HAL open science

Adaptive density estimation under dependence

Irène Gannaz, Olivier Wintenberger

To cite this version:

Irène Gannaz, Olivier Wintenberger. Adaptive density estimation under dependence. ESAIM: Probability and Statistics, 2010, 14, pp.151-172. 10.1051/ps:2008025 . hal-00012077v3

HAL Id: hal-00012077

https://hal.science/hal-00012077v3

Submitted on 15 Oct 2008

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

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

ADAPTIVE DENSITY ESTIMATION UNDER WEAK DEPENDENCE

GANNAZ IRÈNE[(1)] AND WINTENBERGER OLIVIER[(2)]

Abstract. Assume that $(X_t)_{t\in\mathbb{Z}}$ is a real valued time series admitting a common marginal density f with respect to Lebesgue's measure. Donoho *et al.* (1996) propose near-minimax estimators $\widehat{f}_n$ based on thresholding wavelets to estimate f on a compact set in an independent and identically distributed setting. The aim of the present work is to extend these results to general weak dependent contexts. Weak dependence assumptions are expressed as decreasing bounds of covariance terms and are detailed for different examples. The threshold levels in estimators $\widehat{f}_n$ depend on weak dependence properties of the sequence $(X_t)_{t\in\mathbb{Z}}$ through the constant. If these properties are unknown, we propose cross-validation procedures to get new estimators. These procedures are illustrated via simulations of dynamical systems and non causal infinite moving averages. We also discuss the efficiency of our estimators with respect to the decrease of covariances bounds.

1. Introduction

Let $(X_t)_{t\in\mathbb{Z}}$ be a real valued time series admitting a common marginal density f that is compactly supported. The general purpose of this paper is to estimate f by wavelet estimators $\widehat{f}_n$ constructed from n observations $(X_1,\dots,X_n)$. In their seminal paper Donoho *et al.* (1996) (8) showed that projection-like linear estimators are not optimal: introduction of nonlinearity via thresholds of wavelet coefficients is investigated. Wavelets thresholding provides estimators which adapt themselves to the unknown smoothness of f, we refer to Vannucci (1998) (26) for a survey of the use of wavelet bases in density estimation. The present work extends near minimax results of soft and hard-threshold estimators from the independent and identically distributed (iid for short) framework, see Theorem 5 in Donoho *et al.* (1996) (8), to cases where weak dependence between variables occurs.

Our main assumptions give bounds for covariance terms as decreasing sequences which tend to zero when the gap between the past and the future of the time series goes to infinity. In order to give examples satisfying these conditions, we introduce coefficients that give bounds of covariance terms and that are computable for a large class of models. Weak dependent coefficients, introduced by Doukhan and Louhichi (1999) (9), as mixing ones, are well-adapted to that purpose. Using β-mixing coefficients Tribouley and Viennet (1998) (24) proposed minimax estimators with respect to the Mean Integrated Square Error (MISE for short). Comte and Merlevède (2002) (3) obtained near-minimax results using α-mixing coefficients. The loss of a logarithmic factor in the

1991 *Mathematics Subject Classification.* Primary 62G07; Secondary 60G10, 60G99, 62G20.

Key words and phrases. adaptive estimation, cross validation, hard thresholding, near minimax results, nonparametric density estimation, soft thresholding, wavelets, weak dependence.

convergence rate in this last paper is balanced by the generality of the context, as the class of α-mixing models is larger than the one of β-mixing models.

However, α and β-mixing coefficients are not easy to compute for some models and are useless for others; Andrews (1984) (1) proved that the mixing coefficients of the stationary solution of the AR(1) model

$$X_t = \frac{1}{2}\left(X_{t-1} + \xi_t\right), \text{ where } (\xi_t)_{t\in\mathbb{Z}} \text{ iid with a Bernouilli law of parameter } 1/2, \tag{1.1}$$

do not tend to zero as the gap from the past to the future of the time series goes to infinity. Mixing coefficients do not behave nicely in this case as, through a reversion of time, the Markov chain solution of (1.1) is a dynamical system, i.e. $X_{t-1} = T(X_t)$ for some transformation T, namely $T(x) = 2x\mathbb{1}_{0\le x<1/2} + (2x-1)\mathbb{1}_{1/2\le x\le 1}$. So called weak dependence coefficients have been recently developed to deal with such processes, see Dedecker *et al.* (5), Maume-Deschamps (20) and references therein. Introduced by Dedecker and Prieur (2005) in (6), $\widetilde{\phi}$-weak dependence coefficients give sharp bounds on the covariance terms of dynamical systems, such as the stationary solution to (1.1). Using these coefficients, we prove near-minimax results of thresholded wavelet estimators for dynamical systems called expanding maps. To our knowledge, only non adaptive density estimation has been studied in this non-mixing context, see for instance Bosq et Guegan (2), Prieur (22) and Maume-Deschamps (20).

The advantage of our approach is also to treat in one draw many other contexts of dependence. We prove that near-minimaxity still holds for a very large class of models using λ-weak dependence coefficients, defined by Doukhan and Wintenberger, 2007, (14). We pay for generality by adding up conditions on the joint densities of the couples (X_0, X_r) for all $r > 0$. These conditions are not restrictive as it is satisfied for many econometric models such as ARMA, GARCH, ARCH, LARCH, MA models.

The estimation scheme is based on Donoho *et al.*'s procedures developed in (8) for the iid case, and it is adaptive with respect to the regularity of f. Soft and hard-threshold levels $(\lambda_j)_{j_0\le j\le j_1}$ are chosen equal to $K\sqrt{j/n}$ for some $K > 0$. Note that the constant K and the highest resolution level j_1 depend on the weak dependence properties of the observations. If weak dependence properties are known, estimators are near minimax: same rates as in the iid setting are achieved for mean-L^p errors with $1 \le p < \infty$. If weak dependence properties are unknown, we develop cross-validation procedures to approximate threshold levels $\hat{\lambda}_j$ and the highest resolution level $\widehat{j}_1$. We check on simulations that corresponding estimators are adaptive with respect to the regularity of f and to weak dependence properties of $(X_t)_{t\in\mathbb{Z}}$. The order of errors of approximations in the dependence cases are very close to the one in the iid cases.

We believe that we obtain such good results as we work on simulations of processes satisfying our main Assumption **(D)**. This assumption consists of the exponential decay of the covariance terms. We give in this paper some simulations study of dynamical systems that do not satisfy this

Assumption **(D)**. Comparing the behavior of our estimators with the kernel ones show that ours are less efficient when covariance terms do not decrease exponentially fast. We also prove that the error terms of our estimators are unbounded in some cases, due to terms of covariances that decrease too slowly. This is a restriction of our procedure, based on thresholding wavelet coefficients.

The paper is structured as follows. In Section 2, we give notation, we introduce estimation procedures and we formulate weak dependence assumptions. Main results are given in Section 3 and examples of models in Section 4. Cross-validation procedures and their accuracy on simulations are developed in Section 5. Proofs are relegated in the last Section.

2. Preliminaries

2.1. **Notation.** We restrict ourselves to the estimation of a density f which is compactly supported. We suppose that f is supported by $[-B, B]$ for some $B > 0$. For all $p \geq 1$, L^p denotes the space of all functions f such that $\|f\|_p^p = \int |f(x)|^p dx < \infty$.

2.2. **Density estimators.** Throughout the paper, we work within an r-regular orthonormal multiresolution analysis of L^2 (endowed with the usual inner product), associated with a compactly supported scaling function ϕ and a compactly supported mother wavelet ψ. Without loss of generality, we suppose that the support of functions ϕ and ψ is included in an interval $[-A, A]$ for some $A > 0$. Let us recall that ϕ and ψ generate orthonormal basis by dilatations and translations: for a given primary resolution level j_0, the functions $\{\phi_{j,k} : x \mapsto 2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}}$ and $\{\psi_{j,k} : x \mapsto 2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}$ are such that the family

$$\{\phi_{j_0,k},\, k \in \mathbb{Z},\, \psi_{j,k}\, j \geq j_0,\, k \in \mathbb{Z}\}$$

is an orthonormal base of L^2. Any function $f \in L^2$ can thus be decomposed as

$$f = \sum_{k\in\mathbb{Z}} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} \beta_{j,k}\psi_{j,k},$$

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

The nonlinear estimator developed in (8) is defined by the equation

$$\widehat{f}_n = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \gamma_{\lambda_j}(\widehat{\beta}_{j,k})\psi_{j,k},$$

where $\hat{\alpha}_{j,k} = n^{-1}\sum_{i=1}^n \phi_{j,k}(X_i)$ and $\hat{\beta}_{j,k} = n^{-1}\sum_{i=1}^n \psi_{j,k}(X_i)$, and where γ_λ is a threshold function of level λ. The authors consider both hard and soft thresholding functions, corresponding respectively to $\gamma_\lambda(\beta) = \beta\mathbb{1}_{|\beta|>\lambda}$ and $\gamma_\lambda(\beta) = (|\beta| - \lambda)_+\text{sign}(\beta)$. If no distinction is done in the sequel, both hard and soft thresholding estimators are concerned.

Let s, π, r be three positive real numbers satisfying $s + 1/2 - 1/\pi > 0$. We assume that f belongs to the Besov ball $\mathcal{B}^s_{\pi,r}(M_1)$ on the real line, *i.e.* $\|f\|_{s,\pi,r} \leq M_1$ where

$$\|f\|_{s,\pi,r} = |\alpha_{0,0}| + \left(\sum_{j\in\mathbb{N}} \left(2^{j(s\pi+\pi/2-1)} \sum_{k\in\mathbb{Z}} |\beta_{j,k}|^\pi \right)^{r/\pi} \right)^{1/r}.$$

Approximation errors of an estimator f_n are expressed as $\mathbb{E}\|f_n - f\|_p^p$ for $p \geq 1$. Associated minimax rates are the best convergence decrease α of the worst approximation error we may achieve over all estimators f_n:

$$\inf_{f_n} \sup_{f\in\mathcal{B}^s_{\pi,r}(C)} \mathbb{E}\|f_n - f\|_p^p = \mathcal{O}\left(n^{-p\alpha}\right).$$

The minimax rate α is determined in (8) as:

$$\alpha = \begin{cases} \alpha_+ = (s/(1+2s) & \text{if } \epsilon \geq 0, \\ \alpha_- = (s - 1/\pi + 1/p)/(1+2s-2/\pi) & \text{if } \epsilon \leq 0, \end{cases} \qquad \text{where } \epsilon = s\pi - (p-\pi)/2. \tag{2.1}$$

2.3. **Weakly dependent assumption.** Throughout the paper, the symbol δ denotes with no distinction ϕ or ψ and $\widetilde{\delta}_{j,k}(x) = \delta_{j,k}(x) - \mathbb{E}\,\delta_{j,k}(X_0)$ for all integers $j \geq 0$ and k. Define for all positive integers u, v the quantities

$$C^{j,k}_{u,v}(r) = \sup_{\max s_{i+1}-s_i = s_{u+1}-s_u = r} \{|\mathrm{Cov}(\widetilde{\delta}_{j,k}(X_{s_1})\cdots\widetilde{\delta}_{j,k}(X_{s_u}), \widetilde{\delta}_{j,k}(X_{s_{u+1}})\cdots\widetilde{\delta}_{j,k}(X_{s_{u+v}}))|\}. \tag{2.2}$$

Functions ϕ and ψ play a symmetric role through δ in this setting. As stressed in (9), bounds on covariance terms $C^{j,k}_{u,v}(r)$ are useful to extend asymptotic results from the iid case. Now we can state the main assumption of this paper:

(D): There exists a sequence $\rho(r)$ such that for all $r \geq 0$, all indexes j, k, u, v, we have

$$C^{j,k}_{u,v}(r) \leq (u+v+uv)/2\,(2^{j/2}M_2)^{u+v-2}\rho(r) \text{ where } M_2 \text{ is a constant satisfying } \|\delta\|_\infty \leq M_2. \tag{D1}$$

Moreover, there exist real numbers $a, b, C_0 > 0$ depending only on δ, f and on the dependence properties of $(X_t)_{t\in\mathbb{Z}}$ such that

$$\rho(r) \leq C_0 \exp(-ar^b) \text{ for all } r > 0. \tag{D2}$$

In Section 4, we give explicit conditions on the stationary process $(X_t)_{t\in\mathbb{Z}}$ in order that it satisfies Assumption **(D)**. When it is possible, the values of the constants a, b, C_0 and M_2 are given.

3. Main results

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary real valued time series.

3.1. **A useful lemma.** Under similar conditions than Assumption **(D)**, moments inequalities of even orders and Bernstein's type inequalities for the sums $\sum_{i=1}^n \widetilde{\delta}_{j,k}(X_i)$ are respectively given in (9) and (10). The following Lemma recall these inequalities applied on the quantities $\beta_{j,k}$. They remain valid for $\alpha_{j,k}$ as ϕ and ψ play the same role in **(D)**.

Lemma 3.1. *If* **(D)** *holds then for all even integer $q \geq 2$, for all $\lambda \geq 0$ and for j, k such that $0 \leq j \leq \log n$ and $0 \leq k \leq 2^j - 1$:*

$$\mathbb{E}|\widehat{\beta}_{j,k} - \beta_{j,k}|^q \leq C_1 n^{-q/2}, \tag{3.1}$$

$$P\left(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda\right) \leq 2\exp\left(-\frac{n\lambda^2}{C_2 + C_3(2^j/n)^{b/(2(1+2b))}(\sqrt{n}\lambda)^{(2+3b)/(1+2b)}}\right), \tag{3.2}$$

where C_1, C_2 and C_3 are constants depending on q and on the constants of Assumption **(D)***: a, b, C_0 and M_2.*

The proof of this Lemma is given in Section 6.1. Notice that C_3 depends deeply on the dependence context through C_0, see Propositions 4.1 and 4.2 for more details.

3.2. **Near-minimax results of thresholded wavelet estimators.** Following results are extensions to weak dependence settings of Theorem 5 of (8).

Theorem 3.1. *Suppose that $f \in \mathcal{B}^s_{\pi,r}(M_1)$ with $1/\pi < s < N/2$ where N is the regularity of the function ψ. If* **(D)** *holds, then for each $1 \leq p < \infty$ there exists a constant $C(N, p, a, b, C_0, M_1, M_2, A, B)$ such that*

$$\mathbb{E}[\|\widehat{f}_n - f\|_p^p] \leq C \begin{cases} \left(\dfrac{\log n}{n}\right)^{p\alpha} & \text{if } \epsilon \neq 0 \\ \left(\dfrac{\log n}{n}\right)^{p\alpha} (\log n)^{(p/2-(1\wedge\pi)/r)_+} & \text{if } \epsilon = 0 \end{cases}$$

where the minimax rate α and the parameter ϵ are given in (2.1). Here j_0 is chosen as the smallest integer larger than $\log(n)(1+N)^{-1}$, j_1 is the largest integer smaller than $\log(n \log^{-2/b-3}(n))$ and $\lambda_j = K\sqrt{j/n}$ for a sufficiently large constant $K > 0$.

A sketch of the proof of this Theorem is given in Section 6.2.

We refer to (21) for the definition of the parameter N, the regularity of the wavelet function ψ. The condition $s < N/2$ ensures the sparsity of the wavelet coefficients of the density f. This condition is not restrictive as N can be chosen sufficiently large in practice.

The estimators $\widehat{f}_n$ are the same than in the iid case given in (8), except for the highest resolution level j_1 which is smaller here. This restriction is needed in the weak dependence context due to the Bernstein's type inequality (3.2) which is not as sharp as the one in the iid case. But this restriction does not perturb the rate of convergence which is the same as the one obtained in the iid case by (8).

The constant $K > 0$ plays a key role in the asymptotic behavior of $\widehat{f}_n$. From (3.2), we infer that the constant K depends on the parameters a, b, C_0, M_2 through C_1, C_2, C_3 in an intricate way. Then how sufficiently large the constant K must be deeply depends on the dependence structure of observations $(X_1, \dots, X_n)$. Then contrarily to the iid case, we are not able to develop direct procedures based on the observations $(X_1, \dots, X_n)$ that chose a convenient parameter K like in (17). In Section 5, we propose cross validation procedures to determine the threshold levels λ_j when the weak dependence properties of the process $(X_t)_{t\in\mathbb{Z}}$ are unknown.

4. Examples

In this Section, we give examples of models that satisfy Assumption **(D)** using $\widetilde{\phi}$ and λ-weak dependence coefficients introduced respectively in (6) and (14). For each of them, we proceed in two steps: Firstly we give sufficient conditions on these coefficients for ensuring Assumption **(D)** in Propositions 4.1 and 4.2 and secondly we give in Subsections 4.2 and 4.4 examples of models satisfying such conditions.

A Lipschitz function $h : \mathbb{R}^u \to \mathbb{R}$ for some $u \in \mathbb{N}^*$ is a function such that $\mathrm{Lip}\,(h) < \infty$ with

$$\mathrm{Lip}\,(h) \quad = \quad \sup_{(a_1,\dots,a_u)\neq(b_1,\dots,b_u)} \frac{|h(a_1,\dots,a_u) - h(b_1,\dots,b_u)|}{|a_1 - b_1| + \cdots + |a_u - b_u|}.$$

As Lipschitz functions play an important role in weak dependence contexts we restrict ourselves to the cases where $N > 4$. This assumption on the regularity of the wavelet functions implies that ϕ and ψ can be chosen as Lipschitz functions, as established in (4). Note also that $\|f\|_\infty < \infty$ as $f \in \mathcal{B}^s_{\pi,r}(M_1)$, see equation (15) in (8). For convenience, we denote with no distinction $\psi_{j,k}(x) - \beta_{j,k}$ and $\phi_{j,k}(x) - \alpha_{j,k}$ as $\widetilde{\delta}_{j,k}(x)$ for any integers $j \geq 0$ and k.

Weak dependence coefficients is to generalize mixing ones. Let us recall that α-mixing coefficients can be defined in a similar way by two equations:

$$\alpha(r) \quad = \quad \sup_{\substack{1 \leq \ell \\ i+r \leq j_1 \leq \cdots \leq j_\ell \\ \|g\|_\infty \leq 1}} \mathbb{E}|\mathbb{E}\left(g(X_{i_1},\dots,X_{i_u})|\sigma(\{X_j, j \leq i\})\right) - \mathbb{E}\left(g(X_{i_1},\dots,X_{i_u})\right)|,$$

$$\alpha(r) \quad = \quad \sup_{\substack{(u,v)\in\mathbb{N}^*\times\mathbb{N}^* \\ i_1 \leq \cdots \leq i_u \leq i_u + r \leq i_{u+1} \leq \cdots \leq i_{u+v} \\ \|f\|_\infty, \|g\|_\infty \leq 1}} |\mathrm{Cov}(f(X_{i_1},\dots,X_{i_u}), g(X_{i_{u+1}},\dots,X_{i_{u+v}}))|.$$

These coefficients measure the dependence between the past and the future values of the process $(X_t)_{t\in\mathbb{Z}}$ as the gap r between past and future goes to infinity. As these coefficients are often too restrictive, the authors of (7) release them by considering a supremum taken on functions with bounded variations or on Lipschitz bounded functions rather than uniformly bounded functions. These different choices of functions sets lead to different coefficients of weak dependence, namely $\widetilde{\phi}$ and λ respectively. We give hereafter the precise definition of $\widetilde{\phi}$ and λ-weak dependence coefficients.

4.1. **$\widetilde{\phi}$-weak dependence.** Bounded variations functions are defined as follows: let BV and BV_1 denote the sets of functions g supported on $[-A, A]$ satisfying respectively $\|g\|_{BV} < +\infty$ and $\|g\|_{BV} \leq 1$ where

$$\|g\|_{BV} = |g(-A)| + \sup_{n\in\mathbb{N}} \sup_{a_0=-A<a_1<\cdots<a_n=A} \sum_{i=1}^{n} |g(a_i) - g(a_{i-1})|.$$

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-algebra of $\mathcal{A}$ and let $X = (X_1, \ldots, X_v)$, for $v \geq 1$, be a collection of real valued random variables X_i defined on $\mathcal{A}$. We define the coefficient $\widetilde{\phi}$ as it was introduced in (7) by the equation:

$$\widetilde{\phi}(\mathcal{M}, X_1, \ldots, X_v) = \sup_{g_1,\ldots,g_v\in BV_1} \left\| \int \prod_{i=1}^{v} g_i(x_i)\mathbb{P}_{X|\mathcal{M}}(dx) - \int \prod_{i=1}^{v} g_i(x_i)\mathbb{P}_X(dx) \right\|_\infty ,$$

where $dx = (dx_1, \ldots, dx_v)$. The coefficients $\widetilde{\phi}(r)$ are now defined by the equation

$$\widetilde{\phi}(r) = \max_{1\leq\ell} \frac{1}{\ell} \sup_{i+r\leq j_1\leq\cdots\leq j_\ell} \widetilde{\phi}(\sigma(\{X_j; j \leq i\}), X_{j_1}, \ldots, X_{j_\ell}) .$$

These coefficients are multivariate extensions of $\widetilde{\phi}_1(r)$ defined in (6), see also (20). Instead of mixing coefficients, they efficiently treat the dependence structure of dynamical systems and associated Markov chains. A process $(X_t)_{t\in\mathbb{Z}}$ is said to be $\widetilde{\phi}$-weakly dependent if the series $\widetilde{\phi}(r)$ goes to 0 when r goes to infinity, *i.e.* when the gap between observations from the future and observations from the past goes to infinity. Introducing $\widetilde{\phi}$-weakly dependent processes is useful in the present framework due to the following links with the Assumption **(D)**:

Proposition 4.1. *Assume that $(X_t)_{t\in\mathbb{Z}}$ is a process such that there exist $a, b, c > 0$ satisfying*

$$\widetilde{\phi}(r) \leq c^v \exp(-ar^b), \tag{4.1}$$

then Assumption **(D)** *holds with $M_2 = 2(\|\delta\|_\infty + ALip\,\delta)$ and $C_0 = 4c(\|\delta\|_\infty + ALip\,\delta)\|f\|_\infty\|\delta\|_1$.*

The proof of this Proposition is given in Subsection 6.3.

4.2. **Examples of $\widetilde{\phi}$-weakly dependent processes.** Following the work of (7), we give a general class of models where (4.1) is satisfied.

Lemma 4.1. *Assume that $(X_t)_{t\in\mathbb{Z}}$ is a process satisfying the Markov property, taking values in $[0, 1]$ and such that there exist constants $a, b, c > 0$ satisfying, for any functions g, k with $g \in BV_1$ and $\mathbb{E}|k(X_0)| < \infty$, and for all $r \geq 0$,*

$$|Cov(k(X_0), g(X_r))| \leq \mathbb{E}|k(X_0)| \exp(ar^{-b}), \tag{4.2}$$

$$\|\mathbb{E}(g(X_r|X_0 = \cdot)\|_{BV} \leq c, \tag{4.3}$$

then $\widetilde{\phi}_v(r) \leq c^v \exp(-ar^b)$ and the conclusions of Proposition 4.1 follow.

The proof of this Lemma is given in Subsection 6.3. Various examples of processes satisfying conditions of Lemma 4.1 are given in (7). We recall here the case of Markov chains obtained by time reversing expanding maps, as they are extensions of Andrew's example (1.1).

Let us define stationary Markov chains $(X_t)_{t\in\mathbb{Z}}$ associated with dynamical systems through a reversion of the time as non degenerate stationary solutions of the recurrent equation

$$X_t = T^i(X_{t-i}), \qquad \forall t \in \mathbb{Z}, i \in \mathbb{N} \tag{4.4}$$

where $T : [0,1] \to [0,1]$ is a deterministic function.

Remark 1. *For such Markov chains, mixing coefficients are useless. Future values write simply as functions of past values via* (4.4) *and then it is easy to check that $\alpha(r)$ are constant from their definitions. Thus $\alpha(r)$ does not tend to* 0 *and α-mixing coefficients do not evaluate the dependence of such processes.*

A Markov chain $(X_t)_{t\in\mathbb{Z}}$ is associated with an expanding map through a reversion of the time if T satisfies

- *(Regularity)* The function T is differentiable, with a continuous derivate T' and there exists a grid $0 = a_0 \leq a_1 \cdots \leq a_k = 1$ such that $|T'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \ldots, k$.
- *(Expansivity)* For any integer i, let I_i be the set on which the first derivate of T^i, $(T^i)'$, is defined. There exists $a > 0$ and $s > 1$ such that $\inf_{x\in I_i}\{|(T^i)'(x)|\} > as^i$.
- *(Topological mixing)* For any nonempty open sets U, V, there exists $i_0 \geq 1$ such that $T^{-i}(U) \cap V \neq \varnothing$ for all $i \geq i_0$.

Under these three conditions a non degenerate stationary solution $(X_t)_{t\in\mathbb{N}}$ to (4.4) exists and has remarkable properties, see (27) for a nice survey. For instance, the process $(X_t)_{t\in\mathbb{Z}}$ satisfies the conditions of Proposition 4.1 for some $a, c > 0$ and $b = 1$, see (6). Moreover, the marginal distribution is absolutely regular and its distribution belongs to BV. Noticing that $\mathcal{B}^1_{1,1} \subset BV \subset \mathcal{B}^1_{1,\infty}$, see e.g. (8), Theorem 3.1 provides adaptive estimators of the marginal density of $(X_t)_{t\in\mathbb{Z}}$ in that context and extends Theorem 2.2 of (25) where rates of the MISE for non adaptive estimators are given in such context.

4.3. **λ-weak dependence.** The stationary process $(X_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent, as defined in (14), if there exists a sequence of non-negative real numbers $\lambda(r)$ satisfying $\lambda(r) \to 0$ as $r \to \infty$ and such that:

$$\left|\mathrm{Cov}\left(\, h\left(X_{i_1}, \ldots X_{i_u}\right)\,,\, k\left(X_{i_{u+1}}, \ldots, X_{i_{u+v}}\,\right)\right)\right| \leq \\ \left(u\,\|k\|_\infty \mathrm{Lip}(h) \,+\, v\|h\|_\infty\,\mathrm{Lip}(k) + uv\mathrm{Lip}\,(h)\mathrm{Lip}\,(k)\right)\,\lambda(r)$$

for all p-tuples, $(i_1, \ldots, i_p)$ with $i_1 \leq \cdots \leq i_u \leq i_u + r \leq i_{u+1} \leq \cdots \leq i_p$, and for all $h \in \Lambda_u$ and $h \in \Lambda_v$ where

$$\Lambda_u \;=\; \{h : \mathbb{R}^u \to\, \mathbb{R},\ \mathrm{Lip}\,(h) < \infty,\ \|h\|_\infty < \infty\}\,,\ \text{for any } u \leq 1.$$

The λ-weak dependence provides simple bounds of the covariance terms:

$$\left|\mathrm{Cov}\left(\widetilde{\delta}_{j,k}(X_{i_1})\cdots\widetilde{\delta}_{j,k}(X_{i_u}), \widetilde{\delta}_{j,k}(X_{i_{u+1}})\cdots\widetilde{\delta}_{j,k}(X_{i_{u+v}})\right)\right| \leq (u+v+uv)\,\|\widetilde{\delta}_{j,k}\|_\infty^{u+v-2}(\mathrm{Lip}\,\widetilde{\delta}_{j,k})^2\lambda(r).$$

The right hand side term is bounded by $(2^{j/2}2\|\delta\|_\infty)^{u+v-2}\mathrm{Lip}\,\delta^2 2^{3j}\lambda(r)$. At this point, we do not achieve Assumption (D1) as this bound depends on j via an extra term 2^{3j}. An additional assumption is needed to ensure Assumption **(D)**:

(J): The joint densities f_r of (X_0, X_r) exist and are bounded for $r > 0$.

Proposition 4.2. *Assume that $(X_t)_{t\in\mathbb{Z}}$ is a λ-weakly dependent process satisfying* **(J)** *and such that there exist $a', b', c' > 0$ and $a'', b'', c'' > 0$ with:*

$$\lambda(r) \le c' \exp(-a' r^{b'}) \text{ and } \|f_r\|_\infty \le c'' \exp(a'' r^{b''}) \text{ for all } r > 0.$$

If $b'' < b'$ then **(D)** *holds for some $C_0 > 0$ with $M_2 = 2\|\delta\|_\infty$, $a = a'/4$ and $b = b'$.*

The proof of this Proposition is given in Subsection 6.3.

4.4. **Examples of λ-weakly dependent processes.** Firstly we give in this Section two generic λ-weakly dependent models, Bernoulli shifts and random processes with infinite connections. Secondly we detail the conditions of Proposition 4.2 in three more specific models.

Let $H : \mathbb{R}^{\mathbb{Z}} \to [0,1]$ be a measurable function. A Bernoulli shift with innovations ξ_t is defined as

$$X_t = H\left((\xi_{t-i})_{i\in\mathbb{Z}}\right), \qquad t \in \mathbb{Z}.$$

According to (14), such Bernoulli shifts are $\lambda-$weakly dependent with $\lambda(r) \le 2v([r/2])$. If $(\xi_t)_{t\in\mathbb{Z}}$ is iid, $(v(r))_{r>0}$ is a non-increasing sequence satisfying

$$\mathbb{E}\left|H\left(\xi_j, j \in \mathbb{Z}\right) - H\left(\xi'_j, j \in \mathbb{Z}\right)\right| \le v(r) \text{ for all } r > 0,$$

where the iid sequence $(\xi'_j)_{j\in\mathbb{Z}}$ is such that $\xi'_j = \xi_j$ for $|j| \le r$ and ξ'_j independent of ξ_j otherwise. If the weak dependence coefficients $\lambda_\xi(r)$ refer to $(\xi_t)_{t\in\mathbb{Z}}$, we can compute these of $(X_t)_{t\in\mathbb{Z}}$. More precisely, the result in (14) states that if there exists $\ell \ge 0$ such that

$$|H(x) - H(y)| \le b_s(\|z\|^\ell \vee 1)|x_s - y_s|,$$

for some sequence $b_j \ge 0$ satisfying $\sum_j |j| b_j < \infty$ and where $x, y \in \mathbb{R}^{\mathbb{Z}}$ coincide except for the index $s \in \mathbb{Z}$ and $\|x\| = \sup_{i\in\mathbb{Z}} |x_i|$, if $\mathbb{E}|\xi_0|^{m'} < \infty$ for some $m' \ge \ell + 1$ then $(X_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent with

$$\lambda(k) \le c \inf_{r\le[k/2]} \left[\sum_{|j|\ge r} |j| b_j + (2r+1)^2 \lambda_\xi(k-2r)^{\frac{m'-1-\ell}{m'-1+\ell}}\right], \text{ for some } c > 0.$$

Different values of b in Assumption (D2) may arise naturally when realizing the minimum of the equation above, see below for some classical examples.

Another approach is the one of random processes with infinite connections considered in (12). Let $F : [0,1]^{\mathbb{Z}/\{0\}} \times \mathbb{R} \to [0,1]$ be measurable. Under suitable conditions on F, the stationary solution of the equation

$$X_t = F((X_j, j \ne t), \xi_t), \ a.s.,$$

exists and is λ-weakly dependent. We refer the reader to (12) for more details and we end the Section with some specific λ-weakly dependent models.

4.4.1. *Infinite moving average.* A Bernoulli shift is an infinite moving average process if

$$X_t = \sum_{i\in\mathbb{Z}} \alpha_i \xi_{t-i}. \tag{4.5}$$

If ξ_t are iid random variables satisfying $\mathbb{E}|\xi_0| \leq 1$ then $(X_t)_{t\in\mathbb{Z}}$ is λ-weakly dependent with $\lambda(r) \leq 4\sum_{|j|>[r/2]} |a_j|$. If $a_j \leq K\alpha^{|j|}$ for $j \neq 0$, $K > 0$ and $0 < \alpha < 1$ then the condition on λ in Proposition 4.2 holds with $b' = 1$. Then Assumption (D2) is ensured with $b = 1$.

4.4.2. LARCH(∞) *model.* Let $(\xi_t)_{t\in\mathbb{Z}}$ be an iid centered real valued sequence and $a, a_j, j \in \mathbb{N}^*$ be real numbers. LARCH(∞) models are solutions of the recurrence equation

$$X_t = \xi_t \left(a + \sum_{j\neq 0} a_j X_{t-j} \right). \tag{4.6}$$

The stationary solution of (4.6) satisfies the condition on λ in Proposition 4.2 with $b' = 1/2$ if there exists $K, \alpha > 0$ and $\alpha < 1$ such that $a_j \leq K\alpha^{|j|}$ for all $j \neq 0$. Then Assumption (D2) is ensured with $b = 1/2$. See (11) for applications of this model in econometrics.

4.4.3. *Affine model.* Let us consider the stationary solution $(X_t)_{t\in\mathbb{Z}}$ of the equation

$$X_t = M(X_{t-1}, X_{t-2}, \ldots)\xi_t + f(X_{t-1}, X_{t-2}, \ldots),$$

where M and f are both Lipschitz functions. This model contains various time series processes such as ARCH, GARCH, ARMA, ARMA-GARCH, etc. If the ξ_t are iid random variables with a bounded marginal density, then **(J)** holds and the joint densities are uniformly bounded, as stated in the Appendix of (13). Moreover if the functions M and f have exponentially decreasing Lipschitz coefficients, then conditions of Proposition 4.2 hold with $b' = 1/2$, $b'' = 0$ and Assumption (D2) follows with $b = 1/2$.

5. Cross-validation procedures and simulations

The aim of this Section is to evaluate the applicability and the quality of the procedure on simulated data. Even if the estimators are adaptive with respect to the regularity of the density function, a constant appears in the threshold levels that we cannot calibrate if the dependence properties of the observations are unknown. Then we develop a cross-validation scheme in order to apply concretely the estimator on simulated data. We investigate several examples of dependence for the simulated observations that satisfy the convergence result of this paper. We also give counter-examples where the estimators fail to converge near-minimaxly.

5.1. **Cross-validation procedures.** According to Theorem 3.1, let us fix j_0 as the smallest integer larger than $\log(n)(1+N)^{-1}$ and define $\widehat{f}_n^{HTCV}$ and $\widehat{f}_n^{STCV}$ respectively as the hard and soft-thresholding estimators associated with threshold levels $((\widehat{\lambda}_j)_{j_0\leq j\leq j^*})$. Here $j^*=\log_2 n$ and $((\widehat{\lambda}_j)_{j_0\leq j\leq j^*})$ are determined by cross-validation procedures: $\widehat{\lambda}_j=\mathrm{Argmin}_\lambda\, CV_j(\lambda)$ where cross validation criterion are respectively defined by the equations

$$\text{HTCV:}\qquad CV_j(\lambda)=\sum_{k\in\mathbb{Z}}\mathbb{1}_{\{|\widehat{\beta}_{j,k}|\geq\lambda\}}\left[\widehat{\beta}_{j,k}^2-\frac{2}{n(n-1)}\sum_{1\leq i\neq h\leq n}\psi_{j,k}(X_i)\psi_{j,k}(X_h)\right],$$

$$\text{STCV:}\qquad CV_j(\lambda)=\sum_{k\in\mathbb{Z}}\mathbb{1}_{\{|\widehat{\beta}_{j,k}|\geq\lambda\}}\left[\widehat{\beta}_{j,k}^2-\frac{2}{n(n-1)}\sum_{1\leq i\neq h\leq n}\psi_{j,k}(X_i)\psi_{j,k}(X_h)+\lambda^2\right].$$

These criterion are obtained by approximating the coefficients $\beta_{j,k}$ by $\hat{\beta}_{j,k}$ and the products $\beta_{j,k}\beta_{j,k'}$ by $\sum_{h\neq i}\psi_{j,k}(X_i)\psi_{j,k}(X_h)/(n(n-1))$ in the Integrated Square Error.

The estimator $\widehat{j}_1$ is defined as the smallest integer such that $CV_j(\widehat{\lambda}_j)=0$ for all $\widehat{j}_1\leq j\leq j^*$. Notice that cross-validation procedures may consider larger resolution levels than the estimators $\widehat{f}_n$ as j^* is larger than j_1 given in Theorem 3.1.

5.2. **Different dependent samplings satisfying (D).** To illustrate the behavior of this cross-validation scheme, we simulate three different weak-dependence cases with the same marginal absolutely continuous distribution F. The simulations were carried out as follows:

Case 1: Independent observations are given by $X_i=F^{-1}(U_i)$ where the U_i are simulations of independent variables, uniform on [0,1].

Case 2: A $\widetilde{\phi}$-weakly dependent process is obtained by the equation $X_i=F^{-1}(G(Y_i))$ for $i=1,\dots,n$ with $G(x)=2\sqrt{x(1-x)}/\pi$ and $(Y_i)_{i=1,\dots,n}$ given by

$$Y_1=G^{-1}(U_1)$$
$$\text{and, recursively, } Y_i=T^{i-1}(Y_1) \text{ for } 2\leq i\leq n \text{ with } T(x)=4x(1-x).$$

Note that for all $1\leq i\leq n$ the Y_i admits the repartition function G the invariant distribution of T. Moreover the sequence $(Y_1,\dots,Y_n)$ satisfies the assumptions of Proposition 4.1, see (22) for details.

Case 3: A λ-weakly dependent process resulting from the transform $X_i=F^{-1}(G(Y_i))$ of variables $(Y_t)_{t\in\mathbb{Z}}$ which are solution of

$$Y_t=2(Y_{t-1}+Y_{t+1})/5+5\xi_t/21$$

where $(\xi_t)_{t\in\mathbb{Z}}$ is an iid sequence of Bernoulli variables with parameter $1/2$. The stationary solution of this equation admits the representation

$$Y_t = \sum_{j\in\mathbb{Z}} a_j \xi_{t-j},$$

where $a_j = 1/3(1/2)^{|j|}$. This solution belongs to $[0,1]$ and its marginal distribution G is the one of $(U+U'+\xi_0)/3$ where U and U' are independent variables following $\mathcal{U}([0,1])$. For $1 \le i \le n$, the solution Y_i is approximated by $Y_i^{(N)}$ for $1 \le j \le N$ and $j-N \le i \le n+N-j$ where $Y_i^{(j)}$ is generated according to the convergent algorithm given in (12): the initial values $Y_i^{(0)}$ are fixed equal to 0 and, given simulated variables $(\xi_i)_{-N\le i\le n+N}$, we define recursively $Y_i^{(j)} = 2(Y_{i-1}^{(j-1)} + Y_{i+1}^{(j-1)})/5 + 5\xi_i/21$ for $1 \le i \le N$ and $i-N \le t \le n+N-i$. Error of approximation are negligible as decreasing exponentially fast with the parameter N that we fix $N = n$, see Lemma 6 of (12) for more details. Moreover, Proposition 2.1 of (5) ensures that there exists $a, C > 0$ such that $\lambda(r) \le C\exp(-ar)$ for the process $(Y_t)_{t\in\mathbb{Z}}$ and consequently for $(X_t)_{t\in\mathbb{Z}}$.

Two different density functions are considered. The first one is a mixture of a sinus function and a uniform distribution, presenting a discontinuitie, and the second one is a mixture of two gaussian distributions.

5.3. **Comparison of $\widehat{f}_n^{HTCV}$ and $\widehat{f}_n^{STCV}$.** The first density considered is a mixture between a sinus function and a uniform distribution.

The calculations were carried out on MATLAB on a Unix environment. We considered $n = 2^{10}$ observations repeated $M = 500$ times for each of the three weak dependence cases. Once the data simulated we applied cross-validation procedures. The usual DWT algorithm proposed by (19) and implemented in the Wavelab toolbox by Donoho and his collaborators (available on (29)) only gives values of wavelet functions on an equidistant grid. As one needs to compute these values at given data points, we consider an equidistant grid of I points with the number of points I huge with respect to the number of observations n. Then we approximate the values $\psi_{j,k}(X_i)$ by $\psi_{j,k}([X_i I]/I)$ where $[x]$ denotes the closest integer from any real number x. The wavelets used for the decomposition are Daubechies Symmlets with $N = 8$ zero-moments. Notice that another possible scheme is the algorithm of (28) that gives directly the values $\psi_{j,k}(X_i)$. But as it needs much more calculus time it has not been used here for convenience.

In Figures 1 and 2 are represented the estimators $\widehat{f}_n^{HTCV}$ and $\widehat{f}_n^{STCV}$ and the true density function f in different weakly dependent cases. The quality of the estimators is visually good. According to Figures 1 and 2 the weak dependence properties of the simulated data do not seem to affect both procedures of estimation. Density estimators presented in Figures 1 and 2 do not detect the discontinuity in the density. Actually, for any finite data set, not enough simulated values are concentrated around the discontinuity to allow estimators to detect it.

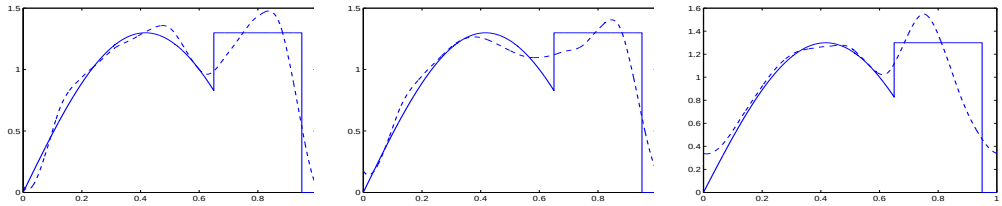

FIGURE 1. Examples of estimators $\widehat{f}_n^{HTCV}$ obtained on 2^{10} observations. The true distribution is represented in dashed lines. Figures from left to right correspond respectively to **Case 1**, **Case 2** and **Case 3**.

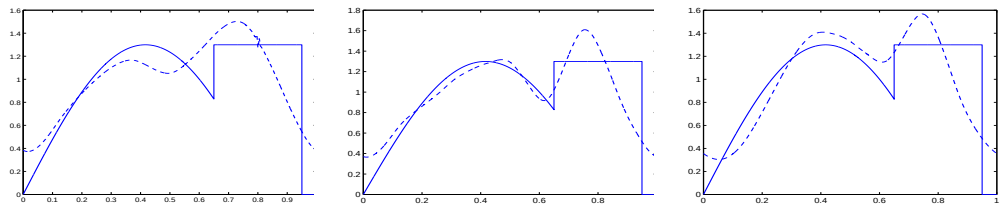

FIGURE 2. Examples of estimators $\widehat{f}_n^{STCV}$ obtained on 2^{10} observations. The true distribution is represented in dashed lines. Figures from left to right correspond respectively to **Case 1**, **Case 2** and **Case 3**.

Table 1 gives approximations by Monte-Carlo method of the MISE. MISE values of the estimators have the same order whereas the weak dependent cases and $\widehat{f}_n^{STCV}$ is preferable in all the cases.

MISE of the estimation			
	Case 1	Case 2	Case 3
HTCV	0.096696	0.077064	0.097193
STCV	0.082934	0.06586	0.097184

TABLE 1. MISE approximated by MC on 500 simulations of samples of size $n = 2^{10}$.

In Figure 3 threshold levels are represented with respect to resolution levels. Their behaviors are similar in all cases: the threshold levels increase with respect to resolution levels. For small resolutions, both HTCV and STCV procedures are close as $\widehat{\lambda}_j^2$ is negligible in $CV_j(\lambda)$. For high resolution it is also the case as $\widehat{\lambda}_j^2$ is big enough to kill almost all $\widehat{\beta}_{j,k}$. Moreover these figures tend to confirm that threshold levels do not depend on weak dependence. Finally remark that the curves do not behave in square root of j as theoretical threshold levels given in Theorem 3.1.

After looking at the threshold levels values, we give in Figure 4 the frequencies of the $\widehat{\beta}_{j,k}$ that are less than $\widehat{\lambda}_j$ with respect to j. As these frequencies are not discretized in two values

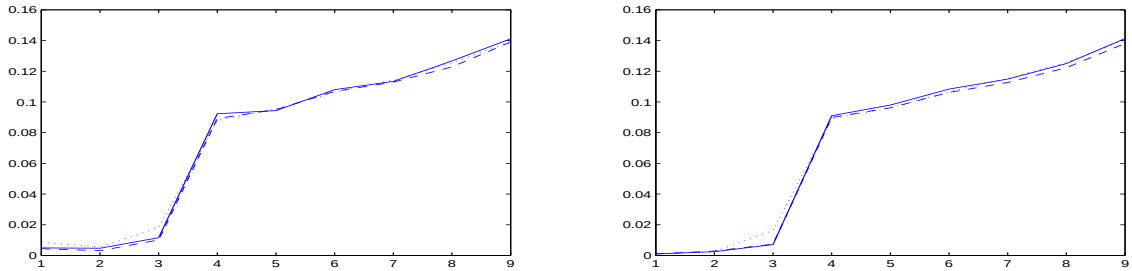

FIGURE 3. Means of the proportions of the threshold levels obtained by cross-validation with respect to the resolution levels for hard-thresholding (left) and soft-thresholding (right). **Case 1** corresponds to the solid line, **Case 2** to the dashed line and **Case 3** to the dotted line.

0 and 1, we can infer that both $\widehat{f}_n^{HTCV}$ and $\widehat{f}_n^{STCV}$ are not equivalent to linear estimators. It is encouraging for both methods as the results in (8) state that linear estimators are not near-minimax. One can also see in these figures that frequencies of effective thresholds are the same among the weak dependence cases.

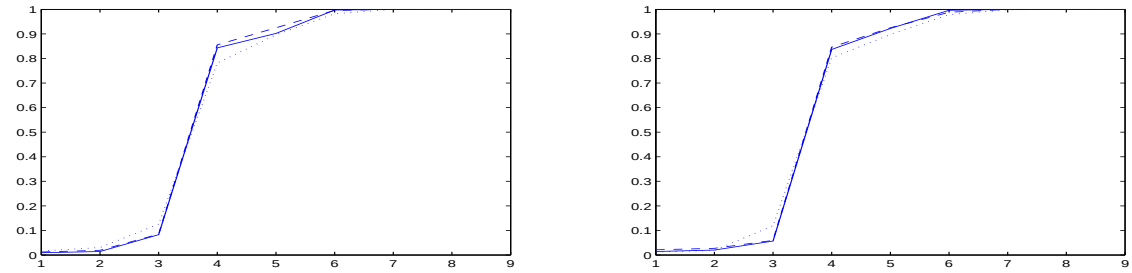

FIGURE 4. Means of the proportions of thresholded coefficients with respect to the resolution levels for hard-thresholding (left) and soft-thresholding (right). **Case 1** corresponds to the solid line, **Case 2** to the dashed line and **Case 3** to the dotted line.

Finally, means of higher resolution levels are given in last Table 2. According to Theorem 3.1, the values of this parameter do depend on the cases of weak dependence. But no significant differences appear on simulations.

Mean of $\widehat{j}_1$			
	Case 1	Case 2	Case 3
HTCV	5.168	5.14	5.13
STCV	5.14	5.04	5.13

TABLE 2. Means of $\widehat{j}_1$ on 500 simulations of $n = 2^{10}$ observations.

5.4. **Comparison with kernel estimators.** As results computed in the last Subsection are systematically better for the STCV estimators than for the HTCV ones, we only present in the sequel results for the STCV estimators. Let us now consider the case of a density function that is a mixture of normal distributions. Like in (16), we compare the quality of the wavelet estimator $\hat{f}^{STCV}$ with linear kernel estimators. The kernel used is Epanechnikov's one and we computed two choices for its width parameter: firstly we consider the width given by the rule of thumb of Matlab, (more precisely, the width is equal to $(q_3 - q_1)/(2 * 0.6745) * (4/(3 * n))^{1/5}$, where q_1 and q_3 denote respectively the first and the third quartile of the empirical distribution) and secondly we consider the width obtained by cross validation on the mean integrated squared error risk.

In Figure 5 are represented the means of wavelet and kernel estimators in the different cases. Like in the last Subsection, there is no visual difference between the different cases of dependence. In this Figure, it appears that the mean of the kernel estimators fails to detect the two modes of the density when the width is choosing according to the rule of thumb. The bandwidth is overestimated in this case. The quality of the wavelet estimators $\widehat{f}_n^{STCV}$ and the kernel estimators with the width from a cross validation procedure are visually equivalent.

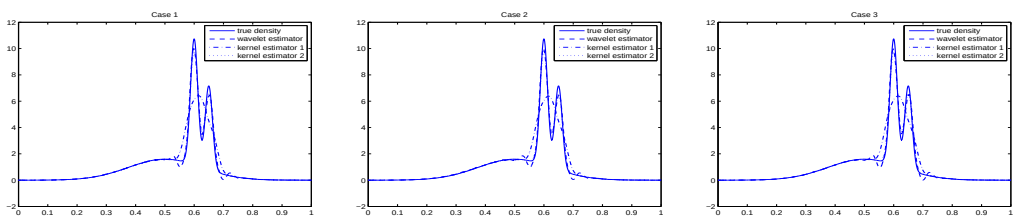

FIGURE 5. Means of estimators $\widehat{f}_n^{STCV}$ obtained from 2^{10} observations on 500 simulations. The mean of the wavelet estimators is represented in dashed lines while the means of the kernel estimators are represented respectively in line with dots for the rule of thumb's width (kernel estimator 1) and in dots for the cross-validation width (kernel estimator 2). Figures from left to right correspond respectively to **Case 1**, **Case 2** and **Case 3**.

To analyse more precisely approximations realized by $\widehat{f}_n^{STCV}$ and by kernel estimators, we represent in Figure 6 the evolution of the mean L^p risk with respect to p for the three estimators in each case of dependence, i.e. $\mathbb{E}(\|g - f\|_p^p)^{1/p}$ with g equals to one of the three estimators. Even if these risks are close to each others, kernel estimator with cross validation bandwidth has the smallest risk for small values $p \leq 4$. Yet, approximations of this kernel estimator clearly get worse with higher values of p, while risks of $\widehat{f}_n^{STCV}$ seem relatively stable for different value of p. Concerning kernel estimator with the width parameter taken according to the rule of thumb, the L^p risk is worse for small values of p but comparable with the one of wavelet estimator for higher values, even if the modes of the density are not detected.

These graphs show that an advantage of $\widehat{f}_n^{STCV}$ is that its mean $\mathbb{L}^p$ risk seems stable for high values of p. Nevertheless, the mean of the L^2 risk is larger than the one of kernel estimators with cross-validation width and the computation time higher. One possible way to improve the quality of approximation for the cross validation procedure may be to consider different levels of thresholding at each resolution level. We do not investigate this axis of research as then the time of computation exploded.

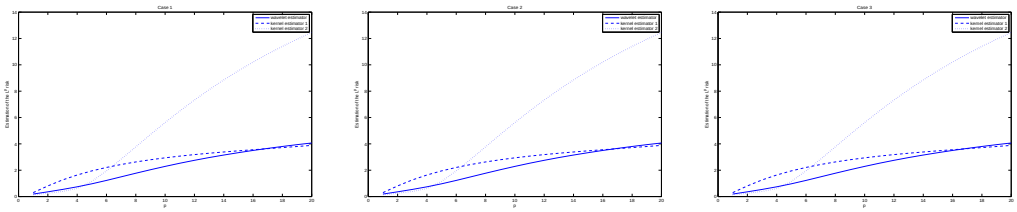

FIGURE 6. Evolution of the L^p risk of estimators $\widehat{f}_n^{STCV}$ and kernel estimators obtained on 2^{10} observations. The wavelet estimator is represented in dashed lines while the kernel estimators are represented respectively in dots line for the rule of thumb width (kernel estimator 1) and in dots for the cross-validation width (kernel estimator 2). Figures from left to right correspond respectively to **Case 1**, **Case 2** and **Case 3**.

5.5. **Different dependent samplings that do not satisfy (D).** In this Subsection, we discuss the necessity of Assumption **(D)**. For this, we study the convergence of the density estimators on some dynamical systems that do not satisfy this assumption. More precisely, we focus on Liverani-Saussol-Vaienti maps, see (18), defined as the solution of $X_t = T^i(X_{t-i})$ with

$$T(x) = \begin{cases} x(1+2^{\alpha'}x^{\alpha'}), & 0 \le x \le 1/2 \text{ for some } 0 < \alpha' < 1 \\ 2x-1, & 1/2 < x \le 1. \end{cases}$$

The process $(X_t)_{t\in\mathbb{Z}}$ is stationary and such that the covariance terms $\mathrm{Cov}(f(X_0), g(X_r))$ are of order $r^{1-1/\alpha'}$, see (30) and refinement in (15). Thus the Assumption **(D)** is not satisfied in this case and we have the non-minimaxity of any thresholded wavelets estimators:

Proposition 5.1. *Suppose that the father wavelet ϕ is such that $\int \phi > 0$ and that the assumptions of Theorem 3.1 are satisfied. If $1 > \alpha' \ge 1/(2\alpha+1)$ with α defined by* (2.1)*, then for the thresholded estimators of the marginal density of the Liverani-Saussol-Vaienti map of index α there exists some $C > 0$ such that:*

$$n^{2\alpha}\mathbb{E}[\|\widehat{f}_n - f\|_2^2] \ge C, \text{ for } n \text{ sufficiently large.}$$

The same result also holds for the cross validation thresholded estimator $\widehat{f}_n^{STCV}$.

The proof of this Proposition is given in Section 6.

To simulate these dynamical systems, we simulate Z_0 according to the Lebesgue measure on $[0,1]$, then we apply recursively T to determine Z_i and finally we set $(X_1,\dots,X_n) = (Z_{n+1},\dots,Z_{2n})$. This approximation of the stationary solution does not affect the study of the convergence rates as the dynamical system Z is ergodic in mean with rate $O(n^{1-1/\alpha'})$, see Theorem 5 of (30).

The analytic expression of the density f is unknown but it is proved to be continuous, locally Lipschitz and to behave like $x^{-\alpha'}$ as $x \to 0$, see (18) and (30). In particular, f is unbounded on $[0,1]$. Then we restrict our study on $[0.01,1]$ where f is bounded. As the true density f is unknown, we compare here the estimators $\widehat{f}_n^{STCV}$ with other estimators. In Figure 7, we plot the

mean of $M = 100$ estimators $\widehat{f}_n^{STCV}$ and Epanechnikov kernel estimators given by Matlab's rule of thumb for 9 different values of α'.

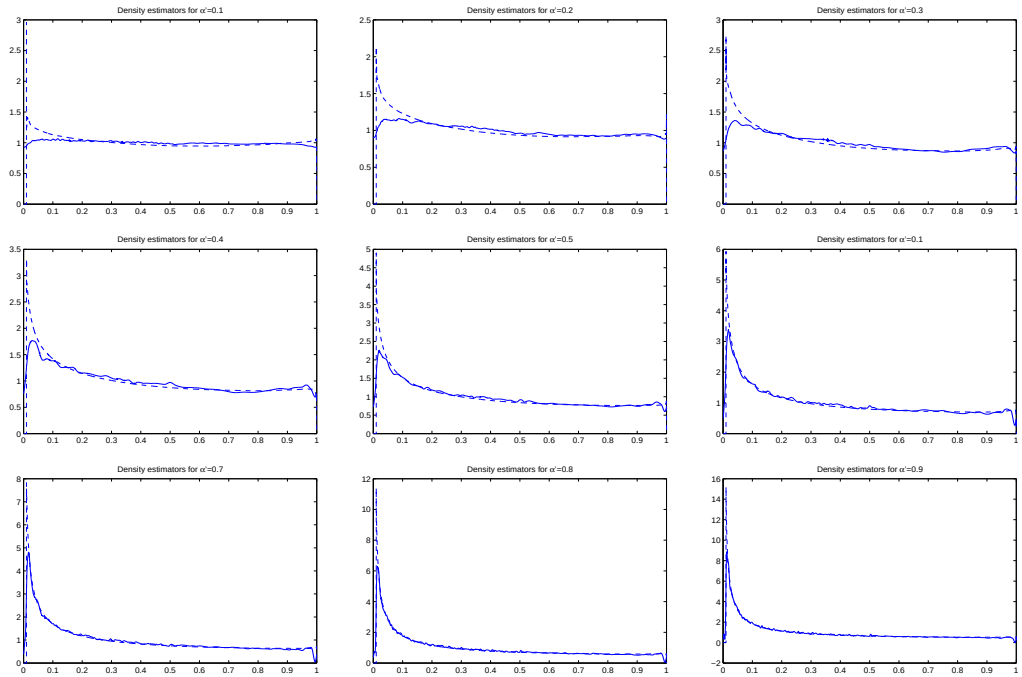

FIGURE 7. Means of the estimators $\widehat{f}_n^{STCV}$ obtained on 2^{10} observations and 500 simulations. Means of the kernel estimators are represented in dashed lines.

Visually, means of both estimators are closed to each other. To detect some difference between the estimators behavior, we decide to compute the moments of order $k = 1, \ldots, 20$ of the estimators integrated on $[0, 1]$:

$$\int_{0.01}^{1} \left(\mathbb{E}\left[g^k(t) \right] \right)^{1/k} dt$$

where the random function g is alternatively $\widehat{f}_n^{STCV}$ or kernel estimators.

For small values $\alpha' \leq 0.02$ and $k \leq 4$, the moments of both estimators have similar values. But as α' growths, all the moments of $\widehat{f}_n^{STCV}$ explode more rapidly than the ones of kernel estimators as k increases. Previous simulations studies show no behavior difference for $\widehat{f}_n^{STCV}$ between independent cases and dependent cases satisfying **(D)**. In dependent cases that do not satisfy **(D)**, the behavior of $\widehat{f}_n^{STCV}$ depends on the decrease rates of covariance terms. On the contrary, the behavior of the kernel estimators with rule of thumb width is more stable when **(D)** is not satisfied.

6. PROOFS

In this Section are collected all the proofs of this paper.

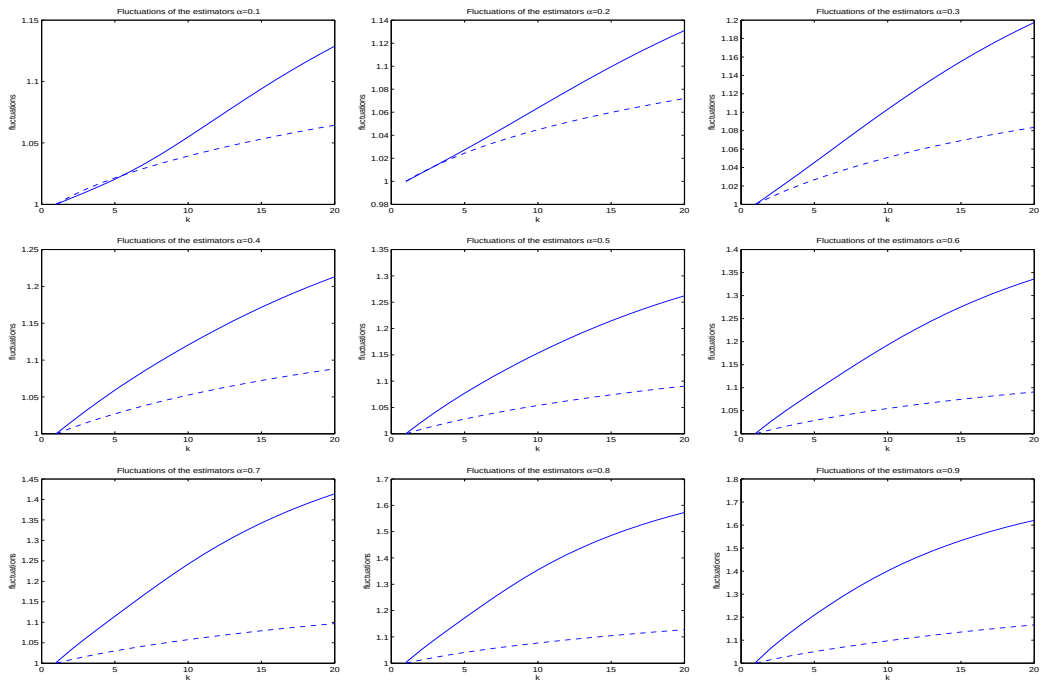

FIGURE 8. Moments of the estimators $\widehat{f}_n^{STCV}$ obtained on 2^{10} observations and 500 simulations. In dashed lines are represented the moments of the kernel estimators.

6.1. **Proofs of Lemma 3.1.** Firstly we give the proof of the inequality (3.1) of Lemma 3.1. This proof is essentially based on Lemmas 6.1 and 6.2. The first one, Lemma 6.1, is simply the Theorem 2 of (9) that we recall here for completeness:

Lemma 6.1. *(Doukhan & Louhichi (1999)) Let $(Z_i)_{1\leq i\leq n}$ be centeres variables.*
Let q be an even integer and $n \geq 2$.
Suppose that for all $p = 2, \dots, q$ and for all $1 \leq s_1 \leq s_p \leq n$ satisfying $\max s_{i+1} - s_i = s_{u+1} - s_u = r$, there exists $V_{p,n}$ such that:

$$n \sum_{r=0}^{n-1} (r+1)^{p-2} Cov(Z_{s_1} \cdots Z_{s_u}, Z_{s_{u+1}} \cdots Z_{s_p}) \leq V_{p,n}.$$

Then, we have

$$\mathbb{E}|\sum_{i=1}^{n} Z_i|^q \leq \frac{(2q-2)!}{(q-1)!} \left\{ V_{2,n}^{q/2} \vee V_{q,n} \right\} , \tag{6.1}$$

We refer the reader to (9) for the proof of this result. We apply this result on $Z_i = \psi_{j,k}(X_i) - \beta_{j,k}$, when **(D)** holds. We first determine the bounds $V_{p,n}$ as under Assumption **(D)** we have:

$$\sum_{r=0}^{n-1} (r+1)^{p-2} \mathrm{Cov}(Z_{s_1} \cdots Z_{s_u}, Z_{s_{u+1}} \cdots Z_{s_p}) \leq \sum_{r=0}^{n-1} (r+1)^{p-2} \rho(r) p^2 (2^{j/2} M_2)^{p-2}.$$

Here we need the following analytic Lemma to bound the quantity $\sum_{r=0}^{\infty} (r+1)^p \rho(r)$:

Lemma 6.2. *If* (D2) *is satisfied, i.e. if* $\rho(r) \leq C_0 e^{-ar^b}$*, then for all integer* p *we have*

$$\sum_{r=0}^{\infty}(r+1)^p \rho(r) \leq C_1 C_2^p (p!)^{1/b}, \tag{6.2}$$

with some constants C_1 *and* C_2 *that are depending on* a *and* b*.*

Applying this result, whose proof is given at the end of this subsection, we directly obtain the new bound:

$$\sum_{r=0}^{n-1}(r+1)^{p-2}\mathrm{Cov}(Z_{s_1}\cdots \leq C_1 p^2 (2^{j/2} M_2 C_2)^{p-2}((p-2)!)^{1/b}$$

We set $V_{p,n} = C_1 p^2 (2^{j/2} M_2 C_2)^{p-2}((p-2)!)^{1/b} n$, and, applying Lemma 6.1, we obtain that:

$$\mathbb{E}|\sum_{i=1}^{n}(Z_i - \mathbb{E}Z_0)|^q \leq \frac{(2q-2)!}{(q-1)!}\left\{(C_1 n)^{q/2}\right\} \vee \left\{C_1 (2^{j/2} M_2 C_2)^{q-2}((q-2)!)^{1/b}\right\}.$$

Dividing by n^{-q} and noticing that $2^{j/2} \leq n$ for $0 \leq j \leq \log n$, we derive that:

$$\mathbb{E}|\widehat{\beta}_{j,k} - \beta_{j,k})|^q \leq n^{q/2}\frac{(2q-2)!}{(q-1)!}\left\{(C_1)^{q/2}\right\} \vee \left\{C_1 (M_2 C_2)^{q-2}((q-2)!)^{1/b}\right\},$$

which corresponds to the inequality (3.1). In particular, for $p = 2$ we have

$$\mathbb{E}|\sum_{i=1}^{n}(Z_i - \mathbb{E}Z_0)|^2 \leq 4C_1 n.$$

Now the inequality (3.2) of Lemma 3.1 is a direct application of Theorem 1 in (10) with $S_n = \sum_{i=1}^{n}(Z_i - \mathbb{E}Z_0)$, $t = n\lambda$, $\nu = 0$, $\mu = 1/b$, $A_n = 4C_1 n$ and $B_n = 2M_2 C_2 2^{(2+b)/b} 2^{j/2}$. We refer the reader to (10) for the definition of the parameters t, ν, μ, A_n and B_n.

Proof of Lemma 6.2. Define $g(x) = (1+x)e^{-ax^b}$ for all $x \leq 0$. Studying its derivative, we can easily see that it exists $x_{a,b}$ such that the function g decreases on $[x_{a,b}, +\infty)$. If we denote $k \geq 1$ the smallest integer greater than $x_{a,b}$, we can infer from (D2) that

$$\begin{aligned}\sum_{r=0}^{\infty}(r+1)^p\rho(r) &\leq C_0\left(\sum_{r=0}^{k-1}(r+1)^p\rho(r) + \int_{k-1}^{\infty}(x+1)^p\exp(-ax^b)dx\right)\\ &\leq C_0\left(C_{a,b} + \int_0^{\infty}(x+1)^p\exp(-ax^b)dx\right).\end{aligned}$$

With a convex inequality on $x \mapsto x^p$, we achieve the bound

$$\sum_{r=0}^{\infty}(r+1)^p\rho(r) \leq C_0\left(C_{a,b} + 2^{p-1}\left(\int_0^{\infty}x^p\exp(-ax^b)dx + \int_0^{\infty}\exp(-ax^b)dx\right)\right).$$

Then, writing $u = ax^b$,

$$\sum_{r=0}^{\infty}(r+1)^p\rho(r) \leq C_0\left(C_{a,b} + 2^{p-1}b^{-1}\left[a^{-(p+1)/b}\Gamma\left(\frac{p+1}{b}\right) + a^{-1/b}\Gamma\left(\frac{1}{b}\right)\right]\right),$$

with Γ defined by $\Gamma(x) = \int_0^\infty u^{x-1}\exp(-u)du$ for all $x > 0$. Let $\lceil x \rceil$ denotes the largest integer smaller than $x > 0$ and note $\overline{\Gamma} = \sup_{x\in]0,1]}\Gamma(x)$. Using the inequalities $\Gamma((p+1)/b) \leq \overline{\Gamma}\lceil(p+1)/b\rceil! \leq \overline{\Gamma}e^{p/b}(p\,!)^{1/b}$ and the fact that this last bound is also available for $\Gamma(1/b)$, we have:

$$\begin{aligned}\sum_{r=0}^{\infty}(r+1)^p\rho(r) &\leq C_0\left(C_{a,b} + 2^{p-1}b^{-1}\left[a^{-(p+1)/b} + a^{-1/b}\right]\overline{\Gamma}e^{p/b}(p\,!)^{1/b}\right)\\ &\leq C_0\left(C_{a,b} + 2^{-1}b^{-1}a^{-1/b}\overline{\Gamma}\right)\left(2e^{1/b}(a^{-1/b}\vee 1)\right)^p(p\,!)^{1/b}.\end{aligned}$$

The Lemma comes immediately by choosing the appropriated constants. □

6.2. **Proof of Theorem 3.1.** The proof of Theorem 3.1 is very similar to the one in the iid case given in (8). In the sequel, C denotes a positive real number that does not depend on n, j nor k. Its value may vary from an equation to another. Let us fix $1 \leq p < \infty$ and consider only the cases where $1 \leq \pi \leq p$. The cases where $\pi > p$ follow from the case $\pi = p$ applying Jensen's inequality on the error term $\mathbb{E}\|\widehat{f}_n - f\|_p^p$. Theorem 3.1 provides the convergence rate for f belonging to the Besov Ball $\mathcal{B}^s_{\pi,r}(M_1)$. In particular $f \in L^2$ and it can be written as:

$$f = \underbrace{\sum_{k=0}^{2_0^j-1}\alpha_{j_0,k}\phi_{j_0,k}}_{E_{j_0}f} + \underbrace{\sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1}\beta_{j,k}\psi_{j,k}}_{D_{j_0}f}.$$

We decompose the estimators $\widehat{f}_n$ of f in the same way:

$$\widehat{f}_n = \underbrace{\sum_{k=0}^{2_0^j-1}\widehat{\alpha}_{j_0,k}\phi_{j_0,k}}_{\widehat{E}_{j_0}f} + \underbrace{\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}\gamma_{\lambda_j}(\widehat{\beta}_{j,k})\psi_{j,k}}_{\widehat{D}_{j_0}f},$$

where the γ_{λ_j} denotes without distinction the soft and hard-threshold function.

Thanks to Minkowski's inequality, the risk of $\widehat{f}_n$ is divided in two terms:

$$\mathbb{E}[\|\widehat{f}_n - f\|_p^p] \leq 2^{p-1}\left(\underbrace{\mathbb{E}[\|\widehat{E}_{j0}f - E_{j0}f\|_p^p]}_{T_1} + \underbrace{\mathbb{E}[\|\widehat{D}_{j_0,j_1}f - D_{j_0}f\|_p^p]}_{T_2}\right).$$

To study the convergence rates of these terms the main tools are the following Lemmas given respectively in (21) and (8). Here for any $p \geq 1$ we denote $\|\cdot\|_{\ell_p}$ the ℓ_p-norm defined by $\|a\|_{\ell_p}^p = \sum_i |a_i|^p$ for any sequence of real number $(a_i)_{i\geq 0}$.

Lemma 6.3. *Let δ denote with no distinction ϕ and ψ. For any $1 \le p \le \infty$, there exists $c_1, c_2 > 0$ such that for all $j \ge 0$ and all sequence $(a_k)_{0 \le k \le 2^j - 1}$ we have*

$$c_1 2^{j(p/2-1)} \|a\|_{\ell_p}^p \le \left\| \sum_{k=0}^{2^j-1} a_k \delta_{j,k} \right\|_p^p \le c_2 2^{j(p/2-1)} \|a\|_{\ell_p}^p.$$

Lemma 6.4. *For any $1 \le p < \infty$ there exists a constant $C > 0$ such that for all $0 \le j^- \le j^+$ and any triangular arrays $(a_{j,k})_{j^- \le j \le j^+, 0 \le k \le 2^j - 1}$ we have*

$$\left\| \sum_{j=j^-}^{j^+} \sum_{k=0}^{2^j-1} a_{j,k} \delta_{j,k} \right\|_p^p \le C \begin{cases} \displaystyle\sum_{j=j^-}^{j^+} 2^{j(p/2-1)} \|a_j\|_{\ell_p}^p, & \text{if } 1 \le p \le 2, \\ \displaystyle\left(\sum_{j=j^-}^{j^+} 2^{j\beta p/(p-2)} \right)^{(p-2)/2} \sum_{j=j^-}^{j^+} 2^{j(p/2-1-\beta p/2)} \|a_j\|_{\ell_p}^p, & \text{if } p > 2. \end{cases}$$

Here β is an arbitrary real number.

6.2.1. *Bias of scale estimation T_1.* Lemma 6.3 for $\delta = \phi$ and $a_k = \widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}$ yields the existence of $C > 0$ such that

$$T_1 = \mathbb{E} \left\| \sum_{k=0}^{2_0^j-1} (\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \phi_{j_0,k} \right\|_p^p \le C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E} |\widehat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^p.$$

Thanks to Lemma 3.1, the term T_1 is bounded by

$$T_1 \le C\, 2^{j_0 p/2} n^{-p/2}.$$

Note that the choice of j_0 in Theorem 3.1 implies that the order of the bound is $(2^{j_0}/n)^{p/2} \le C n^{-pN/(2+2N)}$ negligible compare with $n^{-ps/(1+2s)}$ thanks to the hypothesis $N > 2s$. Note that

$$\frac{\alpha_-}{\alpha_+} = 1 + \frac{\epsilon}{2sp\pi(1 + 2(s - 1/\pi))}, \tag{6.3}$$

in order that if $\epsilon > 0$ then $\alpha_+ < \alpha_-$, if $\epsilon < 0$ then $\alpha_+ > \alpha_-$ and if $\epsilon = 0$ then $\alpha_+ = \alpha_-$. We conclude that for all possible choices of ϵ the term T_1 is negligible.

6.2.2. *Details term T_2.* The proof is based on multiple applications of Lemma 6.4 with $a_{j,k} = \gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}$ and different j^- and j^+. Studying the expectation of the loss, according to Lemma 6.4 and from the linearity of the expectation, the key terms of the bounds are

$$\mathbb{E} \|\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}\|_{\ell_p}^p = \sum_{k=0}^{2^j-1} \mathbb{E} |\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}|^p.$$

The following Lemma gives upper bounds for the terms $\mathbb{E}|\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}|^p$ for any j, k:

Lemma 6.5. *Under the assumptions of Theorem 3.1, there exists a constant $C > 0$ such that for all $j_0 \leq j \leq j_1$ and all $0 \leq k \leq 2^j - 1$ we have*

$$\mathbb{E}|\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}|^p \leq C(\mathbf{B1} \wedge \mathbf{B2} \wedge \mathbf{B3}),$$

where $\mathbf{B1}$=$|\beta_{j,k}|^p$, $\mathbf{B2}$=$|\beta_{j,k}|^\pi \lambda_j^{p-\pi}$ *and* $\mathbf{B3}$=λ_j^p.

Proof. From the definition of the hard and soft-threshold functions γ_{λ_j} we have :

$$|\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}|^p \leq 2^{p-1}\left(|\widehat{\beta}_{j,k} - \beta_{j,k}|^p + |\lambda_j|^p\right)\mathbb{1}_{\{|\widehat{\beta}_{j,k}|>\lambda_j\}} + |\beta_{j,k}|^p \mathbb{1}_{\{|\widehat{\beta}_{j,k}|\leq\lambda_j\}}. \tag{6.4}$$

The idea is to introduce the difference $|\widehat{\beta}_{j,k} - \beta_{j,k}|$ and bound it using Lemma 3.1. More precisely, we have

$$\mathbb{1}_{\{|\widehat{\beta}_{j,k}|>\lambda_j\}} \leq \mathbb{1}_{\{|\widehat{\beta}_{j,k}-\beta_{j,k}|\geq\lambda_j/2\}} + \mathbb{1}_{\{|\beta_{j,k}|\geq\lambda_j/2\}},$$

and the expectation of the first term of the bound in (6.4) is bounded by:

$$\left(\mathbb{E}|\widehat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right)^{1/2}\left(\mathbb{P}(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda_j/2)\right)^{1/2} + |\lambda_j|^p \mathbb{P}(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda_j/2)$$
$$+ (\mathbb{E}|\widehat{\beta}_{j,k} - \beta_{j,k}|^p + |\lambda_j|^p)\mathbb{1}_{\{|\beta_{j,k}|\geq\lambda_j/2\}}.$$

Inequality (3.1) in Lemma 3.1 provides that $\mathbb{E}|\widehat{\beta}_{j,k} - \beta_{j,k}|^p \leq Cn^{-p/2}$ and consequently the expectation terms of the sum are smaller than $|\lambda_j|^p$. Applying (3.2) of Lemma 3.1 with $\lambda = \lambda_j$ satisfying $\lambda_j\sqrt{n} = Kj$, we infer the existence of a constant $c > 0$ such that

$$\left(\mathbb{P}(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda_j/2)\right)^{1/2} \leq C2^{-cKj}$$

from the inequality

$$(2^{j/2}/\sqrt{n})^b \leq (\log n)^{-2b-3} \leq Cj^{-2b-3}.$$

Using $\mathbb{1}_{\{|\widehat{\beta}_{j,k}|\leq\lambda_j\}} \leq \mathbb{1}_{\{|\widehat{\beta}_{j,k}-\beta_{j,k}|\geq\lambda_j\}} + \mathbb{1}_{\{|\beta_{j,k}|\leq 2\lambda_j\}}$, the expectation of the second term of the bound in (6.4) is lower than

$$|\beta_{j,k}|^p\left(\mathbb{P}(|\widehat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda_j) + \mathbb{1}_{\{|\beta_{j,k}|\leq 2\lambda_j\}}\right).$$

As above the probability term is bounded by 2^{-cKj}.

Using all the previous bounds in inequality (6.4) leads to an upper bound for $\mathbb{E}|\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k}|^p$ with the expression $B(p,j,k) + (\lambda_j^p + |\beta_{j,k}|^p)2^{-cKj}$ where

$$B(p,j,k) = \lambda_j^p \mathbb{1}_{\{|\beta_{j,k}|\geq\lambda_j\}} + |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\leq\lambda_j\}}.$$

We investigate three ways for bounding $B(p,j,k)$, using the fact that $\mathbb{1}_{\{a\leq b\}} \leq (b/a)^\alpha$ for all $a, b > 0$ and all $\alpha \geq 0$:

B1: $B(p,j,k) \leq C|\beta_{j,k}|^p$ using the indicator in the first term of the sum,

B2: $B(p,j,k) \leq C|\beta_{j,k}|^{\pi}\lambda_j^{p-\pi}$ with $a=\lambda_j$, $b=|\beta_{j,k}|$ and $\alpha=\pi$ in the first term and $a=|\beta_{j,k}|$, $b=\lambda_j$ and $\alpha=p-\pi$ in the second term of the sum,

B3: $B(p,j,k) \leq C\lambda_j^p$ using the indicator in the second term of the sum.

Thanks to a suitable choice of K it is obvious that $(\lambda_j^p+|\beta_{j,k}|^p)2^{-cKj}$ is smaller than ($\mathbf{B1}\wedge\mathbf{B2}\wedge\mathbf{B3}$). The result of Lemma 6.5 follows. □

The end of the proof of Theorem 3.1 is based on successive uses of **B1**, **B2** and **B3** depending on the resolution levels j.

Let us first consider the highest multi resolution levels $j>j_+$ where **B1** is the most efficient bound given in Lemma 3.1. The integer j_+ will be fixed later. From Minkowski's inequality we obtain the following bound for T_2, up to a constant:

$$\underbrace{\mathbb{E}\Big\|\sum_{j=j_0}^{j_+}\sum_{k=0}^{2^j-1}(\gamma_{\lambda_j}(\widehat{\beta}_{j,k})-\beta_{j,k})\psi_{j,k}\Big\|_p^p}_{T_{21}}+\underbrace{\mathbb{E}\Big\|\sum_{j_+<j\leq j_1}\sum_{k=0}^{2^j-1}\gamma_{\lambda_j}(\widehat{\beta}_{j,k})\psi_{j,k}-\sum_{j_+<j}\sum_{k=0}^{2^j-1}\beta_{j,k}\psi_{j,k}\Big\|_p^p}_{T_{22}}.$$

The term T_{22} is bounded applying Lemma 6.4 with $j^-=j_++1$, $j^+=\infty$ and $a_{j,k}=\gamma_{\lambda_j}(\widehat{\beta}_{j,k})-\beta_{j,k}$ for $j_+<j\leq j_1$ and $0\leq k\leq 2^j-1$, $a_{j,k}=-\beta_{j,k}$ for $j>j_1$ and $0\leq k\leq 2^j-1$. Using Lemma 6.5 with **B1**, we achieve that $\mathbb{E}\|a_j\|_{\ell_p}^p\leq C\|\beta_j\|_{\ell_p}^p$ for all $j>j_+$ and we have:

$$T_{22}\leq C\begin{cases}\displaystyle\sum_{j>j_+}2^{j(p/2-1)}\|\beta_j\|_{\ell_p}^p, & \text{if } 1\leq p\leq 2,\\ \displaystyle\left(\sum_{j>j_+}2^{j\beta p/(p-2)}\right)^{(p-2)/2}\sum_{j>j_+}2^{j(p/2-1-\beta p/2)}\|\beta_j\|_{\ell_p}^p, & \text{if } p>2.\end{cases}$$

The Sobolev inclusion $\mathcal{B}_{\pi,r}^s\subset\mathcal{B}_{p,r}^{s'}\subset\mathcal{B}_{p,\infty}^{s'}$ with $s'=s-1/\pi+1/p$ leads us to choose $\beta=-2s'$. Noting that $\sum_{j>j_+}2^{-j2s'p/(p-2)}\leq C2^{-j_+s'p2/(p-2)}$ we obtain the inequalities:

$$T_{22}\leq C2^{-j_+s'p}\sum_{j=j_+}^{\infty}2^{j(s'p+p/2-1)}\|\beta_j\|_{\ell_p}^p\leq C\|f\|_{s',p,\infty}^p2^{-j_+s'p}.$$

We can choose j_+ as the largest integer such as

$$2^{j_+}\leq\left(\frac{n}{\log n}\right)^{\frac{\alpha}{s'}}.$$

We have to check that $j_+\leq j_1$ for n sufficiently large, *i.e.* $s'>\alpha$. When $\epsilon\leq 0$ we have $\alpha=s'/(1+2(s-1/\pi))$ and $\alpha<s'$ because $s>1/\pi$ by hypothesis. When $\epsilon>0$, equality (6.3) implies that $\alpha_+\leq\alpha_-<s'$ and then obviously $s'>\alpha$ for all the possible values of ϵ. With this choice of j_+, the rate of convergence of T_{22} is then the one stated in the Theorem.

We study the convergence of T_{21} using the bounds **B2** and **B3** given in Lemma 6.5. Let us investigate further these two cases in order to compare their efficiency. On the one hand, using **B2** and noting that $\|f\|_{s,\pi,\infty} \leq \|f\|_{s,\pi,r} < \infty$, we obtain the inequality:

$$\mathbb{E}\|\gamma_{\lambda_j}(\widehat{\beta}_{j.}) - \beta_{j.}\|_{\ell_p}^p \leq C\lambda_j^{p-\pi}\|\beta_j\|_{\ell_\pi}^\pi \leq C\lambda_j^{p-\pi}2^{-j(s\pi+\pi/2-1)}.$$

On the other hand, using **B3** we have directly

$$\mathbb{E}\|\gamma_{\lambda_j}(\widehat{\beta}_{j.}) - \beta_{j.}\|_{\ell_p}^p \leq C2^j\lambda_j^p.$$

The rate of these two upper bounds are equivalent when $\lambda_j^2 = 2^{-j(2s+1)}$. Replacing λ_j by its value, it follows that **B2** is more efficient for $j > j_-$ with j_- the largest integer such that

$$2^{j_-} \leq \left(\frac{n}{\log n}\right)^{\frac{1}{1+2s}}.$$

We check that $j_- \leq j_+$ for all n and all possible values of ϵ as $\alpha = s/(1+2s) > s'/(1+2s)$ if $\epsilon \geq 0$ and $s'/\alpha = 1 + 2s - 2/\pi < 1 + 2s$ if $\epsilon \leq 0$.

We decompose again T_{21} using Minkowski's inequality. It gives the following upper bound, up to a constant:

$$\underbrace{\mathbb{E}\Big\|\sum_{j=j_0}^{j_-}\sum_{k=0}^{2^j-1}(\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k})\psi_{j,k}\Big\|_p^p}_{T_{211}} + \underbrace{\mathbb{E}\Big\|\sum_{j_-<j\leq j_+}\sum_{k=0}^{2^j-1}(\gamma_{\lambda_j}(\widehat{\beta}_{j,k}) - \beta_{j,k})\psi_{j,k}\Big\|_p^p}_{T_{212}}.$$

We control the term T_{211} using **B3** according to the discussion above and applying Lemma 6.4 with $\lambda_j^2 \leq C\log n/n$:

$$T_{211} \leq C(\log n/n)^{p/2}\begin{cases}\displaystyle\sum_{j=j_0}^{j_1} 2^{jp/2}, & \text{if } 1 \leq p \leq 2,\\ \displaystyle\left(\sum_{j_0}^{j_-} 2^{j\beta p/(p-2)}\right)^{(p-2)/2}\sum_{j=j_0}^{j_1} 2^{jp(1-\beta)/2}, & \text{if } p > 2.\end{cases}$$

Let us choose $\beta = 1/2$ in order to obtain the inequalities

$$T_{211} \leq C(\log n/n)^{-p/2}2^{j_-p/2} \leq C(\log n/n)^{p(1-1/(2s+1))/2} \leq C(\log n/n)^{p\alpha} \qquad \text{if } \epsilon \geq 0.$$

This term is negligible if $\epsilon < 0$ using (6.3).

To conclude, it remains to bound T_{212} using Lemma 6.4 and **B2**. We use that $\lambda_j^2 \leq C\log n/n$ and we let appear the symbol $\epsilon = s\pi - (p-\pi)/2$

$$T_{211} \leq C(\log n/n)^{(p-\pi)/2}\begin{cases}\displaystyle\sum_{j_-<j\leq j_+} 2^{-j\epsilon}, & \text{if } 1 \leq p \leq 2,\\ \displaystyle\left(\sum_{j_-<j\leq j_+} 2^{j\beta p/(p-2)}\right)^{(p-2)/2}\sum_{j_-<j\leq j_+} 2^{-j(\epsilon+p\beta/2)}, & \text{if } p > 2.\end{cases}$$

From now we have to distinguish the cases where $\epsilon \neq 0$ to these where $\epsilon = 0$.

- If $\epsilon \neq 0$: Let us take $\beta = -\epsilon/p$. Then if $\epsilon < 0$, we obtain the inequalities:
$$T_{211} \leq C(\log n/n)^{(p-\pi)/2} 2^{-j+\epsilon} \leq C(\log n/n)^{\alpha\epsilon/s'+(p-\pi)/2},$$
and we conclude using the equality $\alpha\epsilon/s' + (p-\pi)/2 = p\alpha$. If $\epsilon > 0$, then
$$T_{211} \leq C(\log n/n)^{(p-\pi)/2} 2^{-j-\epsilon} \leq C(\log n/n)^{p(1-1/(1+2s))/2} \leq C(\log n/n)^{p\alpha}.$$
- If $\epsilon = 0$: Then $p/2 = \pi s + \pi/2$ and as $s\pi > 1$ we deduce that $p > 2$. Moreover we notice that $(p-\pi)/2 = \alpha p$ and thus
$$2^{j(p/2-1)}\mathbb{E}\|\gamma_{\lambda_j}(\widehat{\beta}_{j.}) - \beta_j\|_{\ell_p}^p \leq C\lambda_j^{(p-\pi)/2}\|\beta_j\|_{\ell_\pi}^\pi \leq C\lambda_j^{2\alpha p} 2^{j(\pi s+\pi/2-1)}\|\beta_j\|_{\ell_\pi}^\pi.$$
Let us denote $t_j = 2^{j(\pi s+\pi/2-1)}\sum_{k=0}^{2^j-1}\|\beta_j\|_{\ell_\pi}^\pi$. Notice that from the definition of the Besov norms we have $\sum_j t_j^{r/\pi} \leq C\|f\|_{s,\pi,r}^r$. From Lemma 6.4 with $\beta = 0$ we achieve
$$T_{211} \leq C(\log n/n)^{\alpha p} j_+^{p/2-1} \sum_{j_- < j \leq j_+} t_j, \text{ as } p > 2.$$
If $r \leq \pi$ then $\sum_{j_- < j \leq j_+} t_j \leq C$ and the result of Theorem 3.1 follows. If $\pi < r$ we use Holder's inequality at the powers r/π and $r/(r-\pi)$:
$$\sum_{j=j_0}^{j_+} t_j \leq C\|f\|_{s,\pi,r}^\pi \left(\sum_{j=j_0}^{j_+} t_j^{r/(r-\pi)}\right)^{1-\pi/r} \leq C j_+^{1-\pi/r} \leq C(\log n)^{1-\pi/r}.$$

6.3. **Proofs of results given in Sections 5 and 4.** In this Section are collected the proof of Proposition 4.1, Lemma 4.1, Proposition 4.2 and Propostion 5.1. Denoting with no distinction $\psi_{j,k}(x) - \beta_{j,k}$ and $\phi_{j,k}(x) - \alpha_{j,k}$ as $\widetilde{\delta}_{j,k}(x)$ for any j, k, we collect here some inequalities useful in this Section: $\mathbb{E}|\widetilde{\delta}_{j,k}(X_0)| \leq 2\|f\|_\infty\|\delta\|_1 2^{-j/2}$, $\mathbb{E}|\widetilde{\delta}_{j,k}(X_0)|^2 \leq \|f\|_\infty$, $\|\widetilde{\delta}_{j,k}\|_\infty \leq 2\|\delta\|_\infty 2^{j/2}$, $\mathrm{Lip}\,\widetilde{\delta}_{j,k} \leq \mathrm{Lip}\,\widetilde{\delta} 2^{3j/2}$ and $\|\widetilde{\delta}_{j,k}\|_{BV} \leq (\|\widetilde{\delta}\|_\infty + A\mathrm{Lip}\,\widetilde{\delta})2^{j/2+1}$ for all $j \geq 1$. The last assertion comes from the fact that $\widetilde{\delta}_{j,k}$ is a bounded Lipschitz function supported by $[(-A+k)2^{-j}, (A+k)2^{-j}]$.

Proof of Proposition 4.1. As for any j, k the function $\widetilde{\delta}_{j,k}$ has bounded variations we have
$$\left|\mathrm{Cov}\left(\widetilde{\delta}_{j,k}(X_{s_1})\cdots\widetilde{\delta}_{j,k}(X_{s_u}), \widetilde{\delta}_{j,k}(X_{s_{u+1}})\cdots\widetilde{\delta}_{j,k}(X_{s_{u+v}})\right)\right| \leq v\mathbb{E}\left|\widetilde{\delta}_{j,k}(X_{s_1})\cdots\widetilde{\delta}_{j,k}(X_{s_u})\right| \|\widetilde{\delta}_{j,k}\|_{BV}^v \widetilde{\phi}_v(r).$$
Noticing that $\|\widetilde{\delta}_{j,k}\|_\infty \leq \|\widetilde{\delta}_{j,k}\|_{BV}$, it follows
$$\begin{aligned} C_{u,v}^{j,k}(r) &\leq v\|\widetilde{\delta}_{j,k}\|_{BV}^{u+v-2}\|\widetilde{\delta}_{j,k}\|_{BV}\mathbb{E}|\widetilde{\delta}_{j,k}(X_0)|\widetilde{\phi}_v(r) \\ &\leq (u+v+uv)(2^{j/2+1}c(\|\delta\|_\infty + A\mathrm{Lip}\,\delta))^{u+v-1}2\|f\|_\infty\|\delta\|_1 2^{-j/2}\exp(-ar^b). \end{aligned}$$
Then Proposition 4.1 is proved. □

Proof of Lemma 4.1. The proof is very close to the one given in (7). First notice that (4.3) for $r = 0$ implies that $c \geq 1 = \sup_{g\in BV_1}\|g\|_{BV}$. From (4.2) we infer that $\widetilde{\phi}(\sigma(X_0), X_r) \leq \exp(ar^{-b})$ applying Lemma 4 of (6) on (4.2). From the Markov property we get $\widetilde{\phi}(\sigma(\{X_j, j \leq 0\}), X_r) = \widetilde{\phi}(\sigma(X_0), X_r)$. Now for any $\ell \leq 1$, for all $r \leq i_1 \leq \cdots \leq i_\ell$ consider any $g_{i_j} \in BV_1$ for $1 \leq j \leq \ell$.

Let us prove that we can restrict ourselves to the ℓ-uplets satisfying $i_1 < \cdots < i_\ell$. On the one hand, if $i_j = i_{j'} = r$ then we have $\|\mathbb{E}(g_r(X_r)g_r(X_r)|X_{i_1} = \cdot)\|_{BV} \leq c\|g_r^2\|_{BV}$ from (4.3). As from assumption $\|g_r\|_\infty \leq \|g_r\|_{BV} \leq 1$ then $\|g_r^2\|_{BV} \leq \|g_r\|_{BV} + \|g_r\|_{BV} \leq 2$ and we achieve the bound $\|\mathbb{E}(g_r(X_r)g_r(X_r)|X_{i_1} = \cdot)\|_{BV} \leq 2c$. On the other hand, if $i_j < i_{j'}$ then, from the Markov property, we get the equation

$$\|\mathbb{E}(g_{i_j}(X_{i_j})g_{i_{j'}}(X_{i_{j'}})|X_{i_1} = \cdot)\|_{BV} = \|\mathbb{E}(g_{i_j}(X_{i_j})\mathbb{E}(g_{i_{j'}}(X_{i_{j'}})|X_{i_j})|X_{i_1} = \cdot)\|_{BV}.$$

We proceed in two steps. Firstly from (4.3), we infer thet $x \mapsto g_{i_j,i_{j'}}(x) = \mathbb{E}(g_{i_{j'}}(X_{i_{j'}})|X_{i_j} = x)$ has variations bounded by c. Notice also that $\|g_{i_j,i_{j'}}\|_\infty \leq \|g_{i_{j'}}\|_\infty \leq 1$ and therefore we deduce the bound $\|g_{i_j}g_{i_j,i_{j'}}\|_{BV} \leq \|g_{i_j}\|_{BV} + \|g_{i_j,i_{j'}}\|_{BV} \leq 1+c$. Secondly, using (4.3) on $g_{i_j,i_{j'}}/\|g_{i_j,i_{j'}}\|_{BV}$, we infer that $\|\mathbb{E}(g_{i_j}(X_{i_j})g_{i_{j'}}(X_{i_{j'}})|X_{i_1} = \cdot)\|_{BV} \leq c(1+c) \leq c + c^2$.

As $c \geq 1$, this bound is larger than the one in the case $i_j = i_{j'}$. Now, by straightforward recurrences in the worst cases $i_1 < \cdots < i_\ell$, we get that

$$\|\mathbb{E}(g_{i_1}(X_{i_1}) \cdots g_{i_\ell}(X_{i_\ell})|X_{i_1} = \cdot)\|_{BV} \leq c + \cdots + c^\ell \leq \ell\, c^\ell \quad \text{as } c \leq 1.$$

Then, denoting $g_{i_1,\ldots,c_\ell}(x) = \mathbb{E}(g_{i_1}(X_{i_1}) \cdots g_{i_\ell}(X_{i_\ell})|X_{i_1} = x)$ for all x, we have almost surely the equation

$$\mathbb{E}(g_{i_1}(X_{i_1}) \cdots g_{i_\ell}(X_{i_\ell})|X_0) - \mathbb{E}(g_{i_1}(X_{i_1}) \cdots g_{i_\ell}(X_{i_\ell})) = \mathbb{E}(g_{i_1,\ldots,c_\ell}(X_{i_1})|X_0) - \mathbb{E}(g_{i_1,\ldots,c_\ell}(Xi_1))$$

From the definition of coefficients $\widetilde{\phi}$ we get

$$\|\mathbb{E}(g_{i_1,\ldots,c_\ell}(X_{i_1})|X_0) - \mathbb{E}(g_{i_1,\ldots,c_\ell}(X_{i_1}))\|_\infty \leq \ell c^\ell \widetilde{\phi}(\sigma(\{X_j, j \leq 0\}), X_r).$$

For all $1 \leq \ell \leq v$ this bound holds uniformly for all $g_{i_j} \in BV_1$, $1 \leq j \leq \ell$, using the definition of $\widetilde{\phi}_v(r)$ we conclude that $\widetilde{\phi}_v(r) \leq c^v \widetilde{\phi}(\sigma(\{X_j, j \leq 0\}), X_r) \leq c^v \exp(-ar^b)$. □

Proof of Proposition 4.2. We use the direct bound given in the proof of Lemma 1 of (23) under **(J)**:

$$\left|\mathrm{Cov}\left(\widetilde{\delta}_{j,k}(X_{i_1}) \cdots \widetilde{\delta}_{j,k}(X_{i_u}), \widetilde{\delta}_{j,k}(X_{i_{u+1}}) \cdots \widetilde{\delta}_{j,k}(X_{i_{u+v}})\right)\right| \leq 2^3 (2^{j/2} 2\|\delta\|_\infty)^{u+v-2} \gamma(r)$$

with $\gamma(r) = \mathbb{E}|\widetilde{\delta}_{j,k}(X_0)\widetilde{\delta}_{j,k}(X_r)| \vee (\mathbb{E}|\widetilde{\delta}_{j,k}(X_0)|)^2 \leq (\|f_r\|_\infty \vee 2\|f\|_\infty)\|\delta\|_1^2 2^{-j}$. Noticing that $\gamma \wedge \beta \leq \gamma^{1/4}\beta^{3/4}$ for any positive numbers γ and β, we combine the two bounds on the covariance terms and we infer that (D1) is satisfied with $M_2 = 2\|\delta\|_\infty$ and

$$\rho(r) = 2^{9/2}\|\delta\|_1^{3/2}(\mathrm{Lip}\,\delta)^{1/2}(\|f_r\|_\infty \vee \|f\|_\infty)^{3/4}\lambda(r)^{1/4}.$$

Assumptions of Proposition 4.2 on respective decrease and increase rates of $\lambda(r)$ and $\|f_r\|_\infty$ yield the existence of $C_0 > 0$ such that $\rho(r) \leq C_0 \exp(-a' r^{b'}/4)$. □

Proof of Propostion 5.1. We give the proof for $\widehat{f}_n$ but it also holds for $\widehat{f}_n^{STCV}$. We begin with recalling the result of Corollary 7.1 in (15):

Lemma 6.6 (Gouëzel, 2004)**.** *For any Lipschitz function δ_1, bounded measurable function δ_2 such that $\delta_1, \delta_2 = 0$ in a neighborhood of 0, then for any $0 < \alpha' < 1$, there exists some constant $C > 0$ such that*

$$Cov(\delta_1(X_0), \delta_2(X_r)) \sim C \int \delta_1(x)dx \int \delta_2(x)dx \; r^{1-1/\alpha'} \text{ when } r \to \infty. \tag{6.5}$$

We use the decomposition of $f \in \mathbb{L}^2$ in the orthogonal basis of wavelets functions and we obtain

$$\begin{aligned} \mathbb{E}(\|\widehat{f}_n - f\|_2^2) &\geq \mathbb{E}\| \sum_{k \in S_{j_0}} (\widehat{\alpha}_{j_0-k} - \alpha_{j_0-k})\phi_{j_0,k}\|_2^2 \\ &\geq \sum_{k \in S_{j_0}} \mathbb{E}(\widehat{\alpha}_{j_0-k} - \alpha_{j_0-k})^2. \end{aligned}$$

If we develop $\mathbb{E}(\widehat{\alpha}_{j_0-k} - \alpha_{j_0-k})^2$ using the covariance terms and denoting $\widetilde{\delta}_{j,k}(x) = \phi_{j,k}(x) - \alpha_{j,k}$ for $x \in [0.01, 1]$ and null elsewhere, it comes:

$$\mathbb{E}(\widehat{\alpha}_{j_0-k} - \alpha_{j_0-k})^2 = \frac{1}{n}\mathbb{E}(\widetilde{\delta}_{j,k}(X_i)^2) + 2\sum_{r=1}^{n-1} \frac{n-r}{n^2}\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r)).$$

We want to apply Lemma 6.6 with $\delta_1 = \delta_2 = \widetilde{\delta}_{j_0,k}$. We check easily the assumptions of this Lemma because of the definition of $\widetilde{\delta}_{j_0,k}$, resulting from the fact that we estimate the density on $[0.01, 1]$. Moreover $\int \phi_{j_0,k} = 2^{-j_0/2} \int \phi$ with $\int \phi > 0$ from assumption and then the covariance terms $\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r))$ are equivalent to $C2^{-j_0}\, r^{1-1/\alpha'}$ for some $C > 0$ as r goes to infinity.

Let n_0 be such that for all $n \geq n_0$, $u_{r,n} = \frac{n-r}{n^2}\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r))$ is nonnegative. For some $m > m' > 2$, we decompose the sum of covariance terms, for n sufficiently large, in four sums:

$$\sum_{r=1}^{n-1} \frac{n-r}{n^2}\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r)) = \sum_{r=1}^{n_0} u_{r,n} + \sum_{r=n_0}^{[n/m]} u_{r,n} + \sum_{r=[n/m]}^{[n/m']} u_{r,n} + \sum_{r=[n/m']}^{n-1} u_{r,n},$$

where $[a]$ denotes the integer part of a. The first term goes to 0 with rate n. Then, by definition of n_0, the second and the last terms are nonnegative. Concerning the sum $\sum_{r=[n/m]}^{[n/m']} u_{r,n}$, the summands $u_{r,n}$ are all larger than $(m'-2)/(m'n)\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r))$ that is equivalent to $C2^{-j_0}n^{-1}r^{1-1/\alpha'}$ as $n \to \infty$. The minimax rate α is such that $\alpha < 1/2$ and then by hypothesis $\alpha' \geq 1/(2\alpha+1) > 1/2$. Consequently, when n goes to infinity, the sum $\sum_{r=[n/m]}^{[n/m']} u_{r,n}$ is larger than a partial sum equivalent to $C2^{-j_0}n^{1-1/\alpha'}$ for some $C > 0$ as $n \to \infty$. As we assume $2\alpha \geq 1/\alpha' - 1$ we obtain the existence of some $C > 0$ such that

$$n^{2\alpha}2^{j_0} \sum_{r=1}^{n-1} \frac{n-r}{n^2}\mathrm{Cov}(\widetilde{\delta}_{j_0,k}(X_0), \widetilde{\delta}_{j_0,k}(X_r)) \geq C, \text{ for } n \text{ sufficiently large.}$$

Collecting these facts and using that $|S_{j_0}|$ is up to a constant equal to 2^{j_0}, we obtain that $n^{2\alpha}\mathbb{E}(\|\widehat{f}_n - f\|_2^2)$ is larger than some positive constant and the result of Proposition 5.1 is proved. □

Acknowledgements. We thank Karine Tribouley and Vincent Rivoirard for their helpful advices to one of the author. We also thank referees for having pointed out lapses in the first version of this paper.

References

[1] Andrews, D. Non strong mixing autoregressive processes. *Journal of Applied Probability 21* (1984), 930–934.

[2] Bosq, D., and Guegan, D. Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system. *Statistics & Probability Letters 25* (1995), 201–212.

[3] Comte, F., and Merlevède, F. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM P&S 6* (2002), 211–238.

[4] Daubechies, I. *Ten Lectures on Wavelets*, vol. 61. SIAM press, 1992.

[5] Dedecker, J., Doukhan, P., Lang, G., Leon, J. R., Louhichi, S., and Prieur, C. *Weak Dependence: Models, Theory and Applications.* Springer-Verlag, 2007.

[6] Dedecker, J., and Prieur, C. New dependence coefficients: Examples and applications to statistics. *Probability Theory and Related Fields 132* (2005), 203–235.

[7] Dedecker, J., and Prieur, C. An empirical central limit theorem for dependent sequences. *Stochastic Processes and their Applications 117*, 1 (2007), 121–142.

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

[9] Doukhan, P., and Louhichi, S. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications 84* (1999), 313–342.

[10] Doukhan, P., and Neumann, M. A Bernstein type inequality for times series. *Stochastic Processes and their Applications 117-7* (2007), 878–903.

[11] Doukhan, P., Teyssière, G., and Winant, P. A LARCH(∞) vector valued process. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., vol. 187 of *Lectures Notes in Statistics*, Springer, pp. 245–258.

[12] Doukhan, P., and Truquet, L. A fixed point approach to model random fields. *Alea 2* (2007), 111–132.

[13] Doukhan, P., and Wintenberger, O. Weakly dependent chains with infinite memory. To appear in Stochastic Processes and their Applications.

[14] Doukhan, P., and Wintenberger, O. Invariance principle for new weakly dependent stationary models. *Probabability and Mathematical Statistics 27* (2007), 45–73.

[15] Gouëzel, S. Central limit theorem and stable laws for intermittent maps. *Probability Theory and Related Fields 128* (2004), 82–122.

[16] Hardle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. *Wavelets Approximation and Statistical Applications*, vol. 129. Springer Verlag, Lecture Notes in Statistics, 1998.

[17] Juditsky, A., and Lambert-Lacroix, S. On minimax density estimation on $\mathbb{R}$. *Bernoulli 10*, 2 (2004), 187–220.

[18] Liverani, C., Saussol, B., and Vaienti, S. A probabilistic approach to intermittency. *Ergodic Theory and Dynamical Systems 19* (1999), 671–686.

[19] Mallat, S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 7 (1989), 674–693.

[20] Maume-Deschamps, V. Exponential inequalities and functional estimations for weak dependent data ; applications to dynamical systems. *Stochastics and Dynamics 6* (2006), 535–560.

[21] Meyer, Y. *Wavelets and Operators*. Cambridge University Press, 1992.

[22] Prieur, C. *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques*. PhD thesis, CREST, 2001.

[23] Ragache, N., and Wintenberger, O. Convergence rates for density estimators of weakly dependent time series. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., vol. 187 of *Lectures Notes in Statistics*, Springer, pp. 349–372.

[24] Tribouley, K., and Viennet, G. L_p-adaptive density estimation in a β-mixing framework. *Annales de l'Institut Henri Poincaré, B 34*, 2 (1998), 179–208.

[25] Vanharen, M.-L. Estimation par ondelettes dans les systèmes dynamiques. *Comptes Rendus de l'Académie des Sciences de Paris 342*, 7 (2006), 523–525.

[26] Vannucci, M. Nonparametric density estimation using wavelets. Tech. rep., Texas A and M university, 1998.

[27] Viana, M. Stochastic dynamics of deterministic systems. available at http://w3.impa.br/∼viana, 1997.

[28] Vidakovic, B. Pollen bases and Daubechies-Lagarias algorithm in MATLAB, 2002. available at http://www2.isye.gatech.edu/∼brani/datasoft/DL.pdf.

[29] Wavelab. http://www-stat.stanford.edu/∼wavelab/.

[30] Young, L. Recurrence times and rates of mixing. *Israël Journal of Mathematics 110* (1999), 0021–2172.

(1) Laboratoire Jean Kuntzmann, Grenoble-INP, BP 53, 38041 Grenoble Cedex 9, France

E-mail address: irene.gannaz@imag.fr

(2) SAMOS-MATISSE (Statistique Appliquée et MOdélisation Stochastique) Centre d'Économie de la Sorbonne Université Paris 1 - Panthéon-Sorbonne, CNRS 90, rue de Tolbiac 75634 Paris Cedex 13, France

E-mail address: olivier.wintenberger@univ-paris1.fr