



**HAL**  
open science

# Non linear wavelet density estimation on the real line

Mathieu Sart

► **To cite this version:**

| Mathieu Sart. Non linear wavelet density estimation on the real line. 2023. hal-03915562v4

**HAL Id: hal-03915562**

**<https://hal.science/hal-03915562v4>**

Preprint submitted on 6 Nov 2023

**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.

# NON LINEAR WAVELET DENSITY ESTIMATION ON THE REAL LINE

MATHIEU SART

ABSTRACT. We investigate the problem of optimal density estimation on the real line  $\mathbb{R}$  under  $\mathbb{L}^1$  loss. We carry out a new way to select the important coefficients in some wavelet expansions. We study the resulting estimator when the density is smooth with dominated tails. This assumption is very mild and allows in particular to deal with singularities, spatially inhomogeneous smoothness, and fat tailed distributions.

## 1. INTRODUCTION.

We consider a real valued random variable  $X$  and suppose that its distribution admits a density  $f$  with respect to the Lebesgue measure. Our aim is to estimate  $f$  from the observation of  $n$  independent copies  $X_1, \dots, X_n$  of  $X$ .

The method we use is based on a decomposition of the density in a (bi-orthogonal) wavelet basis. The challenge lies in the choice of the coefficients to be estimated and those to be set to zero. A successful choice leads to an estimator  $\hat{f}$  that is neither over-smoothed nor under-smoothed. A way to check this point mathematically is to consider a class  $\mathcal{F}$  of functions and to compute

$$R_\ell(\hat{f}, \mathcal{F}) = \sup_{f \in \mathcal{F}} \mathbb{E} \left[ \ell(f, \hat{f}) \right].$$

In this formula,  $\ell$  is a loss function that is up to the statistician, for instance  $\ell = d_q^q$  where  $d_q$  is the  $\mathbb{L}^q$  distance. The above maximal risk can then be compared to the minimax risk

$$R_\ell(\mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left[ \ell(f, \hat{f}) \right].$$

Naturally,  $R_\ell(\mathcal{F}) \leq R_\ell(\hat{f}, \mathcal{F})$  and we expect the reverse inequality to be true to within a multiplicative factor. If this factor does not depend on  $n$ ,  $\hat{f}$  is said to be rate optimal (and nearly rate optimal if the factor may grow as  $\log^k n$ ).

A wavelet estimator is usually studied under the assumption that  $f$  belongs to a ball  $\mathfrak{B}_{p,\infty}^\alpha(R)$  of a Besov space. The precise definition of this set is recalled in Section 3.1 below. The parameter  $\alpha$  indicates the regularity of  $f$  whereas  $p$  measures, in some sense, the sparsity of its wavelet expansion. This assumption of regularity is seldom the only one. There are two points to watch out for when assessing the quality of a wavelet estimator. First, we have to examine the precise set  $\mathcal{F} \subset \mathfrak{B}_{p,\infty}^\alpha(R)$  on which it is rate optimal (or at least nearly rate optimal). Second, we have to look at the conditions on  $\alpha$  and  $p$ . The larger  $\mathcal{F}$  is, and the weaker the conditions, the better.

---

*Date:* November, 2023.

*2010 Mathematics Subject Classification.* 62G05, 62G07.

*Key words and phrases.* Besov classes, minimax rates, non-parametric estimation, wavelet methods.

The simplest wavelet estimation method is to keep all the coefficients up to a certain resolution and discard the others. This leads to a linear estimator that is – if the final resolution is properly chosen – rate optimal for the  $\mathbb{L}^q$  loss when  $f$  is a compactly supported density of  $\mathfrak{B}_{p,\infty}^\alpha(R)$  with  $p \geq q$ . The case  $p < q$  is more delicate and can only be solved with non-linear methods, see [DJKP96]. The two main rules that we know of consist in thresholding the coefficients individually or by block. For more details about them, we refer to [Aut06, CC05, DJ96, DJKP96, HKP98, HKPT12, KPT96]. Note, however, that the resulting estimators are usually studied under somewhat restrictive additional assumptions (for instance  $\alpha > 1/p$ , which imposes the continuity of  $f$ ). These additional assumptions can significantly modify the minimax rates. A description of these in the compact case, to within log factors but without superfluous conditions, can be found in [Sar21b]. We would also like to mention an important point: in the 8 references above, 5 of them involve undesirable log factors in the convergence rates. The three exceptions, [CC05, HKP98] and [DJ96], rely on very strong additional assumptions and therefore do not, unfortunately, solve the problem of optimal estimation on a Besov ball, even in the compact case. More generally, we were unable to find in the literature an optimal estimator (based on wavelets or not) when  $p < q$ ,  $\alpha < 1/p$ , and  $f$  compactly supported but not bounded, no matter  $q \in (1, \infty)$ . Log factors separate the upper-bounds from minimax lower-bounds in the best of cases.

In the present manuscript, we pay particular attention to the  $\mathbb{L}^1$  loss. The spatial adaptivity of a non-linear estimator may be revealed by studying its risk when the density is compactly supported and in  $\mathfrak{B}_{p,\infty}^\alpha(R)$  with  $p < 1$ . The optimal estimation rate in this context is  $n^{-\alpha/(2\alpha+1)}$  for all  $\alpha > 1/p - 1$ . The only previous paper we know of that achieves this result is [Bir06a]. His estimator is based on a “theoretical” construction though (the term theoretical means that we do not know how to make his procedure computationally feasible). This rate  $n^{-\alpha/(2\alpha+1)}$  is nevertheless a benchmark that should be reached by a wavelet estimator under the sole condition that  $\alpha$  belongs to an interval of the form  $(1/p - 1, \tau)$ . This restriction on  $\alpha$  ensures the classical characterisation of Besov spaces in terms of wavelet coefficients. We have not been able to find in the literature a local or block thresholded wavelet estimator with this property. This will be the case for our wavelet estimator.

The wavelet approach is less developed in the context of density estimation on the real line  $\mathbb{R}$ . The main papers we know on this subject are [JLL04, RBRTM11]. In both cases, the estimator is term-by-term thresholded. It is also nearly optimal over some balls  $\mathfrak{B}_{p,\infty}^\alpha(R)$  of Besov spaces. More precisely, the first paper deals with  $p = \infty$  and  $q > 1$  whereas the second one covers the cases  $p \geq 1$  and  $q = 2$  (under an additional boundedness assumption when  $p < 2$  though). Other solutions to wavelets have certainly been proposed in the literature to cope with infinite supports. It is beyond the scope of this introduction to describe them all. We simply cite here the papers of [GL11, GL14, Lep13, LW19]. They contain the most general results we know about the minimax rates.

Let us mention that difficulties occur when the estimation is performed on  $\mathbb{R}$  with the  $\mathbb{L}^1$  loss. This distance gives more weight to the estimation errors in the tails of  $f$  than the other  $\mathbb{L}^q$  distances. It follows that a pure regularity assumption – such as  $f \in \mathfrak{B}_{\infty,\infty}^\alpha(R)$  with  $R$  large enough – is not sufficient to ensure the convergence to zero of the minimax risk. A way to bypass this problem is to add a constraint on the tails of  $f$ , see the seminal paper of [GL14]. They show that the optimal estimation rate of a (bounded) function in a Besov ball with  $p \geq 1$  is, within logarithmic factors,  $n^{-\gamma}$  where  $\gamma \in [0, \alpha/(2\alpha + 1)]$  depends on the tails of  $f$ . If we ignore these log factors, it means that the usual rate  $n^{-\alpha/(2\alpha+1)}$  can be recovered when the tails of the distribution are not

too fat. In the present paper, we aim to refine this result. We intend to explore cases they do not cover ( $p < 1$  and  $f$  not bounded for instance), and tackle the tricky question of log factors in the convergence rates.

It is therefore difficult in statistics to perform optimal estimation under smoothness constraints without making assumptions about the density which greatly simplify the mathematics. By “optimal”, we mean that no logarithmic factors are lost. The assumptions we wish to avoid are “ $p \geq q$ ”, “ $f$  bounded”, “ $f$  compactly supported”. None of the aforementioned papers satisfactorily solves this problem, no matter  $q \in [1, \infty)$ . The aim of our work is to shed new light on this subject and provide some answers.

The choice of the  $\mathbb{L}^1$  distance as loss function may induce some difficulties in the non-compact case as explained above, but has a considerable advantage over the other  $\mathbb{L}^q$  distances. Indeed, a density inevitably lies in  $\mathbb{L}^1(\mathbb{R})$  as opposed to  $\mathbb{L}^q(\mathbb{R})$  when  $q \in (1, \infty)$ . We can therefore assess the risk of the estimator even when  $f \notin \mathbb{L}^q(\mathbb{R})$ , and work with collection of densities for which the minimax  $\mathbb{L}^q$  risk does not tend to zero. This is the case for instance when the collection is composed of compactly supported densities of  $\mathfrak{B}_{p,\infty}^{1/p-1/q}(R)$  on  $[0, 1]$  with  $p < q$  and  $R$  large enough. The precise lower minimax bound may be found in [Sar21b]. Here, an optimal estimator converges at the classical rate  $n^{-(1/p-1/q)/(2(1/p-1/q)+1)}$  for the  $\mathbb{L}^1$  loss.

We propose in Section 2 a new estimation method of the density. It is based on a decomposition of  $f$  in a wavelet basis. Our method estimates only some of the coefficients and sets the others to zero. Each coefficient can, a priori, be retained or discarded. It is kept only if its estimation improves a heuristic assessment of the estimator global risk. The latter depends on an error term that is unobserved but that can be controlled with high probability. Thus, the way in which the coefficients are thresholded depends on how this error is controlled. If we control this error too roughly, we end up with a loss of logarithmic factors, see Section 2.2 for more details. Fortunately, the error can be controlled very precisely using appropriate tools from the concentration of measure and empirical processes theories (Talagrand’s inequality, Poissonian inequality for self-bounding functionals, chaining arguments, VC subgraph, universal entropy...). This allows us to get rid of log factors in almost all cases (see below). The downside of our approach is that it leads to lengthy and rather technical proofs.

We study the global  $\mathbb{L}^1$  risk of our estimator in Section 3 under two conditions. The first is that  $f$  belongs to a Besov ball  $\mathfrak{B}_{p,\infty}^\alpha(R)$  with  $p > 0$  and  $\alpha \in ((1/p - 1)_+, \tau)$  where  $\tau$  only depends on the choice of the wavelet basis. The second condition concerns the tails of  $f$ . It is described in detail in Section 3.1.2. To put it simply here, we can say that it allows densities to be fat tailed, i.e. it allows  $f(x)$  to be bounded from above by something of the order of  $|x|^{-1/\theta}$  where  $\theta \in (0, p) \cap (0, 1)$ . We show that our estimator achieves the classical rate  $n^{-\alpha/(2\alpha+1)}$  associated with compactly supported densities when the tails of  $f$  are not too fat, that is when  $\theta < \theta_0$  where  $\theta_0$  depends on  $\alpha, p$ . When  $\theta > \theta_0$ , the rate is rather of the form  $n^{-\gamma}$  for some  $\gamma \in (0, \alpha/(2\alpha + 1))$ . It is still optimal. In the limiting case  $\theta = \theta_0$ , our estimator reaches the rate  $(\log n)^{\gamma'} n^{-\alpha/(2\alpha+1)}$  where  $\gamma'$  depends on  $\alpha, p$ . This logarithmic factor cannot be ruled out, and our estimator is once again rate optimal.

We also examine the risk of our estimator under the sole condition that  $f \in \mathfrak{B}_{p,\infty}^\alpha(R)$  with  $p < 1$  and  $\alpha \in ((1/p - 1)_+, \tau)$ . This assumption turns out to be sufficient to ensure the convergence of our estimator. It is moreover always rate optimal, except possibly when  $\alpha = (1 - p)/(2p - 1)$ , where it could be, in the worst possible scenario, only near optimal. This result contrasts with

what happens when  $p \geq 1$ , where, on the contrary, the minimax risk may not tend to zero. We also investigate the case  $\alpha \in (0, 1/p - 1]$  and show that the condition  $\alpha > (1/p - 1)_+$  is, in some sense, necessary to estimate the density (even if it is compactly supported). A precise formulation of these results may be found in Sections 3.2, 3.3 and 3.4.

For ease of presentation, we have only discussed balls  $\mathfrak{B}_{p,\infty}^\alpha(R)$  of Besov spaces above. However, our bounds are slightly more general. The Besov balls can be replaced in almost all cases by subsets  $\mathcal{WB}_{p,\infty}^\alpha(R)$  of weak Besov spaces without changing the convergence rate of our estimator. These sets are defined in the same section as the Besov balls, i.e. in Section 3.1.

Finally, let us mention that the computational complexity of our procedure is nearly linear in the number  $n$  of observations. We give more details on this point in Section 3.5.

Throughout the paper, we suppose  $n \geq 2$  and denote by  $|A|$  the size of a finite set  $A$ . The letters  $c, C, c', \dots$  stand for quantities that may change from line to line. The proof of our main result is deferred in Section 4. The appendix contains additional proofs as well as some results used in Section 4.

## 2. ESTIMATION PROCEDURE

**2.1. Bi-orthogonal wavelet basis.** Our estimation method relies on a decomposition of the density in a bi-orthogonal wavelet basis. In such a basis, any square integrable function  $f$  takes the form

$$(1) \quad f = \sum_{k \in \mathbb{Z}} \alpha_k \bar{\phi}_k + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k},$$

where for any  $j \geq 0$  and  $k \in \mathbb{Z}$ ,

$$\alpha_k = \int f(x) \phi_k(x) dx, \quad \beta_{j,k} = \int f(x) \psi_{j,k}(x) dx,$$

where for any  $x \in \mathbb{R}$ ,

$$\begin{aligned} \phi_k(x) &= \phi(x - k), & \psi_{j,k}(x) &= 2^{j/2} \psi(2^j x - k), \\ \bar{\phi}_k(x) &= \bar{\phi}(x - k), & \bar{\psi}_{j,k}(x) &= 2^{j/2} \bar{\psi}(2^j x - k), \end{aligned}$$

and where  $\phi, \bar{\phi}, \psi, \bar{\psi}$  stand for dual father and mother wavelets. Equality (1) can be shortened by setting  $\beta_{-1,k} = \alpha_k$ ,  $\psi_{-1,k} = \phi_k$ ,  $\bar{\psi}_{-1,k} = \bar{\phi}_k$ . It then becomes

$$(2) \quad f = \sum_{j=-1}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k}.$$

In the following, we suppose that these four wavelets are bounded and compactly supported. Moreover, we take  $\phi = 1_{[0,1]}$ , and suppose that  $\psi$  is piecewise constant.

The simplest example of bi-orthogonal basis is the Haar basis where  $\bar{\phi} = \phi$  and  $\bar{\psi} = \psi = 1_{[0,1/2]} - 1_{[1/2,1]}$ . This basis is even orthogonal. In general, however,  $\bar{\phi}$  and  $\bar{\psi}$  are not necessarily piecewise constant and can instead be smooth. We refer to [CDF92] for the construction of such a bi-orthogonal basis where  $\bar{\phi}$  and  $\bar{\psi}$  are Hölder continuous with exponent  $\tau$  ( $\tau \in \mathbb{N}^*$  is an arbitrary number to be chosen).

Note that the density  $f$  we estimate is not necessarily in  $\mathbb{L}^2(\mathbb{R})$ . Despite this, the series always converges in  $\mathbb{L}^1$  norm and equality (1) remains valid. This result is based on classical techniques in wavelet analysis, see Appendix H.

**2.2. Outline.** We begin by presenting the underlying ideas of our estimation procedure. It will be described in detail in Section 2.3.

**2.2.1. Heuristic.** We consider a collection  $\mathbf{K} = (K_j)_{j \geq -1}$  of subsets of  $\mathbb{Z}$  and define the linear wavelet estimator

$$\hat{f}_{\mathbf{K}} = \sum_{j=-1}^{\infty} \sum_{k \in K_j} \hat{\beta}_{j,k} \bar{\psi}_{j,k}, \quad \text{where} \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^n \psi_{j,k}(X_i)$$

denotes the empirical version of  $\beta_{j,k}$ .

The quality of this estimator depends on the choice of the collection  $\mathbf{K}$  by the statistician. Ideally,  $\mathbf{K}$  should be chosen in such a way that the risk of the estimator is minimal. Observe that

$$\left\| \hat{f}_{\mathbf{K}} - f \right\|_1 \leq c \sum_{j=-1}^{\infty} \left\{ 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| + \mathcal{B}_j(K_j) + \mathcal{E}_j(K_j) \right\},$$

where  $c$  only depends on the wavelet basis, where

$$\mathcal{B}_j(K_j) = -2^{-j/2} \sum_{k \in K_j} |\beta_{j,k}|,$$

and where

$$(3) \quad \mathcal{E}_j(K_j) = 2^{-j/2} \sum_{k \in K_j} |\hat{\beta}_{j,k} - \beta_{j,k}|$$

represents the error due to the estimation of the coefficients indexed by  $K_j$ .

The term  $\mathcal{B}_j(K_j)$  can be estimated by

$$\hat{\mathcal{B}}_j(K_j) = -2^{-j/2} \sum_{k \in K_j} |\hat{\beta}_{j,k}|.$$

The triangle inequality ensures that

$$|\hat{\mathcal{B}}_j(K_j) - \mathcal{B}_j(K_j)| \leq \mathcal{E}_j(K_j),$$

and hence

$$(4) \quad \left\| \hat{f}_{\mathbf{K}} - f \right\|_1 \leq c \sum_{j=-1}^{\infty} \left\{ 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| + \hat{\mathcal{B}}_j(K_j) + 2\mathcal{E}_j(K_j) \right\}.$$

The key to minimizing the  $\mathbb{L}^1$  risk (or at least the above upper-bound) is therefore to control the error  $\mathcal{E}_j(K_j)$  with high probability. If  $\hat{\mathcal{E}}_j(K_j)$  denotes a known upper-bound of  $\mathcal{E}_j(K_j)$ , minimizing

$$(5) \quad \hat{\mathcal{B}}_j(K_j) + 2\hat{\mathcal{E}}_j(K_j)$$

among all the subsets  $K_j$  of  $\mathbb{Z}$  is a possible strategy to minimize the right-hand side of (4). In other words, we may consider a subset  $\hat{K}_j$  of  $\mathbb{Z}$  minimizing (5). It contains all the indices  $k$  of the coefficients  $\hat{\beta}_{j,k}$  to be kept. We then estimate  $f$  by  $\hat{f}_{\hat{\mathbf{K}}}$  where  $\hat{\mathbf{K}} = (\hat{K}_j)_{j \geq -1}$ . The interpretation

of  $\widehat{\mathbf{K}}$  is therefore relatively simple: it is obtained by minimizing a criterion (5) that aims to estimate the  $\mathbb{L}^1$  error of the estimator.

**2.2.2. Local thresholding.** The point of view we have just discussed is flexible enough to re-discover, and therefore reinterpret, some existing rules of literature. All we have to do is play with  $\widehat{\mathcal{E}}_j(K_j)$ .

For instance, we may choose an estimated error term  $\widehat{\mathcal{E}}_j(K_j)$  of the form

$$(6) \quad \widehat{\mathcal{E}}_j(K_j) = 2^{-1-j/2} \sum_{k \in K_j} \widehat{e}_j(k).$$

The procedure then becomes purely local. It only keeps the coefficients above a given threshold. To be more precise:

$$\widehat{K}_j = \left\{ k \in \mathbb{Z}, |\widehat{\beta}_{j,k}| \geq \widehat{e}_j(k) \right\}.$$

We now propose to reason a little more formally to clarify what  $\widehat{e}_j(k)$  can be. We set the coefficients to zero when  $j$  is large, say  $j \geq \widehat{J}$  for some  $\widehat{J}$ , and minimize the criterion only when  $j$  is smaller. This decision necessarily induces a bias. The larger  $\widehat{J}$  is, the smaller it is. Anyway, a risk bound can be obtained for the resulting estimator on the event  $\mathcal{E}$  defined by

$$“\forall j \leq \widehat{J}, \forall K_j \subset \mathbb{Z}, \widehat{\mathcal{E}}_j(K_j) \geq \mathcal{E}_j(K_j)”.$$

If we require  $\widehat{\mathcal{E}}_j(K_j)$  to be of the form (6), a possible choice for  $\widehat{e}_j(k)$  is

$$(7) \quad \widehat{e}_j(k) = c_1 2^{j/2} \left[ \sqrt{\frac{\log n}{n} \mathbb{E}[\psi^2(2^j X - k)]} + \frac{\log n}{n} \right].$$

This result is valid with probability  $1 - 1/n^c$  for some  $c > 1$ , and all  $j \in \{0, \dots, n\}$ ,  $k \in \mathbb{Z}$  (when  $j = -1$ , replace  $\psi$  by  $\phi$ , and when  $j \geq n$  replace  $\log n$  by  $\log(jn)$ ). We omit the case  $j \geq n$  as the coefficients are usually set to zero in the literature (the classical conditions are even often stronger). Formula (7) can be obtained, for example, by using our Lemma 2 page 20 (with some elementary maths). Naturally, its right-hand side is unknown in practice. The old way out is to replace the expectation by a deterministic upper-bound such as  $c_2 \|f\|_\infty 2^{-j}$ . This gives a threshold of the order of  $\sqrt{\|f\|_\infty \log n/n}$  when  $j$  is not too large, say  $2^j \leq n/\log n$ . It is barely larger than the classic threshold  $\sqrt{\|f\|_\infty j/n}$  proposed in [DJKP96] (see below for more details). When  $f$  is not bounded, or when we do not know an upper-bound on its supremum norm, the foregoing is pointless. An alternative solution is to replace the expectation by a probabilist upper-bound roughly of the order of

$$\frac{1}{n} \sum_{i=1}^n \psi^2(2^j X_i - k).$$

The threshold then becomes random, and the resulting procedure is quite similar to what can be found in literature, see [JLL04, RBRTM11] for instance. The main difference with these two papers is that they estimate the variance of  $\psi(2^j X - k)$  instead of its moment of order 2. But it doesn't matter: in all of the above, the factor  $\log n/n$  appears in the threshold. It therefore seems difficult to avoid in the convergence rates.

It should be noted that the bound (7) is not necessarily the most accurate. In the special case where  $f$  is compactly supported, we only need to control  $\widehat{e}_j(k)$  for all  $k \in \{-A2^j, \dots, A2^j\}$  for

some  $A > 0$  depending on the support (and the wavelets), as the coefficients  $\beta_{j,k}$  are zero for the other  $k$ . We can then take

$$\widehat{e}_j(k) = c_3 2^{j/2} \left[ \sqrt{\frac{j}{n} \mathbb{E}[\psi^2(2^j X - k)]} + \frac{j}{n} \right],$$

see our Lemma 3 (for instance). Bounding  $\mathbb{E}[\psi^2(2^j X - k)]$  from above by  $c_2 \|f\|_\infty 2^{-j}$  then leads to the famous threshold  $\sqrt{\|f\|_\infty j/n}$ , at least when  $j$  is not too large (say  $j2^j \leq n$ ). The procedures we are aware of that are based on this threshold still involve undesirable log factors in the rates though.

**2.2.3. Bounds from empirical process theory.** One way to avoid log factors is to better control the error term  $\mathcal{E}_j(K_j)$ . There are several ways to get a bound that makes the event  $\mathcal{E}$  true with high probability, say  $1 - 1/n^c$  for some  $c > 1$ . For the sake of simplicity, we suppose in this section that  $j \in \{0, \dots, n\}$ .

The simplest bound is as follows:

$$(8) \quad \widehat{\mathcal{E}}_j(K_j) = c_4 \left[ \sqrt{\frac{|K_j|}{n}} + \sqrt{\frac{\log n}{n}} \right],$$

see our Lemma 2. It turns out that correct use of this result already leads to optimal rates in the compact case. We do not go into detail here, as it unfortunately seems too rough to deal with infinite supports.

The first term may indeed be too large. This is especially true when the numbers  $\sigma_{j,k}^2 = \mathbb{E}[\psi^2(2^j X - k)]$  are small. We recall that  $\sigma_{j,k}^2$  somehow rules the estimation error between  $\beta_{j,k}$  and  $\widehat{\beta}_{j,k}$ . A solution to bypass this problem is to group the integers  $k$  for which the  $\sigma_{j,k}^2$  are approximately the same. Let for  $r \in \mathbb{Z}$ ,  $\overline{\mathbb{Z}}_{j,r}$  be the set of  $k$  such that  $\sigma_{j,k}^2$  is of the order of  $2^{-r}$ . We deduce from our Lemmas 2 and 3 that we may take a bound on  $\mathcal{E}_j(K_j)$  when  $K_j \subset \overline{\mathbb{Z}}_{j,r}$  whose leading term is of the order of

$$\frac{|K_j| 2^{-r/2}}{\sqrt{n}} \min \left\{ \sqrt{\log_+ \left( \frac{|\overline{\mathbb{Z}}_{j,r}|}{|K_j|} \right)}, \sqrt{\log_+ \left( \frac{2^r}{|K_j|} \right)} \right\},$$

where  $\log_+(\cdot) = \log(e + \cdot)$ . There are a few simple conditions for the above to be truly leading. Details are omitted.

So, the main idea of our procedure is to try to work with similar errors  $\sigma_{j,k}^2$ , and to juggle the two formulas. We use the first when  $r$  is small, and the second when  $r$  is larger. Naturally, the second formula introduces the troublesome condition  $K_j \subset \overline{\mathbb{Z}}_{j,r}$  as well as  $|\overline{\mathbb{Z}}_{j,r}|$  (the  $\sigma_{j,k}^2$  are unknown). The problematic set  $\overline{\mathbb{Z}}_{j,r}$  can, however, be estimated in some way, or replaced by a larger set. We refer to Section 2.3 for further details.

**2.2.4. About an optimal thresholding procedure.** We can find in the literature a term-by-term thresholding rule that eliminates log factors. We refer here to [DJ96]. It is nevertheless worth noting that it relies on strong assumptions:  $f$  is bounded (with known sup norm) and  $f$  is compactly supported on  $[0, 1]$ . It also requires knowledge of the smoothness of the function to be estimated (the regularity index and the Besov (quasi-) norm). As explained above, our approach



can be simplified in the compact case  $[0, 1]$ . Formula (8) is then sufficient to obtain the right rates of convergence (adaptively and without extra assumptions).

It is also possible to draw on our setup to recover their procedure. All it takes is the right formula for  $\widehat{\mathcal{E}}_j(K_j)$ . For example, we can use the one given by our Lemma 3 (with a set  $\mathcal{C}_{j,r}$  whose cardinality is of the order of  $2^j$  as  $f$  is compactly supported for [DJ96]). By using the rough upper-bound  $\mathbb{E}[\psi^2(2^j X - k)] \leq c_2 \|f\|_\infty 2^{-j}$ , we get:

$$\widehat{\mathcal{E}}_j(K_j) = c_5 \left[ |K_j| \sqrt{2^{-j} \|f\|_\infty \frac{\log_+(2^j/|K_j|)}{n}} + \frac{|K_j| \log_+(2^j/|K_j|)}{n} \right. \\ \left. + \sqrt{2^{-j} \|f\|_\infty |K_j| \frac{\log n}{n}} + \frac{\log n}{n} \right].$$

This equality can be taken on an event of probability  $1 - 1/n^c$ , uniformly for  $j \in \{0, \dots, n\}$  and for all non-empty sets  $K_j \subset \{-A2^j, \dots, A2^j\}$  (the term  $A$  depends on the wavelets).

We recall that the threshold of [DJ96] is applied only when  $j \in \{j_0 + 1, \dots, j_1\}$  where  $j_0$  is of the order of  $\log n$  and where  $2^{j_1}$  is of the order of  $n/\log n$ . By doing elementary maths, we obtain for all  $j \in \{j_0 + 1, \dots, j_1\}$ ,

$$\widehat{\mathcal{E}}_j(K_j) \leq c_6 \left[ (|K_j| + 2^{j_0}) \sqrt{2^{-j} \|f\|_\infty \frac{\log_+(2^j/(|K_j| + 2^{j_0}))}{n}} \right. \\ \left. + \frac{(|K_j| + 2^{j_0}) \log_+(2^j/(|K_j| + 2^{j_0}))}{n} + \sqrt{\|f\|_\infty \frac{\log n}{n}} \right] \\ \leq c_7 \left[ |K_j| 2^{-j/2} \sqrt{\|f\|_\infty \frac{j - j_0}{n}} + 2^{j_0 - j/2} \sqrt{\|f\|_\infty \frac{j - j_0}{n}} + \sqrt{\|f\|_\infty \frac{\log n}{n}} \right].$$

This leads straight to their threshold that is of the order of  $\sqrt{\|f\|_\infty (j - j_0)/n}$  (the last two terms in the right-hand side of this inequality do not depend on  $K_j$  and do not come into play when selecting coefficients by minimizing (5). They do, however, contribute to the estimator risk. This contribution is of the right order of magnitude when  $j_0$  is chosen carefully in line with the properties of  $f$ , see [DJ96] for further details).

**2.2.5. Generalization.** The heuristic of Section 2.2.1 only involves the wavelet coefficients  $\beta_{j,k}$  and their empirical versions  $\widehat{\beta}_{j,k}$ . There is therefore nothing to prevent it from being used in a more general context than that of this paper. For example, we can remain within the density model framework, but assume that the data are mixing instead of independent. The hard part is finding a sharp upper-bound on the estimation error  $\mathcal{E}_j(K_j)$ , which is already lengthy and technical when the data are independent. More generally, we may consider the estimation of the intensity of a random measure. In other words, we observe, for any measurable set  $A \subset \mathbb{R}$ , a random variable  $M(A)$  satisfying

$$\mathbb{E}[M(A)] = \int_A f(x) dx.$$

In this inequality,  $f$  is the target function. It is relatively well known that this setup encompasses different classical statistical frameworks, see [BB09, Aka17, Sar21a] for further details. Section 2.2.1

can then be used with the following definition of empirical wavelet coefficient:

$$\widehat{\beta}_{j,k} = \int \psi_{j,k} dM.$$

The formulas of  $\beta_{j,k}$ ,  $\mathcal{E}_j(K_j)$ , and the general idea of the procedure remain unchanged. What needs to be done is to find a bound  $\widehat{\mathcal{E}}_j(K_j)$  of this error  $\mathcal{E}_j(K_j)$ . It will naturally depends on the statistical framework considered.

**2.3. Estimation procedure.** The aim of this section is to define our estimator  $\widehat{f}$  of the density  $f$ . We refer to Sections 3.2 and 3.3 for the study of its statistical properties and to Section 3.5 for informations on more computational aspects.

We sort the sample in increasing order:  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ . We define the smallest integer  $\widehat{J} \geq 0$  satisfying

$$\min_{1 \leq i \leq n-1} (X_{(i+1)} - X_{(i)}) > 2^{1-\widehat{J}} L_\psi,$$

where  $L_\psi \geq 1$  is such that  $\text{supp } \psi \subset [-L_\psi, L_\psi]$ . For each resolution  $j \in \{-1, \dots, \widehat{J}\}$ , we determine a set  $\widehat{K}_j$  of integers by using an algorithm similar to the one described in Section 2.2.1. This yields our estimator

$$(9) \quad \widehat{f} = \sum_{j=-1}^{\widehat{J}} \sum_{k \in \widehat{K}_j} \widehat{\beta}_{j,k} \bar{\psi}_{j,k}$$

of  $f$ .

The integer  $\widehat{J}$  therefore specifies the highest resolution with which we work. Note that the  $\widehat{\beta}_{j,k}$  are based on very few observations when  $j \geq \widehat{J} + 1$  (at most one). They therefore make little contribution to the estimation of  $f$ .

We define for  $j \geq -1$  and  $k \in \mathbb{Z}$ ,

$$\widehat{\sigma}_{j,k}^2 = \begin{cases} \frac{1}{n} \sum_{i=1}^n \phi^2(X_i - k) & \text{if } j = -1, \\ \frac{1}{n} \sum_{i=1}^n \psi^2(2^j X_i - k) & \text{if } j \geq 0. \end{cases}$$

The role of  $\widehat{\sigma}_{j,k}^2$  is to evaluate empirically the quality of the estimate  $\widehat{\beta}_{j,k}$ .

Let then

$$\widehat{\mathbb{Z}}_j = \{k \in \mathbb{Z}, \widehat{\sigma}_{j,k}^2 \neq 0\}$$

be the set of integers  $k$  for which  $\psi_{j,k}(X_i) \neq 0$  for at least one observation. We define for  $\ell_j \in \mathbb{N}$  and  $r \geq \ell_j + j + 1$ ,

$$\begin{aligned} \widehat{\mathbb{Z}}_{j,r}(\ell_j) &= \{k \in \mathbb{Z}, 2^{-r-1} < \widehat{\sigma}_{j,k}^2 \leq 2^{-r}\} \\ \widehat{\mathbb{Z}}_{j,\ell_j+j}(\ell_j) &= \left\{k \in \mathbb{Z}, \widehat{\sigma}_{j,k}^2 > 2^{-\ell_j-j-1}\right\}. \end{aligned}$$

To each value of  $\ell_j$  corresponds therefore the partition of  $\widehat{\mathbb{Z}}_j$  defined by

$$(10) \quad \widehat{\mathbb{Z}}_j = \bigcup_{r \geq \ell_j + j} \widehat{\mathbb{Z}}_{j,r}(\ell_j).$$

We now explain which coefficients to retain among those indexed by  $\widehat{\mathbb{Z}}_{j,r}(\ell_j)$ . As described in Section 2.2.3, the aim of this reasoning is to group together coefficients whose estimation errors seem similar. The number  $\ell_j$  indicates the point at which formula (8) becomes too rough. By gathering all these coefficients and by choosing  $\ell_j$  appropriately, we obtain the set  $\widehat{K}_j$ .

We now introduce the terms and parameters that appear in our control of the error term. We consider some positive numbers  $\rho_{-1}, \varsigma_{-1}, \rho_0, \varsigma_0 > 0$  that will be specified later on, and set  $\varsigma_j = \varsigma_0$ ,  $\rho_j = \rho_0$  for  $j \geq 1$ . We define the map  $\widehat{\mathcal{E}}_{j,r,\ell_j}(\cdot)$  for  $r = \ell_j + j$  and  $x > 0$  by

$$\widehat{\mathcal{E}}_{j,r,\ell_j}(x) = \sqrt{\frac{x}{n} + x \frac{\log_+(2^{r+1}/x)}{n}} + \sqrt{\frac{\log((j+2)(r+1)n)}{n}} + \frac{\log((j+2)(r+1)n)}{n},$$

where  $\log_+(x) = \log(e+x)$ . When  $r \geq \ell_j + j + 1$ ,  $\widehat{\mathcal{E}}_{j,r,\ell_j}(\cdot)$  is rather defined for  $x > 0$  by

$$\begin{aligned} \widehat{\mathcal{E}}_{j,r,\ell_j}(x) &= x \sqrt{\frac{2^{-r} \log_+(\widehat{\lambda}_{j,r}/x)}{n} + x \frac{\log_+(\widehat{\lambda}_{j,r}/x)}{n}} \\ &\quad + \sqrt{x \frac{2^{-r} \log((j+2)(r+1)n)}{n} + \frac{\log((j+2)(r+1)n)}{n}}, \end{aligned}$$

where

$$\widehat{\lambda}_{j,r} = \min \left\{ (|\widehat{\mathbb{Z}}_j| + 2(2L_\psi + 1)(\log(j+2) + 2 \log n + 1)), 2^{r+1} \right\},$$

if  $2^r > \varsigma_j n / \log((j+2)n)$ , and where

$$\widehat{\lambda}_{j,r} = \min \left\{ |\widehat{\mathbb{Z}}'_{j,r}|, 2^{r+1} \right\}$$

if  $2^r \leq \varsigma_j n / \log((j+2)n)$ . In this equality,

$$\widehat{\mathbb{Z}}'_{j,r} = \{k \in \mathbb{Z}, 2^{-r-2} < \widehat{\sigma}_{j,k}^2 \leq 2^{-r+1}\}.$$

The map  $\widehat{\mathcal{E}}_{j,r,\ell_j}(\cdot)$  is extended by continuity at  $x = 0$ .

We define the criterion  $\gamma_{j,r,\ell_j}(\cdot)$  for  $K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)$  by

$$\gamma_{j,r,\ell_j}(K_{j,r,\ell_j}) = -2^{-j/2} \sum_{k \in K_{j,r,\ell_j}} |\widehat{\beta}_{j,k}| + \rho_j \widehat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|).$$

Thereby, the last term of this equality can be seen as an upper-bound of the error  $\mathcal{E}_j(K_{j,r,\ell_j})$  introduced by (3) in Section 2.2.1. According to the previous heuristics, we pick out  $\widehat{K}_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)$  such that

$$(11) \quad \gamma_{j,r,\ell_j}(\widehat{K}_{j,r,\ell_j}) = \min_{K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)} \gamma_{j,r,\ell_j}(K_{j,r,\ell_j}).$$

Since  $\widehat{\mathbb{Z}}_{j,r}(\ell_j)$  is finite, the set  $\widehat{K}_{j,r,\ell_j}$  does exist. If several sets minimize this criterion,  $\widehat{K}_{j,r,\ell_j}$  denotes any of them.

Moreover, the value of  $\ell_j$  that we will keep in the following is the smallest value  $\hat{\ell}_j$  satisfying

$$(12) \quad \sum_{r=\hat{\ell}_j+j}^{\infty} \gamma_{j,r,\hat{\ell}_j}(\widehat{K}_{j,r,\hat{\ell}_j}) = \min_{\ell_j \in \mathbb{N}} \sum_{r=\ell_j+j}^{\infty} \gamma_{j,r,\ell_j}(\widehat{K}_{j,r,\ell_j}).$$

Such a value does exist (for  $\ell_j$  large enough, the partition given by (10) remains the same).

We finally set

$$\widehat{K}_j = \bigcup_{r=\hat{\ell}_j+j}^{\infty} \widehat{K}_{j,r,\hat{\ell}_j}$$

and define our estimator  $\hat{f}$  by (9).

### 3. THEORETICAL RESULTS.

We study in this section the properties of our estimator  $\hat{f}$ .

**3.1. Classes of functions.** We begin by introducing classes of functions corresponding to assumptions on the density to be estimated.

For this purpose, the following notations will be convenient. We denote for  $p \in (0, +\infty]$  and  $x = (x_k)_{k \in \mathbb{Z}}$  the (quasi)  $\ell^p$  norm of  $x$  by

$$\|x\|_p = \begin{cases} \left( \sum_{k \in \mathbb{Z}} |x_k|^p \right)^{1/p} & \text{if } p < \infty \\ \sup_{k \in \mathbb{Z}} |x_k| & \text{if } p = \infty. \end{cases}$$

The weak (quasi)  $\ell^p$  norm of  $x$  is defined by

$$\|x\|_{p,\infty} = \begin{cases} \sup_{t>0} t \left( \sum_{k \in \mathbb{Z}} 1_{|x_k| \geq t} \right)^{1/p} & \text{if } p < \infty \\ \sup_{k \in \mathbb{Z}} |x_k| & \text{if } p = \infty. \end{cases}$$

We recall that  $\|x\|_{p,\infty} \leq \|x\|_p$  but that the converse is not true in general.

**3.1.1. Besov classes.** The classical Besov spaces  $\mathcal{B}_{p,\infty}^\alpha$  possess a characterisation in terms of wavelets coefficients. It follows from [DJ97] that  $\mathcal{B}_{p,\infty}^\alpha$  may be defined when  $p \in [0, +\infty]$  and  $\alpha \in ((1/p - 1)_+, \tau)$  as the set of functions  $f$  of  $\mathbb{L}^{\max\{p,1\}}(\mathbb{R})$  satisfying  $\|f\|_{\mathcal{B}_{p,\infty}^\alpha} < \infty$  where

$$\|f\|_{\mathcal{B}_{p,\infty}^\alpha} = \sup_{j \geq -1} \left\{ 2^{j(\alpha+1/2-1/p)} \|\beta_{j,\cdot}\|_p \right\}.$$

The value of  $\tau$  is an integer depending on the wavelet basis. It is equal to 1 for the Haar basis. It stands for the smoothness of  $\bar{\phi}$  and  $\bar{\psi}$  when the basis is the bi-orthogonal basis of [CDF92] (see Section 2.1).

We define the Besov ball  $\mathfrak{B}_{p,\infty}^\alpha(R)$  as the set of functions  $f \in \mathcal{B}_{p,\infty}^\alpha$  satisfying  $\|f\|_{\mathcal{B}_{p,\infty}^\alpha} \leq R$ . This set is slightly smaller than the set composed of functions  $f \in \mathcal{B}_{p,\infty}^\alpha$  satisfying

$$\sup_{j \geq 0} \left\{ 2^{j(\alpha+1/2-1/p)} \|\beta_{j,\cdot}\|_p \right\} \leq R.$$

This latter set is denoted by  $\mathcal{B}_{p,\infty}^\alpha(R)$  and is called a strong Besov class. The adjective “strong” is added to avoid any ambiguity with what is below. The difference between a (strong) Besov class and a Besov ball lies therefore in the starting point of the index  $j$ .

We define the weak Besov class  $\mathcal{WB}_{p,\infty}^\alpha(R)$  when  $p$  is finite as the set of functions  $f \in \mathbb{L}^1(\mathbb{R})$  satisfying

$$\sup_{j \geq 0} \left\{ 2^{j(\alpha+1/2-1/p)} \|\beta_{j,\cdot}\|_{p,\infty} \right\} \leq R.$$

When  $p$  is infinite, we set  $\mathcal{WB}_{\infty,\infty}^\alpha(R) = \mathcal{B}_{\infty,\infty}^\alpha(R)$ . The main difference with the strong Besov classes is therefore the use of the (quasi) weak  $\ell^p$  norm in place of the standard  $\ell^p$  (quasi) norm.

**3.1.2. Tail dominance condition.** We define two sets  $\mathcal{T}_\theta(M)$  and  $\mathcal{WT}_\theta(M)$  corresponding to strong and weak conditions on the tails of  $f$ .

We set for  $j \geq 0$  and  $k \in \mathbb{Z}$ ,

$$(13) \quad F_{j,k} = \int_{2^{-j(k-1/2)}}^{2^{-j(k+1/2)}} f(x) dx.$$

We define for  $M > 0$  and  $\theta \in (0, 1)$ , the set  $\mathcal{T}_\theta(M)$  gathering the integrable non-negative functions  $f$  satisfying

$$(14) \quad \sup_{j \geq 0} \left\{ 2^{-j(1-\theta)} \|F_{j,\cdot}\|_\theta^\theta \right\} \leq M.$$

We say that the “strong tail dominance condition” is met when  $f \in \mathcal{T}_\theta(M)$ . This terminology “tail dominance condition” is directly borrowed from [GL14] although their condition differs a little from this one (see the next section for more details).

Inequality (14) can be softened by replacing the (quasi) norm with its weak version. This leads to the set  $\mathcal{WT}_\theta(M)$  and the “weak tail dominance condition”  $f \in \mathcal{WT}_\theta(M)$ . By definition, it contains therefore all the integrable non-negative functions  $f$  such that

$$\sup_{j \geq 0} \left\{ 2^{-j(1-\theta)} \|F_{j,\cdot}\|_{\theta,\infty}^\theta \right\} \leq M.$$

These two conditions are satisfied when there are not too many mass in the tails of the distribution. The parameters  $M$  and  $\theta$  tune this amount of mass allowed. The larger  $\theta$  is, the heavier the tails can be.

The above conditions are defined only when  $\theta \neq 0$ . The limit case  $\theta = 0$  corresponds to compactly supported functions. More precisely, we define  $\mathcal{T}_0(M) = \mathcal{WT}_0(M)$  as the collection of integrable functions  $f$  satisfying

$$\sup_{j \geq 0} \left\{ 2^{-j} |\{k \in \mathbb{Z}, F_{j,k} > 0\}| \right\} \leq M.$$

We show in Appendix G the elementary proposition below.

**Proposition 1.** *The following assertions hold true:*

1. If  $M < 1$ , the set  $\mathcal{T}_\theta(M)$  does not contain any density.
2. Let  $p \in (0, 1)$ ,  $R > 0$ ,  $\alpha \in (1/p - 1, \tau)$ . If  $f \in \mathfrak{B}_{p,\infty}^\alpha(R)$ , then  $f \in \mathcal{T}_p(c_1 R^p)$  where  $c_1$  only depends on the wavelet basis and  $\alpha, p$ . Conversely, if  $f \in \mathfrak{B}_{p,\infty}^\alpha(R) \cap \mathcal{T}_p(R^p)$ , then  $f \in \mathfrak{B}_{p,\infty}^\alpha(c_2 R)$  where  $c_2$  only depends on the wavelet basis and  $\alpha, p$ .
3. If  $f$  is a compactly supported density on  $[-L, L]$ , then  $f$  belongs to  $\mathcal{T}_0(2L + 2)$ .
4. If  $f$  is a density satisfying  $f(x) \leq A^b |x|^{-b}$  for all  $|x| \geq 1$  and some  $A > 0$ ,  $b > 1$ , then  $f$  belongs to  $\mathcal{WT}_{1/b}(M)$  with  $M = c_3(1 + A)$  and some  $c_3$  only depending on  $b$ .

**3.2. Minimax rates.** We evaluate in this section the risk of our estimator  $\hat{f}$  when  $f$  is smooth with dominated tails. More precisely, we consider  $R > 0$ ,  $M \geq 1$ ,  $p \in (0, +\infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ , and

$$\mathcal{F}_{\alpha,p,\theta}(R, M) = \begin{cases} \mathcal{WB}_{p,\infty}^\alpha(R) \cap \mathcal{WT}_\theta(M) & \text{if } p \neq 1 \\ \mathcal{B}_{p,\infty}^\alpha(R) \cap \mathcal{WT}_\theta(M) & \text{if } p = 1. \end{cases}$$

We then give an upper-bound of the maximal risk of our estimator when  $f$  lies in  $\mathcal{F}_{\alpha,p,\theta}(R, M)$ . It involves the following quantities:

$$\begin{aligned} \gamma &= \begin{cases} \alpha/(2\alpha + 1) & \text{if } \theta \leq \alpha/(2\alpha + 1 - 1/p) \\ \alpha(1 - \theta)/(\alpha + 1 - \theta/p) & \text{if } \theta > \alpha/(2\alpha + 1 - 1/p) \end{cases} \\ \nu_n &= \begin{cases} \log n & \text{if } \theta = \alpha/(2\alpha + 1 - 1/p) \text{ and } p \neq 1 \\ (\log n)^{2\gamma} & \text{if } \theta = \alpha/(2\alpha + 1 - 1/p) \text{ and } p = 1 \\ 1 & \text{otherwise} \end{cases} \\ \beta_1 &= \begin{cases} 1/(2\alpha + 1) & \text{if } \theta \leq \alpha/(2\alpha + 1 - 1/p) \\ (1 - \theta)/(1 + \alpha - \theta/p) & \text{if } \theta > \alpha/(2\alpha + 1 - 1/p) \end{cases} \\ \beta_2 &= \begin{cases} (\alpha + 1 - 1/p)/((1 - \theta)(2\alpha + 1)) & \text{if } \theta \leq \alpha/(2\alpha + 1 - 1/p) \\ (\alpha + 1 - 1/p)/(\alpha + 1 - \theta/p) & \text{if } \theta > \alpha/(2\alpha + 1 - 1/p). \end{cases} \end{aligned}$$

Our main result is proved in Section 4 and is as follows:

**Theorem 2.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$ . Then, there exist  $\bar{\varrho}_{-1}, \varsigma_{-1}, \bar{\varrho}_0, \varsigma_0$  and  $n_0$  such that for all  $n \geq n_0$ ,  $\rho_{-1} \geq \bar{\varrho}_{-1}$ ,  $\rho_0 \geq \bar{\varrho}_0$ , our estimator  $\hat{f}$  defined in Section 2.3 satisfies*

$$(15) \quad \sup_{f \in \mathcal{F}_{\alpha,p,\theta}(R, M)} \mathbb{E} \left[ d_1(f, \hat{f}) \right] \leq c \left[ R^{\beta_1} M^{\beta_2} \nu_n + M 1_{\gamma=1-\theta} \right] n^{-\gamma}.$$

Moreover,  $\bar{\varrho}_{-1}, \varsigma_{-1}$  are universal and  $\bar{\varrho}_0, \varsigma_0$  only depends on  $\psi$ . The term  $c$  only depends on  $\alpha, p, \bar{\varphi}, \psi, \bar{\psi}, \theta, \rho_{-1}, \rho_0$ , and  $n_0$  only depends on  $p, \alpha, \theta, R, M$ .

We would like to highlight that the construction of our estimator  $\hat{f}$  does not involve the parameters  $\alpha, p, \theta, R, M$  of the class  $\mathcal{F}_{\alpha,p,\theta}(R, M)$ . They can therefore be unknown.

For pedagogical reasons, let us consider the case where  $f$  is compactly supported in  $[-L, L]$  with  $L \geq 1$ . Proposition 1 entails that  $f \in \mathcal{F}_{\alpha,p,0}(R, 4L)$ . In particular, there is  $n_0$  only depending on  $R, M, L, \alpha, p$  such that for all  $n \geq n_0$ ,

$$\mathbb{E} \left[ d_1(f, \hat{f}) \right] \leq c' R^{1/(2\alpha+1)} L^{(\alpha+1-1/p)/(2\alpha+1)} n^{-\alpha/(2\alpha+1)}.$$

This is the classical estimation rate. It is however attained here under very mild conditions on  $p$  and  $\alpha$ . This result essentially says that our estimator adapts to local variations of the density (the algorithm can increase the number of coefficients  $\beta_{j,k}$  to estimate at the locations where the function varies a lot and on the contrary decrease it at the locations where the function is more flat). Note that there is no requirement that the density be bounded or even in  $\mathbb{L}^2(\mathbb{R})$ . We do not know of any wavelet estimator that achieves the standard rate  $n^{-\alpha/(2\alpha+1)}$  under our assumptions.

In the non compact case, the usual rate of convergence  $n^{-\alpha/(2\alpha+1)}$  applies when  $\theta$  is sufficiently small, that is when the (weak) tail dominance condition is stringent enough. When  $\theta$  is larger, however, the estimation rate deteriorates. It becomes particularly slow when  $\theta$  comes close to 1. We recall that it is not possible to estimate a density on  $\mathbb{R}$  under the sole assumption that  $f$  belongs to a ball  $\mathfrak{B}_{p,\infty}^\alpha(R)$  of a Besov space with  $p \geq 1$  and  $R$  large enough when the loss is the  $\mathbb{L}^1$  norm (see [GL14, IK81, JLL04]).

However, the assumption “ $f$  belongs to a ball of a Besov space” is sufficient to ensure the convergence of our estimator when  $p < 1$  and  $\alpha > 1/p - 1$ . Indeed, our tail dominance condition is fulfilled in this case with  $\theta = p$ . The above theorem can therefore be applied. This gives: for all  $p \in (0, 1)$ ,  $\alpha \in (1/p - 1, \tau)$ ,  $R \geq 1$ , and  $n$  large enough,

$$(16) \quad \sup_{f \in \mathfrak{B}_{p,\infty}^\alpha(R)} \mathbb{E} \left[ d_1(f, \hat{f}) \right] \leq c'' R^{\beta_3} \nu_n n^{-\gamma},$$

where

$$\gamma = \begin{cases} \alpha/(2\alpha + 1) & \text{if } p \in (0, 1/2] \text{ or } \alpha < (1 - p)/(2p - 1) \\ 1 - p & \text{if } p \in (1/2, 1) \text{ and } \alpha \geq (1 - p)/(2p - 1) \end{cases}$$

$$\nu_n = \begin{cases} \log n & \text{if } p \in (1/2, 1) \text{ and } \alpha = (1 - p)/(2p - 1) \\ 1 & \text{otherwise} \end{cases}$$

$$\beta_3 = \begin{cases} \alpha/((2\alpha + 1)(1/p - 1)) & \text{if } p \in (0, 1/2] \text{ or } \alpha < (1 - p)/(2p - 1) \\ p & \text{if } p \in (1/2, 1) \text{ and } \alpha \geq (1 - p)/(2p - 1) \end{cases}$$

and where  $c''$  only depends on  $\alpha, p, \psi, \bar{\psi}, \bar{\phi}, \rho_{-1}, \rho_0$ .

Naturally, the rate (16) can be improved if the tail condition is more stringent, i.e.  $\theta < p$ . It is worth noticing that our Theorem 2 also makes explicitly the dependency in  $M$  and  $R$ . If  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$  with  $\theta = p$  but  $M$  much smaller than  $R^p$ , the factor in front of  $n$  is actually far better than the one shown just above. This explains why we draw the distinction between a Besov class and a Besov ball. The quantity  $\|\beta_{j,\cdot}\|_p$  may play a very different role in the minimax results when  $j = -1$  and when  $j \geq 0$ . The parameter  $R$  not only constrains the spatial variations of  $f$  when  $p < 1$ . It also controls the massiveness of its tails.

As far as we know, only the case  $p \geq 1$  has been studied in the literature of density estimation with infinite support under  $\mathbb{L}^1$  loss. The only papers we are aware of that deal with the subject are the two mentioned below.

First, the authors of [GL14] proposed to estimate the density pointwise. The global risk is then obtained by integrating the pointwise risk. This reasoning has the merit of not depending on a particular loss and of leading to results for all  $\mathbb{L}^q$  losses. The downside is that it may lead to undesirable logarithmic factors in the convergence rates. We do not have any here, except at the boundary. Their tail dominance condition is more or less the same as our strong condition.

Rigorously, it is at least as stringent as our strong condition (Proposition 1 of [CL20]). We do not know whether it is equivalent. Note that they also impose a condition on the supremum norm of  $f$ . Besides, they restricted themselves to balls of Besov spaces and were not interested in weak Besov classes.

Second, wavelets were used in [CL20] to estimate the density under the same tail condition as our strong condition. Unfortunately, the convergence rate of their estimator is slower than ours and is hence not optimal (their exponent in  $n$  is, in absolute value, smaller than ours as soon as  $\theta \neq 0$ ).

Although our result is stated for the  $\mathbb{L}^1$  loss, it can easily be checked that (15) remains true for the distance induced by the Besov norm  $\|\cdot\|_{\mathcal{B}_{1,1}^0}$  defined by

$$\|f\|_{\mathcal{B}_{1,1}^0} = \sum_{j \geq -1} 2^{-j/2} \|\beta_{j,\cdot}\|_1.$$

The proposition below shows the rate (15) is optimal, even when logarithmic factors appear. It also shows that the dependency in  $R$  and  $M$  is the right one. It completes and refines the minimax lower bound of [GL14] and is proved in Appendix E.

**Proposition 3.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in (0, 1) \cap [0, p]$ , and  $\mu_n \in \{1, \nu_n\}$ . There exist some  $R_0, M_0$  such that: for all  $R \geq R_0$ ,  $M \geq M_0$  and all  $n$  large enough,*

$$\inf_{\tilde{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \geq c \left[ R^{\beta_1} M^{\beta_2} \mu_n + M 1_{\gamma=1-\theta} \right] n^{-\gamma},$$

where  $\mathcal{F}$  is a subset of  $\mathcal{B}_{p,\infty}^\alpha(R) \cap \mathcal{WT}_\theta(M)$ , where  $c$  only depends on  $\alpha, p, \theta, \psi, \bar{\psi}, \bar{\phi}$ , and where  $R_0, M_0$  only depend on  $\alpha, p, \theta, \psi, \bar{\psi}, \bar{\phi}$ . Moreover, any density  $f$  of  $\mathcal{F}$  satisfies

$$\sup_{|x| \geq 1} |x|^{1/\theta} f(x) \leq M^{1/\theta}.$$

The set  $\mathcal{F}$  is also included in  $\mathcal{T}_\theta(M)$  when  $\mu_n = 1$ .

This result is a little more precise than a lower minimax bound on  $\mathcal{F}_{\alpha,p,\theta}(R, M)$ . We may for instance observe that the minimax rates on  $\mathcal{WB}_{p,\infty}^\alpha(R) \cap \mathcal{WT}_\theta(M)$  and  $\mathcal{B}_{p,\infty}^\alpha(R) \cap \mathcal{WT}_\theta(M)$  coincide when  $p \neq 1$ . Replacing a strong Besov class with a weak one has no impact on the optimal rates in this case. Similarly, a weak or strong tail condition leads to the same results when  $\theta \neq \alpha/(2\alpha + 1 - 1/p)$ . This is not necessarily true otherwise.

To illustrate this point, let us take  $p = 1$ ,  $\theta = \alpha/(2\alpha + 1 - 1/p) = 1/2$  and  $J$  of the order of  $n^{1/(2\alpha+1)}$ . It is fairly easy to see that the linear estimator

$$\tilde{f} = \sum_{j=-1}^J \sum_{k \in \mathbb{Z}} \hat{\beta}_{j,k} \bar{\psi}_{j,k}$$

achieves the rate  $n^{-\alpha/(2\alpha+1)}$  when  $f \in \mathcal{B}_{1,\infty}^\alpha(R) \cap \mathcal{T}_{1/2}(M)$ . This is faster than the minimax rate on  $\mathcal{B}_{1,\infty}^\alpha(R) \cap \mathcal{WT}_{1/2}(M)$ .

Proposition 3 also ensures the optimality of (16) when  $p < 1$  and  $\alpha \neq (1-p)/(2p-1)$ . We do not know whether the rate is optimal when  $\alpha = (1-p)/(2p-1)$  though (due to the log factor).



**3.3. Minimax rates over a class of fat tailed distributions.** We illustrate here the interest of the weak tail dominance condition compared to its strong version. We consider  $A \geq 1$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $p \in [0, \infty]$ ,  $b > 1$ ,  $b \geq 1/p$ , and the class

$$\mathcal{D}(\alpha, p, R, A, b) = \left\{ f \in \mathcal{B}_{p, \infty}^\alpha(R), \text{ such that } f(x) \leq A^b |x|^{-b} \text{ for all } |x| \geq 1 \right\}.$$

Proposition 1 says that the weak tail dominance condition is met with  $\theta = 1/b$  and  $M = c(1 + A)$ .

We deduce from Theorem 2 that for  $n$  large enough,

$$\sup_{f \in \mathcal{D}(\alpha, p, R, A, b)} \mathbb{E} \left[ d_1(f, \hat{f}) \right] \leq c' \left[ R^{\beta_1} A^{\beta_2} \nu_n + A 1_{\gamma=1-1/b} \right] n^{-\gamma},$$

where

$$\begin{aligned} \gamma &= \begin{cases} \alpha/(2\alpha + 1) & \text{if } b \geq 2 + (1 - 1/p)/\alpha \\ \alpha(1 - 1/b)/(\alpha + 1 - 1/(pb)) & \text{if } b < 2 + (1 - 1/p)/\alpha \end{cases} \\ \nu_n &= \begin{cases} \log n & \text{if } b = 2 + (1 - 1/p)/\alpha \text{ and } p \neq 1 \\ (\log n)^{2\gamma} & \text{if } b = 2 + (1 - 1/p)/\alpha \text{ and } p = 1 \\ 1 & \text{if } b \neq 2 + (1 - 1/p)/\alpha \end{cases} \\ \beta_1 &= \begin{cases} 1/(2\alpha + 1) & \text{if } b \geq 2 + (1 - 1/p)/\alpha \\ (1 - 1/b)/(\alpha + 1 - 1/(pb)) & \text{if } b < 2 + (1 - 1/p)/\alpha \end{cases} \\ \beta_2 &= \begin{cases} (\alpha + 1 - 1/p)/((1 - 1/b)(2\alpha + 1)) & \text{if } b \geq 2 + (1 - 1/p)/\alpha \\ (\alpha + 1 - 1/p)/(\alpha + 1 - 1/(pb)) & \text{if } b < 2 + (1 - 1/p)/\alpha \end{cases} \end{aligned}$$

and where  $c'$  only depends on  $\alpha, p, b, \psi, \bar{\psi}, \bar{\phi}, \rho_{-1}, \rho_0$ .

The parameter  $\gamma$ , which governs the estimation rate of our estimator, depends on  $\alpha$ ,  $b$  and  $p$ . When  $b$  is sufficiently large, we recover the usual rate of convergence. The rate is otherwise slower but still minimax (at least when  $A, R$  are large enough).

Let us now observe that the strong tail dominance condition is not fulfilled for  $\theta = 1/b$  (whatever  $M$ ). We rather have  $\mathcal{D}(\alpha, p, R, A, b) \subset \mathcal{T}_\theta(M_\theta)$  for all  $\theta > 1/b$  and  $M_\theta$  depending on  $\theta$  (and  $b, A$ ). If the theorem were shown only for the strong condition, we could apply it only with values of  $\theta$  larger than  $1/b$ . If done correctly, it gives the right rate of convergence when  $b > 2 + (1 - 1/p)/\alpha$ . However, this causes problems when  $b \leq 2 + (1 - 1/p)/\alpha$  as the exponent then depends on  $\theta$ . Using a value of  $\theta$  larger than  $1/b$  leads to a slower convergence rate.

Note that a bounded and unimodal density  $f$  belongs to  $\mathcal{B}_{1, \infty}^1(R)$  for some  $R$  depending on  $\|f\|_\infty$  and the wavelet basis only. Such a density belongs therefore to  $\mathcal{D}(1, 1, R, A, b)$  if it satisfies  $f(x) \leq Ax^{-b}$  for all  $|x| \geq 1$ . We deduce that  $\hat{f}$  converges to  $f$  at the rate  $n^{-1/3}$  when  $b > 2$ . We thus recover the optimal estimation rate of a bounded unimodal density with compact support although  $f$  can be infinitely supported here.

**3.4. About the condition  $\alpha > 1/p - 1$  when  $p < 1$ .** In the previous sections, we always assumed that  $\alpha > (1/p - 1)_+$ . This condition is empty if  $p \geq 1$  but not otherwise. Note that it is used in the result of [DJ97] to characterize the Besov balls in terms of wavelet coefficients. To explore what happens in the opposite case, we need therefore to redefine these balls.

We consider some  $\alpha, p > 0$  and an arbitrary integer  $r$  larger than  $\alpha$ . We set for  $t, h > 0$ ,

$$\Delta_h^r f(t) = \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} f(t + kh).$$

We define the modulus of smoothness

$$\omega_p(f, x) = \sup_{0 < h \leq x} \left[ \int |\Delta_h^r f(t)|^p dt \right]^{1/p},$$

and define the Besov ball  $\mathfrak{B}_{p,\infty}^\alpha(R)$  as the collection of functions  $f \in \mathbb{L}^p$  such that

$$\|f\|_{\mathfrak{B}_{p,\infty}^\alpha} = \|f\|_p + \sup_{x>0} x^{-\alpha} \omega_p(f, x),$$

is not larger than  $R$ .

The above (quasi) norm depends a priori on the choice of  $r$ . Changing  $r$  leads however to an equivalent (quasi) norm, see [DL93]. It is also equivalent to the (quasi) norm that we defined in Section 3.1.1 when  $\alpha \in ((1/p - 1)_+, \tau)$ . To avoid adding unnecessary notations, we have used the same symbol to designate the Besov ball and the (quasi) norm. However, since the (quasi) norms are not equal, but equivalent, there is a slight ambiguity about what  $\|\cdot\|_{\mathfrak{B}_{p,\infty}^\alpha}$  and  $R$  are. This has no impact on our results though.

The proposition below shows that it is not possible to obtain a convergent estimator for the  $\mathbb{L}^1$  loss under the sole assumption that  $f$  is a compactly supported density of  $\mathfrak{B}_{p,\infty}^\alpha(R)$  when  $\alpha \leq 1/p - 1$ . It is proved in Appendix F.

**Proposition 4.** *Let  $p \in (0, 1)$ ,  $\alpha \in (0, 1/p - 1]$  and  $R > 0$ . Then,*

$$\inf_{\tilde{f}} \sup_{\substack{f \in \mathfrak{B}_{p,\infty}^\alpha(R) \\ \text{supp } f \subset [0,1]}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \geq 1/4.$$

This result can be compared with what exists in the literature of estimation of a compactly supported density under the  $\mathbb{L}^q$  loss with  $q > 1$ . The minimax risk does not tend to 0 either when  $\alpha = 1/p - 1/q$ , see [Sar21b]. However, the optimal estimation rate can be made arbitrarily slow by choosing  $\alpha$  very close to  $1/p - 1/q$  when  $q > 1$ . This phenomenon does not occur here, since the optimal estimation rate is  $n^{-\alpha/(2\alpha+1)}$ , whatever  $p > 0$  and  $\alpha > 1/p - 1$ .

**3.5. Computational complexity.** An estimator is not always derived from a computationally tractable procedure. For example, we have not been able to find in the literature a computationally tractable algorithm that would lead to an optimal estimator in the minimax  $\mathbb{L}^1$  sense when  $f$  is compactly supported on  $[0, 1]$  and in  $\mathfrak{B}_{p,\infty}^\alpha(R)$  with  $p < 1$ . There are admittedly computationally acceptable solutions in the literature to deal with this case, but they seem to be optimal only to within log factors, see [Sar21b]. We recall that the case  $p < 1$  is very different from the case  $p \geq 1$ . The latter is straightforward to solve in the compact case as a simple linear estimator works.

Let us also mention that there are computational difficulties with Kernel estimators. Indeed, the bandwidth should vary with the location so that the estimator adapts to the inhomogeneous smoothness of the density. In this context, procedures that lead to rate optimal estimators have been developed by [Lep15] in the Gaussian white noise model. His solution is based on an algorithm that seems quite difficult to implement though (see his Problem II Section 4.2). Things are a bit simpler when we allow the estimator to be rate optimal within log factors (see [GL14, LW19]).

We show below that the computational complexity of our procedure is nearly linear in the number  $n$  of observations.

We consider some  $j \in \{-1, \dots, \widehat{J}\}$  and assume that the wavelets functions  $\bar{\phi}$ ,  $\bar{\psi}$  and  $\psi$  have been preprocessed. We put  $\psi_j = \psi$  if  $j \geq 0$  and  $\psi_{-1} = \phi$ .

We begin by sorting the observations in increasing order  $X_{(1)} < \dots < X_{(n)}$ . We remark that  $\psi_j(\mathbb{R})$  is finite as  $\psi_j$  is piecewise constant. We consider some  $y \in \psi_j(\mathbb{R}) \setminus \{0\}$ . Then, we find for each  $i \in \{1, \dots, n\}$ , the few integers  $k$  such that  $\psi_j(2^j X_i - k) = y$ . We gather all these elements into a vector of size  $\mathcal{O}(n)$ . By counting in this vector the number of repetitions, we determine the family  $(\widehat{\beta}_{j,k}(y))_{k \in \mathbb{Z}}$  where

$$\widehat{\beta}_{j,k}(y) = \frac{y}{n} \sum_{i=1}^n 1_{\psi_j(2^j X_i - k) = y}.$$

In particular, we deduce

$$\widehat{\beta}_{j,k} = \sum_{y \in \psi_j(\mathbb{R}) \setminus \{0\}} \widehat{\beta}_{j,k}(y).$$

The reasoning is the same for obtaining  $(\widehat{\sigma}_{j,k}^2)_{k \in \mathbb{Z}}$ . So far, the number of calculations performed is at most  $\mathcal{O}(n \log n)$ .

We consider some  $\ell_j \in \mathbb{N}$ ,  $r \geq \ell_j + j$ , find the indices in  $\widehat{\mathbb{Z}}_{j,r}(\ell_j)$  and the size of  $\widehat{\mathbb{Z}}'_{j,r}$ . We sort the wavelet coefficients  $(\widehat{\beta}_{j,k})_{k \in \widehat{\mathbb{Z}}_{j,r}(\ell_j)}$  in descending order of importance:  $|\widehat{\beta}_{j,[1]}| \geq |\widehat{\beta}_{j,[2]}| \geq |\widehat{\beta}_{j,[3]}| \dots$ . Finding  $\widehat{K}_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)$  that minimizes (11) amounts to selecting the  $s$  most important coefficients where  $s$  minimizes

$$-2^{-j/2} \sum_{k=1}^s |\widehat{\beta}_{j,[k]}| + \rho_j \widehat{\mathcal{E}}_{j,r,\ell_j}(s).$$

This set can therefore be built in  $\mathcal{O}(n + |\widehat{\mathbb{Z}}_{j,r}(\ell_j)| \log |\widehat{\mathbb{Z}}_{j,r}(\ell_j)|)$  elementary operations.

Note that the number of  $r$  to consider is at most  $\mathcal{O}(\log n)$ . Moreover, the values of  $\ell_j$  of interest are those between 0 and  $\mathcal{O}(\log n)$  since the partition given by (10) remains the same when  $\ell_j$  is higher. Since  $\sum_{r \geq \ell_j + j} |\widehat{\mathbb{Z}}_{j,r}(\ell_j)| = \mathcal{O}(n)$ , we deduce that all the sets  $\widehat{K}_{j,r,\ell_j}$  (when  $r \geq \ell_j + j$ ,  $\ell_j \in \mathbb{N}$ ,  $j \in \{1, \dots, \widehat{J}\}$  vary) can be obtained in at most  $\mathcal{O}((\widehat{J} + 1)n \log^2 n)$  operations. The computation of  $\widehat{\ell}_j$  is fast as it requires less than  $\mathcal{O}(\log^2 n)$  additional operations. To sum up,  $\mathcal{O}((\widehat{J} + 1)n \log^2 n)$  is the maximal number of operations needed to find all the selected coefficients, that is to find the sets  $(\widehat{K}_j)_{-1 \leq j \leq \widehat{J}}$ .

The computational complexity of our procedure is random but can be bounded from above either with high probability or in expectation. We only present the second possibility. We deduce from Lemma 17 page 29 that  $\mathbb{E}[\widehat{J}] \leq \mathcal{O}(\log n)$  when  $f$  belongs to  $\mathbb{L}^q$  for some  $q > 1$ . This assumption is fulfilled when  $f \in \mathcal{F}_{\alpha,p,\theta}(R,M)$  (see Lemma 25 page 37 if needed). More precisely, we have for all  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$ ,

$$\sup_{f \in \mathcal{F}_{\alpha,p,\theta}(R,M)} \mathbb{E}[\widehat{J}] \leq \mathcal{O}(\log n).$$

In average, the computational complexity is therefore at most  $\mathcal{O}(n \log^3 n)$ .

## 4. PROOF OF THEOREM 2.

We begin by carrying out and proving a non-asymptotic risk bound.

**Theorem 5.** *Suppose that  $f \in \mathbb{L}^q$  for some  $q > 1$ . Then, there exist universal constants  $\bar{\varrho}_{-1, \varsigma_{-1}}$  and terms  $\bar{\varrho}_0, \varsigma_0$  depending on  $\psi$  only such that if  $\rho_{-1} \geq \bar{\varrho}_{-1}$  and  $\rho_0 \geq \bar{\varrho}_0$ , the estimator  $\hat{f}$  defined in Section 2.3 satisfies*

$$(17) \quad \mathbb{E} \left[ \|f - \hat{f}\|_1 \right] \\ \leq c_1 \mathbb{E} \left[ \sum_{j=-1}^{\hat{J}} \inf_{\ell_j \in \mathbb{N}} \sum_{r=\ell_j+j}^{\infty} \inf_{K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j} \left\{ \widehat{B}_{j,r,\ell_j}(K_{j,r,\ell_j}) + \rho_j \widehat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|) \right\} 1_{\mathcal{A}} \right] \\ + c_1 T_1 + c_1 T_2 + c_2 \frac{\log n}{n^2},$$

where

$$(18) \quad \overline{\mathbb{Z}}_j = \{k \in \mathbb{Z}, f_{j,k} \geq 1/n\},$$

$$(19) \quad f_{j,k} = \int f(x) 1_{\text{supp } \psi_{j,k}}(x) dx,$$

$$(20) \quad \widehat{B}_{j,r,\ell_j}(K_{j,r,\ell_j}) = 2^{-j/2} \sum_{k \in (\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j) \setminus K_{j,r,\ell_j}} |\beta_{j,k}|,$$

$$(21) \quad T_1 = \sum_{j=-1}^{\infty} 2^{-j/2} \sum_{k \notin \overline{\mathbb{Z}}_j} |\beta_{j,k}|,$$

$$(22) \quad T_2 = n \sum_{j=-1}^{\infty} 2^{-j/2} \sum_{k \in \overline{\mathbb{Z}}_j} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1},$$

where  $c_1$  only depends on  $\psi, \bar{\psi}, \bar{\phi}$ , and where  $c_2$  only depends on  $\psi, \bar{\psi}, \bar{\phi}, q, \|f\|_q$ . Moreover,  $\mathcal{A}$  is an event on which  $\widehat{J} \leq c_0 \log n$  for some  $c_0$  only depending on  $q, \|f\|_q$ , and on which

$$(23) \quad |\widehat{\mathbb{Z}}_j| \leq 2\mathbb{E}[|\widehat{\mathbb{Z}}_j|] + (7/3)(2L_\psi + 1)(\log(j+2) + 2\log n + 1)$$

for all  $j \geq -1$ . Furthermore, on  $\mathcal{A}$  we have for all  $j \geq -1, k \in \mathbb{Z}$ ,

$$(24) \quad \sigma_{j,k}^2 \leq 2\widehat{\sigma}_{j,k}^2 + c_3 \log((j+2)n)/n,$$

where  $\sigma_{j,k}^2 = \mathbb{E}[\widehat{\sigma}_{j,k}^2]$  and where  $c_3$  only depends on  $\psi$ . We also have on  $\mathcal{A}$ :

$$(25) \quad \widehat{\mathbb{Z}}_{j,r}(\ell_j) \subset \check{\mathbb{Z}}_{j,r} \subset \widehat{\mathbb{Z}}'_{j,r} \subset \check{\mathbb{Z}}'_{j,r}.$$

This embedding is valid for all  $j \geq -1, \ell_j \in \mathbb{N}, r \geq \ell_j + j + 1$  such that

$$2^r \leq \varsigma_j \frac{n}{\log((j+2)n)},$$

and  $\check{\mathbb{Z}}_{j,r}, \check{\mathbb{Z}}'_{j,r}$  are defined by

$$(26) \quad \check{\mathbb{Z}}_{j,r} = \left\{ k \in \mathbb{Z}, \quad 2^{-r-3/2} < \sigma_{j,k}^2 \leq 2^{-r+1/2} \right\},$$

$$(27) \quad \check{\mathbb{Z}}'_{j,r} = \left\{ k \in \mathbb{Z}, \quad 2^{-r-5/2} < \sigma_{j,k}^2 \leq 2^{-r+3/2} \right\}.$$

By convention, the infimum over  $K_{j,r,\ell_j}$  in (17) is equal to 0 if  $\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j = \emptyset$ .

**4.1. Proof of Theorem 5.** For all  $j \geq -1$ , subset  $K \subset \mathbb{Z}$ , we define

$$\widehat{\sigma}_j^2(K) = \sum_{k \in K} \widehat{\sigma}_{j,k}^2,$$

and

$$(28) \quad \sigma_j^2(K) = \mathbb{E} [\widehat{\sigma}_j^2(K)] = \sum_{k \in K} \sigma_{j,k}^2.$$

The proof of the theorem ensues from a succession of lemmas. The first one is classical and is the following:

**Lemma 1.** For all  $K \subset \mathbb{Z}$ ,  $j \geq 0$ , and  $x \in \mathbb{R}$ ,

$$\left( \sum_{k \in K} |\psi(2^j x - k)| \right)^2 \leq c \sum_{k \in K} \psi^2(2^j x - k),$$

where  $c$  only depends on  $\psi$ .

*Proof of Lemma 1.* We use that  $\psi$  is compactly supported and apply Cauchy-Schwarz inequality.  $\square$

**Lemma 2.** Let for all  $j \geq -1$ ,  $\mathcal{C}_j$  be a subset of  $\mathbb{Z}$ . There is an event of probability  $1 - 1/n^4$  on which: for all  $j \geq -1$ , all finite subset  $K_j$  of  $\mathcal{C}_j$ ,

$$(29) \quad 2^{-j/2} \sum_{k \in K_j} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq c_j V_j(K_j, \mathcal{C}_j),$$

where

$$V_j(K_j, \mathcal{C}_j) = \sqrt{\frac{|K_j| \sigma_j^2(K_j) \log_+( \sigma_j^2(\mathcal{C}_j) / \sigma_j^2(K_j) )}{n} + \frac{|K_j| \log_+( n \sigma_j^2(\mathcal{C}_j) / |K_j| )}{n}} \\ + \sqrt{\sigma_j^2(K_j) \frac{\log((j+2)n)}{n} + \frac{\log((j+2)n)}{n}}.$$

Moreover,  $c_j$  is universal if  $j = -1$  and only depends on  $\psi$  if  $j \geq 0$ .

*Proof of Lemma 2.* We only show the lemma when  $j \geq 0$ . The proof when  $j = -1$  is similar (replace  $\psi$  by  $\phi$ ). We consider some  $c_1 > 0$  only depending on  $\psi$  such that

$$\left\| \sum_{k \in \mathbb{Z}} |\psi(2^j \cdot -k)| \right\|_{\infty} \leq 1/c_1.$$

We then consider  $d \geq 1$  and define the at most countable collection

$$\mathcal{F}_j(d) = \left\{ c_1 \sum_{k \in K_j} \psi(2^j \cdot -k), K_j \subset \mathcal{C}_j, |K_j| \leq d \right\}$$

of functions. These functions take values in  $[-1, 1]$  and are piecewise constant on at most  $c_2 d$  pieces where  $c_2$  only depends on  $\psi$ . Therefore,  $\mathcal{F}_j(d)$  is VC subgraph and its dimension is not larger than  $d$ , up to a multiplicative factor depending on  $\psi$  only (see [BB18] for instance).

Let now  $K_j \subset \mathcal{C}_j$  such that  $|K_j| \leq d$ . Elementary computations entail

$$\sum_{k \in K_j} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq 2 \sup_{K'_j \subset K_j} \left| \sum_{k \in K'_j} \left( \widehat{\beta}_{j,k} - \beta_{j,k} \right) \right|.$$

We introduce the map  $\psi_{K'_j}(\cdot)$  defined for  $x \in \mathbb{R}$  by

$$\psi_{K'_j}(x) = c_1 \sum_{k \in K'_j} \psi(2^j x - k).$$

The preceding inequality can then be rewritten as

$$2^{-j/2} \sum_{k \in K_j} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq (2/c_1) \sup_{K'_j \subset K_j} \left| \frac{1}{n} \sum_{i=1}^n \left[ \psi_{K'_j}(X_i) - \mathbb{E}(\psi_{K'_j}(X_i)) \right] \right|.$$

Note that

$$\bar{f}(x) = c_1 \sum_{k \in \mathcal{C}_j} |\psi(2^j x - k)|$$

is an envelope function of  $\mathcal{F}_j(d)$ . Moreover, Lemma 1 gives when  $K'_j \subset K_j$ ,

$$\mathbb{E}[\psi_{K'_j}^2(X)] \leq c_3 \sigma_j^2(K'_j) \leq c_3 \sigma_j^2(K_j),$$

and  $\mathbb{E}[\bar{f}^2(X)] \leq c_3 \sigma_j^2(\mathcal{C}_j)$ . We apply the probabilistic result given by Proposition 6 in Appendix A. We then use that  $x \mapsto x \log_+(a/x)$  is non-decreasing for all  $a > 0$ , and conclude by a union bound.  $\square$

**Lemma 3.** *Let for all  $j \geq -1$ ,  $r \in \mathbb{N}$ ,  $\mathcal{C}_{j,r}$  be a finite subset of  $\mathbb{Z}$ . There is an event of probability  $1 - 1/n^4$  on which: for all  $j \geq -1, r \in \mathbb{N}$ , all subset  $K_{j,r}$  of  $\mathcal{C}_{j,r}$ ,*

$$(30) \quad 2^{-j/2} \sum_{k \in K_{j,r}} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq c_j W_{j,r}(K_{j,r}, \mathcal{C}_{j,r}),$$

where

$$W_{j,r}(K_{j,r}, \mathcal{C}_{j,r}) = \sqrt{\frac{|K_{j,r}| \sigma_j^2(K_{j,r}) \log_+(|\mathcal{C}_{j,r}|/|K_{j,r}|)}{n}} + \frac{|K_{j,r}| \log_+(|\mathcal{C}_{j,r}|/|K_{j,r}|)}{n} \\ + \sqrt{\sigma_j^2(K_{j,r}) \frac{\log((j+2)(r+1)n)}{n}} + \frac{\log((j+2)(r+1)n)}{n}.$$

Moreover,  $c_j$  is universal if  $j = -1$  and only depends on  $\psi$  if  $j \geq 0$ .

*Sketch of the proof of Lemma 3.* The proof is a slight variant to that of Lemma 2. The main difference is that we use (72) in Appendix A in place of (71). We restrict to the case  $j \geq 0$ , and define the collection  $\mathcal{F}_{j,r}(d)$  for  $d \in [1, |\mathcal{C}_{j,r}|]$  by

$$\mathcal{F}_{j,r}(d) = \left\{ c_1 \sum_{k \in K_{j,r}} \psi(2^j \cdot -k), K_{j,r} \subset \mathcal{C}_{j,r}, |K_{j,r}| \leq d \right\}.$$

We have

$$|\mathcal{F}_{j,r}(d)| \leq \sum_{i=0}^d \binom{|\mathcal{C}_{j,r}|}{i}.$$

Proposition 2.5 of [Mas07] entails  $|\mathcal{F}_{j,r}(d)| \leq (e|\mathcal{C}_{j,r}|/d)^d$ . The result follows from Proposition 6 and a union bound.  $\square$

We omit the (easy) proof of the lemma below:

**Lemma 4.** *Let  $f_{j,\cdot} = (f_{j,k})_{k \in \mathbb{Z}}$  be defined by (19). Then, for all  $j \geq -1$ ,  $\|f_{j,\cdot}\|_1 \leq c$ , where  $c$  only depends on  $\psi$  if  $j \geq 0$  and is universal if  $j = -1$ .*

We now state:

**Lemma 5.** *We define for all  $j \geq -1$ ,  $r \in \mathbb{N}$ ,  $m \in \{0, 1, 2\}$ , and all finite subset  $K_{j,r}$  of  $\mathbb{Z}$ ,*

$$\begin{aligned} \mathcal{E}_{j,r}(K_{j,r}, m) = & \sqrt{\frac{|K_{j,r}| \sigma_j^2(K_{j,r}) \log_+(s_{j,r}(K_{j,r}, m))}{n}} + \frac{|K_{j,r}| \log_+(s'_{j,r}(K_{j,r}, m))}{n} \\ & + \sqrt{\sigma_j^2(K_{j,r}) \frac{\log((j+2)(r+1)n)}{n}} + \frac{\log((j+2)(r+1)n)}{n}, \end{aligned}$$

where

$$\begin{aligned} s_{j,r}(K_{j,r}, 0) &= \mathbb{E}[|\widehat{\mathbb{Z}}_j|]/|K_{j,r}| & \text{and} & & s'_{j,r}(K_{j,r}, 0) &= \mathbb{E}[|\widehat{\mathbb{Z}}_j|]/|K_{j,r}| \\ s_{j,r}(K_{j,r}, 1) &= 1/\sigma_j^2(K_{j,r}) & \text{and} & & s'_{j,r}(K_{j,r}, 1) &= n/|K_{j,r}| \\ s_{j,r}(K_{j,r}, 2) &= |\check{\mathbb{Z}}_{j,r}|/|K_{j,r}| & \text{and} & & s'_{j,r}(K_{j,r}, 2) &= |\check{\mathbb{Z}}_{j,r}|/|K_{j,r}|. \end{aligned}$$

Then, there is an event of probability  $1 - 4/n^4$  on which: for all  $j \geq -1$ ,  $r \in \mathbb{N}$ , all finite subset  $K_{j,r}$  of  $\mathbb{Z}$ , and all  $m \in \{0, 1\}$ ,

$$(31) \quad 2^{-j/2} \sum_{k \in K_{j,r}} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq c_j \mathcal{E}_{j,r}(K_{j,r}, m).$$

Moreover, if  $K_{j,r} \subset \check{\mathbb{Z}}_{j,r}$  where  $\check{\mathbb{Z}}_{j,r}$  is defined by (26), the left-hand side of (31) is also not larger than  $c_j \mathcal{E}_{j,r}(K_{j,r}, 2)$ . Here,  $c_j$  is universal if  $j = -1$  and only depends on  $\psi$  if  $j \geq 0$ .

*Proof of Lemma 5.* We only show the lemma when  $|K_{j,r}| \geq 1$ . Note first that  $\sigma_j^2(\mathbb{Z})$  is bounded from above by a numerical value when  $j = -1$  and by a term only depending on  $\psi$  when  $j \geq 0$  (use Lemma 4 and  $\|\psi\|_\infty < \infty$ ). The proof that (31) holds true with  $m = 1$  follows therefore from Lemma 2 (with  $\mathcal{C}_j = \mathbb{Z}$ ). The proof that it is also true with  $m = 2$  is due to Lemma 3 (with  $\mathcal{C}_{j,r} = \check{\mathbb{Z}}_{j,r}$ ).

We now suppose that  $m = 0$ . Let  $g$  be the map defined for  $x \in [0, 1]$  by  $g(x) = 1 - (1 - x)^n$ . Note that  $\widehat{\sigma}_{j,k}^2 > 0$  if and only if  $\psi_{j,k}(X_i) \neq 0$  for some  $i$ . Therefore,

$$(32) \quad \mathbb{E}[|\widehat{\mathbb{Z}}_j|] = \sum_{k \in \mathbb{Z}} g(f_{j,k}),$$

where  $f_{j,k}$  is defined by (19) by  $f_{j,k} = \int_{\text{supp } \psi_{j,k}} f$ . Note that  $g$  is increasing and  $x \mapsto g(x)/x$  is decreasing. Recall that  $\overline{\mathbb{Z}}_j$  is defined in (18). We deduce,

$$\begin{aligned} \mathbb{E}[|\widehat{\mathbb{Z}}_j|] &= \sum_{k \in \overline{\mathbb{Z}}_j} g(f_{j,k}) + \sum_{k \notin \overline{\mathbb{Z}}_j} \frac{g(f_{j,k})}{f_{j,k}} f_{j,k} \\ &\geq |\overline{\mathbb{Z}}_j| g(1/n) + \sum_{k \notin \overline{\mathbb{Z}}_j} \frac{g(1/n)}{1/n} f_{j,k}. \end{aligned}$$

Note that

$$\sigma_j^2(\overline{\mathbb{Z}}_j^c) \leq \max\{1, \|\psi\|_\infty^2\} \sum_{k \notin \overline{\mathbb{Z}}_j} f_{j,k},$$

and hence

$$(33) \quad \mathbb{E}[|\widehat{\mathbb{Z}}_j|] \geq c_1 [|\overline{\mathbb{Z}}_j| + n\sigma_j^2(\overline{\mathbb{Z}}_j^c)],$$

where  $c_1$  only depends on  $\psi$ . We deduce from Lemmas 2 and 3 that with probability  $1 - 2/n^4$ ,

$$2^{-j/2} \sum_{k \in K_{j,r}} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq c_2 [V_j(K_{j,r} \cap \overline{\mathbb{Z}}_j^c, \overline{\mathbb{Z}}_j^c) + W_{j,r}(K_{j,r} \cap \overline{\mathbb{Z}}_j, \overline{\mathbb{Z}}_j)].$$

Elementary computations give

$$V_j(K_{j,r} \cap \overline{\mathbb{Z}}_j^c, \overline{\mathbb{Z}}_j^c) \leq c_3 \left[ \frac{|K_{j,r}| \log_+(n\sigma_j^2(\overline{\mathbb{Z}}_j^c)/|K_{j,r}|)}{n} + \frac{\log((j+2)n)}{n} \right].$$

We conclude using (33).  $\square$

**Lemma 6.** *There is an event of probability  $1 - 4/n^4$  on which: for all  $j \geq -1$ ,  $\varepsilon > 0$ , and finite subset  $K_j$  of  $\mathbb{Z}$ ,*

$$(34) \quad \widehat{\sigma}_j^2(K_j) \leq (1 + \varepsilon)\sigma_j^2(K_j) + c_j \left[ \frac{|K_j| \log_+(\mathbb{E}[|\widehat{\mathbb{Z}}_j|]/|K_j|)}{n} + \frac{\log((j+2)n)}{n} \right],$$

$$(35) \quad \sigma_j^2(K_j) \leq (1 + \varepsilon)\widehat{\sigma}_j^2(K_j) + c_j \left[ \frac{|K_j| \log_+(\mathbb{E}[|\widehat{\mathbb{Z}}_j|]/|K_j|)}{n} + \frac{\log((j+2)n)}{n} \right],$$

where  $c_j$  depends only on  $\varepsilon$  if  $j = -1$  and only depends on  $\varepsilon, \psi$  if  $j \geq 0$ .

*Sketch of the proof of Lemma 6.* We may replace  $\psi$  in the previous proofs by  $\psi^2$ . Hence,

$$\sum_{k \in K_j} |\widehat{\sigma}_{j,k}^2 - \sigma_{j,k}^2| \leq c'_j \mathcal{E}_{j,0}(K_j, 0),$$

where  $c'_j$  is universal if  $j = -1$  and only depends on  $\psi$  if  $j \geq 0$ . We conclude using the elementary inequality  $2\sqrt{xy} \leq \alpha^{-1}x + \alpha y$  valid for all  $\alpha > 0$ .  $\square$

**Lemma 7.** *For all  $j \geq -1$ ,  $\mathbb{E}[|\widehat{\mathbb{Z}}_j|] \leq cn$ , where  $c$  only depends on  $\psi$  when  $j \geq 0$  and is universal if  $j = -1$ .*



*Proof of Lemma 7.* We suppose that  $j \geq 0$  and note that  $\hat{\sigma}_{j,k}^2 > c_1/n$  when  $\hat{\sigma}_{j,k}^2 \neq 0$  as  $\psi$  is piecewise constant. Therefore,

$$\mathbb{E}[|\widehat{\mathbb{Z}}_j|] \leq (n/c_1) \sum_{k \in \mathbb{Z}} \sigma_{j,k}^2.$$

Since  $\psi$  is bounded above,  $\sigma_{j,k}^2 \leq c_2 f_{j,k}$ . We then use Lemma 4.  $\square$

**Lemma 8.** *There exist a universal constant  $\varsigma_{-1}$ , a term  $\varsigma_0$  only depending on  $\psi$  and an event of probability  $1 - 4/n^4$  (the same as that of Lemma 6) on which: for all  $j \geq -1$ ,  $\ell_j \in \mathbb{N}$ ,  $r \geq \ell_j + j + 1$  such that*

$$2^r \leq \varsigma_j \frac{n}{\log((j+2)n)},$$

where  $\varsigma_j = \varsigma_{-1} 1_{j=-1} + \varsigma_0 1_{j \geq 0}$ , the embedding (25) holds true. We also have (24) on this event for all  $j \geq -1$  and  $k \in \mathbb{Z}$ .

*Proof of Lemma 8.* The result follows from Lemma 6 with  $\varepsilon$  small enough and Lemma 7.  $\square$

**Lemma 9.** *We define for all  $j \geq -1$ ,  $r \in \mathbb{N}$ ,  $\ell_j \in \mathbb{N}$ ,  $m \in \{0, 1, 2\}$ , and all finite subset  $K_{j,r,\ell_j}$  of  $\mathbb{Z}$ ,*

$$\begin{aligned} \tilde{\mathcal{E}}_{j,r,\ell_j}(K_{j,r,\ell_j}, m) &= \sqrt{\frac{|K_{j,r,\ell_j}| \hat{\sigma}_j^2(K_{j,r,\ell_j}) \log_+(\tilde{s}_{j,r,\ell_j}(K_{j,r,\ell_j}, m))}{n}} \\ &+ \frac{|K_{j,r,\ell_j}| \log_+(\tilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, m))}{n} + \sqrt{\hat{\sigma}_j^2(K_{j,r,\ell_j}) \frac{\log((j+2)(r+1)n)}{n}} \\ &+ \frac{\log((j+2)(r+1)n)}{n}, \end{aligned}$$

where

$$\begin{aligned} \tilde{s}_{j,r,\ell_j}(K_{j,r,\ell_j}, 0) &= \mathbb{E}[|\widehat{\mathbb{Z}}_j|/|K_{j,r,\ell_j}|] & \text{and} & \quad \tilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, 0) = \mathbb{E}[|\widehat{\mathbb{Z}}_j|/|K_{j,r,\ell_j}|] \\ \tilde{s}_{j,r,\ell_j}(K_{j,r,\ell_j}, 1) &= 1/\hat{\sigma}_j^2(K_{j,r,\ell_j}) & \text{and} & \quad \tilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, 1) = n/|K_{j,r,\ell_j}| \\ \tilde{s}_{j,r,\ell_j}(K_{j,r,\ell_j}, 2) &= |\widehat{\mathbb{Z}}'_{j,r}|/|K_{j,r,\ell_j}| & \text{and} & \quad \tilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, 2) = |\widehat{\mathbb{Z}}'_{j,r}|/|K_{j,r,\ell_j}|. \end{aligned}$$

Then, there is an event of probability  $1 - 8/n^4$  on which: for all  $j \geq -1$ ,  $\ell_j \in \mathbb{N}$ ,  $r \geq \ell_j + j$ , finite subset  $K_{j,r,\ell_j}$  of  $\mathbb{Z}$ , and  $m \in \{0, 1\}$ ,

$$2^{-j/2} \sum_{k \in K_{j,r,\ell_j}} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \leq c_j \tilde{\mathcal{E}}_{j,r,\ell_j}(K_{j,r,\ell_j}, m).$$

Moreover, if  $K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)$  for some  $r \geq \ell_j + j + 1$  and

$$2^r \leq \varsigma_j \frac{n}{\log((j+2)n)},$$

where  $\varsigma_j$  is given by Lemma 8, the inequality holds true with  $m = 2$ . Furthermore,  $c_j$  is universal if  $j = -1$  and only depends on  $\psi$  if  $j \geq 0$ .

*Proof of Lemma 9.* The case  $m = 0$  is merely due to Lemma 5 and (35). The proof when  $m = 2$  follows from Lemma 5, from the inclusions  $\widehat{\mathbb{Z}}_{j,r}(\ell_j) \subset \tilde{\mathbb{Z}}_{j,r} \subset \widehat{\mathbb{Z}}'_{j,r}$  and from the double inequality

$$1/\sqrt{2} \leq \frac{\sigma_j^2(K_{j,r,\ell_j})}{\hat{\sigma}_j^2(K_{j,r,\ell_j})} \leq \sqrt{2}$$

valid for all  $K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j)$ . The case  $m = 1$  essentially follows from elementary computations: let  $\sigma^2$  be the right-hand side of (35) (with  $K_j = K_{j,r,\ell_j}$  and  $\varepsilon = 1$ ). Then, using that  $x \mapsto x \log_+(1/x)$  is non-decreasing,

$$\begin{aligned} \sigma_j^2(K_{j,r,\ell_j}) \log_+(1/\sigma_j^2(K_{j,r,\ell_j})) &\leq \sigma^2 \log_+(1/\sigma^2) \\ &\leq c [\widehat{\sigma}_j^2(K_{j,r,\ell_j}) \log_+(1/\widehat{\sigma}_j^2(K_{j,r,\ell_j})) \\ &\quad + (|K_{j,r,\ell_j}|/n) \log_+(\mathbb{E}[|\widehat{\mathbb{Z}}_j|]/|K_{j,r,\ell_j}|) \log_+(n/|K_{j,r,\ell_j}|) \\ &\quad + (\log((j+2)n)/n) \log_+(n/|K_{j,r,\ell_j}|)]. \end{aligned}$$

Lemma 7 ends the proof.  $\square$

**Lemma 10.** *Lemma 9 holds true (up to an increase of  $c_j$ ) with  $\widetilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, 1)$  replaced by*

$$\widetilde{s}'_{j,r,\ell_j}(K_{j,r,\ell_j}, 1) = 1/\widehat{\sigma}_j^2(K_{j,r,\ell_j}).$$

*Proof of Lemma 10.* We set

$$\begin{aligned} A_j &= \sqrt{\frac{|K_{j,r,\ell_j}| \widehat{\sigma}_j^2(K_{j,r,\ell_j}) \log_+(1/\widehat{\sigma}_j^2(K_{j,r,\ell_j}))}{n}} \\ B_j &= \frac{|K_{j,r,\ell_j}| \log_+(1/\widehat{\sigma}_j^2(K_{j,r,\ell_j}))}{n} \\ B'_j &= \frac{|K_{j,r,\ell_j}| \log_+(n/|K_{j,r,\ell_j}|)}{n}. \end{aligned}$$

We observe that  $x \mapsto x \log_+(1/x)$  is non-decreasing and

$$\sqrt{\log_+(x/\log_+(x))} \geq 0.8 \sqrt{\log_+(x)}.$$

We deduce that if

$$\widehat{\sigma}_j^2(K_{j,r,\ell_j}) \geq \frac{|K_{j,r,\ell_j}| \log_+(n/|K_{j,r,\ell_j}|)}{n},$$

then  $A_j \geq 0.8B'_j$ . If now

$$\widehat{\sigma}_j^2(K_{j,r,\ell_j}) < \frac{|K_{j,r,\ell_j}| \log_+(n/|K_{j,r,\ell_j}|)}{n},$$

then  $B_j \geq 0.8^2 B'_j$ .

We therefore get some  $c > 0$  such that

$$A_j + B'_j \leq c(A_j + B_j),$$

which concludes the proof.  $\square$

**Lemma 11.** *For all  $j \geq -1$ ,  $\xi > 0$ , and probability  $1 - e^{-\xi}$ ,*

$$\mathbb{E}[|\widehat{\mathbb{Z}}_j|] \leq 2|\mathbb{Z}_j| + 2(2L_\psi + 1)\xi.$$

*Proof of Lemma 11.* This result derives from a Poissonian inequality for self-bounding functionals and more precisely from equation (7) of [BLM00]. We set for  $k \in \mathbb{Z}$ ,

$$\begin{aligned} I_j(k) &= \{x \in \mathbb{R}, \psi_{j,k}(x) \neq 0\} \\ &\subset [2^{-j}(k - L_\psi), 2^{-j}(k + L_\psi)], \end{aligned}$$

and remark

$$|\widehat{\mathbb{Z}}_j| = \sum_{k \in \mathbb{Z}} 1_{\exists i \in \{1, \dots, n\}, X_i \in I_j(k)}.$$

We introduce for  $r \in \{1, \dots, n\}$  the random variable

$$|\widehat{\mathbb{Z}}_j(r)| = \sum_{k \in \mathbb{Z}} 1_{\exists i \in \{1, \dots, n\} \setminus \{r\}, X_i \in I_j(k)}.$$

We have  $|\widehat{\mathbb{Z}}_j(r)| \leq |\widehat{\mathbb{Z}}_j|$ . Moreover,

$$\begin{aligned} |\widehat{\mathbb{Z}}_j| - |\widehat{\mathbb{Z}}_j(r)| &= \sum_{k \in \mathbb{Z}} 1_{X_r \in I_j(k)} 1_{\forall i \in \{1, \dots, n\} \setminus \{r\}, X_i \notin I_j(k)} \\ &\leq \sum_{k \in \mathbb{Z}} 1_{X_r \in I_j(k)} \\ &\leq 2L_\psi + 1. \end{aligned}$$

Besides,

$$\begin{aligned} \sum_{r=1}^n \left( |\widehat{\mathbb{Z}}_j| - |\widehat{\mathbb{Z}}_j(r)| \right) &= \sum_{k \in \mathbb{Z}} \sum_{r=1}^n 1_{X_r \in I_j(k)} 1_{\forall i \in \{1, \dots, n\} \setminus \{r\}, X_i \notin I_j(k)} \\ &\leq \sum_{k \in \mathbb{Z}} 1_{\exists i \in \{1, \dots, n\}, X_i \in I_j(k)} \\ &\leq |\widehat{\mathbb{Z}}_j|. \end{aligned}$$

Now, equation (7) of [BLM00] gives for all  $\xi > 0$ , and probability  $1 - e^{-\xi}$ ,

$$\mathbb{E}[|\widehat{\mathbb{Z}}_j|] \leq |\widehat{\mathbb{Z}}_j| + \sqrt{2(2L_\psi + 1)\mathbb{E}[|\widehat{\mathbb{Z}}_j|]\xi}.$$

We conclude by using the elementary inequality  $\sqrt{ab} \leq a/2 + b/2$ . □

**Lemma 12.** *For all  $j \geq -1$ ,  $\xi > 0$ , and probability  $1 - e^{-\xi}$ ,*

$$|\widehat{\mathbb{Z}}_j| \leq 2\mathbb{E}[|\widehat{\mathbb{Z}}_j|] + (7/6)(2L_\psi + 1)\xi.$$

*Proof of Lemma 12.* The proof is similar to that of Lemma 12. We merely use (6) of [BLM00] to get for all  $\xi > 0$ , and probability  $1 - e^{-\xi}$ ,

$$|\widehat{\mathbb{Z}}_j| \leq \mathbb{E}[|\widehat{\mathbb{Z}}_j|] + \sqrt{2(2L_\psi + 1)\mathbb{E}[|\widehat{\mathbb{Z}}_j|]\xi} + 2\frac{2L_\psi + 1}{3}\xi.$$

We conclude as in the preceding proof. □

**Lemma 13.** *With probability  $1 - 1/n^4$ , we have for all  $j \geq -1$ ,*

$$\mathbb{E}[|\widehat{\mathbb{Z}}_j|] \leq 2|\widehat{\mathbb{Z}}_j| + 4(2L_\psi + 1)(\log(j + 2) + 2\log n + 1),$$

and

$$|\widehat{\mathbb{Z}}_j| \leq 2\mathbb{E}[|\widehat{\mathbb{Z}}_j|] + (7/3)(2L_\psi + 1)(\log(j + 2) + 2\log n + 1).$$

*Proof of Lemma 13.* The proof follows from Lemmas 11, 12 and a union bound. □

**Lemma 14.** *There exist a universal constant  $\bar{\varrho}_{-1}$ , a term  $\bar{\varrho}_0$  depending only on  $\psi$ , and an event of probability  $1 - 9/n^4$  on which: for all  $j \geq -1$ ,  $\ell_j \in \mathbb{N}$ ,  $r \geq \ell_j + j$ , and finite subset  $K_{j,r,\ell_j}$  of  $\widehat{\mathbb{Z}}_{j,r}(\ell_j)$ ,*

$$2^{-j/2} \sum_{k \in K_{j,r,\ell_j}} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \leq (\bar{\varrho}_j/2) \widehat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|),$$

where  $\bar{\varrho}_j = \bar{\varrho}_0$  if  $j \geq 1$ . Moreover, (23), (24), and (25) hold true on this event.

*Proof of Lemma 14.* The lemma is a direct result of Lemmas 8, 9, 10 and 13 when  $r \geq \ell_j + j + 1$  (the event of Lemma 8 contains the one of Lemma 9). When  $r = \ell_j + j$ , we use

$$\widehat{\sigma}_j^2(K_{j,r,\ell_j}) \leq \bar{\kappa}_j(2L_\psi + 1),$$

with  $\bar{\kappa}_j = 1_{j=-1} + \|\psi\|_\infty^2 1_{j \geq 0}$ . Moreover,

$$\widehat{\sigma}_j^2(K_{j,r,\ell_j}) \geq 2^{-r-1} |K_{j,r,\ell_j}|.$$

We conclude by using Lemma 9 and the fact that  $x \mapsto x \log_+(1/x)$  is non-decreasing.  $\square$

**Lemma 15.** *There exist terms  $\varsigma_0, \bar{\varrho}_0$  only depending on  $\psi$  and universal constants  $\varsigma_{-1}, \bar{\varrho}_{-1}$  such that if  $\rho_{-1} \geq \bar{\varrho}_{-1}$  and  $\rho_0 \geq \bar{\varrho}_0$ , the estimator  $\hat{f}$  defined by (9) satisfies with probability  $1 - 9/n^4$ : for all  $(\ell_j)_{j \geq -1} \in \mathbb{N}^{\{-1\} \cup \mathbb{N}}$ ,*

$$\begin{aligned} & \left\| f - \hat{f} \right\|_1 \leq c \widehat{T} \\ & + c \sum_{j=-1}^{\widehat{J}} \sum_{r=\ell_j+j}^{\infty} \inf_{K_{j,r,\ell_j} \subset \widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \bar{\mathbb{Z}}_j} \left\{ 2^{-j/2} \sum_{k \in (\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \bar{\mathbb{Z}}_j) \setminus K_{j,r,\ell_j}} |\beta_{j,k}| + \rho_j \widehat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|) \right\}, \end{aligned}$$

where

$$(36) \quad \widehat{T} = \sum_{j=-1}^{\widehat{J}} 2^{-j/2} \sum_{k \notin \bar{\mathbb{Z}}_j} |\beta_{j,k}| + \sum_{j=-1}^{\widehat{J}} 2^{-j/2} \sum_{k \notin \bar{\mathbb{Z}}_j} |\beta_{j,k}| + \sum_{j=\widehat{J}+1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}|,$$

where  $\bar{\mathbb{Z}}_j$  is given by (18) and where  $c$  only depends on  $\bar{\phi}, \bar{\psi}$ . Moreover, (23), (24), and (25) hold true on this event.

*Proof of Lemma 15.* We observe that for all  $j \geq -1$  and  $\ell_j \in \mathbb{N}$ ,

$$(37) \quad \widehat{\mathbb{Z}}_j = \bigcup_{r=\ell_j+j}^{\infty} \widehat{\mathbb{Z}}_{j,r}(\ell_j).$$

In particular,

$$\mathbb{Z} = \bigcup_{r=\ell_j+j}^{\infty} \widehat{\mathbb{Z}}_{j,r}(\ell_j) \cup \widehat{\mathbb{Z}}_j^c.$$

We deduce from (2) that  $\|f - \hat{f}\|_1 \leq c\tilde{T} + c\hat{A}$  where

$$\begin{aligned} \tilde{T} &= \sum_{j=-1}^{\hat{J}} 2^{-j/2} \sum_{k \notin \hat{\mathbb{Z}}_j} |\beta_{j,k}| + \sum_{j=\hat{J}+1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| \\ \hat{A} &= \sum_{j=-1}^{\hat{J}} 2^{-j/2} \sum_{r=\hat{\ell}_j+j}^{\infty} \left\{ \sum_{k \in \hat{\mathbb{Z}}_{j,r}(\hat{\ell}_j) \setminus \hat{K}_{j,r,\hat{\ell}_j}} |\beta_{j,k}| + \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \right\}. \end{aligned}$$

Note that

$$\hat{A} \leq \sum_{j=-1}^{\hat{J}} 2^{-j/2} \left\{ \sum_{r=\hat{\ell}_j+j}^{\infty} \sum_{k \in \hat{\mathbb{Z}}_{j,r}(\hat{\ell}_j)} |\beta_{j,k}| + \sum_{r=\hat{\ell}_j+j}^{\infty} \left\{ - \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} |\beta_{j,k}| + \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \right\} \right\}.$$

The triangle inequality and Lemma 14 entail: on an event of probability  $1 - 9/n^4$ ,

$$\begin{aligned} &-2^{-j/2} \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} |\beta_{j,k}| + 2^{-j/2} \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \\ &\leq -2^{-j/2} \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} |\hat{\beta}_{j,k}| + 2^{1-j/2} \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \\ &\leq -2^{-j/2} \sum_{k \in \hat{K}_{j,r,\hat{\ell}_j}} |\hat{\beta}_{j,k}| + \bar{\rho}_j \hat{\mathcal{E}}_{j,r,\hat{\ell}_j}(|\hat{K}_{j,r,\hat{\ell}_j}|). \end{aligned}$$

By gathering all these results, and by using  $\rho_j \geq \bar{\rho}_j$ ,

$$\hat{A} \leq \sum_{j=-1}^{\hat{J}} 2^{-j/2} \sum_{k \in \hat{\mathbb{Z}}_j} |\beta_{j,k}| + \sum_{j=-1}^{\hat{J}} \sum_{r=\hat{\ell}_j+j}^{\infty} \gamma_{j,r,\hat{\ell}_j}(\hat{K}_{j,r,\hat{\ell}_j}).$$

We use (12), (11), triangle inequality and Lemma 14. This leads to the two following inequalities valid for all  $\ell_j \in \mathbb{N}$ :

$$\begin{aligned} \hat{A} &\leq \sum_{j=-1}^{\hat{J}} 2^{-j/2} \sum_{k \in \hat{\mathbb{Z}}_j} |\beta_{j,k}| \\ &\quad + \sum_{j=-1}^{\hat{J}} \sum_{r=\hat{\ell}_j+j}^{\infty} \inf_{K_{j,r,\ell_j} \subset \hat{\mathbb{Z}}_{j,r}(\ell_j)} \left\{ -2^{-j/2} \sum_{k \in K_{j,r,\ell_j}} |\hat{\beta}_{j,k}| + \rho_j \hat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|) \right\} \\ &\leq \sum_{j=-1}^{\hat{J}} 2^{-j/2} \sum_{k \in \hat{\mathbb{Z}}_j} |\beta_{j,k}| \\ &\quad + \sum_{j=-1}^{\hat{J}} \sum_{r=\hat{\ell}_j+j}^{\infty} \inf_{K_{j,r,\ell_j} \subset \hat{\mathbb{Z}}_{j,r}(\ell_j)} \left\{ -2^{-j/2} \sum_{k \in K_{j,r,\ell_j}} |\beta_{j,k}| + (3/2)\rho_j \hat{\mathcal{E}}_{j,r,\ell_j}(|K_{j,r,\ell_j}|) \right\}. \end{aligned}$$

We use (37) and the triangle inequality to conclude.  $\square$

**Lemma 16.** *Let  $\widehat{T}$  be defined by (36). Then,*

$$\mathbb{E}[\widehat{T}] \leq 3T_1 + 2T_2,$$

where  $T_1$  and  $T_2$  are defined by (21) and (22).

*Proof of Lemma 16.* There is at most one observation  $X_i$  such that  $\psi_{j,k}(X_i) \neq 0$  when  $j \geq \widehat{J} + 1$ . Moreover, no observation  $X_i$  satisfies  $\psi_{j,k}(X_i) \neq 0$  when  $k \notin \widehat{\mathbb{Z}}_j$ , no matter  $j$ . We deduce,

$$\mathbb{E}[\widehat{T}] \leq T_1 + \sum_{j=-1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| \left( (1 - f_{j,k})^n + n f_{j,k} (1 - f_{j,k})^{n-1} \right),$$

where  $f_{j,k}$  is given by (19). We conclude by noticing that  $(1 - f_{j,k})^n + n f_{j,k} (1 - f_{j,k})^{n-1} \leq 2$  if  $f_{j,k} \leq 1/n$  and  $(1 - f_{j,k})^n \leq n f_{j,k} (1 - f_{j,k})^{n-1}$  if  $f_{j,k} > 1/n$ .  $\square$

**Lemma 17.** *Let  $\xi > 0$ . The following assertion holds true with probability  $1 - \xi/n$ : for all  $q > 1$  and  $f \in \mathbb{L}^q$ ,*

$$2^{\widehat{J}} \leq \max \left\{ 1, 8L_\psi \left( n^2 \|f\|_q / \xi \right)^{\frac{q}{q-1}} \right\}.$$

*In particular, for all  $k \geq 1$ ,  $\mathbb{E}[\widehat{J}^k] \leq C \log^k n$  where  $C$  depends on  $k$ ,  $q$ ,  $\|f\|_q$  and  $L_\psi$  only.*

*Proof of Lemma 17.* The proof of this lemma is deferred to Appendix D.  $\square$

*Proof of Theorem 5.* There exists  $c_1$  only depending on  $\bar{\phi}$ ,  $\bar{\psi}$  such that

$$\begin{aligned} \|\hat{f}\|_1 &\leq \frac{c_1}{n} \sum_{i=1}^n \left[ \sum_{k \in \mathbb{Z}} |\phi(2^j X_i - k)| + \sum_{j=0}^{\widehat{J}} \sum_{k \in \mathbb{Z}} |\psi(2^j X_i - k)| \right] \\ &\leq c_2 (\widehat{J} + 2). \end{aligned}$$

We deduce from Lemma 17 an event  $\mathcal{A}_1$  of probability  $1 - 1/n^4$  on which  $\widehat{J} \leq c_0 \log n$ . This lemma also gives  $\mathbb{E}[\widehat{J}^2] \leq c_3 \log^2 n$  and hence  $\mathbb{E}[\|\hat{f}\|_1^2] \leq c_4 \log^2 n$ .

Let  $\mathcal{A}_2$  be the event of probability  $1 - 9/n^4$  that appears in Lemma 15. We set  $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$  and get

$$\mathbb{E} \left[ \|f - \hat{f}\|_1 \right] \leq \mathbb{E} \left[ \|f - \hat{f}\|_1 1_{\mathcal{A}} \right] + \mathbb{E} \left[ \|f - \hat{f}\|_1 1_{\mathcal{A}^c} \right].$$

The first term can be bounded from above by using Lemmas 15 and 16. As to the second term, we use the triangle and Cauchy-Schwarz inequalities to get

$$\begin{aligned} \mathbb{E} \left[ \|f - \hat{f}\|_1 1_{\mathcal{A}^c} \right] &\leq \mathbb{P}(\mathcal{A}^c) + \mathbb{E}[\|\hat{f}\|_1^2]^{1/2} \mathbb{P}(\mathcal{A}^c)^{1/2} \\ &\leq c_5 \frac{\log n}{n^2}. \end{aligned}$$

$\square$

**4.2. Proof of Theorem 2: intermediate lemmas.** We first recall the following result. We refer to [CVNRF15] for its proof (see their Propositions 3.2 and 4.5).

**Lemma 18.** *Let  $\Lambda$  be an at most countable set, and  $x = (x_\lambda)_{\lambda \in \Lambda} \in \mathbb{R}_+^\Lambda$ . For all  $0 < p < q < \infty$ ,*

$$(38) \quad \|x\|_q^q \leq \frac{q}{q-p} \|x\|_\infty^{q-p} \|x\|_{p,\infty}^p.$$

Moreover, for all  $p > 1$ , and finite subset  $\Gamma$  of  $\Lambda$ ,

$$(39) \quad \sum_{\lambda \in \Gamma} x_\lambda \leq \frac{p}{p-1} \|x\|_{p,\infty} |\Gamma|^{1-1/p}.$$

The lemma below is elementary.

**Lemma 19.** *Suppose that  $f \in \mathcal{WT}_\theta(M)$  for some  $\theta \in (0,1)$  and let  $f_{j,\cdot} = (f_{j,k})_{k \in \mathbb{Z}}$  be defined by (19). Then, for all  $j \geq -1$ ,*

$$\|f_{j,\cdot}\|_{\theta,\infty}^\theta \leq cM2^{j(1-\theta)},$$

where  $c$  only depends on  $\psi$  and  $\theta$  if  $j \geq 0$  and only depends on  $\theta$  if  $j = -1$ .

We now show:

**Lemma 20.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0,1) \cap [0,p]$ ,  $R > 0$ ,  $M \geq 1$ . Consider  $v_n > 0$  and define*

$$T_1(v_n) = \sum_{j=-1}^{\infty} 2^{-j/2} \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq v_n}} |\beta_{j,k}|.$$

Then, for all  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$ ,

$$(40) \quad T_1(v_n) \leq c \left[ R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} v_n^{\alpha(1-\theta)/(\alpha+1-\theta/p)} + M v_n^{1-\theta} \right],$$

where  $c$  only depends on  $p, \alpha, \theta, \psi$ .

*Proof of Lemma 20.* We define for  $r \geq 0$ ,  $j \geq -1$ ,

$$\bar{\mathbb{Z}}_{j,r} = \{k \in \mathbb{Z}, 2^{-r-1} < f_{j,k} \leq 2^{-r}\}.$$

Since  $f \in \mathcal{WT}_\theta(M)$ , Lemma 19 implies when  $\theta \in (0,1)$ ,

$$(41) \quad |\bar{\mathbb{Z}}_{j,r}| \leq 2^{(r+1)\theta} \|f_{j,\cdot}\|_{\theta,\infty}^\theta \leq c_1 M 2^{r\theta} 2^{j(1-\theta)},$$

where  $c_1$  only depends on  $\psi$ . We can check that this result remains true when  $\theta = 0$ . We also remark that  $2^{-j/2} |\beta_{j,k}| \leq c_2 f_{j,k}$  as  $\psi$  is bounded and hence

$$T_1(v_n) \leq A + T_1'(v_n),$$

where

$$A = c_2 \sum_{2^r \geq 1/v_n} 2^{-r} |\bar{\mathbb{Z}}_{-1,r}|$$

$$T_1'(v_n) = \sum_{j=0}^{\infty} \sum_{2^r \geq 1/v_n} \sum_{k \in \bar{\mathbb{Z}}_{j,r}} \min \left\{ c_2 2^{-r}, 2^{-j/2} |\beta_{j,k}| \right\}.$$

Note that

$$A \leq c_1 c_2 M 2^{-(1-\theta)} \sum_{2^r \geq 1/v_n} 2^{-r(1-\theta)} \leq c_3 M v_n^{1-\theta}.$$

We now focus on  $T'_1(v_n)$ .

We first suppose that  $p > 1$ . By using (39),

$$2^{-j/2} \sum_{k \in \mathbb{Z}_{j,r}} |\beta_{j,k}| \leq c_4 2^{-j/2} \|\beta_{j,\cdot}\|_{p,\infty} |\overline{\mathbb{Z}}_{j,r}|^{1-1/p},$$

and  $f \in \mathcal{WB}_{p,\infty}^\alpha(R)$ , we get

$$\begin{aligned} T'_1(v_n) &\leq c_5 \sum_{j=0}^{\infty} \sum_{2^r \geq 1/v_n} \min \left\{ 2^{-r} |\overline{\mathbb{Z}}_{j,r}|, R 2^{-j(\alpha+1-1/p)} |\overline{\mathbb{Z}}_{j,r}|^{1-1/p} \right\} \\ &\leq c_6 \sum_{j=0}^{\infty} \sum_{2^r \geq 1/v_n} \min \left\{ M 2^{-(r-j)(1-\theta)}, R M^{1-1/p} 2^{r(1-1/p)\theta} 2^{-j[(1-1/p)\theta+\alpha]} \right\} \\ &\leq c_7 R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} \sum_{2^r \geq 1/v_n} 2^{-r\alpha(1-\theta)/(\alpha+1-\theta/p)} \\ &\leq c_8 R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} v_n^{\alpha(1-\theta)/(\alpha+1-\theta/p)}. \end{aligned}$$

We now suppose  $p \leq 1$  and consider  $j_0 \in \mathbb{N}$ . We have  $T'_1(v_n) \leq T'_{1,1}(v_n) + T'_{1,2}(v_n)$  where

$$\begin{aligned} T'_{1,1}(v_n) &= \sum_{j=0}^{j_0-1} 2^{-j/2} \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq v_n}} |\beta_{j,k}| \\ T'_{1,2}(v_n) &= \sum_{j=j_0}^{\infty} 2^{-j/2} \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq v_n}} |\beta_{j,k}|. \end{aligned}$$

By using  $2^{-j/2} |\beta_{j,k}| \leq c_2 f_{j,k}$ , we get when  $\theta \neq 0$ ,

$$\begin{aligned} T'_{1,1}(v_n) &\leq c_9 \sum_{j=0}^{j_0-1} \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq v_n}} f_{j,k} \\ &\leq c_{10} v_n^{1-\theta} \sum_{j=0}^{j_0-1} \|f_{j,\cdot}\|_{\theta,\infty}^\theta \quad \text{thanks to (38)} \\ &\leq c_{11} v_n^{1-\theta} M 2^{j_0(1-\theta)} \quad \text{thanks to Lemma 19.} \end{aligned}$$

This last inequality remains true when  $\theta = 0$ . Moreover, by using (38) when  $p \neq 1$ ,

$$\begin{aligned} T'_{1,2} &\leq c_{12} v_n^{1-p} \sum_{j=j_0}^{\infty} 2^{-jp/2} \|\beta_{j,\cdot}\|_{p,\infty}^p \\ &\leq c_{13} v_n^{1-p} \sum_{j=j_0}^{\infty} 2^{-jp/2} R^p 2^{-jp(\alpha+1/2-1/p)} \quad \text{as } f \in \mathcal{WB}_{p,\infty}^\alpha(R). \end{aligned}$$



Note that this inequality also holds true when  $p = 1$  and  $f \in \mathcal{B}_{1,\infty}^\alpha(R)$ . Therefore, in both cases,

$$T'_{1,2} \leq c_{14} v_n^{1-p} R^p 2^{-j_0 p (\alpha + 1 - 1/p)}.$$

We conclude by choosing  $j_0$  appropriately.  $\square$

**Lemma 21.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$ . Then, for all  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$ ,*

$$(42) \quad T_1 + T_2 \leq c \left[ R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)} + M n^{-(1-\theta)} \right],$$

where  $T_1$  and  $T_2$  are defined by (21) and (22), and where  $c$  only depends on  $p, \alpha, \theta, \psi$ .

*Proof of Lemma 21.* We focus on  $T_2$  as a bound on  $T_1$  may be obtained via the preceding lemma. We first consider the case  $p \geq 1$ . Since  $f \in \mathcal{WT}_\theta(M)$ , we have when  $\theta \neq 0$ .

$$(43) \quad |\bar{\mathbb{Z}}_j| \leq n^\theta \|f_{j,\cdot}\|_{\theta,\infty}^\theta \leq c_1 M n^\theta 2^{j(1-\theta)},$$

where  $c_1$  only depends on  $\psi$  (Lemma 19). This inequality remains true when  $\theta = 0$ .

We consider  $j_0 \in \mathbb{N}$  and decompose  $T_2$  as  $T_2 = T_{2,1} + T_{2,2}$  where

$$T_{2,1} = n \sum_{j=-1}^{j_0-1} 2^{-j/2} \sum_{k \in \bar{\mathbb{Z}}_j} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1},$$

$$T_{2,2} = n \sum_{j=j_0}^{\infty} 2^{-j/2} \sum_{k \in \bar{\mathbb{Z}}_j} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1}.$$

We use  $2^{-j/2} |\beta_{j,k}| \leq c_2 f_{j,k}$  to get

$$T_{2,1} \leq c_2 n \sum_{j=-1}^{j_0-1} \sum_{k \in \bar{\mathbb{Z}}_j} f_{j,k}^2 (1 - f_{j,k})^{n-1}.$$

Since  $x^2(1-x)^{n-1} \leq c_3/n^2$  for all  $x \in [0, 1]$ ,

$$(44) \quad T_{2,1} \leq \frac{c_4}{n} \sum_{j=-1}^{j_0-1} |\bar{\mathbb{Z}}_j| \leq c_5 M n^{-(1-\theta)} 2^{j_0(1-\theta)} \quad \text{thanks to (43)}.$$

As to  $T_{2,2}$ , we deduce from (39),  $f \in \mathcal{WB}_{p,\infty}^\alpha(R)$ , and (43) that if  $p > 1$ ,

$$2^{-j/2} \sum_{k \in \bar{\mathbb{Z}}_j} |\beta_{j,k}| \leq c_6 R M^{1-1/p} 2^{-j(\alpha+\theta(1-1/p))} n^{(1-1/p)\theta},$$

and hence, using that  $x(1-x)^{n-1} \leq c_7/n$ ,

$$T_{2,2} \leq c_8 R M^{1-1/p} 2^{-j_0(\alpha+\theta(1-1/p))} n^{(1-1/p)\theta}.$$

We finally choose  $j_0$  in a suitable way to conclude the proof when  $p > 1$ . Note that the above reasoning also works with  $p = 1$  if we replace the Lorentz norm  $\|\cdot\|_{1,\infty}$  by the  $\mathbb{L}^1$  norm.

We now turn to the case  $p < 1$ . We write  $T_2 = T_{2,1} + T_{2,2}$  where

$$T_{2,1} = n \sum_{j=-1}^{j_1} 2^{-j/2} \sum_{k \in \mathbb{Z}_j} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1}$$

$$T_{2,2} = n \sum_{j=j_1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{Z}_j} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1}.$$

Note that (44) does not use  $p > 1$  and therefore also holds true when  $p < 1$  (with  $j_0$  replaced by  $j_1$ ). Besides, as  $2^{-j/2} |\beta_{j,k}| \leq c_2 f_{j,k}$  and  $x^2(1-x)^{n-1} \leq c_3/n^2$ ,

$$n 2^{-j/2} |\beta_{j,k}| f_{j,k} (1 - f_{j,k})^{n-1} \leq c_2 n f_{j,k}^2 (1 - f_{j,k})^{n-1} \leq c_9/n.$$

We deduce from (38),

$$T_{2,2} \leq c_{10} n^{-(1-p)} \sum_{j=j_1}^{\infty} \left\| n 2^{-j/2} \beta_{j, \cdot} f_{j, \cdot} (1 - f_{j, \cdot})^{n-1} \right\|_{p, \infty}^p.$$

Yet,  $x(1-x)^{n-1} \leq c_7/n$  for all  $x \in [0, 1]$  and hence,

$$T_{2,2} \leq c_{11} n^{-(1-p)} \sum_{j=j_1}^{\infty} 2^{-jp/2} \|\beta_{j, \cdot}\|_{p, \infty}^p$$

$$\leq c_{12} n^{-(1-p)} R^p 2^{-j_1 p (\alpha + 1 - 1/p)}.$$

It then remains to choose  $j_1$  to conclude. □

**Lemma 22.** For all  $j \geq -1$ ,  $\ell_j \in \mathbb{N}$ ,  $r \geq \ell_j + j$ ,  $\theta \in [0, 1)$ ,  $M \geq 1$ ,  $f \in \mathcal{WT}_\theta(M)$ ,

$$(45) \quad \mathbb{E} \left[ |\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j| \right] \leq c M 2^{r\theta} 2^{j(1-\theta)},$$

where  $c$  only depends on  $\psi$  and  $\theta$ .

*Proof of Lemma 22.* The proof is straightforward when  $\theta = 0$  and we assume therefore that  $\theta > 0$ . We have,

$$\mathbb{E} \left[ |\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j| \right] \leq \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \geq 1/n}} \mathbb{P} [\widehat{\sigma}_{j,k}^2 > 2^{-r-1}].$$

Set  $t = 2^{-r-1}$  and define

$$K_t = \{k \in \mathbb{Z}, \sigma_{j,k}^2 \geq t/2\}.$$

We derive from  $\sigma_{j,k}^2 \leq \max\{1, \|\psi\|_\infty^2\} f_{j,k}$ , Lemma 19, and  $f \in \mathcal{WT}_\theta(M)$ , that  $|K_t| \leq c_1 M 2^{j(1-\theta)} t^{-\theta}$ . Moreover,

$$\mathbb{E} \left[ |\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j| \right] \leq |K_t| + \sum_{\substack{k \notin K_t \\ f_{j,k} \geq 1/n}} \mathbb{P} [\widehat{\sigma}_{j,k}^2 \geq t]$$

$$\leq c_1 M 2^{j(1-\theta)} t^{-\theta} + \sum_{\substack{k \notin K_t \\ f_{j,k} \geq 1/n}} \mathbb{P} [\widehat{\sigma}_{j,k}^2 \geq \sigma_{j,k}^2 + t/2].$$

We use Bennett's inequality (and more precisely equation (2.16) of [Mas07]) to get

$$\mathbb{P} [\widehat{\sigma}_{j,k}^2 \geq \sigma_{j,k}^2 + t/2] \leq \exp [-c_2 n t^2 / (\sigma_{j,k}^2 + t)],$$

where  $c_2$  only depends on  $\|\psi\|_\infty$ . Therefore,

$$\mathbb{E} \left[ |\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j| \right] \leq c_1 2^{j(1-\theta)} M t^{-\theta} + \sum_{\substack{k \notin K_t \\ f_{j,k} \geq 1/n}} e^{-c_3 n t}.$$

Note that the number of  $k$  such that  $f_{j,k} \geq 1/n$  is bounded from above by  $c_4 M n^\theta 2^{j(1-\theta)}$ . Therefore,

$$\mathbb{E} \left[ |\widehat{\mathbb{Z}}_{j,r}(\ell_j) \cap \overline{\mathbb{Z}}_j| \right] \leq c_5 2^{j(1-\theta)} M t^{-\theta} \left[ 1 + (n t)^\theta e^{-c_3 n t} \right].$$

We conclude by remarking that the map  $x \mapsto x^\theta e^{-c_3 x}$  is bounded on  $\mathbb{R}$ .  $\square$

**Lemma 23.** *For all  $j \geq -1$ ,  $\theta \in [0, 1)$ ,  $M \geq 1$ ,  $f \in \mathcal{WT}_\theta(M)$ ,*

$$(46) \quad \mathbb{E} \left[ |\widehat{\mathbb{Z}}_j| \right] \leq c M n^\theta 2^{j(1-\theta)},$$

where  $c$  only depends on  $\psi$  and  $\theta$ .

*Proof of Lemma 23.* The proof is straightforward when  $\theta = 0$  and we assume from now on that  $\theta \in (0, 1)$ . We deduce from Lemma 19,

$$(47) \quad \mathbb{E} \left[ |\widehat{\mathbb{Z}}_j \cap \overline{\mathbb{Z}}_j| \right] \leq |\overline{\mathbb{Z}}_j| \leq c_1 M n^\theta 2^{j(1-\theta)}.$$

We define for  $r \geq 0$ ,  $j \geq -1$ ,

$$\overline{\mathbb{Z}}_{j,r} = \{k \in \mathbb{Z}, 2^{-r-1} < f_{j,k} \leq 2^{-r}\}$$

and use Lemma 19 to get

$$|\overline{\mathbb{Z}}_{j,r}| \leq c_2 M 2^{r\theta} 2^{j(1-\theta)},$$

where  $c_2$  only depends on  $\psi$ . Now,

$$(48) \quad \begin{aligned} \mathbb{E} \left[ |\widehat{\mathbb{Z}}_j \cap \overline{\mathbb{Z}}_j^c| \right] &\leq c_3 n \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq 1/n}} f_{j,k} \\ &\leq c_3 n \sum_{2^r \geq n} |\overline{\mathbb{Z}}_{j,r}| 2^{-r} \\ &\leq c_4 n M 2^{j(1-\theta)} \sum_{2^r \geq n} 2^{-r(1-\theta)} \\ &\leq c_5 M 2^{j(1-\theta)} n^\theta. \end{aligned}$$

We group (47) and (48) together to end the proof.  $\square$

**Lemma 24.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$  and  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$ . Let  $\pi = 1/\min\{1, \min_{x \in \mathbb{R}, \psi(x) \neq 0} \psi^2(x)\}$  and  $\ell$  be an integer such that  $2^\ell \leq M^{1/(1-\theta)} < 2^{\ell+1}$ .*

1. *For all  $2^r > \pi n$  such that  $r \geq \ell + j + 1$ , for some  $j \geq -1$ ,  $\widehat{\mathbb{Z}}_{j,r}(\ell) = \emptyset$ .*
2. *If  $p \geq 1$ , then for all  $j \geq 0$ , and  $r \geq \ell + j$ ,*

$$(49) \quad \mathbb{E} \left[ \widehat{B}_{j,r,\ell}(\emptyset) \right] \leq c R M^{1-1/p} 2^{-j(\alpha-\theta/p+\theta)} 2^{(1-1/p)r\theta}.$$

3. For all  $j \geq -1$ ,  $r = j + \ell$ , and subset  $K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ ,

$$(50) \quad \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) \right] \leq c \left[ \sqrt{\frac{\mathbb{E}[|K_{j,r,\ell}|]}{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

Moreover,

$$(51) \quad \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|) \right] \leq c \left[ \sqrt{\frac{M^{1/(1-\theta)} 2^j}{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

4. For all  $j \geq -1$ ,  $r \geq j + \ell + 1$ ,  $K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ , and  $\theta \neq 0$ ,

$$(52) \quad \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|)1_{\mathcal{A}} \leq c \left[ |K_{j,r,\ell}| \sqrt{\frac{2^{-r} \log_+^2(Mn^\theta 2^{j(1-\theta)})/|K_{j,r,\ell}|}{n}} \right. \\ \left. + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right],$$

where  $\mathcal{A}$  is the event appearing in Theorem 5. Moreover,

$$(53) \quad \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|)1_{\mathcal{A}} \right] \leq c \left[ M2^{r\theta+j(1-\theta)} \sqrt{\frac{2^{-r} \log_+(n2^{-r})}{n}} \right. \\ \left. + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

5. For all  $j \geq -1$ ,  $r \geq j + \ell + 1$ , and  $K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ ,

$$(54) \quad \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) \leq c \left[ |K_{j,r,\ell}| \sqrt{\frac{2^{-r} \log_+^2(2^r/|K_{j,r,\ell}|)}{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

Moreover,

$$(55) \quad \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|) \right] \leq c \left[ M2^{r\theta+j(1-\theta)} \sqrt{\frac{2^{-r} \log_+^2(2^{(r-j)(1-\theta)})/M}{n}} \right. \\ \left. + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

6. For all  $j \geq -1$ ,  $r \geq j + \ell + 1$  such that  $2^r \leq \varsigma_j n / \log((j+2)n)$ ,

$$(56) \quad \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|) \leq c \left[ |\widehat{\mathbb{Z}}'_{j,r}| \sqrt{\frac{2^{-r}}{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

Moreover,

$$(57) \quad \mathbb{E}[|\widehat{\mathbb{Z}}'_{j,r}|1_{\mathcal{A}}] \leq cM2^{r\theta+j(1-\theta)}.$$

In all these inequalities,  $c$  only depends on  $\psi$ ,  $p$ ,  $\theta$ .

*Proof of Lemma 24.* The first point is true because  $\psi$  is piecewise constant. We now assume that  $2^r \leq \pi n$  in the rest of the proof.

We turn to the second point. We have,

$$\widehat{B}_{j,r,\ell}(\emptyset) \leq 2^{-j/2} \sum_{k \in \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} |\beta_{j,k}|.$$

This gives (49) when  $p = 1$  as  $f \in \mathcal{B}_{1,\infty}^\alpha(R)$ . When  $p > 1$ , we use (39) to get

$$\widehat{B}_{j,r,\ell}(\emptyset) \leq c_1 2^{-j/2} \|\beta_{j,\cdot}\|_{p,\infty} |\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|^{1-1/p}.$$

We then take the expectation, apply Jensen's inequality and (45) to get (49).

We now show (50). By using Jensen's inequality,

$$(58) \quad \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) \right] \leq c_2 \left[ \sqrt{\frac{\mathbb{E}[|K_{j,r,\ell}|]}{n}} + \frac{\mathbb{E}[|K_{j,r,\ell}|] \log_+(2^{r+1}/\mathbb{E}[|K_{j,r,\ell}|])}{n} \right. \\ \left. + \sqrt{\frac{\log((j+2)(r+1)n)}{n}} + \frac{\log((j+2)(r+1)n)}{n} \right].$$

Note that  $|K_{j,r,\ell}| \leq |\widehat{\mathbb{Z}}_{j,r}(\ell)|$  and

$$(59) \quad |\widehat{\mathbb{Z}}_{j,r}(\ell)| \leq 2^{r+1} \sum_{k \in \mathbb{Z}} \widehat{\sigma}_{j,k}^2 \leq c_3 2^r.$$

It then follows from the inequality  $2^r \leq \pi n$  and from elementary computations that the second term in (58) is not smaller than the first one, up to a multiplicative factor.

As to (51), we remark that (45) becomes

$$\mathbb{E}[|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|] \leq c_4 M^{1/(1-\theta)} 2^j$$

when  $r = j + \ell$ . We then use (50) with  $K_{j,r,\ell} = \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ .

We now prove (52). Observe that for all  $2^r \leq \pi n$ , and  $r \geq \ell + j + 1$ ,

$$\widehat{\lambda}_{j,r} \leq c_5 \min \left\{ |\widehat{\mathbb{Z}}_j| + \log n, 2^r \right\},$$

where  $c_5$  only depends on the wavelet basis (this uses  $|\widehat{\mathbb{Z}}'_{j,r}| \leq |\widehat{\mathbb{Z}}_j|$  when  $2^r \leq c_j n / \log((j+2)n)$ ). We deduce from (23) that on  $\mathcal{A}$ ,

$$\widehat{\lambda}_{j,r} \leq c_6 (\mathbb{E}[|\widehat{\mathbb{Z}}_j|] + \log n).$$

As  $M \geq 1$ , and  $\theta > 0$ , we deduce from (46),

$$\widehat{\lambda}_{j,r} \leq c_7 M n^\theta 2^{j(1-\theta)},$$

and using (59),

$$(60) \quad \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) 1_{\mathcal{A}} \leq c_8 |K_{j,r,\ell}| \sqrt{\frac{2^{-r} \log_+(M n^\theta 2^{j(1-\theta)}) / |K_{j,r,\ell}|}{n}} \\ + c_8 |K_{j,r,\ell}| \frac{\log_+(M n^\theta 2^{j(1-\theta)}) / |K_{j,r,\ell}|}{n} + c_8 \frac{\log((j+2)(r+1)n)}{\sqrt{n}}.$$

We then use the condition on  $r$  to get (52). As to (53), we apply (60) with  $K_{j,r,\ell} = \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ . We take the expectation, apply Jensen's inequality and (45). This yields

$$\begin{aligned} \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|)1_{\mathcal{A}} \right] &\leq c_9 M 2^{r\theta+j(1-\theta)} \sqrt{\frac{2^{-r} \log_+(n2^{-r})}{n}} \\ &\quad + c_9 M 2^{r\theta+j(1-\theta)} \frac{\log_+(n2^{-r})}{n} + c_9 \frac{\log((j+2)(r+1)n)}{\sqrt{n}}. \end{aligned}$$

We then remark that the second term is not smaller than the first one thanks to the condition on  $r$  (up to a multiplicative factor).

The proof of (54) is merely based on the inequality  $\widehat{\lambda}_{j,r} \leq 2^{r+1}$ , the condition on  $r$ , and on (59). The proof of (55) then follows from Jensen's inequality and (45).

We turn to the proof of (56). Here, we use  $\widehat{\lambda}_{j,r} \leq |\widehat{\mathbb{Z}}'_{j,r}|$ . By doing as in the proof of (59),  $|\widehat{\mathbb{Z}}'_{j,r}| \leq c_{10} 2^r$ . Note also that the maps  $x \mapsto x\sqrt{\log_+(a/x)}$  and  $x \mapsto x\log_+(a/x)$  are increasing. By using moreover  $\widehat{\mathbb{Z}}_{j,r}(\ell) \subset \widehat{\mathbb{Z}}'_{j,r}$ ,

$$\widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|)1_{\mathcal{A}} \leq c_{11} \left[ |\widehat{\mathbb{Z}}'_{j,r}| \sqrt{\frac{2^{-r}}{n}} + \frac{|\widehat{\mathbb{Z}}'_{j,r}|}{n} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

By using the condition on  $r$ , the second term in the above inequality is not larger than the first one, up to a multiplicative factor.

Finally, the proof of (57) comes from the embedding  $\widehat{\mathbb{Z}}'_{j,r} \subset \check{\mathbb{Z}}'_{j,r}$  (see (25)) valid on  $\mathcal{A}$ , and from the following inequalities valid for all  $t > 0$ ,

$$\begin{aligned} |\{k \in \mathbb{Z}, \sigma_{j,k}^2 \geq t\}| &\leq |\{k \in \mathbb{Z}, f_{j,k} \geq c_{12}t\}| \\ &\leq c_{13} M^\theta t^{-\theta} 2^{j(1-\theta)} \end{aligned}$$

thanks to Lemma 19. □

**Lemma 25.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $R > 0$ , and  $f$  be a density of  $\mathcal{WB}_{p,\infty}^\alpha(R)$ . Then, there exist  $C > 0$  and  $q > 1$  such that  $\|f\|_q \leq C$ . Moreover,  $C$  only depends on the wavelet basis,  $p$ ,  $\alpha$  and  $R$ .*

*Proof of Lemma 25.* The proof of this lemma is deferred to Appendix C. □

**4.3. Proof of Theorem 2.** Throughout this section,  $\ell$  denotes an integer such that

$$2^\ell \leq M^{1/(1-\theta)} < 2^{\ell+1}.$$

We now introduce for  $K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j$ ,

$$\widehat{T}_{j,r,\ell}(K_{j,r,\ell}) = \left\{ \widehat{B}_{j,r,\ell}(K_{j,r,\ell}) + \widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) \right\} 1_{\mathcal{A}},$$

where  $\mathcal{A}$  is the event defined in Theorem 5. It follows from Lemmas 21, 24 and 25 that we only need to bound

$$(61) \quad \begin{aligned} & \sum_{\substack{r \geq \ell-1 \\ 2^r \leq \pi n}} \mathbb{E} \left[ \inf_{K_{-1,r,\ell} \subset \widehat{\mathbb{Z}}_{-1,r}(\ell) \cap \overline{\mathbb{Z}}_{-1}} \widehat{T}_{-1,r,\ell}(K_{-1,r,\ell}) \right] \\ & + \sum_{j=0}^{c_0 \log n} \mathbb{E} \left[ \inf_{K_{j,\ell+j,\ell} \subset \widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] \\ & + \sum_{j \geq 0} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right] \end{aligned}$$

from above (we take  $\ell_j = \ell$  for all  $j$ ). Here,  $c_0$  is a factor depending on  $\alpha, p, R$  and the wavelet basis only. We begin by studying the first term of (61).

**Lemma 26.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$  and  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$ . Then,*

$$(62) \quad \begin{aligned} & \sum_{\substack{r \geq \ell-1 \\ 2^r \leq \pi n}} \mathbb{E} \left[ \inf_{K_{-1,r,\ell} \subset \widehat{\mathbb{Z}}_{-1,r}(\ell) \cap \overline{\mathbb{Z}}_{-1}} \widehat{T}_{-1,r,\ell}(K_{-1,r,\ell}) \right] \\ & \leq c \left[ Mn^{-(1-\theta)} [1_{\theta > 1/2} + (\log n)1_{\theta = 1/2}] + M^{1/(2(1-\theta))} n^{-1/2} + (\log^2 n)n^{-1/2} \right], \end{aligned}$$

where  $c$  only depends on  $\psi, \theta, p$ .

*Proof of Lemma 26.* Let  $A$  be the left-hand side of (62). We take  $K_{-1,r,\ell} = \widehat{\mathbb{Z}}_{-1,r}(\ell) \cap \overline{\mathbb{Z}}_{-1}$  and deduce from (51), and (55) that if  $\theta \in [0, 1/2)$ ,

$$\begin{aligned} A & \leq c_1 \left[ \sqrt{\frac{M^{1/(1-\theta)}}{n}} + \frac{\log(\ell n)}{\sqrt{n}} \right] \\ & + c_1 \sum_{\substack{r \geq \ell \\ 2^r \leq \pi n}} \left[ M2^{r(\theta-1/2)} \sqrt{\frac{\log_+(n2^{r(1-\theta)}/M)}{n}} + \frac{\log((r+1)n)}{\sqrt{n}} \right]. \end{aligned}$$

We conclude by applying Lemma 30 in Appendix B. If  $\theta \in (1/2, 1)$ , we rather use (51) and (53) to get

$$\begin{aligned} A & \leq c_2 \left[ \sqrt{\frac{M^{1/(1-\theta)}}{n}} + \frac{\log(\ell n)}{\sqrt{n}} \right] \\ & + c_2 \sum_{\substack{r \geq \ell \\ 2^r \leq \pi n}} \left[ M2^{r(\theta-1/2)} \sqrt{\frac{\log_+(n2^{-r})}{n}} + \frac{\log((r+1)n)}{\sqrt{n}} \right], \end{aligned}$$

and we apply Lemma 30 again. For  $\theta = 1/2$ , we use (56) and (57), to get with  $K_{-1,r,\ell} = \widehat{\mathbb{Z}}_{-1,r}(\ell) \cap \overline{\mathbb{Z}}_{-1}$ ,

$$\sum_{\substack{r \geq \ell \\ 2^r \leq \varsigma_{-1} \pi n / \log n}} \mathbb{E} \left[ \widehat{T}_{-1,r,\ell}(K_{-1,r,\ell}) \right] \leq c_3 \left[ M(\log n) n^{-1/2} + (\log^2 n) n^{-1/2} \right].$$

When  $r$  is higher, we choose  $K_{-1,r,\ell} = \emptyset$  and note:

$$\begin{aligned} \sum_{\substack{r \geq \ell \\ 2^r \geq \varsigma_{-1} \pi n / \log n}} \mathbb{E} \left[ \widehat{T}_{-1,r,\ell}(K_{-1,r,\ell}) \right] &\leq 2^{1/2} \sum_{k \in \mathbb{Z}} |\beta_{-1,k}| \mathbb{P} \left[ \widehat{\sigma}_{-1,k}^2 \leq (\log n) / (\varsigma_{-1} n) \cap \mathcal{A} \right] \\ &\leq 2^{1/2} \sum_{\substack{k \in \mathbb{Z} \\ \sigma_{-1,k}^2 \leq c_4 \log n / n}} |\beta_{-1,k}|. \end{aligned}$$

Therefore, the integers  $k$  are such that  $f_{-1,k} \leq c_5 \log n / n$  and we conclude as in the proof of Lemma 20.  $\square$

The lemma below deals with the second term of (61).

**Lemma 27.** *Let  $p \in (0, \infty]$ ,  $\alpha \in ((1/p - 1)_+, \tau)$ ,  $\theta \in [0, 1) \cap [0, p]$ ,  $R > 0$ ,  $M \geq 1$  and  $f \in \mathcal{F}_{\alpha,p,\theta}(R, M)$ . Then,*

$$\begin{aligned} \sum_{j=0}^{c_0 \log n} \mathbb{E} \left[ \inf_{K_{j,\ell+j,\ell} \subset \widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] \\ \leq c_1 \left[ R^{1/(1+2\alpha)} M^{(\alpha-1/p+1)/((1-\theta)(2\alpha+1))} n^{-\alpha/(2\alpha+1)} \right. \\ \left. + M^{1/(2(1-\theta))} n^{-1/2} \right] + c_2 (\log^2 n) n^{-1/2}, \end{aligned} \tag{63}$$

where  $c_1$  only depends on  $\alpha, p, \theta, \psi$ , and where  $c_2$  only depends on  $\alpha, p, \theta, \psi, R$ .

*Proof of Lemma 27.* We first suppose that  $p \geq 1$ . We have,

$$\begin{aligned} \sum_{j=0}^{c_0 \log n} \mathbb{E} \left[ \inf_{K_{j,\ell+j,\ell} \subset \widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] \\ \leq \sum_{j=0}^{c_0 \log n} \min \left\{ \mathbb{E} \left[ \widehat{T}_{j,\ell+j,\ell}(\emptyset) \right], \mathbb{E} \left[ \widehat{T}_{j,\ell+j,\ell}(\widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \right\}. \end{aligned}$$



We use (49) and (51) to get

$$\begin{aligned}
& \sum_{j=0}^{c_0 \log n} \mathbb{E} \left[ \inf_{K_{j,\ell+j,\ell} \subset \widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] \\
& \leq c_1 \sum_{j=0}^{\infty} \min \left\{ RM^{(1-1/p)/(1-\theta)} 2^{-j\alpha}, \sqrt{\frac{M^{1/(1-\theta)} 2^j}{n}} \right\} + c_2 \frac{\log^2 n}{\sqrt{n}} \\
& \leq c_3 \left[ R^{1/(1+2\alpha)} M^{(\alpha+1-1/p)/((1-\theta)(2\alpha+1))} n^{-\alpha/(2\alpha+1)} + M^{1/(2(1-\theta))} n^{-1/2} \right] \\
& \quad + c_2 (\log^2 n) n^{-1/2}.
\end{aligned}$$

We now suppose  $p < 1$  and consider  $j_0 \geq 0$ . We deduce from (51),

$$\sum_{j=0}^{j_0} \mathbb{E} \left[ \widehat{T}_{j,\ell+j,\ell}(\widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \leq c_4 \sqrt{M^{1/(1-\theta)} \frac{2^{j_0}}{n}} + c_5 \frac{\log^2 n}{\sqrt{n}}.$$

We moreover set  $\eta_j^{1/p-1/2} = R^{-1/2} 2^{(j/2)(\alpha+1/2)} n^{-1/(2p)}$  and

$$K_{j,\ell+j,\ell} = \left\{ k \in \widehat{\mathbb{Z}}_{j,\ell+j}(\ell) \cap \overline{\mathbb{Z}}_j, |\beta_{j,k}| \geq \eta_j \right\}.$$

As  $f \in \mathcal{WB}_{p,\infty}^\alpha(R)$ ,

$$\begin{aligned}
|K_{j,\ell+j,\ell}| & \leq \eta_j^{-p} R^p 2^{-jp(\alpha+1/2-1/p)} \\
& \leq R^{2/(2/p-1)} 2^{-2j(\alpha+1-1/p)/(2/p-1)} n^{1/(2/p-1)}.
\end{aligned}$$

Moreover, (38) leads to

$$\begin{aligned}
\sum_{k \notin K_{j,\ell+j,\ell}} 2^{-j/2} |\beta_{j,k}| & \leq c_6 \eta_j^{1-p} R^p 2^{-jp(\alpha+1/2-1/(2p))} \\
& \leq c_6 R^{1/(2/p-1)} 2^{-j(\alpha+1-1/p)/(2/p-1)} n^{-(1/p-1)/(2/p-1)}.
\end{aligned}$$

Therefore, (50) gives

$$\mathbb{E} \left[ \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] \leq c_7 \left[ n^{-(1/p-1)/(2/p-1)} R^{1/(2/p-1)} 2^{-j(\alpha+1-1/p)/(2/p-1)} + \frac{\log n}{\sqrt{n}} \right]$$

and hence

$$\begin{aligned}
\sum_{j=j_0}^{c_0 \log n} \mathbb{E} \left[ \widehat{T}_{j,\ell+j,\ell}(K_{j,\ell+j,\ell}) \right] & \leq c_8 n^{-(1/p-1)/(2/p-1)} R^{1/(2/p-1)} 2^{-j_0(\alpha+1-1/p)/(2/p-1)} \\
& \quad + c_9 \frac{\log^2 n}{\sqrt{n}}.
\end{aligned}$$

We conclude by choosing  $j_0$  appropriately. □

It then remains to deal with the last term in (61), namely,

$$(64) \quad \sum_{j \geq 0} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right].$$

We distinguish different cases in the sections below.

**4.4. Proof of Theorem 2 when  $p \geq 1$  and  $\theta \neq \alpha/(2\alpha + 1 - 1/p)$ .** To make the proof more concise, we assume temporarily that  $\theta \neq 0$ . We deduce from (49), (53), (55) that for  $c_1$  large enough

$$\begin{aligned} & -\frac{\log^3 n}{\sqrt{n}} + \frac{1}{c_1} \sum_{j=0}^{\infty} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \min \left\{ \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\emptyset) \right], \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \right\} \\ & \leq \sum_{j=0}^{\infty} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \min \left\{ RM^{1-1/p} 2^{-j(\alpha-\theta/p+\theta)} 2^{(1-1/p)r\theta}, M 2^{r\theta+j(1-\theta)} \sqrt{\frac{2^{-r} \log_+(n 2^{-r})}{n}}, \right. \\ & \qquad \qquad \qquad \left. M 2^{r\theta+j(1-\theta)} \sqrt{\frac{2^{-r} \log_+^2(2^{(r-j)(1-\theta)}/M)}{n}} \right\}. \end{aligned}$$

Hence,

$$\begin{aligned} & -\frac{\log^3 n}{\sqrt{n}} + \frac{1}{c_1} \sum_{j=0}^{\infty} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \min \left\{ \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\emptyset) \right], \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \right\} \\ & \leq \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min \left\{ RM^{(1-1/p)/(1-\theta)} 2^{-j\alpha} 2^{(1-1/p)r\theta}, \right. \\ & \qquad \qquad \qquad M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2} \frac{\sqrt{\log_+(n 2^{-r-j} M^{-1/(1-\theta)})}}{\sqrt{n}}, \\ & \qquad \qquad \qquad \left. M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2} \frac{r}{\sqrt{n}} \right\}. \end{aligned}$$

This expression is not larger than  $B_1 + B_2$  where

$$\begin{aligned} B_1 &= RM^{(1-1/p)/(1-\theta)} \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)} \\ \pi R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \leq 2^{j(1+2\alpha)}}} 2^{-j\alpha} 2^{(1-1/p)r\theta} \\ B_2 &= M^{1/(2(1-\theta))} \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)} \\ \pi R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} > 2^{j(1+2\alpha)}}} \min \left\{ 2^{r(\theta-1/2)+j/2} \frac{\sqrt{\log_+(n 2^{-r-j} M^{-1/(1-\theta)})}}{\sqrt{n}}, \right. \\ & \qquad \qquad \qquad \left. 2^{r(\theta-1/2)+j/2} \frac{r}{\sqrt{n}} \right\}. \end{aligned}$$

Note that if  $2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}$  and if  $\pi R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \leq 2^{j(1+2\alpha)}$ , then

$$2^{j\alpha} \geq 2^{-1/2} RM^{(1-1/p)/(1-\theta)} 2^{r(1-\theta/p)}.$$

We deduce,

$$B_1 \leq c_2 R M^{(1-1/p)/(1-\theta)} \sum_{r \geq 1} 2^{(1-1/p)r\theta} \min \left\{ R^{-1} M^{-(1-1/p)/(1-\theta)} 2^{-r(1-\theta/p)}, \left[ R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \right]^{-\alpha/(2\alpha+1)} \right\}.$$

This sum is not larger than  $c_2[B_{1,1} + B_{1,2}]$ , where

$$B_{1,1} = \sum_{r \geq r_0} 2^{-r(1-\theta)}$$

$$B_{1,2} = \sum_{r < r_0} a_r$$

where

$$(65) \quad a_r = R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))} 2^{r \frac{2\alpha+1-1/p}{2\alpha+1} \left[ \theta - \frac{\alpha}{2\alpha+1-1/p} \right]} n^{-\alpha/(2\alpha+1)},$$

and where  $r_0 \geq 1$  is the smallest integer such that

$$2^{r_0(\alpha+1-\theta/p)} \geq R^{-1} M^{-(\alpha+1-1/p)/(1-\theta)} n^\alpha.$$

Note that

$$B_{1,1} \leq c_3 2^{-r_0(1-\theta)}$$

$$\leq c_3 R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)}.$$

As to  $B_{1,2}$ , we remark that  $B_{1,2} = 0$  if  $r_0 = 1$ . We may therefore assume that  $r_0 \geq 2$ , which implies the reverse inequality

$$2^{(r_0-1)(\alpha+1-\theta/p)} \leq R^{-1} M^{-(\alpha+1-1/p)/(1-\theta)} n^\alpha.$$

Note now that the exponent in the sum is negative when  $\theta < \alpha/(2\alpha+1-1/p)$ , and hence

$$B_{1,2} \leq c_4 a_1$$

$$\leq c_5 R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))} n^{-\alpha/(2\alpha+1)}.$$

When  $\theta > \alpha/(2\alpha+1-1/p)$ , the exponent is positive and

$$B_{1,2} \leq c_6 a_{r_0}$$

$$\leq c_7 R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)}.$$

By gathering these results, we obtain the desired bound on  $B_1$ .

We now deal with  $B_2$ . Lemma 30 in Appendix B gives

$$\begin{aligned}
B_2 &\leq c_8 \sum_{r \geq 1} \min \left\{ 2^{-r(1-\theta)}, \right. \\
&\quad M^{1/(2(1-\theta))} 2^{r(\theta-1/2)} \left( R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \right)^{1/2/(2\alpha+1)} \times \\
&\quad \frac{\sqrt{\log_+ \left( n 2^{-r} (R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)})^{-1/(2\alpha+1)} M^{-1/(1-\theta)} \right)}}{\sqrt{n}}, \\
&\quad \left. M^{1/(2(1-\theta))} 2^{r(\theta-1/2)} \left( R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \right)^{1/2/(2\alpha+1)} \frac{r}{\sqrt{n}} \right\} \\
&\leq c_9 \sum_{r \geq 1} \min \left\{ 2^{-r(1-\theta)}, a_r \sqrt{\log_+ C_r}, a_r r \right\} \\
&\leq c_9 [B_{2,1} + B_{2,2}]
\end{aligned}$$

where

$$(66) \quad C_r = R^{-2/(2\alpha+1)} M^{-2(\alpha+1-1/p)/((2\alpha+1)(1-\theta))} n^{2\alpha/(2\alpha+1)} 2^{-2r(\alpha+1-\theta/p)/(2\alpha+1)},$$

and where  $B_{2,1}$  and  $B_{2,2}$  are defined when  $\theta < \alpha/(2\alpha + 1 - 1/p)$  by

$$\begin{aligned}
B_{2,1} &= \sum_{r=1}^{\infty} r a_r \\
B_{2,2} &= 0.
\end{aligned}$$

When  $\theta > \alpha/(2\alpha + 1 - 1/p)$ , we rather set

$$B_{2,1} = \sum_{r \geq r_0} 2^{-r(1-\theta)}$$

and

$$B_{2,2} = \sum_{r < r_0} a_r \sqrt{\log_+ C_r}.$$

When  $\theta < \alpha/(2\alpha + 1 - 1/p)$ ,

$$\begin{aligned}
B_{2,1} &\leq c_{10} a_1 \\
&\leq c_{11} R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))} n^{-\alpha/(2\alpha+1)}.
\end{aligned}$$

When  $\theta > \alpha/(2\alpha + 1 - 1/p)$ , the sum  $B_{2,1}$  is equal to  $B_{1,1}$  and has already been bounded, see the above. Moreover, we deduce from Lemma 30,

$$\begin{aligned}
B_{2,2} &\leq c_{12} a_{r_0} \sqrt{\log_+(C_{r_0})} \\
&\leq c_{13} R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)}.
\end{aligned}$$

It then remains to group all these results to obtain the wished bound on  $B_2$ .

Let us now remark that the condition  $\theta \neq 0$  was made in the proof in order to use (53) when  $\theta > \alpha/(2\alpha + 1 - 1/p)$ . It is not necessary when  $\theta < \alpha/(2\alpha + 1 - 1/p)$ , which is obviously the case when  $\theta = 0$ . The proof remains therefore valid when  $\theta = 0$ .

**4.5. Proof of Theorem 2 when  $p > 1$  and  $\theta = \alpha/(2\alpha + 1 - 1/p)$ .** The previous proof can easily be adapted to deal with  $\theta = \alpha/(2\alpha + 1 - 1/p)$ . However, it would lead to additional logarithmic factors. A slight refinement can be obtained thanks to (56). We deduce from Lemma 24 that for  $c_1$  large enough,

$$\begin{aligned} & -\frac{\log^3 n}{\sqrt{n}} + \frac{1}{c_1} \sum_{j=0}^{\infty} \sum_{\substack{r \geq \ell + j + 1 \\ 2^r \leq \pi n \\ 2^r \leq \varsigma_j n / \log((j+2)n)}} \min \left\{ \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\emptyset) \right], \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \right\} \\ & \leq \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)} \\ 2^j \leq 2\varsigma_j(n/\log n) 2^{-r} M^{-1/(1-\theta)}}} \min \left\{ RM^{(1-1/p)/(1-\theta)} 2^{-j\alpha} 2^{(1-1/p)r\theta}, \frac{M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2}}{\sqrt{n}} \right\}. \end{aligned}$$

We may now bound the right-hand side of this inequality by  $B_1 + B_2$  where  $B_1$  has been defined in the preceding section by

$$B_1 = RM^{(1-1/p)/(1-\theta)} \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)} \\ R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} \leq 2^{j(1+2\alpha)}} 2^{-j\alpha} 2^{(1-1/p)r\theta}$$

and where

$$B_2 = M^{1/(2(1-\theta))} \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)} \\ R^2 M^{-(2/p-1)/(1-\theta)} n 2^{r(1-2\theta/p)} > 2^{j(1+2\alpha)}} \frac{2^{r(\theta-1/2)+j/2}}{\sqrt{n}}$$

corresponds to the definition of the preceding section, up to log factors. Similar computations then lead to

$$B_1 + B_2 \leq c_2 R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))} n^{-\alpha/(2\alpha+1)} \log n.$$

It then remains to deal with

$$\begin{aligned} & \sum_{j \geq 0} \sum_{\substack{r \geq \ell + j + 1 \\ 2^r \leq \pi n \\ 2^r > \varsigma_j n / \log((j+2)n)}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right] \\ & \leq \sum_{j \geq 0} \sum_{\substack{r \geq \ell + j + 1 \\ 2^r > \varsigma_j n / \log n}} \mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\emptyset) \right] \\ & \leq \sum_{j \geq 0} 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| \mathbb{P} \left[ \widehat{\sigma}_{j,k}^2 \leq (\log n)/(\varsigma_j n) \cap \mathcal{A} \right] \\ & \leq \sum_{j \geq 0} 2^{-j/2} \sum_{\substack{k \in \mathbb{Z} \\ \sigma_{j,k}^2 \leq c_3 \log n/n}} |\beta_{j,k}| \\ & \leq \sum_{j \geq 0} 2^{-j/2} \sum_{\substack{k \in \mathbb{Z} \\ f_{j,k} \leq c_4 \log n/n}} |\beta_{j,k}|. \end{aligned}$$

We conclude by applying Lemma 20.

**4.6. Proof of Theorem 2 when  $p = 1$  and  $\theta = \alpha/(2\alpha + 1 - 1/p)$ .** Since  $f \in \mathcal{B}_{1,\infty}^\alpha(R)$ ,

$$\begin{aligned} \sum_{r \geq \ell+j+1} \widehat{B}_{j,r,\ell}(\emptyset) &\leq 2^{-j/2} \sum_{r \geq \ell+j+1} \sum_{k \in \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} |\beta_{j,k}| \\ &\leq 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| \\ (67) \qquad \qquad \qquad &\leq R2^{-j\alpha}. \end{aligned}$$

Let us remark that the boundary  $\theta = \alpha/(2\alpha + 1 - 1/p)$  is  $\theta = 1/2$  here. We deduce from (56) and (57) that if  $2^r \leq \varsigma_j n / \log((j+2)n)$ ,

$$\mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|) 1_{\mathcal{A}} \right] \leq c_1 \left[ \frac{M2^{j/2}}{\sqrt{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

Therefore,

$$(68) \quad \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \varsigma_j n / \log((j+2)n)}} \mathbb{E} \left[ \widehat{\mathcal{E}}_{j,r,\ell}(|\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j|) 1_{\mathcal{A}} \right] \leq c_2 \left[ \frac{M2^{j/2} \log n}{\sqrt{n}} + \frac{\log^2 n}{n} \right].$$

We put (68) and (67) together and conclude the proof as in the preceding section.

**4.7. Proof of Theorem 2 when  $p < 1$  and  $\theta \neq \alpha/(2\alpha + 1 - 1/p)$ .** As in Section 4.4, we first suppose that  $\theta \neq 0$ . We define for  $r \geq \ell + j + 1$ ,

$$K_{j,r,\ell} = \left\{ k \in \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j, |\beta_{j,k}| \geq 2^{-(r-j)/2} n^{-1/2} \right\}.$$

As  $f \in \mathcal{WB}_{p,\infty}^\alpha(R)$  and (38), we get  $|K_{j,r,\ell}| \leq k_{j,r}$  and

$$(69) \quad \sum_{k \notin K_{j,r,\ell}} 2^{-j/2} |\beta_{j,k}| \leq c_1 k_{j,r} 2^{-r/2} n^{-1/2}$$

where  $k_{j,r} = R^p n^{p/2} 2^{-jp(\alpha+1-1/p)} 2^{rp/2}$ .

We consider  $r$  and  $j$  such that  $2^r \leq \pi n$ , and  $j \leq r - \ell - 1$ . We deduce from (52), (54), and the conditions on  $r, j$ ,

$$\begin{aligned} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) &\leq c_2 k_{j,r} 2^{-r/2} n^{-1/2} \log_+ \left( M n^{\theta-p/2} 2^{-rp/2} 2^{jp(\alpha+1-\theta/p)} / R^p \right) + c_3 (\log n) n^{-1/2} \\ \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) &\leq c_4 k_{j,r} 2^{-r/2} n^{-1/2} \log_+ \left( n^{-p/2} 2^{r(1-p/2)} 2^{jp(\alpha+1-1/p)} / R^p \right) + c_5 (\log n) n^{-1/2}. \end{aligned}$$

We derive from these two inequalities, from (53) and (55),

$$\begin{aligned} & - \frac{\log^3 n}{\sqrt{n}} + \frac{1}{c_6} \sum_{j \geq 0} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right] \\ & \leq \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, A'_{j,r}, B_{j,r}, B'_{j,r}\} \end{aligned}$$

where

$$\begin{aligned}
A_{j,r} &= R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-jp(\alpha-1/(2p)+1/2)} \\
&\quad \times \log_+ \left( M^{1-p/(2(1-\theta))} n^{\theta-p/2} 2^{-rp/2} 2^{jp(\alpha+1/2-\theta/p)} / R^p \right) \\
A'_{j,r} &= R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-jp(\alpha-1/(2p)+1/2)} \\
&\quad \times \log_+ \left( M^{(1-p/2)/(1-\theta)} n^{-p/2} 2^{r(1-p/2)} 2^{jp(\alpha+1/2)} / R^p \right) \\
B_{j,r} &= M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2} \sqrt{\frac{\log_+ (n 2^{-r-j} M^{-1/(1-\theta)})}{n}} \\
B'_{j,r} &= M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2} \frac{r}{\sqrt{n}}.
\end{aligned}$$

Let  $j_r$  be the smallest (possibly negative) integer such that

$$R^2 M^{-(2/p-1)/(1-\theta)} n 2^{-r(2\theta/p-1)} \leq 2^{j_r(2\alpha+1)}.$$

Lemma 30 entails when  $j_r \geq 0$ :

$$\begin{aligned}
&\sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, B_{j,r}\} \\
&\leq c_7 \left\{ R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-j_r p(\alpha-1/(2p)+1/2)} \right. \\
&\quad \times \log_+ \left( M^{1-p/(2(1-\theta))} n^{\theta-p/2} 2^{-rp/2} 2^{j_r p(\alpha+1/2-\theta/p)} / R^p \right) \\
&\quad \left. + M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j_r/2} \sqrt{\frac{\log_+ (n 2^{-r-j_r} M^{-1/(1-\theta)})}{n}} \right\} \\
&\leq c_8 a_r \log_+ C_r,
\end{aligned}$$

where  $a_r$  and  $C_r$  have been defined in Section 4.4 by (65) and (66). The same results holds true when  $j_r < 0$  since then

$$\begin{aligned}
&\sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, B_{j,r}\} \\
&\leq \sum_{j \geq 0} A_{j+j_r,r} \\
&\leq c_9 R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-j_r p(\alpha-1/(2p)+1/2)} \\
&\quad \times \log_+ \left( M^{1-p/(2(1-\theta))} n^{\theta-p/2} 2^{-rp/2} 2^{j_r p(\alpha+1/2-\theta/p)} / R^p \right) \\
&\leq c_{10} a_r \log_+ C_r.
\end{aligned}$$

Likewise, by supposing without loss of generality that  $j_r \geq 0$ ,

$$\begin{aligned}
& \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A'_{j,r}, B'_{j,r}\} \\
& \leq c_{11} \left\{ R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-j_r p(\alpha-1/(2p)+1/2)} \right. \\
& \quad \times \log_+ \left( M^{(1-p/2)/(1-\theta)} n^{-p/2} 2^{r(1-p/2)} 2^{j_r p(\alpha+1/2)} / R^p \right) \\
& \quad \left. + M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j_r/2} \frac{r}{\sqrt{n}} \right\} \\
& \leq c_{12} r a_r.
\end{aligned}$$

We also have,

$$\sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} B_{j,r} \leq c_{13} 2^{-r(1-\theta)}.$$

Therefore,

$$\sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, A'_{j,r}, B_{j,r}, B'_{j,r}\} \leq c_{14} \sum_{r=1}^{\infty} \min \left\{ a_r \log_+ C_r, r a_r, 2^{-r(1-\theta)} \right\}.$$

We conclude the proof as in the end of Section 4.4 (by noticing, as previously, that the reasoning does not need (52) and (53) when  $\theta < \alpha/(2\alpha + 1 - 1/p)$  and is therefore also valid when  $\theta = 0$ ).

**4.8. Proof of Theorem 2 when  $p < 1$  and  $\theta = \alpha/(2\alpha + 1 - 1/p)$ .** We define the same set  $K_{j,r,\ell}$  and the same real number  $k_{j,r}$  as in the preceding section. When  $2^r \leq \zeta_j n / \log((j+2)n)$ ,  $\widehat{\lambda}_{j,r} \leq |\widehat{\mathbb{Z}}'_{j,r}|$  and hence

$$\begin{aligned}
\widehat{\mathcal{E}}_{j,r,\ell}(|K_{j,r,\ell}|) &= \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \\
&\leq c_1 k_{j,r} \left[ \sqrt{\frac{2^{-r} \log_+ \left( |\widehat{\mathbb{Z}}'_{j,r}| / k_{j,r} \right)}{n}} + \frac{\log_+ \left( |\widehat{\mathbb{Z}}'_{j,r}| / k_{j,r} \right)}{n} \right] \\
&\leq c_2 k_{j,r} \sqrt{\frac{2^{-r} \log_+^2 \left( |\widehat{\mathbb{Z}}'_{j,r}| / k_{j,r} \right)}{n}}.
\end{aligned}$$

Jensen's inequality, (57) and (69) imply

$$\begin{aligned}
\mathbb{E} \left[ \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right] &= \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \\
&\leq c_3 k_{j,r} \sqrt{\frac{2^{-r} \log_+^2 \left( MR^{-p} 2^{r(\theta-p/2)} n^{-p/2} 2^{j_r p(\alpha+1-\theta/p)} \right)}{n}}.
\end{aligned}$$



Note that (56) and (57) entail

$$\mathbb{E} \left[ \widehat{T}_{j,r,\ell}(\widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j) \right] \leq c_4 \left[ \frac{M2^{r(\theta-1/2)+j(1-\theta)}}{\sqrt{n}} + \frac{\log((j+2)(r+1)n)}{\sqrt{n}} \right].$$

We deduce,

$$\begin{aligned} & -\frac{\log^3 n}{\sqrt{n}} + \frac{1}{c_5} \sum_{j \geq 0} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n \\ 2^r \leq \varsigma_j n / \log((j+2)n)}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right] \\ & \leq \sum_{r \geq 1} \sum_{2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}} \min\{A_{j,r}, B_{j,r}\}, \end{aligned}$$

where

$$\begin{aligned} A_{j,r} &= R^p M^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-r(1-p)/2} 2^{-jp(\alpha-1/(2p)+1/2)} \\ & \quad \times \log_+ \left( M^{(1-p/2)/(1-\theta)} R^{-p} 2^{r(\theta-p/2)} n^{-p/2} 2^{jp(\alpha+1/2)} \right) \\ B_{j,r} &= M^{1/(2(1-\theta))} 2^{r(\theta-1/2)+j/2} \sqrt{1/n}. \end{aligned}$$

By doing as in the preceding section,

$$\sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, B_{j,r}\} \leq c_6 n^{-\alpha/(2\alpha+1)} R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))}.$$

Moreover, this sum is equal to 0 if  $2^r > 2\pi n M^{-1/(1-\theta)}$ . Thus,

$$\begin{aligned} & \sum_{r \geq 1} \sum_{\substack{j \geq 0 \\ 2^j \leq \pi n 2^{1-r} M^{-1/(1-\theta)}}} \min\{A_{j,r}, B_{j,r}\} \\ & \leq c_7 (\log n) n^{-\alpha/(2\alpha+1)} R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((2\alpha+1)(1-\theta))}. \end{aligned}$$

Note finally that we may bound

$$\sum_{j \geq 0} \sum_{\substack{r \geq \ell+j+1 \\ 2^r \leq \pi n \\ 2^r > \varsigma_j n / \log((j+2)n)}} \mathbb{E} \left[ \inf_{K_{j,r,\ell} \subset \widehat{\mathbb{Z}}_{j,r}(\ell) \cap \overline{\mathbb{Z}}_j} \widehat{T}_{j,r,\ell}(K_{j,r,\ell}) \right]$$

from above as we did at the end of Section 4.5.

## REFERENCES

- [Aka17] Nathalie Akakpo. Multivariate intensity estimation via hyperbolic wavelet selection. *Journal of Multivariate Analysis*, 161:32–57, 2017.
- [Aut06] F Autin. Maxiset for density estimation on  $\mathbb{R}$ . *Mathematical Methods of Statistics*, 15(2):123–145, 2006.
- [BB09] Yannick Baraud and Lucien Birgé. Estimating the intensity of a random measure by histogram type estimators. *Probability Theory and Related Fields*, 143:239–284, 2009.

- [BB18] Yannick Baraud and Lucien Birgé. Rho-estimators revisited: General theory and applications. *The Annals of Statistics*, 46(6B):3767–3804, 2018.
- [Bir06a] Lucien Birgé. Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Annales de l’IHP Probabilités et statistiques*, 42(3):273–325, 2006.
- [Bir06b] Lucien Birgé. Statistical estimation with model selection. *Indagationes Mathematicae*, 17(4):497–537, 2006.
- [BLM00] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. A sharp concentration inequality with applications. *Random Structures & Algorithms*, 16(3):277–292, 2000.
- [CC05] Eric Chicken and T Tony Cai. Block thresholding for density estimation: local and global adaptivity. *Journal of Multivariate Analysis*, 95(1):76–106, 2005.
- [CDF92] Albert Cohen, Ingrid Daubechies, and J-C Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 45(5):485–560, 1992.
- [CL20] Kaikai Cao and Youming Liu. Uncompactly supported density estimation with  $l^1$  risk. *Communications on Pure & Applied Analysis*, 19(8):4007, 2020.
- [CVNRF15] René Erlin Castillo, Fabio Andrés Vallejo Narvaez, and Julio C. Ramos Fernández. Multiplication and composition operators on weak  $l_p$  spaces. *Bulletin of the Malaysian Mathematical Sciences Society*, 38(3):927–973, Jul 2015.
- [DJ96] Bernard Delyon and Anatoli Juditsky. On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, 3(3):215–228, 1996.
- [DJ97] Bernard Delyon and Anatoli Juditsky. On the computation of wavelet coefficients. *Journal of Approximation Theory*, 88(1):47–79, 1997.
- [DJKP96] David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996.
- [DJW18] John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.
- [DL93] Ronald A DeVore and George G Lorentz. *Constructive approximation*, volume 303. Springer Science & Business Media, 1993.
- [GK06] Evarist Giné and Vladimir Koltchinskii. Concentration inequalities and asymptotic results for ratio type empirical processes. *The Annals of Probability*, 34(3):1143–1216, 2006.
- [GL11] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.
- [GL14] Alexander Goldenshluger and Oleg Lepski. On adaptive minimax density estimation on  $\mathbb{R}^d$ . *Probability Theory and Related Fields*, 159(3):479–543, 2014.
- [HKP98] Peter Hall, Gérard Kerkyacharian, and Dominique Picard. Block threshold rules for curve estimation using kernel and wavelet methods. *The Annals of Statistics*, 26(3):922–942, 1998.
- [HKPT12] Wolfgang Härdle, Gérard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129. Springer Science & Business Media, 2012.
- [IK81] IA Ibragimov and RZ Khasminski. More on estimation of the density of a distribution. *Journal of Soviet Mathematics*, 108(194):198, 1981.

- [JLL04] Anatoli Juditsky and Sophie Lambert-Lacroix. On minimax density estimation on  $\mathbb{R}$ . *Bernoulli*, 10(2):187–220, 2004.
- [KPT96] Gérard Kerkycharian, Dominique Picard, and Karine Tribouley. Lp adaptive density estimation. *Bernoulli*, 2(3):229–247, 1996.
- [Lep13] Oleg Lepski. Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *The Annals of Statistics*, 41(2):1005–1034, 2013.
- [Lep15] Oleg Lepski. Adaptive estimation over anisotropic functional classes via oracle approach. *The Annals of Statistics*, 43(3):1178–1242, 2015.
- [LW19] Oleg Lepski and Thomas Willer. Oracle inequalities and adaptive estimation in the convolution structure density model. *The Annals of Statistics*, 47(1):233–287, 2019.
- [Mas07] Pascal Massart. *Concentration inequalities and model selection: Ecole d’Eté de Probabilités de Saint-Flour XXXIII-2003*. Springer, 2007.
- [RBRTM11] Patricia Reynaud-Bouret, Vincent Rivoirard, and Christine Tuleau-Malot. Adaptive density estimation: a curse of support? *Journal of Statistical Planning and Inference*, 141(1):115–139, 2011.
- [Sar21a] Mathieu Sart. Estimating a density, a hazard rate, and a transition intensity via the  $\rho$ -estimation method. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 57(1):195–249, 2021.
- [Sar21b] Mathieu Sart. Minimax bounds for besov classes in density estimation. *Electronic Journal of Statistics*, 15(1):3184–3216, 2021.
- [vdV13] Wellner van der Vaart. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 2013.
- [Yu97] Bin Yu. Assouad, fano, and le cam. *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435, 1997.

## A. A PROBABILITY INEQUALITY.

The following proposition is based on standard results in empirical processes. It uses the notion of VC subgraph classes, see [vdV13] for their definitions and properties.

**Proposition 6.** *Let  $\mathcal{F}$  be an at most countable VC subgraph class of functions  $f$  defined on  $\mathbb{R}$  and with values in  $[-1, 1]$ . We suppose that the VC dimension of  $\mathcal{F}$  is not larger than  $d \geq 1$ . We consider  $\bar{f}$  such that  $|f(x)| \leq \bar{f}(x) \leq 1$  for all  $x \in \mathbb{R}$ . We further consider a map  $\sigma^2(\cdot)$  defined on  $\mathcal{F} \cup \{\bar{f}\}$  and satisfying  $\sigma^2(f) \geq \mathbb{E}[f^2(X)]$  for all  $f \in \mathcal{F} \cup \{\bar{f}\}$ .*

*Then, there exists for all  $\xi > 0$  an event of probability  $1 - e^{-\xi}$  on which: for all  $f \in \mathcal{F}$ ,*

$$(70) \quad \frac{1}{n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| \leq C \min\{R_1(f), R_2(f)\},$$

where

$$(71) \quad R_1(f) = \sqrt{\frac{d}{n} \sigma^2(f) \log_+(\sigma^2(\bar{f})/\sigma^2(f))} + \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d) + \sqrt{\sigma^2(f) \frac{\xi + \log n}{n}} + \frac{\xi + \log n}{n},$$

$$(72) \quad R_2(f) = \sqrt{\frac{1}{n} \sigma^2(f) \log_+(|\mathcal{F}|)} + \frac{1}{n} \log_+(|\mathcal{F}|) + \sqrt{\sigma^2(f) \frac{\xi + \log n}{n}} + \frac{\xi + \log n}{n},$$

and where  $C$  is a numerical value. In the above inequality,  $R_2(f) = +\infty$  if  $\mathcal{F}$  is infinite.

*Proof of Proposition 6.* We need the two following lemmas (see Theorem 3.1 of [GK06] for the first, and [Mas07] for the second).

**Lemma 28.** *Let  $\mathcal{F}$  be an at most countable VC subgraph class of functions  $f$  defined on  $\mathbb{R}$  and with values in  $[-1, 1]$ . We suppose that the VC dimension of  $\mathcal{F}$  is not larger than  $d \geq 1$ . We consider a map  $\bar{f}$  such that  $|f(x)| \leq \bar{f}(x) \leq 1$  for all  $x \in \mathbb{R}$ . Let  $\sigma^2(\bar{f})$  be a real number such that  $\sigma^2(\bar{f}) \geq \mathbb{E}[\bar{f}^2(X)]$ . Let moreover  $\sigma^2 > 0$  be such that  $\sup_{f \in \mathcal{F}} \mathbb{E}[f^2(X)] \leq \sigma^2$ . Then,*

$$\mathbb{E} \left[ \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| \right] \leq C \left[ \sqrt{\frac{d}{n} \sigma^2 \log_+(\sigma^2(\bar{f})/\sigma^2)} + \frac{d}{n} \log_+(\sigma^2(\bar{f})/\sigma^2) \right],$$

where  $C$  is a numerical value.

**Lemma 29.** *Let  $\mathcal{F}$  be a finite class of functions  $f$  defined on  $\mathbb{R}$  and with values in  $[-1, 1]$ . Let  $\sigma^2 > 0$  such that  $\sup_{f \in \mathcal{F}} \mathbb{E}[f^2(X)] \leq \sigma^2$ . Then,*

$$\mathbb{E} \left[ \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| \right] \leq C \left[ \sqrt{\frac{1}{n} \sigma^2 \log_+(|\mathcal{F}|)} + \frac{1}{n} \log_+(|\mathcal{F}|) \right],$$

where  $C$  is a numerical value.

We only prove the inequality with  $R_1(f)$ , the proof with  $R_2(f)$  is similar (use Lemma 29 instead of Lemma 28 and set  $j_0$  below as the smallest integer such that  $2^{j_0-1} \geq n$ ). We may assume without loss of generality that  $d \leq n$  and  $\sigma^2(f) \in (0, 1]$  for all  $f \in \mathcal{F}$ . Let  $j_0$  be the smallest integer such that  $2^{j_0-1} \geq n/d$ . We define for  $j \in [1, j_0 - 1]$ ,

$$\mathcal{F}_j = \{f \in \mathcal{F}, 2^{-j} < \sigma^2(f) \leq 2^{-j+1}\},$$

and set

$$\mathcal{F}_{j_0} = \{f \in \mathcal{F}, \sigma^2(f) \leq 2^{-j_0+1}\}.$$

Let then for all  $j \in [1, j_0]$ ,

$$\begin{aligned} \sigma_j^2 &= 2^{-j+1} \\ Z_j &= \sup_{f \in \mathcal{F}_j} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right|. \end{aligned}$$

We consider some  $\xi_0 > 0$  and deduce from Talagrand's inequality (see the second equation on page 170 of [Mas07]), and from the elementary inequalities  $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ ,  $2\sqrt{xy} \leq x+y$ , that on an event  $\Omega_j(\xi_0)$  of probability  $1 - e^{-\xi_0}$ ,

$$Z_j \leq c \left[ \mathbb{E}[Z_j] + \sqrt{(\sigma_j^2/n)\xi_0} + \xi_0/n \right],$$

where  $c$  is universal. We now set for all  $\sigma^2 > 0$ ,

$$A(\sigma^2) = \sqrt{\frac{d}{n}\sigma^2 \log_+(\sigma^2(\bar{f})/\sigma^2)} + \frac{d}{n} \log_+(\sigma^2(\bar{f})/\sigma^2).$$

Any  $\sigma^2(f)$  belongs to  $[(1/2)\sigma_j^2, \sigma_j^2]$  when  $f \in \mathcal{F}_j$  with  $j \leq j_0 - 1$ . We deduce from Lemma 28 that for all such  $j$ ,

$$\begin{aligned} \mathbb{E}[Z_j] &\leq CA(\sigma_j^2) \\ &\leq C \left[ \sqrt{2} \sqrt{\frac{d}{n}\sigma^2(f) \log_+(\sigma^2(\bar{f})/\sigma^2(f))} + \frac{d}{n} \log_+(2n\sigma^2(\bar{f})/d) \right]. \end{aligned}$$

We deduce that on  $\Omega_j(\xi_0)$ : for all  $f \in \mathcal{F}_j$ ,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| &\leq C' \left[ \sqrt{\frac{d}{n}\sigma^2(f) \log_+(\sigma^2(\bar{f})/\sigma^2(f))} \right. \\ &\quad \left. + \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d) + \sqrt{(\sigma^2(f)/n)\xi_0} + \xi_0/n \right], \end{aligned}$$

where  $C'$  is universal. When  $j = j_0$ , we rather have

$$\begin{aligned} \mathbb{E}[Z_j] &\leq CA(d/n) \\ &\leq C \left[ \frac{d}{n} \sqrt{\log_+(n\sigma^2(\bar{f})/d)} + \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d) \right] \\ &\leq 2C \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d). \end{aligned}$$

Hence, on  $\Omega_{j_0}(\xi_0)$ : for all  $f \in \mathcal{F}_{j_0}$ ,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]) \right| &\leq C'' \left[ \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d) + \sqrt{(d/n^2)\xi_0} + \xi_0/n \right] \\ &\leq C''' \left[ \frac{d}{n} \log_+(n\sigma^2(\bar{f})/d) + \xi_0/n \right], \end{aligned}$$

where  $C'''$  is universal. We now set  $\xi_0 = \xi + \log j_0 < \xi + 1.4 + \log n$  and conclude using a union bound.  $\square$

## B. AN ELEMENTARY LEMMA

**Lemma 30.** For all  $a_1, a_2, b > 0$ ,  $k \in [0, 1]$ , and  $r_0 \geq 1$ ,

$$\sum_{r=1}^{r_0} 2^{ra_1} \log_+^k (b2^{-ra_2}) \leq c2^{r_0 a_1} \log_+^k (b2^{-r_0 a_2}),$$

and

$$\sum_{r=r_0}^{\infty} 2^{-ra_1} \log_+^k (b2^{ra_2}) \leq c2^{-r_0 a_1} \log_+^k (b2^{r_0 a_2}),$$

where  $c$  only depends on  $a_1, a_2, k$ .

*Proof of Lemma 30.* We only show the first inequality. The proof of the second inequality follows the same line. When  $k = 1$ , we write

$$\begin{aligned} & \sum_{r=1}^{r_0} 2^{ra_1} \log_+ (b2^{-ra_2}) \\ (73) \quad & \leq c_1 \left[ \sum_{\substack{1 \leq r \leq r_0 \\ 2^{(r+1)a_2} \geq be^{-a_2/(a_1+1)}}} 2^{ra_1} + \sum_{\substack{1 \leq r \leq r_0 \\ 2^{(r+1)a_2} < be^{-a_2/(a_1+1)}}} 2^{ra_1} \log (b2^{-ra_2}) \right] \\ & \leq c_2 \left[ 2^{r_0 a_1} + \sum_{\substack{1 \leq r \leq r_0 \\ 2^{(r+1)a_2} < be^{-a_2/(a_1+1)}}} 2^{-r} f(2^{-r}) \right], \end{aligned}$$

where  $f$  denotes the map defined for  $x > 0$  by  $f(x) = x^{-a_1-1} \log(bx^{a_2})$ . Let  $r_1$  be the largest integer such that  $2^{(r_1+1)a_2} < be^{-a_2/(a_1+1)}$ . The proof when  $k = 1$  is complete if  $r_1 \leq 0$  and we assume from now on that  $r_1 \geq 1$ . We set  $r_2 = \min\{r_0, r_1\}$ . Since  $f$  is decreasing and non negative when  $bx^{a_2} \geq e^{a_2/(a_1+1)}$ , we get

$$\begin{aligned} \sum_{\substack{1 \leq r \leq r_0 \\ 2^{(r+1)a_2} < be^{-a_2/(a_1+1)}}} 2^{-r} f(2^{-r}) & \leq 2 \sum_{r=1}^{r_2} \int_{2^{-r-1}}^{2^{-r}} f(x) dx \\ & \leq 2 \int_{2^{-r_2-1}}^{\infty} f(x) dx \\ & \leq c_3 2^{r_2 a_1} \log_+ (b2^{-a_2 r_2}) \\ & \leq c_4 2^{r_0 a_1} \log_+ (b2^{-a_2 r_0}). \end{aligned}$$

By putting this result in (73), we get the lemma when  $k = 1$ . The proof when  $k \neq 1$  then follows from Hölder inequality:

$$\begin{aligned}
\sum_{r=1}^{r_0} 2^{ra_1} \log_+^k (b2^{-ra_2}) &= \sum_{r=1}^{r_0} 2^{rka_1} \log_+^k (b2^{-ra_2}) 2^{r(1-k)a_1} \\
&\leq \left[ \sum_{r=1}^{r_0} 2^{ra_1} \log_+ (b2^{-ra_2}) \right]^k \left[ \sum_{r=1}^{r_0} 2^{ra_1} \right]^{1-k} \\
&\leq c_5 2^{r_0 k a_1} \log_+^k (b2^{-r_0 a_2}) \times 2^{r_0(1-k)a_1} \\
&\leq c_5 2^{r_0 a_1} \log_+^k (b2^{-r_0 a_2}).
\end{aligned}$$

□

### C. PROOF OF LEMMA 25

*Proof of Lemma 25.* Suppose that  $p < \infty$  and consider an arbitrary real number  $q$  in  $(\max\{1, p\}, p(\alpha + 1))$ . We use (38) to get

$$\|\beta_{j,\cdot}\|_q^q \leq \frac{q}{q-p} \|\beta_{j,\cdot}\|_\infty^{q-p} \cdot \|\beta_{j,\cdot}\|_{p,\infty}^p.$$

Moreover, as  $\psi$  is bounded,  $\|\beta_{j,\cdot}\|_\infty \leq c_1 2^{j/2}$ . Therefore, using  $f \in \mathcal{WB}_{p,\infty}^\alpha(R)$ , we obtain for all  $j \geq 0$

$$\|\beta_{j,\cdot}\|_q^q \leq c_2 2^{j(1-p(\alpha+1)+q/2)},$$

where  $c_2$  only depends on  $\psi, q, R$ . We deduce,

$$(74) \quad \sum_{j=0}^{\infty} 2^{j(q/2-1)} \sum_{k \in \mathbb{Z}} |\beta_{j,k}|^q < \infty.$$

Suppose that  $p = \infty$  and take  $q > 1$ . Then, using that  $|\beta_{j,k}| \leq c_3 2^{j/2} f_{j,k}$ ,  $\sum_{k \in \mathbb{Z}} f_{j,k} \leq c_4$  (Lemma 4), and  $\|\beta_{j,\cdot}\|_\infty \leq c_5 2^{-j(\alpha+1/2)}$ ,

$$\|\beta_{j,\cdot}\|_q^q \leq c_6 2^{j(1+\alpha-(1/2+\alpha)q)},$$

and the sum (74) is finite.

Note that  $|\alpha_k| \leq 1$  and  $\sum_{k \in \mathbb{Z}} |\alpha_k| \leq 1$  and hence  $\|\beta_{-1,\cdot}\|_q^q \leq 1$ . Now, for all  $j \geq 0$ ,

$$\begin{aligned}
\left\| \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k} \right\|_q^q &= 2^{j(q/2-1)} \int \left| \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}(t-k) \right|^q dt \\
&\leq 2^{j(q/2-1)} \int \left| \sum_{k \in \mathbb{Z}} |\beta_{j,k}| |\bar{\psi}(t-k)|^{1/q} |\bar{\psi}(t-k)|^{1-1/q} \right|^q dt \\
&\leq 2^{j(q/2-1)} \int \left( \sum_{k \in \mathbb{Z}} |\beta_{j,k}|^q |\bar{\psi}(t-k)| \right) \left( \sum_{k \in \mathbb{Z}} |\bar{\psi}(t-k)| \right)^{q-1} dt \\
&\leq 2^{j(q/2-1)} \int \left( \sum_{k \in \mathbb{Z}} |\beta_{j,k}|^q |\bar{\psi}(t-k)| \right) \left\| \sum_{k \in \mathbb{Z}} |\bar{\psi}(t-k)| \right\|_\infty^{q-1} dt
\end{aligned}$$

$$\leq C2^{j(q/2-1)} \sum_{k \in \mathbb{Z}} |\beta_{j,k}|^q.$$

The same result holds true when  $j = -1$ . Let

$$\pi_J(f) = \sum_{j=-1}^J \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k}.$$

The above ensures that  $(\pi_J(f))_{J \geq 1}$  is a Cauchy sequence in  $(\mathbb{L}^q, \|\cdot\|_q)$  and converges therefore to a map that must be  $f$ . In particular,

$$\|f\|_q \leq \sum_{j=-1}^{\infty} \left\| \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k} \right\|_q,$$

hence the result.  $\square$

#### D. PROOF OF LEMMA 17

*Proof of Lemma 17.* For all  $u > 0$ ,

$$(75) \quad \mathbb{P} \left[ \min_{1 \leq i \leq n-1} (X_{(i+1)} - X_{(i)}) \leq u \right] \leq n \sup_{1 \leq i \leq n-1} \mathbb{P} [X_{(i+1)} - X_{(i)} \leq u].$$

The density of  $X_{(i+1)} - X_{(i)}$  is given for  $x \geq 0$  by

$$\varphi_i(x) = \frac{n!}{(i-1)!(n-i-1)!} \int_{\mathbb{R}} F(t)^{i-1} (1-F(t+x))^{n-i-1} f(t) f(t+x) dt,$$

where  $F$  denotes the cumulative distribution function of  $X$ . We have for all  $u > 0$ ,

$$\begin{aligned} & \mathbb{P} [X_{(i+1)} - X_{(i)} \leq u] \\ &= \frac{n!}{(i-1)!(n-i-1)!} \int_{\mathbb{R}} F(t)^{i-1} f(t) \left[ \int_0^u (1-F(t+x))^{n-i-1} f(t+x) dx \right] dt \\ &= \frac{n!}{(i-1)!(n-i-1)!} \mathbb{E} [F(X_1)^{i-1} (1-F(X_2))^{n-i-1} 1_{X_1 \leq X_2 \leq X_1+u}] \\ &\leq \frac{n!}{(i-1)!(n-i-1)!} \mathbb{E} [F(X_2)^{i-1} (1-F(X_2))^{n-i-1} 1_{X_1 \leq X_2 \leq X_1+u}] \\ &\leq \frac{n!}{(i-1)!(n-i-1)!} \mathbb{E} \left[ F(X_2)^{i-1} (1-F(X_2))^{n-i-1} \int_{X_2-u}^{X_2} f(t) dt \right]. \end{aligned}$$

By using Hölder inequality,

$$\int_{X_2-u}^{X_2} f(t) dt \leq \|f\|_q u^{1-1/q}.$$

Moreover,  $F(X_2)$  obeys to a uniform distribution on  $[0, 1]$  as  $F$  is continuous and hence

$$\mathbb{E} [F(X_2)^{i-1} (1-F(X_2))^{n-i-1}] = \frac{(i-1)!(n-i-1)!}{n!}.$$

We use (75) to get

$$\mathbb{P} \left[ \min_{1 \leq i \leq n-1} (X_{(i+1)} - X_{(i)}) \leq u \right] \leq n \|f\|_q u^{1-1/q}.$$



We conclude by setting  $u$  such that  $n\|f\|_q u^{1-1/q} = \xi/n$  and by using

$$2^{\hat{J}} \leq \max \left\{ 1, 4L_\psi \min_{1 \leq i \leq n-1} (X_{(i+1)} - X_{(i)})^{-1} \right\}.$$

□

### E. PROOF OF PROPOSITION 3

We need the following version of Assouad's famous lemma. We do not prove this version here as it can easily be derived from mild modifications of the classical proof (see [Bir06b] for instance). We also refer to [Yu97, DJW18] for an almost identical version.

**Lemma 31.** *Let  $D \geq 1$  and  $(f_\delta)_{\delta \in \mathcal{D}}$  be a family of densities indexed by  $\mathcal{D} = \{0, 1\}^D$ . Let  $\kappa_1, \dots, \kappa_D$  be  $D$  positive numbers and  $\Delta$  be the distance defined for all  $\delta, \delta' \in \mathcal{D}$  by*

$$\Delta(\delta, \delta') = \sum_{j=1}^D \kappa_j |\delta_j - \delta'_j|.$$

*For all  $j \in \{1, \dots, D\}$ , and for all  $\delta, \delta' \in \mathcal{D}$  satisfying  $\delta_j \neq \delta'_j$  and  $\delta_k = \delta'_k$  for all  $k \neq j$ , we suppose that*

$$(76) \quad h^2(f_\delta, f_{\delta'}) \leq 1/(2n),$$

*where  $h$  is the Hellinger distance defined by*

$$h^2(f_\delta, f_{\delta'}) = \frac{1}{2} \int \left( \sqrt{f_\delta(x)} - \sqrt{f_{\delta'}(x)} \right)^2 dx.$$

*Then, whatever the estimator  $\tilde{\delta} \in \mathcal{D}$ ,*

$$\sup_{\delta \in \mathcal{D}} \mathbb{E} \left[ \Delta(\tilde{\delta}, \delta) \right] \geq c \sum_{j=1}^D \kappa_j,$$

*where  $c > 0$  is universal.*

Since Proposition 3 is written for  $R$  and  $M$  large enough, we only need to build a subset  $\mathcal{F}$  of  $\mathfrak{B}_{p,\infty}^\alpha(C_1 R)$  whose densities  $f$  satisfy

$$\sup_{|x| \geq 1} |x|^{1/\theta} f(x) \leq C_3 M^{1/\theta}.$$

When  $\mu_n = 1$ , we must also have  $f \in \mathcal{T}_\theta(C_2 M)$ . Here,  $C_1, C_2, C_3 > 0$  are positive terms only depending on  $\alpha, p, \theta$  and the wavelet basis.

We now introduce some notations. In the following,  $\zeta \in \mathfrak{B}_{p,\infty}^\alpha(R)$  is a compactly supported density on  $(-1, 0)$  bounded by 1. Likewise,  $\varphi \in \mathfrak{B}_{p,\infty}^\alpha(\rho)$  stands for a compactly supported density on  $(1/4, 2)$ , bounded from above by 1 and such that  $\inf_{x \in [1/2, 1]} \varphi(x) \geq 1/4$ . The symbol  $\rho$  refers to any number that ensures the existence of  $\varphi$ . Finally,  $L$  is an arbitrary number larger than 1 such that  $\text{supp } \bar{\psi}$  and  $\text{supp } \bar{\phi}$  are both included in  $(-2^L, 2^L)$ .

**E.1. Case  $\mu_n = 1$ .** Let  $j \geq -1$  and  $j_0 \geq 0$  be two integers such that  $2^{j_0+j-L} \geq 12$ . Let  $\underline{k} \geq 1$  be the smallest integer satisfying  $1 + 2\underline{k} \geq 2^{j_0+j-L-1}$ , and  $\bar{k} \geq 1$  be the largest integer such that  $4\bar{k} + 2\underline{k} + 1 \leq 2^{j_0+j-L}$ . We endow  $\mathcal{D} = \{0, 1\}^{\bar{k}}$  with the Hamming distance  $\Delta$  defined for  $\delta, \delta' \in \{0, 1\}^{\bar{k}}$  by

$$\Delta(\delta, \delta') = \sum_{k=1}^{\bar{k}} |\delta_k - \delta'_k|.$$

We consider  $b > 0$  and define  $h_\delta(x)$  for  $\delta \in \mathcal{D}$ ,  $x \in \mathbb{R}$  by

$$h_\delta(x) = b \left[ \sum_{k=1}^{\bar{k}} \delta_k \bar{\psi}_{j, 2^{L+1}(k+\underline{k})}(x) + \sum_{k=1}^{\bar{k}} (1 - \delta_k) \bar{\psi}_{j, 2^{L+1}(k+\underline{k}+\bar{k})}(x) \right].$$

We also set

$$g(x) = M^{1/\theta} 2^{-j_0/\theta} \varphi(2^{-j_0} x)$$

and

$$q = M^{1/\theta} 2^{j_0(1-1/\theta)} + b\bar{k} 1_{j=-1}.$$

We then define for  $\delta \in \mathcal{D}$  and  $x \in \mathbb{R}$ ,

$$f_\delta(x) = (1 - q)\zeta(x) + g(x) + h_\delta(x).$$

We show after the present proof:

**Lemma 32.** *Suppose that the parameters  $j \geq -1$ ,  $j_0 \geq 0$ ,  $b > 0$  are chosen in such a way that  $2^{j_0+j-L} \geq 12$ , and such that*

$$(77) \quad b 2^{j_0/p} 2^{j(\alpha+1/2)} 1_{j \geq 0} \leq a_1 R$$

$$(78) \quad M^{1/\theta} 2^{j_0(1/p-1/\theta-\alpha)} \leq a_2 R$$

$$(79) \quad b 2^{j/2} 2^{j_0/\theta} \leq a_3 M^{1/\theta}$$

$$(80) \quad 2^{j_0/\theta} b^2 \leq a_4 M^{1/\theta} n^{-1}$$

$$(81) \quad q \leq 1.$$

*In the above conditions,  $a_1, a_2, a_3, a_4$  are suitable terms depending only on  $p, \theta, \varphi$  and the wavelet basis.*

*Then,  $\mathcal{F} = \{f_\delta, \delta \in \mathcal{D}\}$  is a collection of densities included in  $\mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)$ . Any density  $f_\delta$  satisfies*

$$\sup_{|x| \geq 1} |x|^{1/\theta} f_\delta(x) \leq C_3 M^{1/\theta}.$$

*Moreover,*

$$(82) \quad \inf_{\tilde{f}} \sup_{f \in \mathcal{F}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \geq c b 2^{j_0+j/2},$$

*where  $c$  only depends on  $p, \theta, \varphi$  and the wavelet basis.*

We first prove when  $\theta < p$

$$(83) \quad \inf_{\tilde{f}} \sup_{f \in \mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \\ \geq c R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)}.$$

For this, we define  $j_0$  as the smallest integer such that

$$2^{j_0 \left[ \frac{1}{p} - \frac{\alpha+1}{\theta} \right]} \leq \frac{R}{M^{\frac{\alpha+1}{\theta}} n^\alpha}.$$

Note that  $j_0$  tends to infinity when  $n$  grows up. Therefore the condition  $2^{j_0+j-L} \geq 12$  is satisfied when  $n$  is large enough. We define  $c_1$  and  $c_2$  small enough so that  $c_1^{\alpha+1} c_2 \leq a_1$ ,  $c_1 c_2^2 \leq a_4$ ,  $c_1 c_2 \leq a_3$ . Since  $\theta < p$ , we may take  $n$  large enough and define the largest integer  $j \geq 0$  such that

$$2^j \leq c_1 M^{1/\theta} n 2^{-j_0/\theta}.$$

We define

$$b = c_2 2^{j/2} n^{-1}.$$

We may check that (77), (78), (79) and (80) are fulfilled. Note that  $q$  tends to 0 when  $n$  grows up and is therefore smaller than 1 when  $n$  is large enough. We then deduce (83) from (82).

When  $\theta = p$ , the above reasoning works when  $R^p \geq M/c_1^{\alpha p}$ . This condition ensures the existence of  $j \geq 0$ . It is worth mentioning that the lower bound

$$R^{(1-\theta)/(1+\alpha-\theta/p)} M^{(\alpha+1-1/p)/(\alpha+1-\theta/p)} n^{-\alpha(1-\theta)/(\alpha+1-\theta/p)}$$

follows from

$$(84) \quad \inf_{\tilde{f} \in \mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)} \sup_{f \in \mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \geq c M 1_{\gamma=1-\theta} n^{-\gamma}$$

when  $\theta = p$  and  $R^p < M/c_1^{\alpha p}$ . The proof of which is given below.

We now show (84). We set  $j = -1$ , consider  $c_1, c_2$  such that  $c_2 \leq a_4$  and  $\sqrt{c_1 c_2/2} \leq a_3$ , define  $j_0$  as the largest integer such that

$$2^{j_0/\theta} \leq c_1 M^{1/\theta} n,$$

and  $b^2$  such that

$$2^{j_0/\theta} b^2 = c_2 M^{1/\theta} n^{-1}.$$

We may check that the conditions of the lemma are met hence the result.

We finally prove

$$\begin{aligned} & \inf_{\tilde{f} \in \mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)} \sup_{f \in \mathcal{B}_{p,\infty}^\alpha(C_1 R) \cap \mathcal{T}_\theta(C_2 M)} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \\ & \geq c R^{1/(2\alpha+1)} M^{(\alpha+1-1/p)/((1-\theta)(2\alpha+1))} n^{-\alpha/(2\alpha+1)}. \end{aligned}$$

We define  $j_0 \geq 0$  such that  $2^{j_0-1} \leq c_1^{\theta/(1-\theta)} M^{1/(1-\theta)} \leq 2^{j_0}$  where  $c_1 = 2 \cdot 15^{-1/\theta}$ . We define  $j$  as the smallest integer such that

$$M^{-1/\theta} R^2 2^{-j(2\alpha+1)} \leq n^{-1} 2^{j_0(2/p-1/\theta)}.$$

Since  $j$  tends to  $+\infty$  when  $n$  grows up, the condition  $2^{j_0+j-L} \geq 12$  is fulfilled when  $n$  is large enough. Besides,  $q \leq 1/2$  and (78) holds true if  $R_0, M_0$  are large enough. Moreover, we set

$$b = c_2 R 2^{-j_0/p} 2^{-j(\alpha+1/2)}$$

where  $c_2 = \min \{a_1, \sqrt{a_4}\}$ . We conclude by applying Lemma 32 as above.  $\square$

*Proof of Lemma 32.* We begin by showing two lemmas relating to the properties of  $h_\delta$  and  $g$ .

**Lemma 33.** For all  $\delta \in \mathcal{D}$ , the map  $h_\delta$  is compactly supported on  $[2^{j_0-1}, 2^{j_0}]$  and belongs to

$$\mathcal{B}_{p,\infty}^\alpha \left( b\bar{k}^{1/p} 2^{j(\alpha+1/2-1/p)} 1_{j \geq 0} \right).$$

Moreover,  $|h_\delta| \in \mathcal{T}_\theta (c_1 b^\theta 2^{j\theta/2} 2^{j_0})$ ,

$$(85) \quad \int h_\delta(x) dx = b\bar{k} 1_{j=-1}$$

and

$$\|h_\delta\|_\infty \leq c_2 b 2^{j/2}.$$

We therefore also have

$$\sup_{|x| \geq 1} |x|^{1/\theta} |h_\delta(x)| \leq c_3 b 2^{j/2} 2^{j_0/\theta}.$$

Here,  $c_1, c_2, c_3$  only depend on the wavelet basis and  $\theta$ .

*Sketch of proof of Lemma 33.* We remark that the supports of  $\bar{\psi}_{j,2^{L+1}k}$  and  $\bar{\psi}_{j,2^{L+1}k'}$  are disjoint when  $k \neq k'$ . Therefore,

$$\|h_\delta\|_\infty \leq b \sup_{k \in \mathbb{Z}} \|\bar{\psi}_{j,2^{L+1}k}\|_\infty \leq b \max\{\|\bar{\phi}\|_\infty, 2^{j/2} \|\bar{\psi}\|_\infty\}.$$

We then deduce from Lemma 2.1 of [CL20],

$$|h_\delta| \in \mathcal{T}_\theta \left( \|h_\delta\|_\infty^\theta (2^{j_0+1} + 1) \right),$$

which shows  $|h_\delta| \in \mathcal{T}_\theta (c_1 b^\theta 2^{(j/2)\theta} 2^{j_0})$ .

We get (85) by noticing that  $\int \bar{\phi} = 1$  and  $\int \bar{\psi} = 0$  as the wavelet basis is bi-orthogonal and that  $\phi = 1_{[0,1]}$ .  $\square$

**Lemma 34.** The map  $g$  is compactly supported on  $[2^{j_0-2}, 2^{j_0+2}]$ . It satisfies

$$\int g(x) dx = M^{1/\theta} 2^{j_0(1-1/\theta)}.$$

Besides, for all  $x \in [2^{j_0-1}, 2^{j_0}]$ ,

$$g(x) \geq 4^{-1} M^{1/\theta} 2^{-j_0/\theta}.$$

It belongs to

$$\mathcal{B}_{p,\infty}^\alpha \left( c_4 M^{1/\theta} 2^{j_0(1/p-1/\theta-\alpha)} \right) \cap \mathcal{T}_\theta(c_5 M),$$

and satisfies

$$\sup_{|x| \geq 1} |x|^{1/\theta} |g(x)| \leq c_6 M^{1/\theta}.$$

Here,  $c_4, c_5, c_6$  only depend on  $\varphi, \theta$

*Sketch of the proof of Lemma 34.* We only prove that  $g$  belongs to the Besov class. The wavelet coefficient of  $g$  is denoted for  $j' \geq 0$  and  $k' \in \mathbb{Z}$  by

$$\beta'_{j',k'} = \int g(x) \psi_{j',k'}(x) dx.$$

Yet,

$$\beta'_{j',k'} = M^{1/\theta} 2^{(1/2-1/\theta)j_0} \int \varphi(x) \psi_{j'+j_0,k'}(x) \, dx.$$

In particular,

$$\|\beta'_{j',\cdot}\|_p \leq \rho M^{1/\theta} 2^{j_0(1/p-1/\theta-\alpha)} 2^{-j'(\alpha+1/2-1/p)}$$

hence the result.  $\square$

We now turn to the proof of Lemma 32. By choosing  $a_1$  and  $a_2$  appropriately, we deduce from (77) and (78) that  $g$  and  $h_\delta$  lie in  $\mathcal{B}_{p,\infty}^\alpha(R)$ . We deduce from (81) and from the (quasi) triangle inequality for the (quasi) norms  $\|\cdot\|_p$  that  $f_\delta \in \mathcal{B}_{p,\infty}^\alpha(c_7R)$ . Note that (79) yields  $g(x) \geq 2\|h_\delta\|_\infty$  for all  $x \in [2^{j_0-1}, 2^{j_0}]$  if  $a_3$  is suitably taken. In particular,  $f_\delta$  is non-negative and is therefore a density. Moreover, (79) implies  $|h_\delta| \in \mathcal{T}_\theta(M)$  and hence  $f_\delta \in \mathcal{T}_\theta(c_8M)$ . Besides,

$$\sup_{|x| \geq 1} |x|^{1/\theta} f_\delta(x) \leq c_9 M^{1/\theta}.$$

We now observe that  $f_\delta(x) = (1-q)\zeta(x) + g(x)$  for all  $x \notin [2^{j_0-1}, 2^{j_0}]$ . And, for all  $x \in [2^{j_0-1}, 2^{j_0}]$ , we have

$$f_\delta(x) \geq 8^{-1} M^{1/\theta} 2^{-j_0/\theta}.$$

Consider now  $j \in \{1, \dots, \bar{k}\}$  and  $\delta, \delta' \in \mathcal{D}$  satisfying  $\delta_j \neq \delta'_j$  and  $\delta_k = \delta'_k$  for all  $k \neq j$ . Then,

$$\begin{aligned} h^2(f_\delta, f_{\delta'}) &\leq \frac{1}{8} \int \frac{(f_\delta(x) - f_{\delta'}(x))^2}{\min\{f_\delta(x), f_{\delta'}(x)\}} \, dx \\ &\leq M^{-1/\theta} 2^{j_0/\theta} \int (h_\delta(x) - h_{\delta'}(x))^2 \, dx \\ &\leq 2M^{-1/\theta} 2^{j_0/\theta} b^2. \end{aligned}$$

We may therefore fulfil condition (76) by choosing  $a_4$  thanks to (80). Lemma 31 then entails: any estimator  $\tilde{\delta} \in \mathcal{D}$  satisfies

$$\sup_{\delta \in \mathcal{D}} \mathbb{E} \left[ \Delta(\tilde{\delta}, \delta) \right] \geq c\bar{k}.$$

We have for all  $\delta, \delta' \in \mathcal{D}$ ,

$$\begin{aligned} d_1(f_\delta, f_{\delta'}) &= d_1(h_\delta, h_{\delta'}) \\ &= c_{10} b 2^{-j/2} \Delta(\delta, \delta') \end{aligned}$$

as the supports of  $\bar{\psi}_{j,2^{L+1}k}$  are disjoint when  $k$  varies.

Let now  $\tilde{f}$  be an estimator of  $f$  and  $\tilde{\delta} \in \mathcal{D}$  such that

$$d_1(\tilde{f}, f_{\tilde{\delta}}) = \inf_{\delta \in \mathcal{D}} d_1(\tilde{f}, f_\delta).$$

We have,

$$\begin{aligned} \sup_{\delta \in \mathcal{D}} \mathbb{E} \left[ d_1(f_\delta, \tilde{f}) \right] &\geq \frac{1}{2} \sup_{\delta \in \mathcal{D}} \mathbb{E} \left[ d_1(f_\delta, f_{\tilde{\delta}}) \right] \\ &\geq c_{11} b 2^{-j/2} \bar{k}. \end{aligned}$$

We use the definition of  $\bar{k}$  to conclude.  $\square$

**E.2. Case**  $\mu_n = \nu_n$ ,  $p \neq 1$ ,  $\theta = \alpha/(2\alpha + 1 - 1/p)$ . When  $p > 1$ , we consider two arbitrary numbers  $s_1 < s_2$  in

$$\left( \frac{1}{2\alpha + 1}, \frac{1}{2\alpha + 1} \left[ 1 - \frac{1 - 1/p}{1 - 1/\theta} \right] \right).$$

When  $p < 1$ ,  $s_1 < s_2$  are arbitrary numbers lying in

$$\left( \frac{1}{2\alpha + 1} \left[ 1 - \frac{1 - 1/p}{1 - 1/\theta} \right], \frac{1}{2\alpha + 1} \right).$$

We define the largest integers  $j_0, j_1$  such that  $2^{j_0} \leq n^{s_1}$  and  $2^{j_1} \leq n^{s_2}$ . We set

$$\begin{aligned} a_1^{1/p-1} &= R^{2\alpha/(2\alpha+1)} M^{-(1+2\alpha-1/p)/(2\alpha+1)}, \\ a_2 &= R^{1/(2\alpha+1)} M^{(1+2\alpha-1/p)/(2\alpha+1)}. \end{aligned}$$

We define for all  $j \in \{j_0, \dots, j_1\}$  the largest integer  $\ell_j \geq 0$  such that

$$2^{\ell_j} \leq a_1 2^{j\alpha/(1-1/p)} n^{-\alpha/((2\alpha+1)(1-1/p))}.$$

We define when  $p < 1$  the smallest integer  $t \geq 1$  satisfying  $2^{\alpha t/(1-1/p)} < 1/4$ . When  $p > 1$ ,  $t$  is rather defined as the smallest integer such that  $2^{\alpha t/(1-1/p)} > 4$ . We consider the subset

$$J = \{j_0 + kt, k \in \mathbb{N}, k \leq (j_1 - j_0)/t\}$$

of  $\{j_0, \dots, j_1\}$  and observe that the intervals  $[2^{\ell_j-1}, 2^{\ell_j}]$  are disjoint when  $j$  varies in  $J$ . Besides, as  $j_1 - j_0$  is of the order of  $\log n$ ,  $J$  is non-empty when  $n$  is large enough.

We define for  $n$  large enough the smallest integer  $\underline{k}_j \geq 1$  satisfying  $1 + 2\underline{k}_j \geq 2^{\ell_j+j-L-1}$ , and the largest integer  $\bar{k}_j \geq 1$  satisfying  $2\bar{k}_j + 2\underline{k}_j + 1 \leq 2^{\ell_j+j-L}$ .

We consider  $a > 0$  to be specified later on, and set for  $j \in J$ ,  $x \in \mathbb{R}$ ,

$$\begin{aligned} g_j(x) &= aM^{1/\theta} 2^{-\ell_j/\theta} \varphi(2^{-\ell_j} x) \\ g(x) &= \sum_{j \in J} g_j(x). \end{aligned}$$

We show:

**Lemma 35.** *For all  $\varepsilon > 0$ , there is  $n_0$  such that  $g$  lies in  $\mathcal{B}_{p,\infty}^\alpha(\varepsilon)$  if  $n \geq n_0$ . Moreover,  $g$  satisfies*

$$(86) \quad \sup_{x \in \mathbb{R}} |x|^{1/\theta} g(x) \leq 2^{2+1/\theta} aM^{1/\theta}.$$

*Proof of Lemma 35.* We first show that  $g$  belongs to the Besov class. We denote for  $j' \geq 0$ ,  $k' \in \mathbb{Z}$ , the wavelet coefficient of  $g$  by

$$\beta'_{j',k'} = \int g(x) \psi_{j',k'}(x) dx.$$

It is equal to

$$\beta'_{j',k'} = aM^{1/\theta} \sum_{j \in J} 2^{-\ell_j/\theta} 2^{\ell_j/2} \int \varphi(x) \psi_{j'+\ell_j,k'}(x) dx.$$

Minkowski's integral inequality yields when  $p \geq 1$  is finite:

$$\begin{aligned} \|\beta'_{j', \cdot}\|_p &\leq aM^{1/\theta} \sum_{j \in J} 2^{-\ell_j/\theta} 2^{\ell_j/2} \left[ \sum_{k' \in \mathbb{Z}} \left| \int \varphi(x) \psi_{j'+\ell_j, k'}(x) \, dx \right|^p \right]^{1/p} \\ &\leq C'_1 aM^{1/\theta} \left[ \sum_{j \in J} 2^{-\ell_j(\alpha+1/\theta-1/p)} \right] 2^{-j'(\alpha+1/2-1/p)}, \end{aligned}$$

as  $\varphi \in \mathfrak{B}_{p, \infty}^\alpha(\rho)$  and hence  $g \in \mathcal{B}_{p, \infty}^\alpha(\varepsilon)$  when  $n$  is large enough. This result is also true when  $p = \infty$ . When  $p < 1$ , we merely use the inequality  $(\sum_k x_k)^p \leq \sum_k x_k^p$  to get

$$\begin{aligned} \|\beta'_{j', \cdot}\|_p^p &\leq \left(aM^{1/\theta}\right)^p \sum_{j \in J} 2^{-\ell_j p/\theta} 2^{\ell_j p/2} \sum_{k' \in \mathbb{Z}} \left| \int \varphi(x) \psi_{j'+\ell_j, k'}(x) \, dx \right|^p \\ &\leq C'_2 \left(aM^{1/\theta}\right)^p \left[ \sum_{j \in J} 2^{-\ell_j p(\alpha+1/\theta-1/p)} \right] 2^{-j' p(\alpha+1/2-1/p)}, \end{aligned}$$

which leads to the same result.

We now turn to the proof of (86). Since  $g$  is compactly supported on  $(0, +\infty)$  we may suppose that  $x > 0$ . It is straightforward that  $\varphi(x) \leq 2^{1/\theta} x^{-1/\theta} 1_{(1/4, 2)}(x)$  for all  $x > 0$ . We deduce,

$$\begin{aligned} x^{1/\theta} g(x) &\leq aM^{1/\theta} \sum_{j \in J} \left(2^{-\ell_j} x\right)^{1/\theta} \varphi(2^{-\ell_j} x) \\ &\leq a2^{1/\theta} M^{1/\theta} \sum_{j \in J} 1_{[2^{\ell_j-2}, 2^{\ell_j+2}]}(x). \end{aligned}$$

Since the intervals  $[2^{\ell_j-1}, 2^{\ell_j}]$  are disjoint, the sum is not larger than 4, hence the result.  $\square$

We put  $D = \sum_{j \in J} \bar{k}_j$ , and identify the set

$$\mathcal{D} = \left\{ (\delta_{j,k})_{j \in J, k \in \{1, \dots, \bar{k}_j\}}, \delta_{j,k} \in \{0, 1\} \right\},$$

with  $\{0, 1\}^D$ .

Let now for all  $j \in J$ ,

$$b_j = a_2 n^{-\alpha/(2\alpha+1)} 2^{-\ell_j} 2^{-j/2}.$$

We also set for all  $x \in \mathbb{R}$ , and  $\delta \in \mathcal{D}$ ,

$$\begin{aligned} h_{\delta, j}(x) &= b_j \sum_{k=1}^{\bar{k}_j} \delta_{j,k} \bar{\psi}_{j, 2^{L+1}(k+\underline{k}_j)}(x), \\ h_\delta(x) &= \sum_{j \in J} h_{\delta, j}(x). \end{aligned}$$

We show:

**Lemma 36.** *For all  $j \in J$ ,  $\text{supp } h_{\delta, j} \subset [2^{\ell_j-1}, 2^{\ell_j}]$ . Moreover,  $h_\delta$  belongs to  $\mathcal{B}_{p, \infty}^\alpha(c_1 R)$ , and for all  $n \geq n_0$ ,  $2|h_\delta(x)| \leq g(x)$ . Here,  $c_1$  only depends on the wavelet basis and  $\alpha, p$ .*

*Sketch of the proof of Lemma 36.* We only show that  $2|h_\delta(x)| \leq g(x)$ . Since the supports of  $h_{\delta,j}$  are disjoint when  $j$  varies, we have

$$\begin{aligned} \frac{\sup_{x \in [2^{\ell_j-1}, 2^{\ell_j}]} |h_\delta(x)|}{\inf_{x \in [2^{\ell_j-1}, 2^{\ell_j}]} g(x)} &\leq \frac{\sup_{x \in [2^{\ell_j-1}, 2^{\ell_j}]} |h_{\delta,j}(x)|}{\inf_{x \in [2^{\ell_j-1}, 2^{\ell_j}]} g_j(x)} \\ &\leq 4 \frac{\|\bar{\psi}\|_\infty b_j 2^{j/2}}{aM^{1/\theta} 2^{-\ell_j/\theta}}. \end{aligned}$$

We conclude by noting that the right-hand side of this inequality tends to 0 with  $n$  (uniformly in  $j$ ).  $\square$

We introduce

$$q = \int (g(x) + h_\delta(x)) \, dx.$$

This term does not depend on  $\delta$  as the integrate of  $h_\delta$  is zero. It is equal to

$$q = aM^{1/\theta} \sum_{j \in J} 2^{\ell_j(1-1/\theta)},$$

and tends therefore to 0 when  $n$  grows up. We then consider the density  $f_\delta$  defined for all  $\delta \in \mathcal{D}$  and  $x \in \mathbb{R}$ ,

$$\begin{aligned} f_\delta(x) &= (1-q)\zeta(x) + g(x) + h_\delta(x) \\ &= (1-q)\zeta(x) + \sum_{j \in J} \{g_j(x) + h_{\delta,j}(x)\}. \end{aligned}$$

The two preceding lemmas ensure  $f_\delta \in \mathcal{B}_{p,\infty}^\alpha(c_2R)$  and

$$\sup_{x \in \mathbb{R}} |x|^{1/\theta} f_\delta(x) \leq c_3 M^{1/\theta}.$$

We set  $\mathcal{F} = \{f_\delta, \delta \in \mathcal{D}\}$ , and endow  $\mathcal{D}$  with the distance  $\Delta$  defined for  $\delta, \delta' \in \mathcal{D}$  by

$$\Delta(\delta, \delta') = \sum_{j \in J} 2^{-j/2} b_j \sum_{k=1}^{\bar{k}_j} |\delta_{j,k} - \delta'_{j,k}|.$$

Note that the supports of  $\bar{\psi}_{j,2^{L+1}(k+\bar{k}_j)}$  are pairwise disjoint when  $k \in \{1, \dots, \bar{k}_j\}$  and  $j \in J$ . In particular, for all  $\delta, \delta' \in \mathcal{D}$ ,

$$\|f_\delta - f_{\delta'}\|_1 = \left( \int |\bar{\psi}| \right) \Delta(\delta, \delta').$$

Consider now an estimator  $\tilde{f}$  of  $f$  and define  $\tilde{\delta} \in \mathcal{D}$  such that

$$d_1(f_{\tilde{\delta}}, \tilde{f}) = \inf_{\delta \in \mathcal{D}} d_1(f_\delta, \tilde{f}).$$

The triangle inequality entails

$$(87) \quad \sup_{f \in \mathcal{F}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] \geq (1/2) \sup_{\delta \in \mathcal{D}} \mathbb{E} \left[ d_1(f_\delta, f_{\tilde{\delta}}) \right].$$



We now check that the assumptions of Lemma 31 are met. We consider  $\delta, \delta' \in \mathcal{D}$ ,  $j \in J$ ,  $k \in \{1, \dots, \bar{k}_j\}$ , and suppose that  $\delta_{j',k'} = \delta'_{j',k'}$  for all  $(j', k') \neq (j, k)$ . We have,

$$\begin{aligned} h^2(f_\delta, f_{\delta'}) &\leq \frac{2^{\ell_j/\theta}}{2aM^{1/\theta}} \int (f_\delta(x) - f_{\delta'}(x))^2 dx \\ &\leq c_4 \frac{2^{\ell_j/\theta}}{aM^{1/\theta}} b_j^2 \\ &\leq c_4 a_2^2 \frac{2^{\ell_j(1/\theta-2)}}{aM^{1/\theta}} n^{-2\alpha/(2\alpha+1)} 2^{-j}, \\ &\leq c_5 a^{-1} n^{-1}. \end{aligned}$$

The right-hand side of this inequality is smaller than  $1/(2n)$  if  $a$  is large enough. We deduce from Lemma 31 and (87) that any estimator  $\tilde{f}$  satisfies

$$\begin{aligned} \sup_{f \in \mathcal{F}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] &\geq c_6 \sum_{j \in J} 2^{-j/2} b_j \bar{k}_j, \\ &\geq c_7 a_2 |J| n^{-\alpha/(2\alpha+1)}. \end{aligned}$$

We conclude by noticing that  $|J| \geq c_8 \log n$  where  $c_8 > 0$  only depends on  $\alpha, p, s_1, s_2$ .  $\square$

**E.3. Case**  $\mu_n = \nu_n$ ,  $p = 1$ ,  $\theta = \alpha/(2\alpha + 1 - 1/p)$ . We consider  $a > 0$  and define the largest  $j$  such that

$$2^j \leq a^{-1/(2\alpha+1)} R^{2/(2\alpha+1)} M^{-2/(1+2\alpha)} n^{1/(2\alpha+1)} (\log n)^{-2/(2\alpha+1)}.$$

We introduce two arbitrary numbers  $s_1 < s_2$  in  $(0, \alpha/(4\alpha + 2))$ . We define the smallest integer  $r_0$  and the largest integer  $r_1$  satisfying

$$n^{s_1} \leq 2^{r_0} \leq 2^{r_1} \leq n^{s_2}.$$

We define for all  $r \in \{r_0, \dots, r_1\}$ , the smallest integer  $\underline{k}_r \geq 1$  satisfying  $1 + 2\underline{k}_r \geq 2^{2r+j-L-1}$ , and the largest integer  $\bar{k}_r \geq 1$  satisfying  $2\bar{k}_r + 2\underline{k}_r + 1 \leq 2^{2r+j-L}$ .

We set for all  $x \in \mathbb{R}$ ,

$$\begin{aligned} g_r(x) &= aM^2 2^{-4r} \varphi(2^{-2r} x) \\ g(x) &= \sum_{r=r_0}^{r_1} g_r(x), \end{aligned}$$

where  $\varphi$  is the density introduced at the beginning of Section E. The lemma below gathers the properties of  $g$  and is proved as Lemma 35.

**Lemma 37.** *For all  $\varepsilon > 0$ , there is  $n_0$  such that  $g$  belongs to  $\mathcal{B}_{p,\infty}^\alpha(\varepsilon)$  when  $n \geq n_0$ . It moreover satisfies*

$$\sup_{x \in \mathbb{R}} x^2 g(x) \leq 16aM^2.$$

Let

$$\mathcal{D} = \left\{ (\delta_{k,r})_{\substack{k \in \{1, \dots, \bar{k}_r\}, \\ r \in \{r_0, \dots, r_1\}}}, \delta_{k,r} \in \{0, 1\} \right\},$$

which we identify with  $\{0, 1\}^D$  where  $D = \sum_{r=r_0}^{r_1} \bar{k}_r$ .

Moreover, we define for all  $x \in \mathbb{R}$ ,  $r \in \{r_0, \dots, r_1\}$ ,  $\delta \in \mathcal{D}$ ,

$$h_{\delta,r}(x) = b_r \sum_{k=1}^{\bar{k}_r} \delta_{k,r} \bar{\psi}_{j,2^{L+1}(k+\underline{k}_r)}(x),$$

where

$$b_r = R^{1/(2\alpha+1)} M^{2\alpha/(2\alpha+1)} 2^{-j/2} 2^{-2r} (\log n)^{-1/(2\alpha+1)} n^{-\alpha/(2\alpha+1)}.$$

We then set

$$h_\delta(x) = \sum_{r=r_0}^{r_1} h_{\delta,r}(x)$$

and have:

**Lemma 38.** *For all  $r \in \{r_0, \dots, r_1\}$ ,  $\text{supp } h_{\delta,r} \subset [2^{2r-1}, 2^{2r}]$ . Moreover,  $h_\delta$  belongs to  $\mathcal{B}_{p,\infty}^\alpha(c_1 R)$ , and satisfies  $2|h_\delta(x)| \leq g(x)$  for all  $x \in \mathbb{R}$  and  $n \geq n_0$ . Here,  $c_1$  only depends on the wavelet basis and  $\alpha, a$ .*

We introduce

$$\begin{aligned} q &= \int (g(x) + h_\delta(x)) \, dx \\ &= aM^2 \sum_{r=r_0}^{r_1} 2^{-2r}. \end{aligned}$$

We then set for all  $\delta \in \mathcal{D}$  and  $x \in \mathbb{R}$ ,

$$f_\delta(x) = (1 - q)\zeta(x) + g(x) + h_\delta(x).$$

As  $q$  tends to 0 when  $n$  grows up, it is smaller than 1 if  $n$  is large enough and  $f_\delta$  is a density.

We gather all the  $f_\delta$  in the set  $\mathcal{F} = \{f_\delta, \delta \in \mathcal{D}\}$ , and endow it with the distance  $\Delta$  defined for  $\delta, \delta' \in \mathcal{D}$  by

$$\Delta(\delta, \delta') = \sum_{r=r_0}^{r_1} b_r \sum_{k=1}^{\bar{k}_r} |\delta_{k,r} - \delta'_{k,r}|.$$

We have for all  $\delta, \delta' \in \mathcal{D}$ ,

$$\|f_\delta - f_{\delta'}\|_1 = c_2 2^{-j/2} \Delta(\delta, \delta')$$

as the supports of  $\bar{\psi}_{j,2^{L+1}(k+\underline{k}_r)}$  are pairwise disjoint when  $k \in \{1, \dots, \bar{k}_r\}$  and  $r \in \{r_0, \dots, r_1\}$ .

We consider  $\delta, \delta' \in \mathcal{D}$ ,  $r \in \{r_0, \dots, r_1\}$ ,  $k \in \{1, \dots, \bar{k}_r\}$ , and suppose that  $\delta_{k',r'} = \delta'_{k',r'}$  for all  $(k', r') \neq (k, r)$ . We have,

$$\begin{aligned} h^2(f_\delta, f_{\delta'}) &\leq a^{-1} M^{-2} 2^{4r} \int (f_\delta(x) - f_{\delta'}(x))^2 \, dx \\ &\leq c_3 a^{-1} M^{-2} 2^{4r} b_r^2 \\ &\leq c_4 a^{-1} R^{2/(2\alpha+1)} M^{-2/(2\alpha+1)} (\log n)^{-2/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} 2^{-j} \\ &\leq c_5 a^{-2\alpha/(2\alpha+1)} n^{-1}. \end{aligned}$$

The right-hand side of this inequality is smaller than  $1/(2n)$  if  $a$  is large enough.

We deduce from Assouad's Lemma 31 that any estimator  $\tilde{f}$  satisfies

$$\begin{aligned} \sup_{f \in \mathcal{F}} \mathbb{E} \left[ d_1(f, \tilde{f}) \right] &\geq c_6 2^{-j/2} \sum_{r=r_0}^{r_1} b_r \bar{k}_r, \\ &\geq c_7 (r_1 - r_0 + 1) R^{1/(2\alpha+1)} M^{2\alpha/(2\alpha+1)} (\log n)^{-1/(2\alpha+1)} n^{-\alpha/(2\alpha+1)}. \end{aligned}$$

It then remains to notice that  $r_1 - r_0 + 1$  is of the order of  $\log n$ .  $\square$

## F. PROOF OF PROPOSITION 4

We only need to prove the proposition when  $\alpha = 1/p - 1$ , what we assume below. We define  $j_0 \geq 2$  as the smallest integer such that  $(n^p + 1)2^{-j_0} \leq 1/4$  and consider  $j_1 \geq j_0$ .

We set for  $j \in \{j_0, \dots, j_1\}$ ,  $k_j = 2(n^p + 1)(2^{j-j_0} - 1)$  and

$$K_j = \{k_j, k_j + 1, k_j + 2, \dots, k_j + n^p\}.$$

We introduce for  $k \in \mathbb{Z}$ ,  $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ . Note that  $I_{j,k} \subset [0, 1/2)$  for all  $j$  between  $j_0$  and  $j_1$  and all  $k \in K_j$ . Moreover,  $I_{j,k} \cap I_{j',k'} = \emptyset$  if  $j \neq j'$ , no matter  $(k, k') \in K_j \times K_{j'}$ .

We set  $D = (j_1 - j_0 + 1)(n^p + 1)$  and write the elements  $\delta$  of  $\{0, 1\}^D$  as  $\delta = (\delta_{j,k})_{j \in \{j_0, \dots, j_1\}, k \in K_j}$ . We endow  $\{0, 1\}^D$  with the Hamming distance defined for all  $\delta, \delta' \in \{0, 1\}^D$  by

$$\Delta(\delta, \delta') = \sum_{j=j_0}^{j_1} \sum_{k \in K_j} |\delta_{j,k} - \delta'_{j,k}|.$$

We set for all  $\delta \in \{0, 1\}^D$  and  $x \in \mathbb{R}$ ,

$$\varphi_\delta(x) = \frac{1}{D} \sum_{j=j_0}^{j_1} 2^j \sum_{k \in K_j} \delta_{j,k} 1_{I_{j,k}}(x).$$

We show below after the present proof:

**Lemma 39.** *For all  $\varepsilon > 0$ ,  $D$  large enough, and  $\delta \in \{0, 1\}^D$ ,  $\varphi_\delta$  belongs to  $\mathfrak{B}_{p,\infty}^\alpha(\varepsilon)$ .*

We define for  $x \in \mathbb{R}$ ,

$$f_\delta(x) = \varphi_\delta(x) + \varphi_{1-\delta}(x - 1/2).$$

Thereby,  $f_\delta$  is a compactly supported density on  $[0, 1]$  lying in  $\mathfrak{B}_{p,\infty}^\alpha(\mathbb{R})$  if  $D$  is large enough. Moreover, for all  $\delta, \delta' \in \{0, 1\}^D$ ,

$$d_1(f_\delta, f_{\delta'}) = \frac{2}{D} \Delta(\delta, \delta').$$

Recall that the square  $h^2$  of the Hellinger distance is not larger than half of  $d_1$ . We conclude by applying Assouad's lemma and by choosing  $j_1$  large enough.  $\square$

*Proof of Lemma 39.* We have for all  $h > 0$ ,

$$\begin{aligned}
\int |\varphi_\delta(t+h) - \varphi_\delta(t)|^p dt &\leq \frac{1}{D^p} \sum_{j=j_0}^{j_1} 2^{jp} \sum_{k \in K_j} \int |1_{I_{j,k}}(t+h) - 1_{I_{j,k}}(t)|^p dt \\
&\leq \frac{2}{D^p} \sum_{j=j_0}^{j_1} 2^{jp} \sum_{k \in K_j} \min\{2^{-j}, h\} \\
&\leq \frac{2(n^p + 1)}{D^p} \sum_{j=j_0}^{j_1} 2^{jp} \min\{2^{-j}, h\} \\
&\leq c \frac{n^p + 1}{D^p} h^{1-p} \\
&\leq c \frac{n^p + 1}{D^p} h^{\alpha p}.
\end{aligned}$$

Consider now some odd number  $r$  larger than  $\alpha$  and note that

$$2\Delta_h^r(\varphi_\delta)(t) = \sum_{k=0}^r \binom{r}{k} (-1)^k [\varphi_\delta(t+kh) - \varphi_\delta(t+(r-k)h)].$$

The above entails that for all  $D$  large enough,

$$\int |\Delta_h^r(\varphi_\delta)(t)|^p dt \leq (\varepsilon/2)^p h^{\alpha p}.$$

By noticing that

$$\|\varphi_\delta\|_p^p \leq \frac{n^p + 1}{D^p} \sum_{j=j_0}^{j_1} 2^{-j(1-p)} \leq (\varepsilon/2)^p$$

when  $D$  is large enough, we conclude that  $\varphi_\delta$  belongs to  $\mathfrak{B}_{p,\infty}^\alpha(\varepsilon)$ .  $\square$

## G. PROOF OF PROPOSITION 1.

If  $f$  is a density,  $F_{j,k} \leq 1$  and

$$\sum_{k \in \mathbb{Z}} F_{j,k} = 1.$$

We deduce from the elementary inequality

$$\left( \sum_{k \in \mathbb{Z}} F_{j,k}^\theta \right)^{1/\theta} \geq \sum_{k \in \mathbb{Z}} F_{j,k}$$

that  $M$  must satisfy  $M \geq 1$  if  $f \in \mathcal{T}_\theta(M)$ . This proves the first assertion.

We now show the first part of the second point. We deduce from (2) that for all  $j \geq 0$  and  $k \in \mathbb{Z}$ ,

$$F_{j,k} = \sum_{j'=-1}^{\infty} \sum_{k' \in \mathbb{Z}} \beta_{j',k'} \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} \bar{\psi}_{j',k'}.$$

Let  $\bar{L} > 0$  such that  $[-\bar{L}, \bar{L}]$  contains the supports of  $\bar{\phi}$  and  $\bar{\psi}$ . The above integral is zero if  $k \notin K_{j',k'}$  where

$$K_{j',k'} = \left[ -1/2 + 2^{j-\max\{j',0\}} (k' - \bar{L}), 1/2 + 2^{j-\max\{j',0\}} (k' + \bar{L}) \right].$$

We deduce

$$F_{j,k} \leq c_1 \sum_{j'=-1}^{\infty} \sum_{k' \in \mathbb{Z}} |\beta_{j',k'}| 1_{k \in K_{j',k'}} 2^{-\max\{j-j'/2, j'/2\}},$$

where  $c_1$  depends on  $\bar{\phi}, \bar{\psi}$  only. Now, using the same elementary inequality as previously,

$$\begin{aligned} \sum_{k \in \mathbb{Z}} F_{j,k}^p &\leq c_1^p \sum_{k \in \mathbb{Z}} \sum_{j'=-1}^{\infty} \sum_{k' \in \mathbb{Z}} |\beta_{j',k'}|^p 1_{k \in K_{j',k'}} 2^{-p \max\{j-j'/2, j'/2\}}, \\ &\leq c_1^p \sum_{j'=-1}^{\infty} \sum_{k' \in \mathbb{Z}} |K_{j',k'}| |\beta_{j',k'}|^p 2^{-p \max\{j-j'/2, j'/2\}}, \\ &\leq c_2 \sum_{j'=-1}^{\infty} \sum_{k' \in \mathbb{Z}} |\beta_{j',k'}|^p \left[ 2^{j(1-p)+j'(p/2-1)} + 2^{-j'p/2} \right] \\ &\leq c_2 R^p \sum_{j'=-1}^{\infty} \left[ 2^{j(1-p)} 2^{-j'p\alpha} + 2^{-j'p(\alpha+1-1/p)} \right] \\ &\leq c_3 R^p 2^{j(1-p)}, \end{aligned}$$

which gives the first part of the second point.

The proof of the second part of the second point, as well as the proof of the third point is quite easy, and we move directly to the proof of the fourth point. For all  $k \notin [-2^j - 1/2, 2^j + 1/2]$ ,

$$F_{j,k} \leq A^b \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} |x|^{-b} dx.$$

In particular, for all  $k \geq 2^j + 1/2$ ,

$$F_{j,k} \leq A^b \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} [2^{-j}(k-1/2)]^{-b} dx \leq A^b 2^{-j(1-b)} [(k-1/2)]^{-b}$$

and for all  $k \leq -2^j - 1/2$ ,

$$F_{j,k} \leq A^b 2^{-j(1-b)} [(-k-1/2)]^{-b}.$$

The number of  $k \notin [-2^j - 1/2, 2^j + 1/2]$  such that  $F_{j,k} \geq t$  is therefore not larger than

$$2 \left[ A t^{-1/b} 2^{-j(1/b-1)} + 1 \right].$$

Moreover, the number of  $k \in [-2^j - 1/2, 2^j + 1/2]$  such that  $F_{j,k} \geq t$  is not larger than

$$\begin{aligned} t^{-1/b} \sum_{k \in [-2^j - 1/2, 2^j + 1/2]} F_{j,k}^{1/b} &\leq t^{-1/b} \left( \sum_{k \in \mathbb{Z}} F_{j,k} \right)^{1/b} [2^{j+1} + 2]^{1-1/b} \\ &\leq t^{-1/b} [2^{j+1} + 2]^{1-1/b}. \end{aligned}$$

Now,

$$\begin{aligned} \sup_{t>0} \left\{ t^{1/b} \sum_{k \in \mathbb{Z}} 1_{F_{j,k} \geq t} \right\} &= \sup_{t \in (0,1]} \left\{ t^{1/b} \sum_{k \in \mathbb{Z}} 1_{F_{j,k} \geq t} \right\} \\ &\leq (2^{j+1} + 2)^{1-1/b} + 2 \left( A 2^{-j(1/b-1)} + 1 \right), \end{aligned}$$

which concludes the proof.  $\square$

### H. PROOF OF EQUALITY (1) IN $(\mathbb{L}^1(\mathbb{R}), d_1)$ .

The following arguments come mainly from [HKPT12]. We introduce the Kernel  $K$  defined for  $x, y \in \mathbb{R}$  by

$$K(x, y) = \sum_{k \in \mathbb{Z}} \bar{\phi}(x - k) \phi(y - k).$$

We put for  $J \geq 0$ ,  $K_J(x, y) = 2^J K(2^J x, 2^J y)$  and

$$K_J f(x) = \int_{\mathbb{R}} K_J(x, y) f(y) dy.$$

We also set

$$K'_J f = \sum_{k \in \mathbb{Z}} \alpha_k \bar{\phi}_k + \sum_{j=0}^J \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k}.$$

When  $f \in \mathbb{L}^2(\mathbb{R})$ ,  $K_J f$  is the (oblic) projection of  $f$  on the space spanned by the basis  $(\bar{\phi}_{J,k})_{k \in \mathbb{Z}}$  where  $\bar{\phi}_{J,k}(x) = 2^{J/2} \bar{\phi}(2^J x - k)$ . Therefore,  $K_J f = K'_J f$  for all  $f \in \mathbb{L}^2(\mathbb{R})$ .

Since  $\phi$  and  $\bar{\phi}$  are two compactly supported bounded functions, there exists an integrable function  $F$  such that  $|K(x, y)| \leq F(x - y)$  for all  $x, y \in \mathbb{R}$ . This entails that  $K_J$  is continuous in  $(\mathbb{L}^1(\mathbb{R}), d_1)$ . The same thing holds true for  $K'_J$  and hence, by using a density argument,  $K_J f = K'_J f$  for all  $f \in \mathbb{L}^1(\mathbb{R})$ .

Since  $K_J f \rightarrow f$  in  $(\mathbb{L}^2(\mathbb{R}), d_2)$ , we have for all  $f \in \mathbb{L}^2(\mathbb{R})$

$$\int_{\mathbb{R}} \left( \int_{\mathbb{R}} K(x 2^J, y) dy - 1 \right)^2 f^2(x) dx \rightarrow 0 \quad \text{when } J \rightarrow +\infty.$$

We deduce (see Lemma 8.4 of [HKPT12]) that  $\int_{\mathbb{R}} K(x, y) dy = 1$  for all  $x \in \mathbb{R}$ .

Therefore, we have for all  $f \in \mathbb{L}^1(\mathbb{R})$ ,

$$\begin{aligned} \|K_J f - f\|_1 &\leq \int_{\mathbb{R}} |K_J(x, y)| |f(y) - f(x)| dx dy \\ &\leq \int_{\mathbb{R}} F(t) \left( \int_{\mathbb{R}} |f(x + 2^{-J} t) - f(x)| dx \right) dt. \end{aligned}$$

This entails  $K_J f \rightarrow f$  in  $(\mathbb{L}^1(\mathbb{R}), d_1)$ .  $\square$

UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE, CNRS, INSTITUT CAMILLE JORDAN UMR 5208, F-42023, SAINT-ÉTIENNE, FRANCE.

*Email address:* mathieu.sart@univ-st-etienne.fr