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

On the mean integrated squared error of a plug-in estimator for convolutions

Christophe Chesneau · Fabien Navarro

Received:

Abstract The nonparametric estimation of convolutions is considered. Using the mean integrated squared error, we explore the performance of a plug-in estimator under mild assumptions on the model. We illustrate these general results via wavelet hard thresholding estimators for two different density estimation problems. In particular, we prove that they attain fast rates of convergence for a wide class of unknown functions. Simulation results illustrate the theory.

Keywords Convolutions estimation · Plug-in estimator · Wavelets · Hard thresholding · Rates of convergence.

2000 Mathematics Subject Classification 62G07, 62G20.

1 Problem statement

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, f be an unknown function related to n random variables $Z_1, \ldots, Z_n$ and $m \geq 2$ be a fixed integer. We aim to estimate the "m-convolutions"

$$g(x) = \bigstar_m f(x) = \underbrace{(f \star \ldots \star f)}_{m}(x) \\ = \int \ldots \int f(x - u_2 - \ldots - u_m) f(u_2) \ldots f(u_m) du_2 \ldots du_m \qquad (1)$$

from $Z_1, \ldots, Z_n$.

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

Fabien Navarro
Laboratoire de Mathématiques Nicolas Oresme and GREYC CNRS-ENSICAEN-Univ. de Caen, ENSICAEN, 14050 Caen Cedex, France. E-mail: Fabien.Navarro@greyc.ensicaen.fr

The most standard case concerns the density estimation problem where $Z_1, \ldots, Z_n$ are *i.i.d.* with common unknown density f and g is the density of $S = \sum_{i=1}^{m} Z_i$. A detailed application in the field of health insurance can be found in Panjer and Willmot (1992). Methods and results can be found in Frees (1994), Saavedra and Cao (2000), Ahmad and Fan (2001), Ahmad and Mugdadi (2003), Prakasa Rao (2004), Schick and Wefelmeyer (2004, 2007), Du and Schick (2007) and Giné and Mason (2007). In particular, Saavedra and Cao (2000) have introduced the natural plug-in estimator $\hat{g} = \bigstar_m \hat{f}$, where $\hat{f}$ denotes a kernel estimator.

In this study, considering the general form of the problem, we provide a contribution to the approach of Saavedra and Cao (2000). We investigate the global estimation of g via

$$\hat{g}(x) = \bigstar_m \hat{f}(x), \tag{2}$$

where $\hat{f}$ denotes an estimator for f (not necessarily the kernel one). First of all, under various assumptions on f and $\hat{f}$, we determine sharp upper bounds for the Mean Integrated Squared Error (MISE) of $\hat{g}$. These bounds depend on the $\mathbb{L}_p$-risk of $\hat{f}$ and show how the performances of $\hat{g}$ and $\hat{f}$ are associated.

Then we adapt our theory to the wavelet hard thresholding estimators. We are interested on such estimators because they achieve a high degree of adaptivity and capability of handling the singularities of the unknown function. We refer to e.g. Antoniadis (1997), Härdle *et al.* (1998) and Vidakovic (1999) for the details and discussions on their advantages over traditional methods. Adopting the minimax approach under the MISE and over a wide class of unknown functions (the Besov balls), we determine their rates of convergence. These results are applied to two different density estimation problems: the standard density one (described above) and the deconvolution density one. For each of these problems, the obtained rates of convergence are the standard "near optimal" ones corresponding to $m = 1$. A comprehensive simulation study supports our theoretical findings. In particular, for n large enough, we show that the practical performance of our wavelet estimator compares favorably to the standard ones (i.e. those of Frees (1994) and Saavedra and Cao (2000)) for a wide variety of functions.

The paper is organized as follows. Two general results are presented in Section 2. Section 3 describes our wavelet hard thresholding methodology and the obtained rates of convergence. Applications of our theory and simulation results are presented in (1) Sections 4 for the standard density estimation, (2) Section 5 for the deconvolution density estimation. Technical proofs are given in Section 6.

2 Two general results

Under different assumptions on f and $\hat{f}$, Theorems 1 and 2 investigate the influence of $\hat{f}$ on the performance of $\hat{g}$. The benchmark is the MISE for $\hat{g}$ i.e. $\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right)$ (it is assumed that $\hat{g}$, g and the associated MISE exist).

Theorem 1 *Consider the estimation problem and notations of Section 1. Suppose that there exist two constants* $C_1 > 0$ *and* $C_2 > 0$ *such that*

$$\int |f(x)|dx \leq C_1, \qquad \int |\hat{f}(x)|dx \leq C_2. \tag{3}$$

Then

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C_3 \mathbb{E}\left(\int |\hat{f}(x) - f(x)|^2 dx\right),$$

where $C_3 = (C_1 + C_2)^{2m-2}$.

The proof of Theorem 1 is driven by the definition of the MISE and uses the Parseval theorem combines with some properties of the continuous Fourier transform. Note that it completely differs to (Saavedra and Cao 2000, Proof of Theorem 4) which consists in developing the Mean Squared Error (MSE) of the considered estimator, then derive MISE results.

Theorem 1 shows that, under (3), $\hat{f}$ and $\hat{g}$ attain the same rates of convergence under the MISE. For a wide variety of models, the assumption (3) on $\hat{f}$ is often satisfied by kernel-type estimators (see e.g. (Tsybakov 2004, Chapter 1, Section 1.4)). However, it excludes a wide family of series expansion estimators (as, for instance, the wavelet estimators).

For these reasons, under other assumptions, Theorem 2 below provides an alternative.

Theorem 2 *Consider the estimation problem and notations of Section 1. Suppose that* $\operatorname{supp} f \subseteq [-T, T]$, *where* $T > 0$ *is a fixed constant, and there exist two constants* $C_1 > 0$ *and* $C_2 > 0$ *such that*

$$\int |f(x)|dx \leq C_1, \qquad \mathbb{E}\left(\int |\hat{f}(x) - f(x)|^{4m-4} dx\right) \leq C_2. \tag{4}$$

Then

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C_3 \sqrt{\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^4 dx\right)},$$

where $C_3 = \sqrt{2^{4m-4} T \left((2T)^{4m-5} C_2 + 2^{4m-4} C_1^{4m-4}\right)}$.

The proof of Theorem 2 is based on the one of Theorem 1 and some technical inequalities related to the $\mathbb{L}_p$-norms.

The next section applies this theorem to wavelet hard thresholding estimators and determine rates of convergence for the considered $\hat{g}$.

3 Wavelet estimators

Let $N \geq 1$ be an integer, and ϕ and ψ be the initial wavelet functions of the Daubechies wavelets *dbN*. These functions have the particularity to be compactly supported and $\mathcal{C}^\nu$ where ν is an integer depending on N.

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 and a set of consecutive integers Λ_j with a length proportional to 2^j such that, for any integer $\ell \geq \tau$, the collection

$$\mathcal{B} = \{\phi_{\ell,k}(.),\ k \in \Lambda_\ell;\ \psi_{j,k}(.);\ j \in \mathbb{N} - \{0, \ldots, \ell - 1\},\ k \in \Lambda_j\},$$

is an orthonormal basis of $\mathbb{L}^2([-T,T]) = \{h : [-T,T] \to \mathbb{R};\ \int_{-T}^{T} h^2(x)dx < \infty\}$. We refer to Cohen *et al.* (1993) and Mallat (2009).

Suppose that $f \in \mathbb{L}^2([-T,T])$. Then, for any integer $\ell \geq \tau$, f can be expanded on $\mathcal{B}$ as

$$f(x) = \sum_{k \in \Lambda_\ell} \alpha_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k}\psi_{j,k}(x),$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the (unknown) wavelet coefficients of f defined by

$$\alpha_{j,k} = \int f(x)\phi_{j,k}(x)dx, \qquad \beta_{j,k} = \int f(x)\psi_{j,k}(x)dx. \tag{5}$$

Considering the general estimation problem described in Section 1, let $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ be estimators of $\alpha_{j,k}$ and $\beta_{j,k}$ respectively.

We suppose that there exist three constants $C > 0$, $\kappa > 0$ and $\delta > 0$ such that $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ satisfy, for any $j \in \{\tau, \ldots, j_1\}$,

$$\mathbb{E}(|\hat{\alpha}_{j,k} - \alpha_{j,k}|^{\upsilon}) \leq C 2^{\upsilon\delta j} \left(\frac{\ln n}{n}\right)^{\upsilon/2}, \tag{6}$$

$$\mathbb{E}(|\hat{\beta}_{j,k} - \beta_{j,k}|^{\upsilon}) \leq C 2^{\upsilon\delta j} \left(\frac{\ln n}{n}\right)^{\upsilon/2} \tag{7}$$

and

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \frac{\kappa}{2} 2^{\delta j} \sqrt{\frac{\ln n}{n}}\right) \leq C \left(\frac{\ln n}{n}\right)^2, \tag{8}$$

where $\upsilon = 4m - 4 \geq 4$ and j_1 is the integer satisfying $(n/\ln n)^{1/(2\delta+1)} < 2^{j_1+1} \leq 2(n/\ln n)^{1/(2\delta+1)}$.

Then we define the hard thresholding estimator $\hat{f}$ by

$$\hat{f}(x) = \sum_{k\in\Lambda_\tau} \hat{\alpha}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k\in\Lambda_j} \hat{\beta}_{j,k}\mathbf{1}_{\left\{|\hat{\beta}_{j,k}|\geq\kappa 2^{\delta j}\sqrt{\ln n/n}\right\}}\psi_{j,k}(x), \qquad (9)$$

where, for any random event $\mathcal{A}$, $\mathbf{1}_{\mathcal{A}}$ is the indicator function on $\mathcal{A}$,

The idea of the hard thresholding rule in (9) is to only estimate the "large" unknown wavelet coefficients of f which contain its main characteristics. Details can be found in e.g. Antoniadis (1997), Härdle *et al.* (1998) and Vidakovic (1999).

To explore the asymptotic performance of $\hat{f}$ (and, a fortiori, $\hat{g}$), we need some smoothness assumptions on f. In this study, we suppose that f belongs to a Besov balls $B^s_{p,r}(M)$ with $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$. In terms of wavelet coefficients, it means that there exists a constant $M^* > 0$ (depending on M) such that (5) satisfy

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

In this expression, s is a smoothness parameter and p and r are norm parameters. Besov balls contain the Hölder and Sobolev balls. See e.g. Meyer (1992) and (Härdle *et al.* 1998, Chapter 9).

Theorem 3 below determines sharp rates of convergence for (2) defined with (9) under the MISE over Besov balls.

Theorem 3 *Consider the estimation problem and notations of Section 1. Suppose that* $\mathrm{supp}\, f \subseteq [-T, T]$*, where* $T > 0$ *is a fixed constant, and there exists a constant* $C > 0$ *such that* $\int |f(x)|^{4m-4}dx \leq C$*. Let* $\hat{g}$ *be* (2) *with* $\hat{f}$ *defined by* (9)*.*

Then, for any $r \geq 1$*, any* $\{p \geq 4$ *and* $s > 0\}$ *or any* $\{p \in [1, 4)$ *and* $s > \max((2\delta+1)/p, (1/p)(4-p)(\delta+1/2)\}$*, there exists a constant* $C > 0$ *such that*

$$\sup_{f\in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C\left(\frac{\ln n}{n}\right)^{2s/(2s+2\delta+1)}.$$

The proof of Theorem 3 uses Theorem 2 and a result on the rates of convergence of $\hat{f}$ under the $\mathbb{L}_p$-risk over Besov balls proved by Kerkyacharian and Picard (2000).

Mention that, in the same framework, $(\ln n/n)^{2s/(2s+2\delta+1)}$ is the rate of convergence attained by $\hat{f}$. In this sense, the asymptotic properties of the MISE of $\hat{g}$ and $\hat{f}$ are similar. Theorem 3 provides a theoretical contribution to this intuitive idea.

The next section is devoted to some applications of this result.

4 Application I: the density model

4.1 Upper bound

We observe n *i.i.d.* random variables $Z_1, \ldots, Z_n$ with common unknown density f. For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^{m} Z_i$ and g be the density of S. The goal is to estimate g from $Z_1, \ldots, Z_n$.

As mentioned in Section 1, such a problem has already been considered with kernel-type estimators and various settings by e.g. Frees (1994), Saavedra and Cao (2000), Ahmad and Fan (2001), Ahmad and Mugdadi (2003), Schick and Wefelmeyer (2004, 2007) and Du and Schick (2007).

Proposition 1 below investigates the rates of convergence of $\hat{g}$ (2) constructed from a specific wavelet hard thresholding estimator $\hat{f}$ under the MISE over Besov balls.

Proposition 1 *Consider the standard density model and the associated notations. Suppose that* $\operatorname{supp} f \subseteq [-T, T]$*, where* $T > 0$ *is a fixed constant, and there exists a constant* $C > 0$ *such that* $\sup_{x \in \mathbb{R}} f(x) \leq C$.

Let $\hat{g}$ *be* (2) *and* $\hat{f}$ *be as in* (9) *with* $\delta = 0$,

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{j,k}(Z_i), \qquad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(Z_i). \tag{10}$$

Then, for any $r \geq 1$*, any* $\{p \geq 4$ *and* $s > 0\}$ *or any* $\{p \in [1, 4)$ *and* $s > \max(1/p, (1/(2p))(4 - p)\}$*, there exists a constant* $C > 0$ *such that*

$$\sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

Remark that $(\ln n/n)^{2s/(2s+1)}$ is the "near optimal" rates of convergence in the minimax sense for the standard density estimation problem. See (Donoho *et al.* 1996, Theorems 2 and 3).

4.2 Simulation study

The original observations were generated from an *i.i.d.* sample of random variables $Z_1, \ldots, Z_n$ with a density supported in $[-T, T]$ (see Fig. 1).

In the following simulation study, we have first analyze the performances of our adaptive wavelet estimation procedure on a family of normal mixture densities.

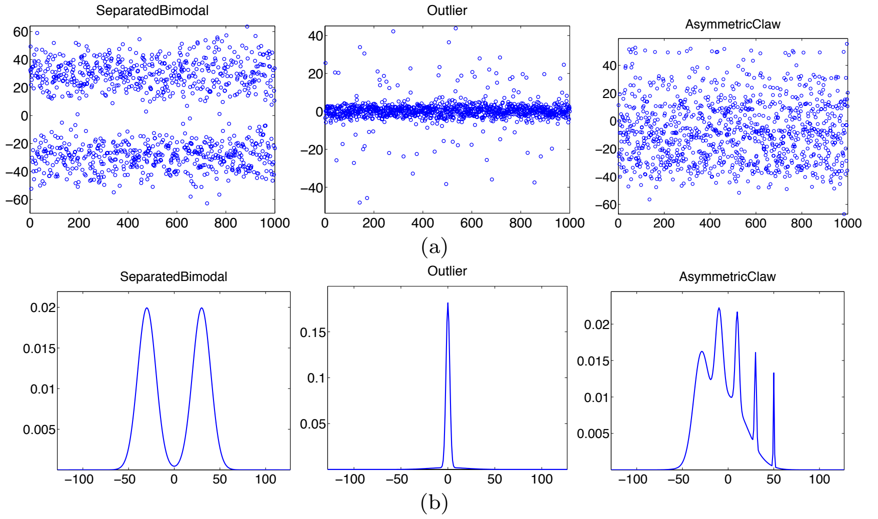
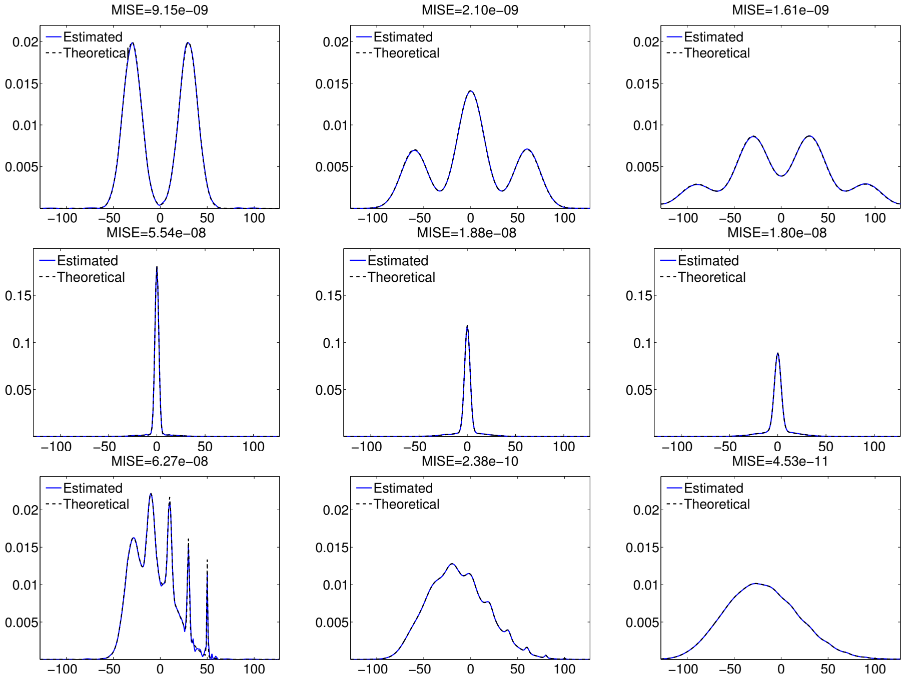


Fig. 1 Original (a) observations and (b) densities from $n = 1000$ samples $Z_1, \dots, Z_n$; MixtGauss (left), Claw (middle) and Discrete Comb (right).

$$\begin{aligned} \mathit{SeparatedBimodal}: &\ \frac{1}{2}\mathcal{N}\left(-1/3, (1/2)^2\right) + \frac{1}{2}\mathcal{N}\left(1/2, (3/2)^2\right); \\ \mathit{Outlier}: &\ \frac{1}{10}\mathcal{N}(0,1) + \frac{9}{10}\mathcal{N}\left(0, (1/10)^2\right); \\ \mathit{AsymmetricClaw}: &\ \frac{1}{2}\mathcal{N}(0,1) + \sum_{l=-2}^{2} \frac{2^{1-l}}{31}\mathcal{N}\left(l + 1/2, (2^{-l}/10)^2\right). \end{aligned}$$

These densities exhibit various behaviours going from smooth to non homogeneous. They were first introduced by Marron and Wand (1992). Note that AsymmetricClaw represents a strongly multimodal density and will be hard to estimate in full with a small sample size. The SeparatedBimodal density was used to illustrate the performances of our estimator for smooth density.

Since our estimation method is adaptive, we have chosen a predetermined threshold κ (universal thresholding, see e.g. Donoho *et al.* (1996)) for all the tests and the Symmlet wavelet with 6 vanishing moments was used throughout all experiments. The index of the highest resolution space was chosen to be $J = 8$ so that a total of $M = 256$ empirical coefficients were computed for $l = -M/2, \dots, M/2 - 1$. For each density, $n = 10^4$ independent samples of size $M = 256$ (i.e. $T = M/2$) were generated and the MISE was approximated as an average of the Integrated Squared Error (ISE) over 10 replications. The results are depicted in Fig. 2 for $m = 1$, $m = 2$ and $m = 3$ respectively. One can see that our adaptive hard thresholding estimator is very effective to estimate each of the nine densities. It has good adaptiveness properties and can recover smooth density such as the SeparatedBimodal as well as non homogeneous density such as the AsymmetricClaw density.



(a): $m = 1$. (b): $m = 2$. (c): $m = 3$.

Fig. 2 Original (dashed black) and estimated density (solid blue) from 10 replications using our wavelet hard thresholding estimator $\hat{g}$ from $n = 10^4$ samples $Z_1, \dots, Z_n$. From top to bottom SeparatedBimodal, Outlier, AsymmetricClaw for (a): $m = 1$, (b): $m = 2$, (c): $m = 3$.

Then we have compared the performance of our adaptive wavelet estimator to those of two different kernel-based estimators. The first one, presented in Saavedra and Cao (2000), is based on convolving kernel density estimators: $\hat{g} = \bigstar_m \hat{f}$, where $\hat{f}$ denotes a kernel estimator. The other one, introduced by Frees (1994), is the Frees type local U-statistic estimator defined as follow

$$\hat{g}(x) = \frac{1}{\binom{n}{m} b} \sum_{(n,m)} K\left(\frac{x - h(Z_{i_1}, \dots, Z_{i_m})}{b}\right), \tag{11}$$

where b is the bandwidth or smoothing parameter, K is a kernel function and $\sum_{(n,m)}$ denotes summation over all $\binom{n}{m}$ subsets. We have focused here on the interesting case where $h(Z_1, \dots, Z_m) = \sum_{i=1}^{m} Z_i$ (see Frees (1994) for applications). These two estimators are based on convolving kernel density estimators and for $m = 1$ they correspond to the classical kernel estimator $\hat{f}$

$$\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - Z_i}{b}\right). \tag{12}$$

In the sequel, we name the estimator of Saavedra and Cao (2000) by 'Kernel', the one of Frees (1994) by 'Frees' and our estimator by 'Wavelet'.

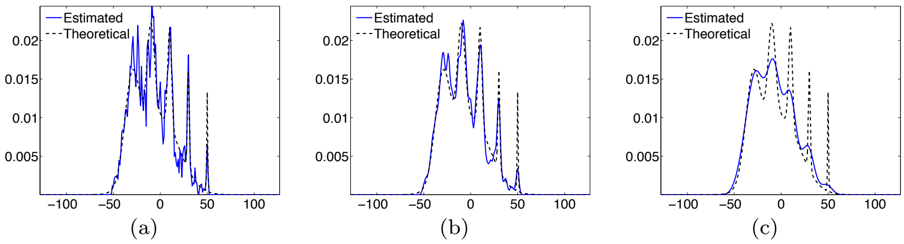


Fig. 3 Original (dashed black) and estimated density (solid blue) using different Gaussian kernel density estimators from $n = 10^3$ samples $Z_1, \ldots, Z_n$ with bandwidths (a): $b = 0.5$. (b): $b = 1.32$. (c): $b = 5$.

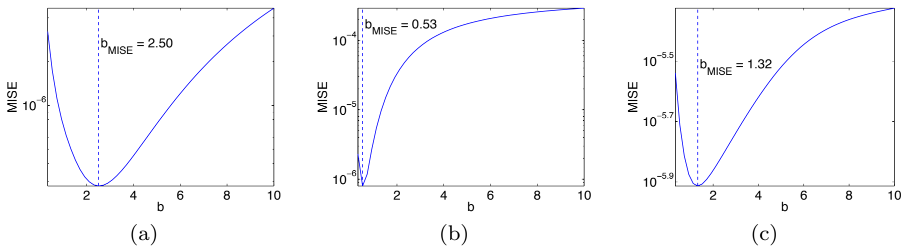


Fig. 4 MISE values as a function of b (from 1 replication) for (a): SeparatedBimodal. (b): Outlier. (c): AsymmetricClaw from $n = 10^3$ samples. The vertical dashed line represents the minimizer of the MISE.

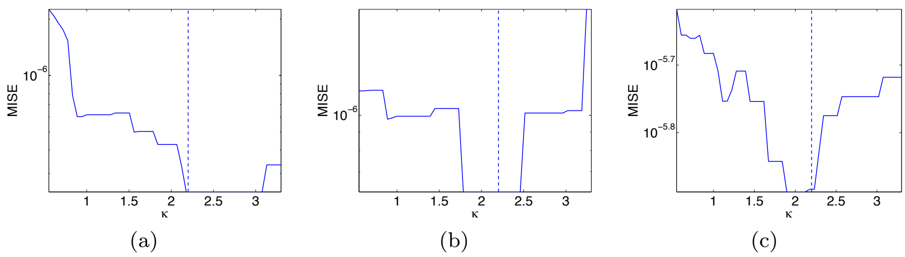


Fig. 5 MISE values as a function of the threshold parameter κ (from 1 replication) for (a): SeparatedBimodal. (b): Outlier. (c): AsymmetricClaw from $n = 1000$ samples. The vertical dashed line represents our predetermined threshold.

In the case of *i.i.d.* random variables, the choice of the kernel is not crucial for density estimation. However, it is well known that the choice of the bandwidth is very important. To illustrate this basic problem of the kernel estimate, we showed three different kernel density estimators with bandwidths $b = 0.5$, $b = 1.32$ and $b = 5$ in Fig. 3. When the bandwidth is too large ($b = 5$), the modes of the AsymmetricClaw density are not well recovered. On the other hand, the smaller bandwidth estimate ($b = 0.5$) correctly identified all the modes, but the result is under smoothed since it contains too much spurious artifacts. In comparison to our adaptive hard thresholded wavelet estimator of Fig. 2 (bottom left), the kernel estimates are disadvantageous. Indeed, the wavelet density estimate captures all of the modes of the AsymmetricClaw.

1.0e-06× *SeparatedBimodal*, $m = 2$

n	10	20	50	100
Wavelet	21.55	8.307	3.492	2.424
Kernel	6.815	4.730	1.612	1.221
Frees	7.334	4.491	1.416	0.983
b_{MISE}				
Kernel	10.15	7.27	6.21	5.30
Frees	11.97	8.48	6.66	4.09

1.0e-06× *SeparatedBimodal*, $m = 3$

n	10	20	50	100
Wavelet	14.67	5.707	2.733	2.471
Kernel	4.572	3.331	1.088	0.839
Frees	5.953	3.336	0.971	0.676
b_{MISE}				
Kernel	10.15	7.27	6.21	5.30
Frees	14.55	10.30	7.57	3.93

1.0e-05× *Outlier*, $m = 2$

n	10	20	50	100
Wavelet	5.360	4.783	1.525	1.157
Kernel	3.367	3.181	1.334	1.218
Frees	3.219	2.954	1.256	1.168
b_{MISE}				
Kernel	1.51	1.51	1.05	0.60
Frees	2.12	1.81	1.21	0.75

1.0e-05× *Outlier*, $m = 3$

n	10	20	50	100
Wavelet	6.274	3.940	1.685	1.232
Kernel	3.131	2.882	1.437	1.243
Frees	3.204	2.753	1.382	1.203
b_{MISE}				
Kernel	1.51	1.51	1.05	0.60
Frees	3.03	1.96	1.51	0.90

1.0e-06× *AsymmetricClaw*, $m = 2$

n	10	20	50	100
Wavelet	4.106	5.539	1.108	1.206
Kernel	1.354	3.927	0.917	1.109
Frees	1.245	3.442	0.866	0.918
b_{MISE}				
Kernel	16.54	12.26	8.61	6.71
Frees	20.03	19.24	12.74	14.17

1.0e-06× *AsymmetricClaw*, $m = 3$

n	10	20	50	100
Wavelet	3.838	6.288	1.189	1.242
Kernel	1.416	4.584	0.937	1.126
Frees	1.703	4.387	0.875	0.955
b_{MISE}				
Kernel	16.54	12.26	8.61	6.71
Frees	23.37	22.73	15.12	18.13

Table 1 MISE from 10 replications for each method ('Kernel', 'Frees' and 'Wavelet' denote the estimators of Saavedra and Cao (2000), Frees (1994) and our respectively). From top to bottom SeparatedBimodal, Outlier, AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right).

Many procedures of bandwidth selection for density estimation have been developed in the literature, but here, we have been focused on a global bandwidth selection based on minimizing the MISE of $\hat{g}$ for both kernel-based estimators (details can be found in e.g. Mugdadi and Ahmad (2004) where several bandwidth selection procedures were used). Fig. 4 shows how the MISE depend on the bandwidth for the three densities for $n = 10^3$. To illustrate the adaptiveness properties of our estimator we display the empirical MISE as a function of the threshold κ in Fig. 5, where the vertical dashed line represents the threshold parameter that we used throughout these simulations. One can see that the minimum of the MISE occurs around this threshold for the three densities and this also applies to other samples sizes (see Fig. 6).

We evaluated the three procedures on small to medium sample. Each method was applied to $n = 10, 20, 50, 100$ data points of each of the densities and we have investigated the binning into $M = 256$ binpoints. The Gaussian kernel with an optimal bandwidth (i.e. which minimizes the MISE) was used through this experiment for both kernel-based methods. The MISE from 10 replications are tabulated in Table 1. Frees method outperforms in almost all cases our and Saavedra and Cao (2000) methods in this example, especially for the smooth density (SeparatedBimodal). Note that we intentionally chose n small sample to make the comparison between these three methods. Indeed,

$m = 2$

n	10	20	50	100
Wavelet	0.01	0.01	0.01	0.02
Kernel	0.04	0.04	0.04	0.05
Frees	0.17	0.22	0.53	1.38

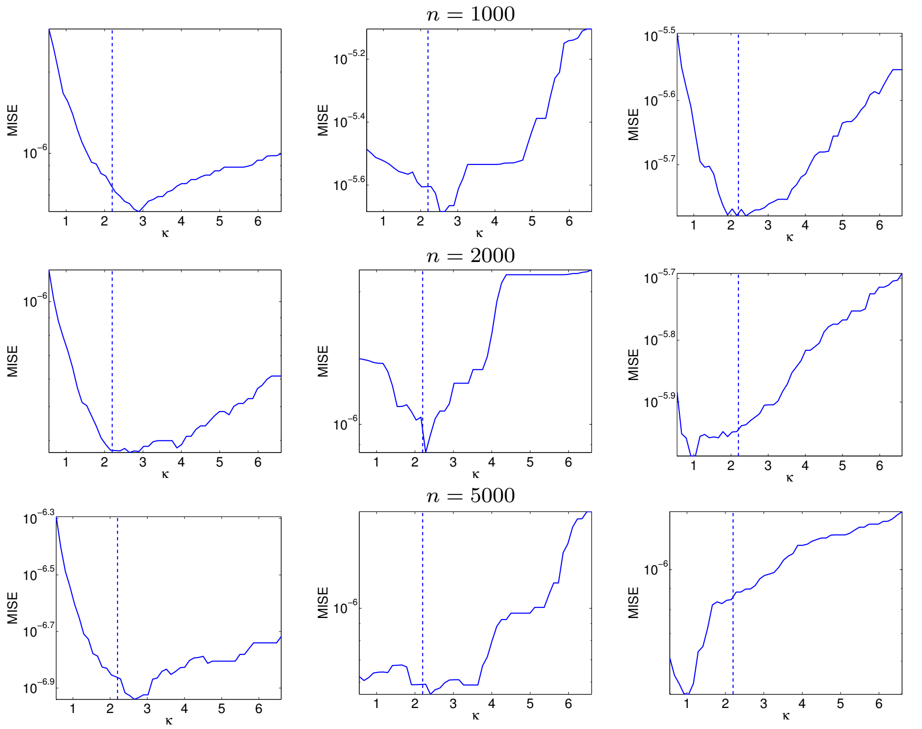
$m = 3$

n	10	20	50	100	200
Wavelet	0.01	0.01	0.01	0.02	0.02
Kernel	0.04	0.04	0.04	0.05	0.05
Frees	0.37	1.30	9.62	143	2037

Table 2 Execution times in seconds for $m = 2$ and $m = 3$ (from only one replication). The algorithms were run under Matlab with an Intel Core 2 duo 3.06GHz CPU, 4Gb RAM.

even if the Frees estimator provided accurate estimation, competitive to our estimator, the computational cost is very expensive and it would be technically impossible to compute it on a large sample. Table 2 reports the average execution times in seconds for $m = 2$ and $m = 3$ (i.e. $h(Z_1, Z_2) = Z_1 + Z_2$ and $h(Z_1, Z_2, Z_3) = Z_1 + Z_2 + Z_3$). For the Frees estimator, the computational cost increases dramatically as far as the sampling parameter n increases and during the computation to estimate the density of the sum of more than two *i.i.d.* random variables (see Table 2: $m = 3$). From a practical point of view, unlike Frees estimator, our method can easily be computed for $m > 3$. This is an obvious consequence of the $\binom{n}{m}$ evaluations of a kernel associated to the usual kernel density estimation procedure. Further economies would be possible if resampling methodology were available. Note that the additional computational burden of the convolution product is marginal for both methods based on such approach (due to $\mathcal{O}(n \ln_2 n)$ computational complexity of the FFT algorithm). Our procedure has a much lower computational cost than Frees and is comparable to Kernel as can be seen from Table 2.

We conclude this section by a comparison to the natural kernel plug-in estimator of Saavedra and Cao (2000) on larger samples ($n = 1000, 2000, 5000$ and $M = 256$ binpoints). The practical effectiveness of several classical kernels are also investigated. The values of the MISE were calculated from 10 replications and tabulated in Table 3. Our wavelet method is slightly better than the Kernel one in almost all cases but none of them clearly outperforms the others for all tests densities and all sample. For the AsymmetricClaw, the estimator of Saavedra and Cao (2000) marginally outperforms our adaptive wavelet estimator (see Fig. 6:(c)) . Furthermore, it is obvious that in all the cases the MISE is decreasing as the sample size is increasing. For each kernel type the results are very close with a slight advantage to the Epanechnikov kernel estimate for large sample. Without any prior smoothness knowledge on the unknown density, our adaptive estimator provides very competitive results in comparison of all three kernels used to these sample.



(a) (b) (c)

Fig. 6 MISE values as a function of the threshold parameter κ from 10 replications for (a): SeparatedBimodal. (b): Outlier. (c): AsymmetricClaw from top to bottom $n = 1000, 2000, 5000$ samples. The vertical dashed line represents our predetermined threshold.

1.0e-08× *SeparatedBimodal*, $m = 2$

n	1000	2000	5000
Wavelet			
Hard	11.513	4.563	2.491
Kernel			
Normal	13.519	6.434	4.188
Epanechnikov	13.414	6.467	4.085
Biweight	13.301	6.452	4.125
b_{MISE}			
Normal	3.18	2.59	2.26
Epanechnikov	7.03	5.81	4.98
Biweight	8.27	6.86	5.93

1.0e-08× *SeparatedBimodal*, $m = 3$

n	1000	2000	5000
Wavelet			
Hard	8.538	3.029	1.749
Kernel			
Normal	9.916	4.442	2.830
Epanechnikov	9.842	4.458	2.765
Biweight	9.781	4.449	2.790
b_{MISE}			
Normal	3.18	2.59	2.26
Epanechnikov	7.03	5.81	4.98
Biweight	8.27	6.86	5.93

1.0e-07× *Outlier*, $m = 2$

n	1000	2000	5000
Wavelet			
Hard	8.702	2.669	1.333
Kernel			
Normal	10.180	3.460	2.856
Epanechnikov	9.401	3.068	1.521
Biweight	9.642	3.097	1.662
b_{MISE}			
Normal	0.58	0.52	0.50
Epanechnikov	1.28	1.08	0.77
Biweight	1.51	1.31	0.93

1.0e-07× *Outlier*, $m = 3$

n	1000	2000	5000
Wavelet			
Hard	8.251	2.523	1.501
Kernel			
Normal	8.667	2.892	2.505
Epanechnikov	8.148	2.650	1.595
Biweight	8.311	2.648	1.730
b_{MISE}			
Normal	0.58	0.52	0.50
Epanechnikov	1.28	1.08	0.77
Biweight	1.51	1.31	0.93

1.0e-08× *AsymmetricClaw*, $m = 2$

n	1000	2000	5000
Wavelet			
Hard	6.819	3.493	2.172
Kernel			
Normal	6.200	3.342	2.153
Epanechnikov	6.239	3.363	2.077
Biweight	6.227	3.352	2.163
b_{MISE}			
Normal	1.41	0.95	0.66
Epanechnikov	2.90	1.93	1.28
Biweight	3.51	2.34	1.55

1.0e-08× *AsymmetricClaw*, $m = 3$

n	1000	2000	5000
Wavelet			
Hard	5.292	3.128	1.732
Kernel			
Normal	5.224	3.103	1.726
Epanechnikov	5.292	3.128	1.732
Biweight	5.222	3.108	1.737
b_{MISE}			
Normal	1.41	0.95	0.66
Epanechnikov	2.90	1.93	1.28
Biweight	3.51	2.34	1.55

Table 3 MISE from 10 replications. From top to bottom SeparatedBimodal, Outlier and AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right). MISE optimal bandwidth b_{MISE}.

5 Application II: the deconvolution density model

5.1 Upper bound

We observe n *i.i.d.* random variables $Z_1, \ldots, Z_n$ where, for any $i \in \{1, \ldots, n\}$,

$$Z_i = X_i + \epsilon_i, \tag{13}$$

$X_1, \ldots, X_n$ are *i.i.d.* random variables and $\epsilon_1, \ldots, \epsilon_n$ are *i.i.d.* random variables. Classically, $X_1, \ldots, X_n$ are measurements of some characteristic of interest contaminated by noise represented by $\epsilon_1, \ldots, \epsilon_n$. For any $i \in \{1, \ldots, n\}$, X_i and ϵ_i are independent. The density of X_1 is unknown and denoted f, whereas the one of ϵ_1 is known and denoted h. For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^{m} X_i$ and g be the density of S. The goal is to estimate g from $Z_1, \ldots, Z_n$. This problem can be viewed as a generalization of the standard deconvolution density one which corresponds to $m = 1$. See e.g. Caroll and Hall (1988), Fan (1991), Fan and Liu (1997), Pensky and Vidakovic (1999), Fan and Koo (2002), Butucea and Matias (2005), Comte *et al.* (2006), Delaigle and Gijbels (2006) and Lacour (2006). However, to the best of our knowledge, the general problem i.e. with $m \geq 2$ is a new challenge.

Proposition 2 below investigates the rates of convergence of $\hat{g}$ (2) constructed from a specific wavelet hard thresholding estimator $\hat{f}$ under the MISE over Besov balls.

Proposition 2 *Consider* (13) *and he associated notations. We define the Fourier transform of an integrable function* u *by* $\mathcal{F}(u)(x) = \int_{-\infty}^{\infty} u(y)e^{-ixy}dy$, $x \in \mathbb{R}$. *The notation* $\overline{\cdot}$ *will be used for the complex conjugate.*

Suppose that $\operatorname{supp} f \subseteq [-T, T]$, *where* $T > 0$ *is a fixed constant, and there exist three constants* $C > 0$, $c > 0$ *and* $\delta > 1$ *such that*

$$\sup_{x \in \mathbb{R}} h(x) \leq C, \qquad |\mathcal{F}(h)(x)| \geq \frac{c}{(1 + x^2)^{\delta/2}}, \qquad x \in \mathbb{R}. \tag{14}$$

Let $\hat{g}$ *be* (2) *with* $\hat{f}$ *as in* (9) *with*

$$\hat{\alpha}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\phi_{j,k})(x)}}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx \tag{15}$$

and

$$\hat{\beta}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\psi_{j,k})(x)}}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx. \tag{16}$$

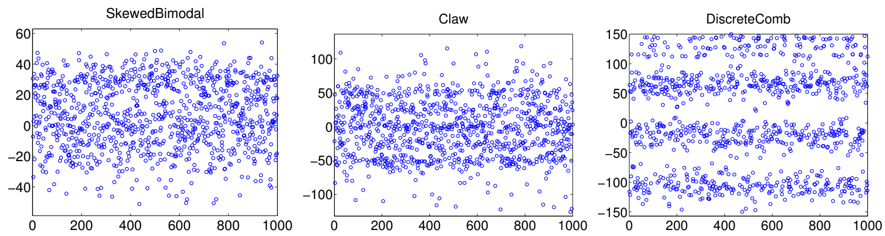
Then, for any $r \geq 1$, *any* $\{p \geq 4$ *and* $s > 0\}$ *or any* $\{p \in [1, 4)$ *and* $s > \max((2\delta+1)/p, (1/p)(4-p)(\delta+1/2)\}$, *there exists a constant* $C > 0$ *such that*

$$\sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C \left(\frac{\ln n}{n}\right)^{2s/(2s+2\delta+1)}.$$

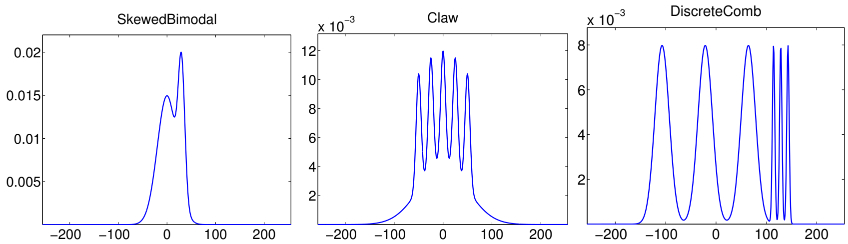
The rate of convergence $(\ln n/n)^{2s/(2s+2\delta+1)}$ corresponds to the "near optimal" one in the minimax sense when $m = 1$. See (Fan and Koo 2002, Theorem 2).

5.2 Simulation study

In this simulation, $n = 10^4$ samples $Z_1, \ldots, Z_n$ were generated according to model (13) and we considered Laplace errors (which respect the standard



(a)



(b)

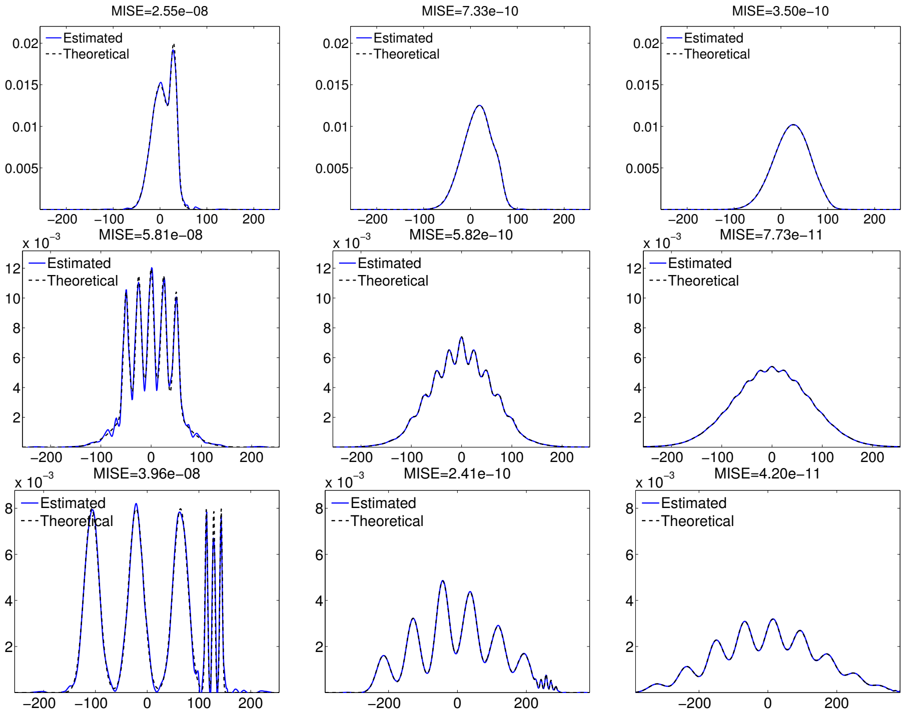
Fig. 7 Noisy samples $Z_1, \ldots, Z_n$ according to (13) from $n = 1000$, where $\epsilon_i \sim \mathcal{L}(0, 1)$. SkewedBimodal (left), Claw (middle) and DiscreteComb (right).

ordinary smooth assumption). The data sets used in this deconvolution study are also normal mixture densities

$$\begin{aligned}
\textit{SkewedBimodal}: &\ \frac{3}{4}\mathcal{N}(0,1) + \frac{1}{4}\mathcal{N}\left(3/2, (1/3)^2\right);\\
\textit{Claw}: &\ \frac{1}{2}\mathcal{N}(0,1) + \sum_{l=0}^{4} \frac{1}{10}\mathcal{N}\left(l/2-1, (1/2)^2\right);\\
\textit{DiscreteComb}: &\ \sum_{l=0}^{2} \frac{2}{7}\mathcal{N}\left((12l-15)/7, (2/7)^2\right) + \sum_{l=8}^{10} \frac{1}{21}\mathcal{N}\left(2l/7, (1/21)^2\right).
\end{aligned}$$

These densities exhibit different types of smoothness going from smooth to non homogeneous densities. They were initially introduced by Marron and Wand (1992). Claw and DiscreteComb represent strongly multimodal densities and will be hard to estimate them in full with classic methods. The Skewed-Bimodal density was used to illustrate the performances of our estimator for smooth density.

Fig. 8 shows the results of $\hat{g}$ for $m = 1$, $m = 2$ and $m = 3$ respectively. Clearly, for these nine densities, our adaptive hard thresholding estimator is very effective.



(a): $m = 1$. (b): $m = 2$. (c): $m = 3$.

Fig. 8 Original (dashed black) and estimated density (solid blue) from 10 replications using our wavelet hard thresholding estimator $\hat{g}$ from $n = 10^4$ samples $Z_1, \ldots, Z_n$ generated according to (13), where $h(x) = (1/2)e^{-|x|}$, $x \in \mathbb{R}$. (a): $m = 1$. (b): $m = 2$. (c): $m = 3$. (note that (14) is satisfied with $\delta = 2$).

Conclusion and perspectives

The agreement of our simulations with our theoretical findings show the relevance of our estimator in the context of two classical density estimation problems. The practical comparisons to state-of-the art methods such as the estimator of Frees (1994) or the one of Saavedra and Cao (2000) have demonstrated the usefulness and the efficiency of adaptive thresholding methods in estimating densities of the sum of random variables. It would be interesting to include both theoretical and practical comparisons with other wavelet thresholding estimators as the block thresholding one (see e.g. Cai (1999) and Chesneau *et al.* (2010)). Another interesting perspective would be to extend our results to a multidimensional setting. These aspects need further investigations that we leave for a future work.

6 Proofs

Lemma 1 *For any positive integer m and $(u,v)\in\mathbb{C}^2$, we have*

$$|u^m - v^m| \le |u-v|(|u|+|v|)^{m-1}.$$

Proof of Lemma 1. The factor theorem yields

$$u^m - v^m = (u-v)\sum_{k=0}^{m-1} v^k u^{(m-1)-k}.$$

It follows from the triangular inequality, $\binom{m-1}{k}\ge 1$, $k\in\{0,\dots,m-1\}$, and the binomial theorem that

$$\begin{aligned}|u^m - v^m| &\le |u-v|\sum_{k=0}^{m-1}|v|^k|u|^{(m-1)-k}\\ &\le |u-v|\sum_{k=0}^{m-1}\binom{m-1}{k}|v|^k|u|^{(m-1)-k} = |u-v|(|u|+|v|)^{m-1}.\end{aligned}$$

Lemma 1 is proved.

□

Proof of Theorem 1. Let us define the Fourier transform of an integrable function u by

$$\mathcal{F}(u)(y) = \int_{-\infty}^{\infty} u(x)e^{-iyx}dx, \qquad y\in\mathbb{R}.$$

The Parseval theorem, $\mathcal{F}(\hat{g})(y) = (\mathcal{F}(\hat{f})(y))^m$ and $\mathcal{F}(g)(y) = (\mathcal{F}(f)(y))^m$ yield

$$\begin{aligned}&\int |\hat{g}(x)-g(x)|^2dx = \frac{1}{2\pi}\int |\mathcal{F}(\hat{g}-g)(y)|^2dy\\ &= \frac{1}{2\pi}\int |\mathcal{F}(\hat{g})(y)-\mathcal{F}(g)(y)|^2dy = \frac{1}{2\pi}\int |(\mathcal{F}(\hat{f})(y))^m - (\mathcal{F}(f)(y))^m|^2dy.\end{aligned}$$

Applying Lemma 1 with $u=\mathcal{F}(\hat{f})(y)$ and $v=\mathcal{F}(f)(y)$, by (3), $|\mathcal{F}(f)(y)|\le \int|f(x)|dx\le C_1$, $|\mathcal{F}(\hat{f})(y)|\le\int|\hat{f}(x)|dx\le C_2$, and using again the Parseval theorem, we obtain

$$\begin{aligned}&\int |\hat{g}(x)-g(x)|^2dx\\ &\le \frac{1}{2\pi}\int |\mathcal{F}(\hat{f})(y)-\mathcal{F}(f)(y)|^2\left(|\mathcal{F}(\hat{f})(y)|+|\mathcal{F}(f)(y)|\right)^{2m-2}dy\\ &\le (C_1+C_2)^{2m-2}\frac{1}{2\pi}\int|\mathcal{F}(\hat{f}-f)(y)|^2dy = C_3\int|\hat{f}(x)-f(x)|^2dx.\end{aligned} \tag{17}$$

Therefore

$$\mathbb{E}\left(\int |\hat{g}(x)-g(x)|^2dx\right) \leq C_3\mathbb{E}\left(\int |\hat{f}(x)-f(x)|^2dx\right).$$

Theorem 1 is proved.

□

Proof of Theorem 2. Proceeding as in (17), using the triangular inequality, $|\mathcal{F}(\hat{f}-f)(y)| \leq \int |\hat{f}(x)-f(x)|dx$, by (4), $|\mathcal{F}(f)(y)| \leq \int |f(x)|dx \leq C_1$ and the Parseval theorem, we have

$$\begin{aligned}
&\int |\hat{g}(x)-g(x)|^2dx \\
&\leq \frac{1}{2\pi}\int |\mathcal{F}(\hat{f})(y)-\mathcal{F}(f)(y)|^2\left(|\mathcal{F}(\hat{f})(y)|+|\mathcal{F}(f)(y)|\right)^{2m-2}dy \\
&\leq \frac{1}{2\pi}\int |\mathcal{F}(\hat{f})(y)-\mathcal{F}(f)(y)|^2\left(|\mathcal{F}(\hat{f})(y)-\mathcal{F}(f)(y)|+2|\mathcal{F}(f)(y)|\right)^{2m-2}dy \\
&= \frac{1}{2\pi}\int |\mathcal{F}(\hat{f}-f)(y)|^2\left(|\mathcal{F}(\hat{f}-f)(y)|+2|\mathcal{F}(f)(y)|\right)^{2m-2}dy \\
&\leq \left(\int |\hat{f}(x)-f(x)|dx+2C_1\right)^{2m-2}\frac{1}{2\pi}\int |\mathcal{F}(\hat{f}-f)(y)|^2dy \\
&= A\times B,
\end{aligned}$$

where

$$A=\left(\int |\hat{f}(x)-f(x)|dx+2C_1\right)^{2m-2}, \qquad B=\int |\hat{f}(x)-f(x)|^2dx.$$

Hence, by the Cauchy-Schwarz inequality,

$$\mathbb{E}\left(\int |\hat{g}(x)-g(x)|^2dx\right) \leq \left(\mathbb{E}(A^2)\right)^{1/2}\left(\mathbb{E}(B^2)\right)^{1/2}.$$

Let us now bound $\mathbb{E}(A^2)$ and $\mathbb{E}(B^2)$.

Using $|x+y|^a \leq 2^{a-1}(|x|^a+|y|^a)$, $(x,y)\in\mathbb{R}^2$, $a\geq 1$, the Hölder inequality and supp $f\subseteq[-T,T]$, we have

$$\begin{aligned}
A^2 &\leq 2^{4m-5}\left(\left(\int |\hat{f}(x)-f(x)|dx\right)^{4m-4}+2^{4m-4}C_1^{4m-4}\right) \\
&\leq 2^{4m-5}\left((2T)^{4m-5}\int |\hat{f}(x)-f(x)|^{4m-4}dx+2^{4m-4}C_1^{4m-4}\right).
\end{aligned}$$

So, by (4),

$$\begin{aligned}
\mathbb{E}(A^2) &\leq 2^{4m-5}\left((2T)^{4m-5}\mathbb{E}\left(\int |\hat{f}(x)-f(x)|^{4m-4}dx\right)+2^{4m-4}C_1^{4m-4}\right) \\
&\leq 2^{4m-5}\left((2T)^{4m-5}C_2+2^{4m-4}C_1^{4m-4}\right).
\end{aligned}$$

It follows from the Hölder inequality and $\operatorname{supp} f \subseteq [-T, T]$ that

$$B \leq \sqrt{2T}\sqrt{\int |\hat{f}(x) - f(x)|^4 dx}.$$

Therefore

$$\mathbb{E}(B^2) \leq 2T\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^4 dx\right).$$

Hence

$$\mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx\right) \leq C_3\sqrt{\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^4 dx\right)},$$

where

$$C_3 = \sqrt{2^{4m-4}T\left((2T)^{4m-5}C_2 + 2^{4m-4}C_1^{4m-4}\right)}.$$

Theorem 2 is proved.

□

Proof of Theorem 3. We aim to apply Theorem 2. Let us investigate the assumption (4).

Since $\operatorname{supp} f \subseteq [-T, T]$ and $\int |f(x)|^{4m-4} dx \leq C$, the Cauchy-Schwarz inequality yields the first condition.

Observe that

$$|\hat{\beta}_{j,k}\mathbf{1}_{\left\{|\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j}\sqrt{\ln n/n}\right\}} - \beta_{j,k}| \leq |\hat{\beta}_{j,k} - \beta_{j,k}| + |\beta_{j,k}|,$$

It follows from (8) that

$$\begin{aligned}
&\mathbb{E}\left(|\hat{\beta}_{j,k}\mathbf{1}_{\left\{|\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j}\sqrt{\ln n/n}\right\}} - \beta_{j,k}|^{4m-4}\right) \\
&\leq 2^{4m-5}\left(\mathbb{E}(|\hat{\beta}_{j,k} - \beta_{j,k}|^{4m-4}) + |\beta_{j,k}|^{4m-4}\right) \\
&\leq C\left(2^{(4m-4)\delta j}\left(\frac{\ln n}{n}\right)^{2m-2} + |\beta_{j,k}|^{4m-4}\right).
\end{aligned}$$

Using this inequality, (6), (7) and arguing similarly to (Kerkyacharian and Picard 2000, Theorem 5.1), we obtain

$$\begin{aligned}
&\mathbb{E}\left(\int |\hat{f}(x) - f(x)|^{4m-4} dx\right) \\
&\leq C\left(\left(\frac{\ln n}{n}\right)^{2m-2} 2^{(1+2\delta)j_1(2m-2)} + \int |f(x)|^{4m-4} dx\right) \leq C.
\end{aligned}$$

We now need a consequence of (Kerkyacharian and Picard 2000, Theorem 5.1) formulated below (see also (Johnstone *et al.* 2004, Proposition 1)).

Theorem 4 (Kerkyacharian and Picard (2000)) *Let $\hat{f}$ be* (9) *(under* (6), (7) *and* (8)*). Then, for any $\theta > 1$, any $r \geq 1$, any $\{p \geq \theta$ and $s > 0\}$ or any $\{p \in [1, \theta)$ and $s > \max((2\delta+1)/p, (1/p)(\theta - p)(\delta + 1/2)\}$, there exists a constant $C > 0$ such that*

$$\sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{f}(x) - f(x)|^\theta dx \right) \leq C \left(\frac{\ln n}{n} \right)^{\theta s/(2s+2\delta+1)}.$$

Thanks to Theorem 4 with $\theta = 4$, we have

$$\sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{f}(x) - f(x)|^4 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{4s/(2s+2\delta+1)}.$$

It follows from Theorem 2 that

$$\begin{aligned} \sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{g}(x) - g(x)|^2 dx \right) &\leq C \sqrt{\sup_{f \in B^s_{p,r}(M)} \mathbb{E}\left(\int |\hat{f}(x) - f(x)|^4 dx \right)} \\ &\leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}. \end{aligned}$$

This ends the proof of Theorem 3.

□

Proof of Proposition 1. Thanks to (Donoho *et al.* 1996, Subsection 5.1.1, (16) and (17)), under the assumptions $\operatorname{supp} f \in [-T, T]$, the estimators $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ (10) satisfy (6), (7) and (8) with $\delta = 0$. The rest of the proof follows from Theorem 3.

□

Proof of Proposition 2. In a similar fashion to (Fan and Koo 2002, E. Proof of Theorem 7), under the assumptions $\operatorname{supp} f \subseteq [-T, T]$ and (14), we prove that the estimators $\hat{\alpha}_{j,k}$ (15) and $\hat{\beta}_{j,k}$ (16) satisfy (6), (7) and (8) with the same δ. We obtain the desired result via Theorem 3.

□

Acknowledgements: This work is supported by ANR grant NatImages, ANR-08-EMER-009. We thanks Jalal Fadili for his suggestions which lead to the improved version of the paper.

References

Ahmad, I.A. and Fan, Y. (2001). Optimal bandwidth for kernel density estimators of functions of observations. *Statist. Probab. Lett.*, 51, (3), 245-251.

Ahmad, I.A. and Mugdadi, A.R. (2003). Analysis of kernel density estimation of functions of random variables. *J. Nonparametric Statistics*, 15, 579-605.

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

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model. *Bernoulli*, 11, **2**, 309-340.

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

Caroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83, 1184-1186.

Chesneau C., Fadili M.J. and Starck J.-L. (2010). Stein Block Thresholding For Image Denoising , *Applied and Computational Harmonic Analysis*, 28, 1, 67-88.

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.

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *The Canadian Journal of Statistics*, 34, 431-452.

Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica*, 16, 773 -788.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics*, 24, 2, 508-539.

Du, J. and Schick, A. (2007). Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Systems*, 2, 67-87.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.*, 19, 1257-1272.

Fan, J. and Liu, Y. (1997). A note on asymptotic normality for deconvolution kernel density estimators. *Sankhya*, 59, 138-141.

Fan, J. and Koo, J.Y. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, 48, 734-747.

Frees, E. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89, 17-525.

Giné, E. and Mason, D.M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, 35, 1105-1145.

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. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, 9, 2, 283-345.

Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris Ser. I Math.*, 342 (11), 877-882.

Mallat, S. (2009). A wavelet tour of signal processing. Elsevier/ Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.

Marron, J.S. and Wand, M.P. (1992). Exact Mean Integrated Squared Error. *The Annals of Statistics*, 20, 2, 712-736.

Meyer, Y. (1992). Wavelets and Operators. *Cambridge University Press*, Cambridge.

Mugdadi, A.R. and Ahmad, I. (2004). A Bandwidth Selection for Kernel Density Estimation of Functions of Random Variables. *Computational Statistics and Data Analysis*, 47,1, 49-62.

Panjer, H.H. and Willmot, G.E. (1992). Insurance Risk Models. Society of Actuaries, Schaumburg.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27, 2033–2053.

Prakasa Rao, B.L.S. (2004). Moment inequalities for supremum of empirical processes of U-statistic structure and application to density estimation, *J.Iran. Statist. Soc.*, 3, 59-68.

Johnstone, I., Kerkyacharian, G., Picard, D., Raimondo, M., 2004. Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society, Serie B*, 66, (3), 547-573.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, 28, 799-815.

Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, 16, 925-935.

Schick, A. and Wefelmeyer, W. (2007). Root n consistent density estimators of convolutions in weighted $\mathbb{L}_1$-norms. *J. Statist. Plann. Inference*, 137, 1765-1774.

Tsybakov, A. (2004). *Introduction à l'estimation nonparamétrique*. Springer Verlag, Berlin.
Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc., New York, 384 pp.