1)^2\lambda_n^2}{2\left(\theta^2 n z_n + 2n(\kappa/2-1)\lambda_n\gamma_n/3\right)}\right).$$
Since
$$\lambda_n\gamma_n = \theta\sqrt{\frac{z_n \ln(n/z_n)}{n}}\theta\sqrt{\frac{nz_n}{\ln(n/z_n)}} = \theta^2 z_n, \qquad \lambda_n^2 = \theta^2 \frac{z_n \ln(n/z_n)}{n},$$
we have, for any $\kappa \ge 8/3 + 2 + 2\sqrt{16/9+4}$,
$$S \le 2\exp\left(-\frac{(\kappa/2-1)^2 \ln(n/z_n)}{2\left(1+2(\kappa/2-1)/3\right)}\right) = 2\left(\frac{n}{z_n}\right)^{-\frac{(\kappa/2-1)^2}{2(1+2(\kappa/2-1)/3)}} \le 2\left(\frac{z_n}{n}\right)^2.$$
□

6.2 Proofs of the main results

Proof of Theorem 1. We expand the function f_ν on $\mathcal{B}$ as
$$f_\nu(x) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x).$$
We have
$$\widehat{f}^L(x) - f_\nu(x) = \sum_{k=0}^{2^{j_0}-1} \left(\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}\right)\phi_{j_0,k}(x) - \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x).$$
Hence
$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^L(x) - f_\nu(x)\right)^2 dx\right) = A + B,$$
where
$$A = \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left(\left(\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}\right)^2\right), \qquad B = \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}^2.$$
Proposition 1 gives
$$A \le 2^{j_0}\frac{z_n}{n} \le \left(\frac{z_n}{n}\right)^{2s/(2s+1)}.$$
Since $p \ge 2$, we have $B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$. Hence
$$B \le C2^{-2j_0 s} \le C\left(\frac{z_n}{n}\right)^{2s/(2s+1)}.$$

So

$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^L(x) - f_\nu(x)\right)^2 dx\right) \le C\left(\frac{z_n}{n}\right)^{2s/(2s+1)}.$$

The proof of Theorem 1 is complete.

□

Proof of Theorem 2. We expand the function f_ν on $\mathcal{B}$ as

$$f_\nu(x) = \sum_{k=0}^{2^\tau-1} \alpha_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x).$$

We have

$$\begin{aligned}
&\widehat{f}^H(x) - f_\nu(x)\\
&= \sum_{k=0}^{2^\tau-1} (\widehat{\alpha}_{\tau,k} - \alpha_{\tau,k})\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \left(\widehat{\beta}_{j,k}\mathbf{1}_{\left\{|\widehat{\beta}_{j,k}|\ge\kappa\lambda_n\right\}} - \beta_{j,k}\right)\psi_{j,k}(x)\\
&- \sum_{j=j_1+1}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}(x).
\end{aligned}$$

Hence

$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^H(x) - f_\nu(x)\right)^2 dx\right) = R + S + T, \tag{23}$$

where

$$R = \sum_{k=0}^{2^\tau-1} \mathbb{E}\left((\widehat{\alpha}_{\tau,k} - \alpha_{\tau,k})^2\right), \qquad S = \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \mathbb{E}\left(\left(\widehat{\beta}_{j,k}\mathbf{1}_{\left\{|\widehat{\beta}_{j,k}|\ge\kappa\lambda_n\right\}} - \beta_{j,k}\right)^2\right)$$

and

$$T = \sum_{j=j_1+1}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}^2.$$

Let us bound R, T and S, in turn.

By Proposition 1 and the inequalities: $z_n < n/e$, $z_n \ln(n/z_n) < n$ and $2s/(2s+1) < 1$, we have

$$R \le C\frac{z_n}{n} \le C\frac{z_n \ln(n/z_n)}{n} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \tag{24}$$

For $r \geq 1$ and $p \geq 2$, we have $B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$. Using $z_n < n/e$, $z_n \ln(n/z_n) < n$ and $2s/(2s+1) < 2s$, we obtain

$$T \leq C \sum_{j=j_1+1}^{\infty} 2^{-2js} \leq C 2^{-2j_1 s} \leq C \left(\frac{n}{z_n}\right)^{-2s} \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s} \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}.$$

For $r \geq 1$ and $p \in [1, 2)$, we have $B^s_{p,r}(M) \subseteq B^{s+1/2-1/p}_{2,\infty}(M)$. Since $s > 1/p$, we have $s + 1/2 - 1/p > s/(2s+1)$. So, by $z_n < n/e$ and $z_n \ln(n/z_n) < n$, we have

$$T \leq C \sum_{j=j_1+1}^{\infty} 2^{-2j(s+1/2-1/p)} \leq C 2^{-2j_1(s+1/2-1/p)} \leq C \left(\frac{n}{z_n}\right)^{-2(s+1/2-1/p)} \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2(s+1/2-1/p)} \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}.$$

Hence, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$, we have

$$T \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \tag{25}$$

The term S can be decomposed as

$$S = S_1 + S_2 + S_3 + S_4, \tag{26}$$

where

$$S_1 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2 \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}| \geq \kappa\lambda_n\right\}} \mathbf{1}_{\{|\beta_{j,k}| < \kappa\lambda_n/2\}}\right),$$

$$S_2 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2 \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}| \geq \kappa\lambda_n\right\}} \mathbf{1}_{\{|\beta_{j,k}| \geq \kappa\lambda_n/2\}}\right),$$

$$S_3 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(\beta^2_{j,k} \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}| < \kappa\lambda_n\right\}} \mathbf{1}_{\{|\beta_{j,k}| \geq 2\kappa\lambda_n\}}\right)$$

and

$$S_4 = \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(\beta^2_{j,k} \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}| < \kappa\lambda_n\right\}} \mathbf{1}_{\{|\beta_{j,k}| < 2\kappa\lambda_n\}}\right).$$

Upper bounds for S_1 and S_3. We have

$$\left\{|\widehat{\beta}_{j,k}| < \kappa\lambda_n,\ |\beta_{j,k}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa\lambda_n/2\right\},$$

$$\left\{|\widehat{\beta}_{j,k}| \geq \kappa\lambda_n,\ |\beta_{j,k}| < \kappa\lambda_n/2\right\} \subseteq \left\{|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa\lambda_n/2\right\}$$

and

$$\left\{|\widehat{\beta}_{j,k}| < \kappa\lambda_n,\ |\beta_{j,k}| \geq 2\kappa\lambda_n\right\} \subseteq \left\{|\beta_{j,k}| \leq 2|\widehat{\beta}_{j,k} - \beta_{j,k}|\right\}.$$

So

$$\max(S_1, S_3) \leq C \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2 \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}-\beta_{j,k}|>\kappa\lambda_n/2\right\}}\right).$$

It follows from the Cauchy-Schwarz inequality and Propositions 2 and 3 that

$$\begin{aligned}
&\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2 \mathbf{1}_{\left\{|\widehat{\beta}_{j,k}-\beta_{j,k}|>\kappa\lambda_n/2\right\}}\right) \\
&\leq \left(\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right)\right)^{1/2} \left(\mathbb{P}\left(|\widehat{\beta}_{j,k} - \beta_{j,k}| > \kappa\lambda_n/2\right)\right)^{1/2} \\
&\leq C\frac{z_n^2 \ln(n/z_n)}{n^2}.
\end{aligned}$$

Hence, using $z_n < n/e$, $z_n \ln(n/z_n) < n$ and $2s/(2s+1) < 1$, we have

$$\begin{aligned}
\max(S_1, S_3) &\leq C\frac{z_n^2 \ln(n/z_n)}{n^2} \sum_{j=\tau}^{j_1} 2^j \leq C\frac{z_n^2 \ln(n/z_n)}{n^2} 2^{j_1} \\
&\leq C\frac{z_n \ln(n/z_n)}{n} \leq C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \qquad (27)
\end{aligned}$$

Upper bound for S_2. Using Proposition 2 and the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^2\right) \leq \left(\mathbb{E}\left(\left(\widehat{\beta}_{j,k} - \beta_{j,k}\right)^4\right)\right)^{1/2} \leq C\frac{z_n \ln(n/z_n)}{n}.$$

Hence

$$S_2 \leq C\frac{z_n \ln(n/z_n)}{n} \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}|>\kappa\lambda_n/2\}}.$$

Let j_2 be the integer defined by

$$\frac{1}{2}\left(\frac{n}{z_n \ln(n/z_n)}\right)^{1/(2s+1)} < 2^{j_2} \leq \left(\frac{n}{z_n \ln(n/z_n)}\right)^{1/(2s+1)}. \qquad (28)$$

We have

$$S_2 \leq S_{2,1} + S_{2,2},$$

where

$$S_{2,1} = C\frac{z_n \ln(n/z_n)}{n} \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}|>\kappa\lambda_n/2\}}$$

and

$$S_{2,2} = C\frac{z_n \ln(n/z_n)}{n} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \mathbf{1}_{\{|\beta_{j,k}|>\kappa\lambda_n/2\}}.$$

We have

$$S_{2,1} \le C\frac{z_n \ln(n/z_n)}{n} \sum_{j=\tau}^{j_2} 2^j \le C\frac{z_n \ln(n/z_n)}{n} 2^{j_2} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}.$$

For $r \ge 1$ and $p \ge 2$, since $B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$,

$$\begin{aligned} S_{2,2} &\le C\frac{z_n \ln(n/z_n)}{n\lambda_n^2} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \le C \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \le C2^{-2j_2 s} \\ &\le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

For $r \ge 1$, $p \in [1,2)$ and $s > 1/p$, since $B^s_{p,r}(M) \subseteq B^{s+1/2-1/p}_{2,\infty}(M)$ and $(2s+1)(2-p)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} S_{2,2} &\le C\frac{z_n \ln(n/z_n)}{n\lambda_n^p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \ge 1$, $\{p \ge 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$,

$$S_2 \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \tag{29}$$

Upper bound for S_4. We have

$$S_4 \le \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}|<2\kappa\lambda_n\}}.$$

Let j_2 be the integer (28). We have

$$S_4 \le S_{4,1} + S_{4,2},$$

where

$$S_{4,1} = \sum_{j=\tau}^{j_2} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}|<2\kappa\lambda_n\}}, \qquad S_{4,2} = \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \mathbf{1}_{\{|\beta_{j,k}|<2\kappa\lambda_n\}}.$$

We have

$$S_{4,1} \le C\lambda_n^2 \sum_{j=\tau}^{j_2} 2^j \le C \frac{z_n \ln(n/z_n)}{n} 2^{j_2} \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}.$$

For $r \ge 1$ and $p \ge 2$, since $B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$, we have

$$S_{4,2} \le \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k}^2 \le C \sum_{j=j_2+1}^{\infty} 2^{-2js} \le C2^{-2j_2 s} \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}.$$

For $r \ge 1$, $p \in [1,2)$ and $s > 1/p$, since $B^s_{p,r}(M) \subseteq B^{s+1/2-1/p}_{2,\infty}(M)$ and $(2-p)(2s+1)/2 + (s+1/2-1/p)p = 2s$, we have

$$\begin{aligned} S_{4,2} &\le C\lambda_n^{2-p} \sum_{j=j_2+1}^{j_1} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \\ &\le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{(2-p)/2} \sum_{j=j_2+1}^{\infty} 2^{-j(s+1/2-1/p)p} \\ &\le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{(2-p)/2} 2^{-j_2(s+1/2-1/p)p} \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}. \end{aligned}$$

So, for $r \ge 1$, $\{p \ge 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$,

$$S_4 \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}. \tag{30}$$

It follows from (26), (27), (29) and (30) that

$$S \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}. \tag{31}$$

Combining (23), (24), (25) and (31), we have, for $r \ge 1$, $\{p \ge 2$ and $s > 0\}$ or $\{p \in [1,2)$ and $s > 1/p\}$,

$$\mathbb{E} \left(\int_0^1 \left(\widehat{f}^H(x) - f_\nu(x) \right)^2 dx \right) \le C \left(\frac{z_n \ln(n/z_n)}{n} \right)^{2s/(2s+1)}.$$

The proof of Theorem 2 is complete.

□

Acknowledgment. This work is supported by ANR grant NatImages, ANR-08-EMER-009.

References

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion), *Journal of the Italian Statistical Society*, Series B, **6**, 97-144.

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statist. Probab. Lett.*, **35**, 225-232.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898-924.

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

Cai, T. and Brown, L. D. (1999). Wavelet estimation for samples with random uniform design, *Statistics and Probability Letters*, **42**, 313-321.

Chesneau, C. (2007). Wavelet block thresholding for samples with random design: a minimax approach under the $\mathbb{L}^p$ risk, *Electronic Journal of Statistics*, **1**, 331-346.

Chicken, E. (2003). Block thresholding and wavelet estimation for nonequispaced samples, *Journal of Statistical Planning and Inference*, **116**, 113-129.

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.

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. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, **90**, 432, 1200-1224.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage, *The Annals of Statistics*, **26**, 879-921.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia (with discussion), *Journal of the Royal Statistical Society* Serie B, **57**, 301-369.

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

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets, *Bernoulli*, **10** (6), 1053-1105.

Maiboroda, R. E. (1996). Estimators of components of a mixture with varying concentrations, *Ukrain. Mat. Zh.*, **48**, 4, 562-566.

Meyer, Y. (1990). *Ondelettes et Opérateurs*, Hermann, Paris.

Pensky, M. and Vidakovic, B. (2001). On non-equally spaced wavelet regression, *Ann. Inst. Statist. Math.*, **53**, 681–690.

Petrov, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

Pham Ngoc, T. M. (2009). Regression in random design and Bayesian warped wavelets estimators, *Electronic Journal of Statistics*, **3**, 1084-1112.

Pokhyl'ko, D. (2005). Wavelet estimators of a density constructed from observations of a mixture, *Theor. Prob. and Math. Statist.*, 70, 135-145.

Prakasa Rao, B. L. S. (1983). Nonparametric functional estimation, Academic Press, Orlando.

Prakasa Rao, B. L. S. (1999). Nonparametric functional estimation: an overview, *Asymptotics, Nonparametrics and Time Series*, Ed. Subir Ghosh, 461-509, Marcel Dekker Inc. New York.

Prakasa Rao, B. L. S. (2010). Wavelet linear estimation for derivatives of a density from observations of mixtures with varying mixing proportions, *Indian Journal of Pure and Applied Mathematics*, 41, **1**, 275-291.

Rosenthal, H. P. (1970). On the subspaces of $\mathbb{L}^p$ ($p \geq 2$) spanned by sequences of independent random variables, *Israel Journal of Mathematics*, **8**, pp. 273-303.

Tsybakov, A. B. (2004). *Introduction a l'estimation non-paramtrique*, Springer.

Zhang, S. and Zheng, Z. (1999). Nonlinear wavelet estimation of regression function with random design, *Science in China Series A: Mathematics*, **42**, 8, 825-833.