Adaptive wavelet estimation of a function in an indirect regression model

Christophe Chesneau, Jalal M. Fadili

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

Adaptive wavelet estimation of a function in an indirect regression model

Christophe Chesneau · Jalal Fadili

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 mean integrated squared error over Besov balls, we prove that it attains a sharp rate of convergence. Simulation results are reported to support our theoretical findings.

Keywords indirect nonparametric regression · rate of convergence · Besov balls · wavelets · hard thresholding.

2000 Mathematics Subject Classification 62G07, 62G20.

1 Introduction

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 each having a known distribution such that, for any $i \in \{1, \ldots, n\}$, the set of possible values of V_i is

$$v_i \in \{1, \ldots, m\}, \qquad m \in \mathbb{N}^* .$$

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

Jalal Fadili
GREYC CNRS UMR 6072, Image Processing Group, ENSICAEN 14050 Caen Cedex, France. E-mail: Jalal.Fadili@greyc.ensicaen.fr

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 uniformly distributed on $[0, 1]$ and $\xi_1, \ldots, \xi_n$ are n i.i.d. unobserved random variables with finite first and second moments, i.e.

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

The distribution of ξ_1 may be unknown. We suppose that $V_1, \ldots, V_n, X_1, \ldots, X_n$, $\xi_1, \ldots, \xi_n$ are mutually independent. The primary objective pursued in this paper is to estimate f_ν, for a fixed $\nu \in \{1, \ldots, m\}$, from $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Examples of application Model (1) is rather general and has many potential applications. Here we describe some examples that fall within the scope (1), and for which our estimator can have practical usefulness. In general, one can think of recovery problems and inverse problems in signal and image processing with missing or partially/uncertainly observed data, such as in computerized tomography, sensor networks, etc.

For instance, sensor network is a collection of spatially distributed autonomous sensors intended to measure and monitor physical phenomena at diverse locations, e.g. temperature, humidity, pressure, wind direction and speed, chemical concentrations, pollutant levels and vital body functions; see e.g. [1] for an overview. In a sensor network, every sensor node is also equipped with a transceiver which receives commands from a central computer and transmits data to that computer. Sensor networks are encountered in several applications which include industrial monitoring, video surveillance, traffic, medical and weather monitoring, etc. In this sensor network example, and assuming for simplicity that each sensor records only one physical parameter, we can think of the function f_d, $d \in \{1, \cdots, m\}$, as the physical parameter at sensor d, and X_i as the recording time. Given that the measurements gathering process is centralized, only one sensor information is collected at a time. The problem now is that given n noisy versions Y_i and recording times X_i of the physical parameter from non-necessarily identified sensors (i.e. unknown V_i), the goal is to recover the parameter profile f_ν at any sensor $\nu \in \{1, \cdots, d\}$. The noise in the observations Y_i can be due to measurement noise or to faulty sensors. In this setting, the distribution of V_i is typically dictated by the spatial configuration, and other parameters such as the reliability of a sensor.

To estimate f_ν, various methods can be investigated (kernel methods, spline methods, etc.) (see e.g. [23,24] and [27] for extensive overview). In this study, we focus our attention on wavelet-based 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, etc.) 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 [2] and [15].

When model (1) is considered with $V_1 = \ldots = V_n = 1$, it becomes the classical nonparametric regression model. In this case, to estimate $f_1 = f$, several wavelet methods have been designed. There is an extensive literature on the subject, see e.g. [11,12,14,13], [10], [3], [6], [29], [4,5], [19], [8], [16], [7] and [21]. However, to the best of our knowledge, there is no adaptive wavelet estimator for f_ν in the general model (1).

Contributions In this paper, we design and study an adaptive wavelet estimator for f_ν that relies on the hard thresholding rule in the wavelet domain. It has the originality to combine an "observation thresholding technique" introduced by [10] with some technical tools that account for the distribution of $V_1, \ldots, V_n$. Moreover, we evaluate its performance via the MISE over Besov balls. Under mild assumptions, to be specified and discussed in Section 2, we prove that our estimator attains a sharp rate of convergence: it is the one attained by the best nonadaptive linear wavelet estimator (the one which minimizes the MISE) up to a logarithmic factor. We also report some simulation results to illustrate the potential applicability of the estimator and to support our theoretical findings.

Paper organization The paper is organized as follows. Assumptions on the model and some notations are introduced in Section 2. Section 3 provides a brief description of wavelet bases on $[0, 1]$ and Besov balls, focusing only on essential ingredients relevant to our work. The estimators are presented in Section 4. The main results are stated in Section 5. Conclusions and perspectives are drawn in Section 6 and Section 7 is devoted to the proofs.

2 Model assumptions

In the sequel, $a(i)$ is the i-th entry of a vector a. We use the notation $\langle a, b \rangle_n = \frac{1}{n} \sum_{i=1}^n a(i)b(i)$ for the normalized euclidean inner product in $\mathbb{R}^n$, and $\|\cdot\|_n$ the associated norm.

Additional assumptions on the model (1) are as follows.

Assumption on $(f_d)_{d\in\{1,\ldots,m\}}$**.** We suppose that the collection of functions f_d is uniformly bounded, i.e. $\exists\ C_* > 0$ such that

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

Assumptions on $(V_i)_{i\in\{1,\ldots,n\}}$**.** Recall that $V_1, \ldots, V_n$ are assumed unobserved. However for any $i \in \{1, \ldots, n\}$, we suppose that the following probabilities are known

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

We also suppose that the Gram matrix

$$\Gamma_n = \frac{1}{n} W^T W = (\langle w_k, w_\ell \rangle_n)_{(k,\ell)\in\{1,\ldots,m\}^2}$$

is (symmetric) positive-definite, or equivalently that the matrix of probabilities $W = (w_1, \cdots, w_m) \in [0,1]^{n\times m}$ is full column rank.
For the considered ν (the one which refers to the estimation of f_ν) and any $i \in \{1, \ldots, n\}$, we set

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

where $M^n_{\nu,k}$ denotes the minor (ν, k) of the matrix Γ_n.
To get the gist of (3) and the importance of positive-definiteness of Γ_n, it is useful to view the vector $a_\nu = (a_\nu(1), \ldots, a_\nu(n))^T$ as the solution of the following (strictly convex) quadratic program, i.e. a quadratic objective with linear (here orthogonormality) constraints

$$\min_{b\in\mathbb{R}^n} \|b\|_n^2 \quad \text{such that} \quad \langle w_d, b\rangle_n = \delta_{\nu,d}, \text{ for } d \in \{1, \cdots, m\} \tag{4}$$

where $\delta_{\nu,d}$ is the Kronecker delta. Using the Lagrange multipliers, it is easy to see that the unique minimizer of (4) is given by

$$a_\nu = W\Gamma_n^{-1}\Delta_\nu, \tag{5}$$

where Δ_ν is a vector of zeros except at its ν-th entry. Using the cofactors of Γ_n to get its inverse, we recover (3). Positive-definiteness of Γ_n is important for (5) to make sense, otherwise a_ν would not be uniquely defined.
In a nutshell, $a_\nu \in \mathrm{Span}(w_k, k \in \{1, \cdots, m\})$ is the dual vector of minimal norm, i.e. a_ν correlates perfectly with the proper row ν of the matrix W, otherwise the inner products are zero. In the context of mixture density estimation, [17] showed that a_ν is the minimal risk weight vector to be used for the empirical measure constructed from the observations to yield an unbiased estimator of the ν-th distribution in the mixture. See [17,22, 25] for further technical details.

If V_i were observed along with (X_i, Y_i), then only observations (X_i, Y_i) corresponding to $V_i = \nu$ should be involved in the estimator. But in our setting, V_i are unobserved, and in this case, a careful decision should be made based upon all available observations to incorporate them in the estimator by "weighting" them wisely using the prior probabilities $w_k(i)$. In view of the above discussion on a_ν, it appears natural to construct such a decision using this vector. We therefore let

$$z_n = \|a_\nu\|_n^2 \tag{6}$$

where it is supposed that $z_n < n/e$. This upper-bound is not restrictive and it can be shown that a sufficient condition for it to hold is that

$\max_{k,i} w_k(i) > \sqrt{e/n}$ which is reasonable. The wavelet hard thresholding estimator that we will describe in Section 4 will explicitly involve z_n, hence a_ν.

3 Wavelets and Besov balls

Wavelet basis. Let $N \in \mathbb{N}^*$, and ϕ and ψ be respectively the father and mother wavelet functions of the Daubechies family db_N. Denote the scaled and translated versions

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

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

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

(with an appropriate treatment at the boundaries) forms an orthonormal basis of $\mathbb{L}^2([0,1])$, the set of square-integrable functions on the interval $[0,1]$. The interested reader may refer to [9] for futher details.
In turn, 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 given through the inner product that equips $\mathbb{L}^2([0,1])$

$$\alpha_{j,k} = \int_0^1 h(x)\phi_{j,k}(x)dx, \quad \beta_{j,k} = \int_0^1 h(x)\psi_{j,k}(x)dx. \tag{7}$$

Besov balls. Here, instead of the original definition of Besov spaces through the modulus of continuity, we will focus on the now classical definition of the Besov norm of a function through a sequence space norm on its wavelet coefficients. More precisely, 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[1] (7) obey

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

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 [18].

[1] The wavelet is assumed to have a sufficient number of vanishing moments.

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 $k \in \{0, \ldots, 2^j - 1\}$,

- $\alpha_{j,k} = \int_0^1 f_\nu(x)\phi_{j,k}(x)dx$ are estimated by

$$\widehat{\alpha}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} a_\nu(i) Y_i \phi_{j,k}(X_i), \tag{8}$$

- $\beta_{j,k} = \int_0^1 f_\nu(x)\psi_{j,k}(x)dx$ are estimated by

$$\widehat{\beta}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Z_{i,j,k} \mathbf{1}_{\{|Z_{i,j,k}| \leq \gamma_n\}}, \tag{9}$$

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

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

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

$$\gamma_n = \theta \sqrt{\frac{n z_n}{\ln(n/z_n)}}, \tag{10}$$

  z_n is defined by (6), $\theta = \sqrt{C_*^2 + \mathbb{E}(\xi_1^2)}$ and C_* is the one in (2). The value of θ allows to upper-bound the mean squared-error in the estimates of the scaling and wavelet coefficients $\widehat{\alpha}_{j,k}$ and $\widehat{\beta}_{j,k}$; see the proofs of Proposition 1 and 2, and more precisely (18) and (23).

Remark 1 It is worth mentioning 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 $(1/n)\sum_{i=1}^{n} Z_{i,j,k}$ is an unbiased estimator of $\beta_{j,k}$. See the proofs of Proposition 1 and 2 in Section 7, and more precisely (15) and (20).

Remark 2 The "observations thresholding technique" used in (9) has been firstly introduced by [10] 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

- $(a_\nu(i))_{i\in\{1,\ldots,n\}}$ and a fortiori the distributions of $V_1, \ldots, V_n$ (only $z_n < n/e$ is required),
- $\xi_1, \ldots, \xi_n$ (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

$$\widehat{f}^L(x) = \sum_{k=0}^{2^{j_0}-1} \widehat{\alpha}_{j_0,k}\phi_{j_0,k}(x), \tag{11}$$

where $\widehat{\alpha}_{j,k}$ is given by (8) 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

$$\widehat{f}^H(x) = \sum_{k=0}^{2^\tau-1} \widehat{\alpha}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \widehat{\beta}_{j,k}\mathbf{1}_{\left\{|\widehat{\beta}_{j,k}|\geq\kappa\lambda_n\right\}}\psi_{j,k}(x), \tag{12}$$

where $\widehat{\alpha}_{j,k}$ is defined by (8), $\widehat{\beta}_{j,k}$ by (9), 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

$$\lambda_n = \theta\sqrt{\frac{z_n \ln(n/z_n)}{n}}. \tag{13}$$

The bound on κ comes from the Bernstein concentration inequality, see Lemma 2.
Further details on the hard thresholding wavelet estimator for the standard nonparametric regression model can be found for instance in the seminal work of [11,12,14] as well as in [10].
Note that the choice of γ_n in (10) depends on λ_n in (13): we have $\lambda_n = \theta^2 z_n/\gamma_n$. The definitions of γ_n and λ_n are based on theoretical considerations that will be clarified shortly. These considerations allow our estimator to attain a sharp convergence rate on the MISE.

5 Results

Theorem 1 (Convergence rate of $\widehat{f}^L$) *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 $\widehat{f}^L$ as defined by* (11). *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 (8) and (9), and a suitable decomposition of the MISE.

Since the common distribution of $\xi_1, \ldots, \xi_n$ is unknown (a priori), we can not apply the standard lower bound theorems to prove that the rate of convergence $v_n = (z_n/n)^{2s/(2s+1)}$ is the optimal one (in the minimax sense) for (1) (most of these theorems can be found in [27, Chapter 2]). However, since $\widehat{f}^L$ is constructed to be the nonadaptive linear estimator which optimizes the MISE, assuming the smoothness of f_ν is known, our benchmark will be v_n.

One may 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 ($= n^{-2s/(2s+1)}$), which is the optimal (minimax) convergence rate (see [27]).

We now turn to the rate of the nonlinear wavelet hard thresholding estimator.

Theorem 2 (Convergence rate of $\widehat{f}^H$) *Consider* (1) *under the assumptions of Section* 2. *Let* $\widehat{f}^H$ *as defined by* (12). *Then there exists a constant* $C > 0$ *such that*

$$\sup_{f_\nu \in B^s_{p,r}(M)} \mathbb{E}\left(\int_0^1 \left(\widehat{f}^H(x) - f_\nu(x)\right)^2 dx\right) \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)},$$

with $r \geq 1$, $\{p \geq 2$ *and* $s > 0\}$ *or* $\{p \in [1, 2)$ *and* $s > 1/p\}$.

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

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

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 as the standard hard thresholding estimator for the classical nonparametric regression model (see [11,12,14]). The latter is known to be optimal in the minimax sense up to a logarithmic term.

6 Simulation results

In this simulation, $n = 4096$ observed data samples (Y_i, X_i) were generated according to model (1), where X_i were equi-spaced in $[0, 1]$ with $X_1 = 0$ and $X_n = 1$, and $\xi_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$ with $\sigma = 0.01$. We have used three piece-wise regular test functions exhibiting different degrees of smoothness, and we have chosen arbitrarily $C_* = 1$. These functions are widely used in the non-linear wavelet estimation literature. The V_i's were sampled randomly in $\{1, 2, 3\}$ with probabilities $w_d(i)$ such that each function was randomly observed third of the time on $[0, 1]$. We used the Daubechies db_3 wavelet and our test code was based on [28].

The results are depicted in Fig. 1. It can be clearly seen that our adaptive hard thresholding estimator is very effective to estimate each of the three test functions. The recovered wavelet coefficients are also shown where most of the irreagularities are captured in the estimated coefficients. In the figure, we also display the indicators of the true indices (those of the first 20 samples) for each test function, i.e. 1 if sample i is selected from function $\nu \in \{1, 2, 3\}$ and 0 otherwise. The corresponding weight vector a_ν is shown, and it can be seen that a_ν fullfills its expected role by wisely weighting the appropriate observations.

7 Conclusion and perspectives

In this work, an adaptive wavelet hard thresholding estimator was constructed to estimate an arbitrary function f_ν from the sophisticated regression model (1). Under mild assumptions on the noise and the V_i's, it was proved that it attains a sharp rate of convergence over a wide class of functions belonging to Besov spaces.

There are several perspectives that rise naturally from this work:

- It would be interesting to investigate the estimation of f_ν in (1) when the design point X_1 has a more complex distribution beyond the random uniform one. In this case, the warped wavelet basis introduced in the nonparametric regression estimation by [16] could be a promising tool to attack this problem.
- Another important extension would be to consider the case where the distributions of $V_1, \ldots, V_n$ are unknown, which is the case in many practical situations.
- A last point would be to try to improve the estimation of f_ν (e.g. by removing the extra logarithmic term). The block-thresholding rule named BlockJS developed in wavelet estimation by [4, 5] seems to be a good candidate.

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

8 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.

8.1 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 (7) of f_ν and $\widehat{\alpha}_{j,k}$ be as in (8). Then there exists a constant $C > 0$ such that*

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

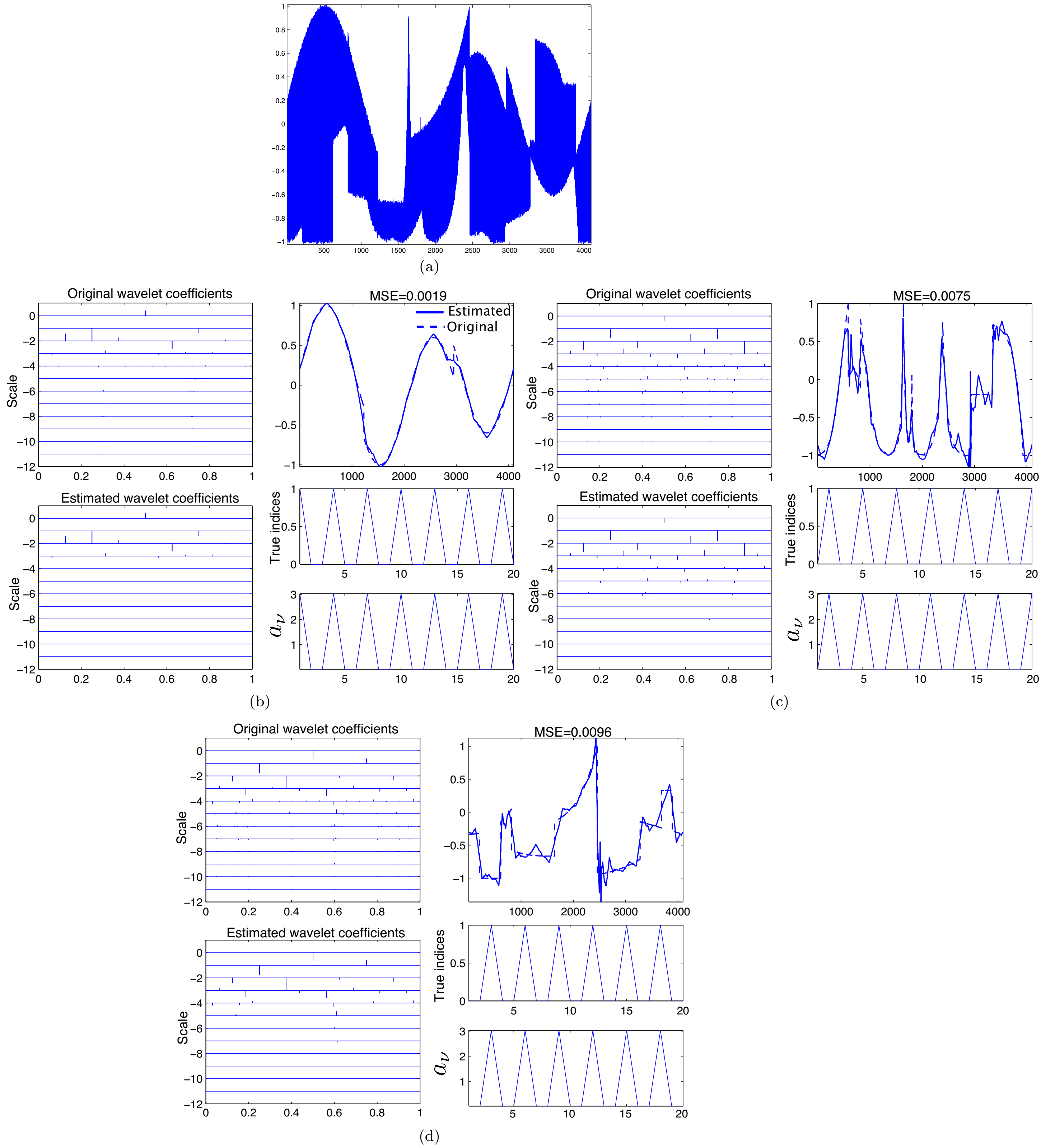

Fig. 1 Estimated functions using our adaptive wavelet hard thresholding from $n = 4096$ noisy observations with $\xi_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$, and X_i are equi-spaced in $[0, 1]$. In this experiment, we used $m = 3$ irregular test functions with different degrees of smoothness. (a): noisy observations. (b)-(d): estimated functions.

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

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

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

$$\begin{aligned}\mathbb{E}(W_{i,j,k}) &= \mathbb{E}(a_\nu(i) Y_i \phi_{j,k}(X_i)) = \mathbb{E}(a_\nu(i)(f_{V_i}(X_i) + \xi_i)\phi_{j,k}(X_i)) \\ &= a_\nu(i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}(X_i)) + a_\nu(i)\mathbb{E}(\xi_i)\mathbb{E}(\phi_{j,k}(X_i)) \\ &= a_\nu(i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}(X_i)) \\ &= a_\nu(i) \sum_{d=1}^{m} w_d(i) \int_0^1 f_d(x)\phi_{j,k}(x)dx. \end{aligned} \tag{14}$$

It follows from (14) and (4) that

$$\begin{aligned}\mathbb{E}(\widehat{\alpha}_{j,k}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(W_{i,j,k}) = \frac{1}{n}\sum_{i=1}^{n}\left(a_\nu(i)\sum_{d=1}^{m} w_d(i)\int_0^1 f_d(x)\phi_{j,k}(x)dx\right) \\ &= \sum_{d=1}^{m}\int_0^1 f_d(x)\phi_{j,k}(x)dx \left(\frac{1}{n}\sum_{i=1}^{n} a_\nu(i) w_d(i)\right) \\ &= \int_0^1 f_\nu(x)\phi_{j,k}(x)dx = \alpha_{j,k}. \end{aligned} \tag{15}$$

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

$$\begin{aligned}\mathbb{E}\left((\widehat{\alpha}_{j,k} - \alpha_{j,k})^2\right) &= \mathbb{V}(\widehat{\alpha}_{j,k}) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} W_{i,j,k}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(W_{i,j,k}) \\ &\le \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left(W_{i,j,k}^2\right). \end{aligned} \tag{16}$$

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

$$\mathbb{E}\left(W_{i,j,k}^2\right) = \mathbb{E}(a_\nu^2(i) Y_i^2 \phi_{j,k}^2(X_i)) = a_\nu^2(i)\mathbb{E}\left((f_{V_i}(X_i) + \xi_i)^2 \phi_{j,k}^2(X_i)\right). \tag{17}$$

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), $\sup_{d\in\{1,\dots,m\}}\sup_{x\in[0,1]}|f_d(x)| \le C_*$, we have

$$\begin{aligned}&\mathbb{E}\left((f_{V_i}(X_i) + \xi_i)^2 \phi_{j,k}^2(X_i)\right) \\ &= \mathbb{E}\left(f_{V_i}^2(X_i)\phi_{j,k}^2(X_i)\right) + 2\mathbb{E}(\xi_i)\mathbb{E}(f_{V_i}(X_i)\phi_{j,k}^2(X_i)) + \mathbb{E}\left(\xi_i^2\right)\mathbb{E}\left(\phi_{j,k}^2(X_i)\right) \\ &= \mathbb{E}\left(f_{V_i}^2(X_i)\phi_{j,k}^2(X_i)\right) + \mathbb{E}\left(\xi_1^2\right) \le C_*^2\mathbb{E}\left(\phi_{j,k}^2(X_i)\right) + \mathbb{E}\left(\xi_1^2\right) \\ &= C_*^2 + \mathbb{E}\left(\xi_1^2\right) = \theta^2. \end{aligned} \tag{18}$$

Putting (17) and (18) together, we obtain

$$\mathbb{E}\left(W_{i,j,k}^2\right) \le \theta^2 a_\nu^2(i). \tag{19}$$

It follows from (16) and (19) that

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

The proof of Proposition 1 is complete.

□

Proposition 2 *For any integer $j \ge \tau$ and any $k \in \{0,\dots,2^j-1\}$, let $\beta_{j,k}$ be the wavelet coefficient* (7) *of f_ν and $\widehat{\beta}_{j,k}$ be as in* (9)*. Then there exists a constant $C>0$ such that*

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

Proof of Proposition 2. Taking ψ instead of ϕ in (15), we obtain

$$\begin{aligned}\beta_{j,k} &= \int_0^1 f_\nu(x)\psi_{j,k}(x)dx = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(Z_{i,j,k}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|>\gamma_n\}}). \qquad (20)\end{aligned}$$

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

$$\mathbb{E}\left(\left(\widehat{\beta}_{j,k}-\beta_{j,k}\right)^4\right) \le 8(A+B), \qquad (21)$$

where

$$A = \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}} - \mathbb{E}(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}})\right)\right)^4\right)$$

and

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

Let us bound A and B, in turn.

Upper bound for A. Let us present the Rosenthal inequality (see [26]).

Lemma 1 (Rosenthal's inequality) *Let $n\in\mathbb{N}^*$, $p\ge 2$ and $U_1,\dots,U_n$ 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*

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

Set, for any $i \in \{1, \ldots, n\}$,

$$U_{i,j,k} = Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}| \leq \gamma_n\}} - \mathbb{E}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}| \leq \gamma_n\}}\right).$$

Then, for any $i \in \{1, \ldots, n\}$, we have $\mathbb{E}(U_{i,j,k}) = 0$ and using (19) (with ψ instead of ϕ), for any $b \in \{2, 4\}$,

$$\mathbb{E}\left(U_{i,j,k}^b\right) \leq 2^b \mathbb{E}\left(Z_{i,j,k}^b \mathbf{1}_{\{|Z_{i,j,k}| \leq \gamma_n\}}\right) \leq 2^b \gamma_n^{b-2} \mathbb{E}\left(Z_{i,j,k}^2\right) \leq 2^b \gamma_n^{b-2} \theta^2 a_\nu^2(i).$$

It follows from the Rosenthal inequality and $z_n < n/e$ that

$$\begin{aligned} A &= \frac{1}{n^4}\mathbb{E}\left(\left(\sum_{i=1}^n U_{i,j,k}\right)^4\right) \leq C\frac{1}{n^4}\max\left(\sum_{i=1}^n \mathbb{E}\left(U_{i,j,k}^4\right), \left(\sum_{i=1}^n \mathbb{E}\left(U_{i,j,k}^2\right)\right)^2\right) \\ &\leq C\frac{1}{n^4}\max\left(\gamma_n^2 \sum_{i=1}^n a_\nu^2(i), \left(\sum_{i=1}^n a_\nu^2(i)\right)^2\right) \\ &= C\frac{1}{n^4}\max\left(\frac{n^2}{\ln(n/z_n)} z_n^2, n^2 z_n^2\right) = C\frac{z_n^2}{n^2}. \end{aligned} \tag{22}$$

Upper bound for B. Using again (19) (with ψ instead of ϕ), for any $i \in \{1, \ldots, n\}$, we obtain

$$\begin{aligned} \mathbb{E}\left(|Z_{i,j,k}|\mathbf{1}_{\{|Z_{i,j,k}| > \gamma_n\}}\right) &\leq \frac{\mathbb{E}(Z_{i,j,k}^2)}{\gamma_n} \leq \frac{1}{\theta}\sqrt{\frac{\ln(n/z_n)}{nz_n}}\theta^2 a_\nu^2(i) \\ &= \theta\sqrt{\frac{\ln(n/z_n)}{nz_n}} a_\nu^2(i). \end{aligned}$$

Therefore

$$\begin{aligned} B &= \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}(|Z_{i,j,k}|\mathbf{1}_{\{|Z_{i,j,k}| > \gamma_n\}})\right)^4 \leq \theta^4 \frac{(\ln(n/z_n))^2}{n^2 z_n^2}\left(\frac{1}{n}\sum_{i=1}^n a_\nu^2(i)\right)^4 \\ &= \theta^4 \frac{(\ln(n/z_n))^2}{n^2 z_n^2} z_n^4 = \theta^4 \frac{(z_n \ln(n/z_n))^2}{n^2}. \end{aligned} \tag{23}$$

Combining (21), (22) and (23) and using $z_n < n/e$, we have

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

This completes the proof of Proposition 2.

□

Proposition 3 *For any integer $j \geq \tau$ and any $k \in \{0, \ldots, 2^j - 1\}$, let $\beta_{j,k}$ be the wavelet coefficient (7) of f_ν, $\widehat{\beta}_{j,k}$ be (9) and λ_n be as in (13). Then, for any $\kappa \geq 8/3 + 2 + 2\sqrt{16/9 + 4}$,*

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

Proof of Proposition 3. By (20) we have

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

Using (19) (with ψ instead of ϕ), we obtain

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

Hence

$$\mathbb{P}\left(|\widehat{\beta}_{j,k}-\beta_{j,k}|\ge\kappa\lambda_n/2\right) \le \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}-\mathbb{E}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}\right)\right)\right|\ge(\kappa/2-1)\lambda_n\right). \tag{24}$$

Now we need the Bernstein inequality presented in the lemma below (see [20]).

Lemma 2 (Bernstein's inequality) *Let $n\in\mathbb{N}^*$ and $U_1,\ldots,U_n$ be n zero mean independent random variables such that there exists a constant $M>0$ satisfying* $\sup_{i\in\{1,\ldots,n\}}|U_i|\le M<\infty$. *Then, for any $\lambda>0$,*

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

Set, for any $i\in\{1,\ldots,n\}$,

$$U_{i,j,k}=Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}-\mathbb{E}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}\right).$$

Then, for any $i\in\{1,\ldots,n\}$, we have $\mathbb{E}(U_{i,j,k})=0$,

$$|U_{i,j,k}|\le|Z_{i,j,k}|\,\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}+\mathbb{E}\left(|Z_{i,j,k}|\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}\right)\le 2\gamma_n$$

and, using again (19) (with ψ instead of ϕ),

$$\mathbb{E}\left(U_{i,j,k}^2\right)=\mathbb{V}\left(Z_{i,j,k}\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}\right)\le\mathbb{E}\left(Z_{i,j,k}^2\mathbf{1}_{\{|Z_{i,j,k}|\le\gamma_n\}}\right)\le\mathbb{E}\left(Z_{i,j,k}^2\right)\le\theta^2 a_\nu^2(i).$$

So

$$\sum_{i=1}^{n}\mathbb{E}\left(U_{i,j,k}^2\right)\le\theta^2\sum_{i=1}^{n}a_\nu^2(i)=\theta^2 n z_n.$$

It follows from the Bernstein inequality that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_{i,j,k}\right| \geq n(\kappa/2-1)\lambda_n\right) \leq 2\exp\left(-\frac{n^2(\kappa/2-1)^2\lambda_n^2}{2\left(\theta^2 n z_n + 2n(\kappa/2-1)\lambda_n\gamma_n/3\right)}\right). \qquad (25)$$

Remark that

$$\lambda_n\gamma_n = \theta\sqrt{\frac{z_n\ln(n/z_n)}{n}}\theta\sqrt{\frac{nz_n}{\ln(n/z_n)}} = \theta^2 z_n, \qquad \lambda_n^2 = \theta^2\frac{z_n\ln(n/z_n)}{n}.$$

Putting (24) and (25) together, for any $\kappa \geq 8/3+2+2\sqrt{16/9+4}$, we have

$$\mathbb{P}\left(|\widehat{\beta}_{j,k}-\beta_{j,k}| \geq \kappa\lambda_n/2\right) \leq 2\exp\left(-\frac{(\kappa/2-1)^2\ln(n/z_n)}{2\left(1+2(\kappa/2-1)/3\right)}\right) = 2\left(\frac{n}{z_n}\right)^{-\frac{(\kappa/2-1)^2}{2(1+2(\kappa/2-1)/3)}} \leq 2\left(\frac{z_n}{n}\right)^2.$$

This ends the proof of Proposition 3.

□

8.2 Proofs of the main results

Proof of Theorem 1. We expand the function f_ν on $\mathcal{B}$ as

$$f_\nu(x) = \sum_{k=0}^{2^{j_0}-1}\alpha_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}\psi_{j,k}(x),$$

where

$$\alpha_{j_0,k} = \int_0^1 f_\nu(x)\phi_{j_0,k}(x)dx, \qquad \beta_{j,k} = \int_0^1 f_\nu(x)\psi_{j,k}(x)dx.$$

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

$$\mathbb{E}\left(\int_0^1\left(\widehat{f}^L(x)-f_\nu(x)\right)^2dx\right) = A+B,$$

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

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

Since $p \geq 2$, we have $B^s_{p,r}(M) \subseteq B^s_{2,\infty}(M)$. Hence

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

So

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

The proof of Theorem 1 is complete.

□

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

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

where

$$\alpha_{\tau,k} = \int_0^1 f_\nu(x) \phi_{\tau,k}(x) dx, \qquad \beta_{j,k} = \int_0^1 f_\nu(x) \psi_{j,k}(x) dx.$$

We have

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

Hence

$$\mathbb{E}\left(\int_0^1 \left(\widehat{f}^H(x) - f_\nu(x)\right)^2 dx\right) = R + S + T, \tag{26}$$

where

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

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

$$R \le C\frac{z_n}{n} \le C\frac{z_n \ln(n/z_n)}{n} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \quad (27)$$

For $r \ge 1$ and $p \ge 2$, 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

$$\begin{aligned} T &\le C\sum_{j=j_1+1}^{\infty} 2^{-2js} \le C2^{-2j_1 s} \le C\left(\frac{n}{z_n}\right)^{-2s} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s} \\ &\le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

For $r \ge 1$ 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

$$\begin{aligned} T &\le C\sum_{j=j_1+1}^{\infty} 2^{-2j(s+1/2-1/p)} \le C2^{-2j_1(s+1/2-1/p)} \\ &\le C\left(\frac{n}{z_n}\right)^{-2(s+1/2-1/p)} \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2(s+1/2-1/p)} \\ &\le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

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

$$T \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \quad (28)$$

The term S can be decomposed as

$$S = S_1 + S_2 + S_3 + S_4, \quad (29)$$

where

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

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

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

and

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

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

$$\begin{aligned}
\max(S_1, S_3) &\leq C \frac{z_n^2 \ln(n/z_n)}{n^2} \sum_{j=\tau}^{j_1} 2^j \leq C \frac{z_n^2 \ln(n/z_n)}{n^2} 2^{j_1} \\
&\leq C \frac{z_n \ln(n/z_n)}{n} \leq C \left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \qquad (30)
\end{aligned}$$

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

$$\frac{1}{2}\left(\frac{n}{z_n\ln(n/z_n)}\right)^{1/(2s+1)} < 2^{j_2} \le \left(\frac{n}{z_n\ln(n/z_n)}\right)^{1/(2s+1)}. \qquad (31)$$

We have

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

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 $B^s_{p,r}(M)\subseteq B^s_{2,\infty}(M)$,

$$\begin{aligned} S_{2,2} &\le C\frac{z_n\ln(n/z_n)}{n\lambda_n^2}\sum_{j=j_2+1}^{j_1}\sum_{k=0}^{2^j-1}\beta_{j,k}^2 \le C\sum_{j=j_2+1}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}^2 \le \sum_{j=j_2+1}^{\infty}2^{-2js} \\ &\le C2^{-2j_2s} \le C\left(\frac{z_n\ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \end{aligned}$$

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 $s>1/p\}$,

$$S_2 \le C\left(\frac{z_n\ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \qquad (32)$$

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 (31). 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

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

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 $s > 1/p\}$,

$$S_4 \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \tag{33}$$

It follows from (29), (30), (32) and (33) that

$$S \le C\left(\frac{z_n \ln(n/z_n)}{n}\right)^{2s/(2s+1)}. \tag{34}$$

Combining (26), (27), (28) and (34), we have, for $r \geq 1$, $\{p \geq 2$ and $s > 0\}$ or $\{p \in [1, 2)$ and $s > 1/p\}$,

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

This ends the proof of Theorem 2.

□

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

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Communications Magazine pp. 102–114 (2002)
2. Antoniadis, A.: Wavelets in statistics: a review (with discussion). Journal of the Italian Statistical Society, Series B **6**, 97–144 (1997)
3. Antoniadis, A., Gregoire, G., Vial, P.: Random design wavelet curve smoothing. Statist. Probab. Lett. **35**, 225–232 (1997)
4. Cai, T.: Adaptive wavelet estimation: a block thresholding and oracle inequality approach. Annals of Statistics **27**, 898–924 (1999)
5. Cai, T.: On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. Statistica Sinica **12**(4), 1241–1273 (2002)
6. Cai, T., Brown, L.D.: Wavelet estimation for samples with random uniform design. Statist. Probab. Lett. **42**, 313–321 (1999)
7. Chesneau, C.: Wavelet block thresholding for samples with random design: a minimax approach under the $\mathbb{L}^p$ risk. Electronic Journal of Statistics **1**, 331–346 (2007)
8. Chicken, E.: Block thresholding and wavelet estimation for nonequispaced samples. Journal of Statistical Planning and Inference **116**, 113–129 (2003)
9. Cohen, A., Daubechies, I., Jawerth, B., Vial, P.: Wavelets on the interval and fast wavelet transforms. Applied and Computational Harmonic Analysis **24**(1), 54–81 (1993)
10. Delyon, B., Juditsky, A.: On minimax wavelet estimators. Applied Computational Harmonic Analysis **3**, 215–228 (1996)
11. Donoho, D., Johnstone, I.: Ideal spatial adaptation via wavelet shrinkage. Biometrika **81** (1994). 425–455
12. Donoho, D., Johnstone, I.: Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Association **90** (1995). 1200–1224
13. Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D.: Wavelet shrinkage: asymptopia? Journal of the Royal Statistical Society Series B **57** (1995). 301–369
14. Donoho, D., Johnstone, I.M.: Minimax estimation via wavelet shrinkage. Ann. Statist. (1998)
15. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.B.: Wavelets, Approximation and Statistical Applications. Lecture Notes in Statistics. Springer, New York (1998)
16. Kerkyacharian, G., Picard, D.: Regression in random design and warped wavelets,. Bernoulli **10**(6), 1053–1105 (2004)
17. Maiboroda, R.E.: Estimates of components of a mixture with varying concentrations. Ukrain. Mat. Zh. **48**(4), 562–566 (1996)
18. Meyer, Y.: Wavelet and Operators. Cambridge University Press (1992)
19. Pensky, M., Vidakovic, B.: On non-equally spaced wavelet regression. Ann. Inst. Statist. Math. **53**, 681–690 (2001)
20. Petrov, V.V.: Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford: Clarendon Press (1995)
21. Pham Ngoc, T.M.: Regression in random design and bayesian warped wavelets estimators. Electronic Journal of Statistics **3**, 1084–1112 (2009)

22. Pokhyl'ko, D.: Wavelet estimators of a density constructed from observations of a mixture. Theor. Prob. and Math. Statist. **70**, 135–145 (2005)
23. Prakasa Rao, B.L.S.: Nonparametric functional estimation. Academic Press, Orlando (1983)
24. Prakasa Rao, B.L.S.: Nonparametric functional estimation: an overview, chap. Asymptotics, Nonparametrics and Time Series, pp. 461–509. Marcel Dekker Inc., New York (1999)
25. Prakasa Rao, B.L.S.: 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 (2010)
26. Rosenthal, H.P.: On the subspaces of $\mathbb{L}^p$ ($p \geq 2$) spanned by sequences of independent random variables. srael Journal of Mathematics **8**, 273–303 (1970)
27. Tsybakov, A.: Introduction à l'estimation nonparametrique. springer New York (2004)
28. Wavelab 802: Wavelab toolbox. http://www-stat.stanford.edu/~wavelab (2001)
29. Zhang, S., Zheng, Z.: Nonlinear wavelet estimation of regression function with random design. Science in China Series A: Mathematics **42**(8), 825–833 (1999)