Adaptive and optimal estimation under shape and smoothness constraints

Mathieu Sart

ADAPTIVE AND OPTIMAL ESTIMATION UNDER SHAPE AND SMOOTHNESS CONSTRAINTS

MATHIEU SART

Abstract. We consider the problem of estimating the density f of a real valued random variable. We present a fully adaptive procedure based on a wavelet series expansion of f. We study the $\mathbb{L}^1$ risk of our estimator when f satisfies a mild smoothness condition on $\mathbb{R}$, i.e f belongs to a (weak) Besov class. Its tails are assumed to be s-monotone and may decay as a power law. We show the importance of considering the negative resolutions of the wavelet expansion to optimally estimate a fat tailed density. We also reveal new minimax rates. In particular, we explain when the shape constraint improves the estimation rate of a smooth density.

1. Introduction

We observe n independent and identically distributed real valued random variables $X_1, \dots, X_n$. We suppose that these variables admit a density f with respect to the Lebesgue measure on $\mathbb{R}$. The aim of this paper is to estimate f under shape and smoothness constraints. More precisely, we suppose that f is smooth on $\mathbb{R}$ and its tails are s-monotone.

We recall that a 1-monotone function is a non-increasing function on $(0, +\infty)$. When $s \geq 2$, a function f is said to be s-monotone if it admits derivatives up to order $s-2$ with the following properties: for all $\ell \in \{0, \dots, s-2\}$, $(-1)^\ell f^{(\ell)}$ is non-negative, non-decreasing, and convex on $(0, +\infty)$. In this sentence, $f^{(\ell)}$ denotes the ℓ^{th} derivative of f (with the convention $f^{(0)} = f$). It is more common in the literature to speak of "k-monotone function". We prefer to choose the symbol s in place of k to keep this letter for later.

The problem of estimating a s-monotone density with support in $[0, +\infty)$ can be solved using the maximum likelihood method. When $s = 1$, it leads to the so-called Grenander estimator. We refer to the book [GJ14] for its properties and for numerous bibliographical references. Results for density estimation under a condition of s-monotony with $s \geq 2$ can be found in [Bal04, Gao08, GW09, BFW13], see also the references therein.

The simplest solution to generalize the concept of s-monotony when supp $f \not\subset [0, +\infty)$ is probably to assume that f is unimodal and to put the constraint on either side of the mode. Difficulties arise when the mode is unknown because maximizing the likelihood leads to singularities. We refer to [Weg70, Wan95, BF96, Bir97] for some examples of techniques that address this issue. It should be noted, however, that the larger s, the more severe the constraint. When $s \geq 2$, such a function is necessarily irregular in the vicinity of the mode, which is rarely the case for the "usual"

Date: November, 2024.

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

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

distributions. This is why we prefer to put the shape constraint on the tails of f. More formally, we suppose that there are a and b such that $f(a-\cdot)$ and $f(b+\cdot)$ are s-monotone on $(0,+\infty)$.

We supplement this assumption by supposing that f satisfies a smoothness constraint on the whole line $\mathbb{R}$. The only properties of f that are assumed on $[a,b]$ are therefore properties of regularity. The conditions on the tails, on the other hand, are stronger: they are both regular and s-monotone. Note that we can always assume $a=b$ so that the two assumptions apply to $\mathbb{R}$, but this is not necessary.

The smoothness condition we consider is very mild and allows, for example, the regularity of f to be spatially inhomogeneous. These last words mean that f is globally smooth, but strong local variations, discontinuities or even singularities are possible. More formally, we suppose that f belongs to a Besov ball $\mathfrak{B}^{\alpha}_{p,\infty}(R)$, a strong Besov class $\mathcal{B}^{\alpha}_{p,\infty}(R)$ or a weak Besov class $\mathcal{WB}^{\alpha}_{p,\infty}(R)$. The parameter α indicates the global regularity and p the way it is measured. Roughly speaking, the smaller p is, the more spatially inhomogeneous the regularity can be. By way of example, the functions in these sets are Hölder continuous with exponent α when $p=\infty$ and α is not an integer. Conversely, these sets contain densities with discontinuities, and this even when α is large if p is sufficiently small ($p\leq 1/\alpha$ works). In the present paper, we may deal with all values of α,p satisfying $p>0$ and $\alpha\in((1/p-1)_+,\tau)$. In this interval, τ denotes a parameter only depending on the wavelets. This restriction on α,p is very weak and, in some sense, unavoidable. It allows f to be unbounded and to satisfy $\inf_{q\geq q_0}\|f\|_q=+\infty$ for some $q_0>1$ possibly close to 1. The definitions of the above Besov sets may be found in Section 3.1. More details on the notion of inhomogeneous spatial regularity are given in Section 3.2.

We may estimate a smooth density on $\mathbb{R}$ by using a Kernel or a wavelet estimator. In the first case, the whole point is to determine the bandwidth. A fixed bandwidth can be obtained by using a global Lepski rule [Lep15]. If we want the estimator to adapt to the irregularities of f, it is more interesting to let the bandwidth vary with x. A local Lepski rule can then be used, see [GL14, LW19]. As for the wavelet methods, the difficulty lies in deciding which coefficients to keep. Different methods exist in the literature, and we refer to [DJKP96, DJ96, KPT96, HKP98, JLL04, CC05, RBRTM11, HKPT12, Sar24b] for more details. It is beyond the scope of this introduction to describe all the procedures and all the theoretical results of the literature. We simply mention here that convergence rates have been established for Kernel and wavelet estimators when f lies in one the Besov sets above. They are, however, often proved under additional assumptions (such as f compactly supported, f bounded or on α,p) and often involve undesirable logs factors. A summary of the optimal rates (up to logs) in the compact case can be found in [Sar21]. For the non-compact case, we rather refer to [GL14, Lep15] for Kernel estimators and [JLL04, RBRTM11, Sar24b] for wavelets estimators.

Although we have made a distinction between shape and smoothness constraints, these are not unrelated. For instance, a function of bounded variation lies in $\mathfrak{B}^{1}_{1,\infty}(R)$ if R is suitably chosen, see Lemma 9.2 of [DL93]. A unimodal density f necessarily belongs to $\mathcal{WB}^{\alpha}_{p,\infty}(R)$ if it is s-monotone on both sides of the mode (for some values of p,α,R, and if it satisfies $\|f\|_q<\infty$ for some $q>1$, see our Proposition 3 below). We may therefore see the shape constraints as a special case of our regularity assumptions. However, as we will explain later, the results in the two cases do not always coincide.

In the literature, several authors have addressed the issue of estimating a smooth unimodal function. Among the solutions proposed are the possibility of regularizing a shape constrained estimator, or optimizing over a class of shape constrained densities a smoothed version of a classical criterion (such as the log likelihood or the least squares). For more details on these two ideas, we refer to [HH02, VDVVDL03, HK05, LM17]. Alternatively, we may construct a spline estimator that automatically fulfils the shape constraint. This point is supported by the works of [BF96, Mey12, Che20]. The papers we have just mentioned only concern density estimation. We refer to the nice review paper [DL18] for a description of what is done in literature in other statistical frameworks.

An interesting question is whether or not combining the two types of constraint can result in faster convergence rates than those achievable under these same constraints, but not mixed. A theoretical study for the Hellinger loss was carried out in [Sar24a] on this subject. In the compact case, the estimation rate of a smooth function with exponent $\alpha > 1$ cannot be improved by assuming it to be unimodal. The situation is reversed when the density is fat tailed: a faster convergence rate can always be obtained by combining the two assumptions. Things can be more complex for other losses though. In the present paper, we solve this question for the $\mathbb{L}^1$ loss. The superiority of this loss function over the other $\mathbb{L}^q$ losses is due to the fact that f may have an infinite $\mathbb{L}^q$ norm.

To sum up, we model the properties of f through three conditions. We describe them in a rigorous way in Section 3. Here is a rough version.

1. The first condition requires f to be in a strong or weak Besov class $\mathcal{B}^{\alpha}_{p,\infty}(R)$ or $\mathcal{WB}^{\alpha}_{p,\infty}(R)$.
2. The second condition assumes that $f(a-\cdot)$ and $f(b+\cdot)$ are s-monotone on $(0,+\infty)$.
3. The third condition supposes that f is fat-tailed, that is $f(x)$ is at most of the order of $|x|^{-1/\theta}$ for some $\theta \in (0,1)$ when $x \notin [a,b]$.

The above parameters $p, \alpha, s, \theta, a, b, R$ are not assumed to be known.

We propose to use a wavelet estimator to estimate f. For each resolution, we rely on the selection rule of [Sar24b] to determine the coefficients to estimate. The others are set to zero. The main advantage of this rule is that it eliminates log factors in the convergence rates in almost all cases.

In [Sar24b], but this is also true in many preceding papers, the thresholding rule is only applied to non-negative resolutions. We show here that such a strategy necessarily leads to a non-optimal estimator when the three conditions above are met and when θ is large enough, that is when the tails of f tend too slowly to 0. We circumvent this difficulty by thresholding some coefficients in the negative resolutions. All this is explained in detail in Section 2.

We study the global $\mathbb{L}^1$ risk of the resulting estimator $\hat{f}$ under the previous three conditions and compare it to a minimax lower bound. We carry out these new results in Section 4. Here is an overview.

When the shape level s is smaller than the smoothness index α, the rate we get depends on the values of α, s, p, θ. We show that it is possible to split the set $\{(p,\theta),\, p \in (1/(\alpha+1), +\infty], \theta \in (0,1)\}$ into several parts. In some of these parts, the optimal convergence rate is the same with or without shape constraint. In other, however, the convergence rate of $\hat{f}$ is faster than the minimax rate without the shape constraint. The gain can be very substantial. By way of example, we may make the optimal estimation rate without the shape constraint arbitrarily slow by choosing θ close enough to 1. This phenomenon does not occur when the shape constraint is assumed. So, depending

on the values of p, θ, s, α, a shape constraint may, or may not, improve the estimation rate of a smooth density. It should be noted, however, that our rates are always faster than $n^{-s/(2s+1)}$. Consequently, a smoothness condition of level $\alpha > s$ improves the expected estimation rate of a shape constrained density.

We are also interested in the opposite case $s \geq \alpha$. We then show that our estimator attains the standard estimation rate $n^{-\alpha/(2\alpha+1)}$, and this, whatever $\theta \in (0,1)$. The interest of the shape constraint here is that it makes it possible to achieve the estimation rate of the compact case, even when the tails of f tend very slowly to 0.

Finally, we assume f unimodal and set the shape constraint on either side of the mode. We then show that our smoothness assumption is automatically fulfilled with $\alpha = s$ if f belongs to $\mathbb{L}^q$ for some $q \in (1, +\infty]$. The above therefore applies and leads to the convergence rate $n^{-s/(2s+1)}$ for $\hat{f}$, and this, whatever $\theta \in (0,1)$.

Throughout the paper, we suppose $n \geq 3$ and denote by $|A|$ the size of a finite set A. The letters $c, C, c_1, \dots$ denote quantities that may change from line to line.

2. Estimation procedure

2.1. Wavelet estimators. We consider $\tau \in \mathbb{N}^\star$ and a bi-orthogonal wavelet basis with the following properties:

- The father wavelet is $\phi = \mathbf{1}_{[0,1]}$.
- The mother wavelet ψ is piecewise constant.
- The duals $\bar{\phi}$ and $\bar{\psi}$ of ϕ and ψ are compactly supported.
- The duals $\bar{\phi}$ and $\bar{\psi}$ are Hölder continuous with exponent τ.
- The mother wavelet ψ satisfies $\int_{\mathbb{R}} x^k \psi(x)\,\mathrm{d}x = 0$ for all $k \in \{0, \dots, \tau - 1\}$.

We refer to [CDF92] for a construction of such a basis.

We then have the following expansion of f: for all $J_0 \in \mathbb{Z}$,

$$f = \sum_{k \in \mathbb{Z}} \alpha_{J_0,k} \bar{\phi}_{J_0,k} + \sum_{j=J_0}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k} \bar{\psi}_{j,k}, \tag{1}$$

where for any $j \geq J_0$ and $k \in \mathbb{Z}$,

$$\alpha_{J_0,k} = \int f(x) \phi_{J_0,k}(x)\,\mathrm{d}x, \quad \beta_{j,k} = \int f(x) \psi_{j,k}(x)\,\mathrm{d}x,$$

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

$$\phi_{J_0,k}(x) = 2^{J_0/2} \phi(2^{J_0} x - k), \quad \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k),$$
$$\bar{\phi}_{J_0,k}(x) = 2^{J_0/2} \bar{\phi}(2^{J_0} x - k), \quad \bar{\psi}_{j,k}(x) = 2^{j/2} \bar{\psi}(2^j x - k).$$

This formula may serve as a starting point for defining a wavelet estimator. The idea is to replace part of the unknown coefficients $\alpha_{J_0,k}$, and $\beta_{j,k}$ by their empirical versions:

$$\widehat{\alpha}_{J_0,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{J_0,k}(X_i) \quad \text{and} \quad \widehat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} \psi_{j,k}(X_i).$$

The others are treated differently to reduce the variance of the estimator. For simplicity, they are systematically set to zero in this paper. We therefore only focus on estimators of the form

$$\hat{f} = \sum_{k\in\widehat{K}_{J_0,\alpha}} \widehat{\alpha}_{J_0,k}\bar{\phi}_{0,k} + \sum_{j=J_0}^{\infty}\sum_{k\in\widehat{K}_{j,\beta}} \widehat{\beta}_{j,k}\bar{\psi}_{j,k}, \tag{2}$$

where $\widehat{K}_{J_0,\alpha}$ and $\widehat{K}_{j,\beta}$ are finite sets. Determining these sets adequately, using only the data and as little information as possible on f is a difficult problem. But as we will see in the upcoming section, the choice of J_0 should not be overlooked either.

2.2. About the initial level. The classical solution to assess the quality of a wavelet estimator is to study its risk under the assumption that the target is smooth. This notion of "smoothness" generally corresponds to the condition "f belongs to a Besov ball $\mathfrak{B}^{\alpha}_{p,\infty}(R)$" (see Section 3.1 for its definition). The latter imposes a constraint on the size of the coefficients $\alpha_{0,k}$ and $\beta_{j,k}$, $j \geq 0$, $k \in \mathbb{Z}$. Many of them are therefore small (in absolute value) and smaller than the estimation error we would make if we estimated them. It is then best not to estimate them.

It is worth noticing that this constraint only relates to non-negative indexes j. It is therefore natural to start the expansion of f at the level $J_0 = 0$. However, other assumptions, such as some about the shape of the density, give information about all the coefficients $\beta_{j,k}$, not just those for which $j \geq 0$. Not taking this information into account can lead to a non-optimal estimator.

More precisely, the proposition below gives a lower bound on the risk of a wavelet estimator when $J_0 = 0$. It applies regardless of the method used to select the coefficients, i.e. to define the sets $\widehat{K}_{0,\alpha}$, $\widehat{K}_{j,\beta}$. It is proved in Section 5.1.

Proposition 1. *Let $\widehat{K}_{0,\alpha}$ be a (possibly random) subset of $\mathbb{Z}$, and for each $j \geq 0$, $\widehat{K}_{j,\beta}$ be a (possibly random) subset of $\mathbb{Z}$. Consider $k_0 \in \mathbb{N}^\star$ and suppose that f satisfies*

$$c_1|k|^{-1/\theta} \leq f(x) \leq c_2|k|^{-1/\theta} \tag{3}$$

for all $|k| \geq k_0$, $x \in [k-1/2, k+1/2]$ and some $\theta \in (0,1)$, $c_1, c_2 > 0$.

Then, the estimator $\hat{f}$ defined by (2) with $J_0 = 0$ satisfies for all n large enough,

$$\mathbb{E}\left[d_1(f,\hat{f})\right] \geq c_3 n^{-(1-\theta)}, \tag{4}$$

where c_3 is positive and only depends on c_1, c_2, θ.

The rate $n^{-(1-\theta)}$ therefore appears to be an absolute limit for the wavelet methods when $J_0 = 0$. This bound applies for densities satisfying (3), that is for densities f for which $f(x)$ is of the order of $|x|^{-1/\theta}$ when $|x|$ is large.

Unfortunately, the optimal rate of convergence under shape and smoothness constraints may be much faster than $n^{-(1-\theta)}$. For instance, the estimator we propose in this paper estimates f at the rate $\nu_n n^{-\min\{\alpha/(2\alpha+1),(\alpha+\theta)/(\alpha+\theta+2\alpha\theta+2)\}}$ when f is, moreover, unimodal and Hölder continuous with exponent $\alpha > 1$. In this formula, $\nu_n = 1$ if $\theta \neq \alpha/(2\alpha+1)$. We refer to Theorem 6 for more details. This rate is faster than $n^{-(1-\theta)}$ when θ is close enough to 1 and more precisely when $\theta > \theta'$ where $\theta' \in (1/2, 2/3)$ is defined by

$$\theta' = \big(\alpha + (\alpha^2 + 12\alpha + 12)^{1/2} - 2\big)/(2(2\alpha+1)).$$

2.3. Selecting the coefficients. We consider two integers $\widehat{J}_0$ and $\widehat{J}_1$ to be specified later on. They indicate the first and last scale levels. Compared to the previous sections, we have added a hat on J_0 because this parameter may depend on the data.

For each $j \in \{\widehat{J}_0, \dots, \widehat{J}_1\}$, we apply the method of [Sar24b] to determine which coefficients are estimated and which are not. We recall the basic ideas. We define for all $k \in \mathbb{Z}$,

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

We gather all the $k \in \mathbb{Z}$ for which $\widehat{\sigma}_{j,k}^2 \neq 0$ in the set

$$\widehat{\mathbb{Z}}_j = \left\{k \in \mathbb{Z},\, \widehat{\sigma}_{j,k}^2 > 0\right\}. \tag{5}$$

Note that $\widehat{\beta}_{j,k} = 0$ when $k \notin \widehat{\mathbb{Z}}_j$. We may therefore only focus on the k in $\widehat{\mathbb{Z}}_j$ to determine the coefficients to estimate.

We define the criterion $\mathrm{crit}_j(\cdot)$ for all $K_j \subset \widehat{\mathbb{Z}}_j$ by

$$\mathrm{crit}_j(K_j) = -2^{-j/2}\sum_{k\in K_j} |\widehat{\beta}_{j,k}| + \kappa \widehat{\mathcal{E}}_j(K_j), \tag{6}$$

where κ and $\widehat{\mathcal{E}}_j(K_j)$ are to be specified. The first term in this inequality stands for the $\mathbb{L}^1$ norm of the estimated coefficients (up to the minus sign). The more coefficients we estimate, the smaller this term is. The second term represents (an upper bound on) the estimation error of all these coefficients and works in the opposite direction.

We then select the set $\widehat{K}_j$ that achieves the best compromise between these two terms. More precisely, we define $\widehat{K}_j$ as any subset of $\widehat{\mathbb{Z}}_j$ such that

$$\mathrm{crit}_j(\widehat{K}_j) = \min_{K_j \subset \widehat{\mathbb{Z}}_j} \mathrm{crit}_j(K_j). \tag{7}$$

This set $\widehat{K}_j$ does exist as $\widehat{\mathbb{Z}}_j$ is finite.

We finally consider the estimator

$$\hat{f} = \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{\widehat{J}_0,k}\bar{\phi}_{\widehat{J}_0,k} + \sum_{j=\widehat{J}_0}^{\widehat{J}_1}\sum_{k\in\widehat{K}_j} \widehat{\beta}_{j,k}\bar{\psi}_{j,k} \tag{8}$$

of f.

2.4. Bounding the error. The preceding criterion depends heavily on the choice we make for $\widehat{\mathcal{E}}_j(K_j)$. We define it here.

We consider $\varsigma > 0$ to be specified later on. We define for all $j \in \mathbb{Z}$, the largest integer $\bar{r}_{n,j} \geq 0$ such that $2^{-\bar{r}_{n,j}} \geq \min\{1, \varsigma\log(n(|j|+1))/n\}$. We consider for all $r \geq 0$,

$$\begin{aligned}
\widehat{\mathbb{Z}}_{j,r,-} &= \left\{k \in \mathbb{Z},\, \widehat{\sigma}_{j,k}^2 > \|\psi\|_\infty^2 2^{-r}\right\}\\
\widehat{\mathbb{Z}}_{j,r} &= \left\{k \in \mathbb{Z},\, \|\psi\|_\infty^2 2^{-r} < \widehat{\sigma}_{j,k}^2 \leq \|\psi\|_\infty^2 2^{-r+1}\right\}\\
\widehat{\mathbb{Z}}'_{j,r} &= \left\{k \in \mathbb{Z},\, \|\psi\|_\infty^2 2^{-r-2} < \widehat{\sigma}_{j,k}^2 \leq \|\psi\|_\infty^2 2^{-r+3}\right\}
\end{aligned}$$

and put

$$\widehat{\lambda}_{j,r} = \begin{cases} \min\left\{|\widehat{\mathbb{Z}}'_{j,r}|, 2^r\right\} & \text{if } r \le \bar{r}_{n,j} \\ \min\left\{|\widehat{\mathbb{Z}}_j| + \log((1+|j|)n), 2^r\right\} & \text{if } r > \bar{r}_{n,j} \end{cases}$$

where $\widehat{\mathbb{Z}}_j$ is given by (5).

We introduce the maps $\widehat{\mathcal{E}}_{j,r,+}(\cdot)$ and $\widehat{\mathcal{E}}_{j,r}(\cdot)$ for $x \ge 0$ by

$$\widehat{\mathcal{E}}_{j,r,+}(x) = \sqrt{\frac{x}{n}} + \sqrt{\frac{\log((|j|+1)(1+r)) + \log n}{n}}$$

$$\widehat{\mathcal{E}}_{j,r}(x) = x\sqrt{\frac{2^{-r}\log_+\left(|\widehat{\lambda}_{j,r}|/x\right)}{n}} + x\frac{\log_+\left(|\widehat{\lambda}_{j,r}|/x\right)}{n} + \sqrt{x\frac{2^{-r}\log((|j|+1)(1+r)n)}{n}} + \frac{\log((|j|+1)(1+r)n)}{n}.$$

In these formulas, the notation $\log_+(\cdot) = \log(e+\cdot)$ and the convention $0 \times \log_+(\cdot/0) = 0$ apply. We finally set for all $K_j \subset \widehat{\mathbb{Z}}_j$,

$$\widehat{\mathcal{E}}_j(K_j) = \inf_{r_j \ge 0}\left\{\widehat{\mathcal{E}}_{j,r_j,-}(|K_j \cap \widehat{\mathbb{Z}}_{j,r_j,-}|) + \sum_{r=r_j+1}^{\infty}\widehat{\mathcal{E}}_{j,r}(|K_j \cap \widehat{\mathbb{Z}}_{j,r}|)\right\}. \tag{9}$$

The differences between this formula and that of [Sar24b] are mainly due to the fact that j can be negative here.

We would like to mention that the infimum is actually taken over a finite number of integers, and the same thing holds true for the sum. This comes from the fact that $\widehat{\mathbb{Z}}_{j,r} = \widehat{\mathbb{Z}}_j^c$ when r is too large. Moreover, finding the set $\widehat{K}_j$ using (7) and (9) is not computationally difficult. We refer to Section 3.5 of [Sar24b] for more details on this aspect.

2.5. First and last resolution. We now turn our attention to $\widehat{J}_0$ and $\widehat{J}_1$.

Since we estimate all the coefficients associated with the father wavelet, we propose to define $\widehat{J}_0$ small enough so that all the $\widehat{\alpha}_{\widehat{J}_0,k}$ are zero except for a small number of them. Very precisely, we consider $\varrho > 0$ and set

$$\widehat{\mathbb{Z}}_j(\alpha) = \{k \in \mathbb{Z},\, \widehat{\alpha}_{j,k} \ne 0\}. \tag{10}$$

We then put

$$\widehat{J}_0 = \max\left\{j \le 0,\, |\widehat{\mathbb{Z}}_j(\alpha)| \le \varrho\log((1+|j|)n)\right\}. \tag{11}$$

Note that $\widehat{J}_0$ is well defined as the cardinal of $\widehat{\mathbb{Z}}_j(\alpha)$ is at most of the order of n.

The value of $\widehat{J}_1$ is defined as in [Sar24b]. We introduce an integer $L_{\mathrm{wav}} \in \mathbb{N}^\star$ so that $[-L_{\mathrm{wav}}, L_{\mathrm{wav}}]$ contains the support of the four wavelets $\phi, \bar{\phi}, \psi, \bar{\psi}$. We sort the sample $X_{(1)} < \cdots < X_{(n)}$ in

increasing order and define $\widehat{J}_1 \geq 0$ as

$$\widehat{J}_1 = \min_{J\geq 0}\left\{\min_{1\leq i\leq n-1}\left(X_{(i+1)} - X_{(i)}\right) \geq L_{\mathrm{wav}}2^{1-J}\right\}. \tag{12}$$

The heuristic behind this definition is that the estimators of $\beta_{j,k}$ are based on at most one data when $j \geq \widehat{J}_1 + 1$ and are therefore likely poor.

The definition of our estimator $\hat{f}$ is now complete. It still depends on the three parameters $\varsigma, \varrho, \kappa$. They will be specified in Section 4.

3. Classes of functions

We carry out in this section several classes of functions to model our assumptions on the density. There are three kind of conditions we wish to consider. The first one is to guarantee the regularity of f on the real line. The second ensures a particular shape of its tails (they may be monotone, convex...). The last that they are dominated by the inverse of a polynomial function.

3.1. Smoothness condition. We define for $p \in (0, +\infty]$ and $\alpha \in ((1/p-1)_+, \tau)$ the Besov space $\mathcal{B}^\alpha_{p,\infty}$ as the set of functions $f \in \mathbb{L}^{\max\{p,1\}}(\mathbb{R})$ satisfying $\|f\|_{\mathcal{B}^\alpha_{p,\infty}} < \infty$ where

$$\|f\|_{\mathcal{B}^\alpha_{p,\infty}} = \|\alpha_{0,\cdot}\|_p + \sup_{j\geq 0}\left\{2^{j(\alpha+1/2-1/p)}\|\beta_{j,\cdot}\|_p\right\}. \tag{13}$$

The classical Besov ball $\mathfrak{B}^\alpha_{p,\infty}(R)$ is then defined for $R > 0$ by

$$\mathfrak{B}^\alpha_{p,\infty}(R) = \left\{f \in \mathcal{B}^\alpha_{p,\infty},\ \|f\|_{\mathcal{B}^\alpha_{p,\infty}} \leq R\right\}.$$

When $p < 1$, the (quasi) $\mathbb{L}^p$ norm of f is smaller than $\|f\|_{\mathcal{B}^\alpha_{p,\infty}}$, up to a multiplicative factor. In particular, a density whose tails are of the order of $|x|^{-1/\theta}$ does not belong to $\mathfrak{B}^\alpha_{p,\infty}(R)$ when $\theta \geq p$, even if it is very regular. So when we say that $f \in \mathfrak{B}^\alpha_{p,\infty}(R)$, we are not just saying that f is smooth. We also put a condition on its tails.

A solution to decouple the smoothness condition from the tail constraint is not to take into account the coefficients associated with the father wavelet. This brings us to the notion of strong Besov class:

$$\mathcal{B}^\alpha_{p,\infty}(R) = \left\{f \in \mathbb{L}^1(\mathbb{R}),\ \forall j \geq 0,\ \|\beta_{j,\cdot}\|_p \leq R2^{-j(\alpha+1/2-1/p)}\right\}.$$

This set $\mathcal{B}^\alpha_{p,\infty}(R)$ is most useful when $p < 1$. In this case, $\mathcal{B}^\alpha_{p,\infty}(R)$ contains densities that are not in $\mathfrak{B}^\alpha_{p,\infty}(R)$. When $p \geq 1$, the double inequality $\|\alpha_{0,\cdot}\|_p \leq \|\alpha_{0,\cdot}\|_1 \leq c\|f\|_1$ holds true, where c is a term only depending on the wavelets. The densities of $\mathcal{B}^\alpha_{p,\infty}(R)$ then also belong to $\mathfrak{B}^\alpha_{p,\infty}(c' + R)$.

It is possible to weaken the condition $f \in \mathcal{B}^\alpha_{p,\infty}(R)$ by replacing the ℓ^p (quasi) norm $\|\cdot\|_p$ by its weak version $\|\cdot\|_{p,\infty}$. The latter is defined for $p \in (0, +\infty)$ and $x = (x_k)_{k\in\mathbb{Z}}$ by

$$\|x\|_{p,\infty} = \sup_{t>0}\ t\left(\sum_{k\in\mathbb{Z}}\mathbf{1}_{|x_k|\geq t}\right)^{1/p}.$$

When $p=\infty$, $\|x\|_{p,\infty}=\|x\|_\infty$. We recall that $\|x\|_{p,\infty}\le\|x\|_p$ for all $x\in\mathbb{R}^{\mathbb{Z}}$. The weak Besov class then corresponds to

$$\mathcal{WB}^\alpha_{p,\infty}(R)=\left\{f\in\mathbb{L}^1(\mathbb{R}),\,\forall j\ge 0,\ \|\beta_{j,\cdot}\|_{p,\infty}\le R2^{-j(\alpha+1/2-1/p)}\right\}.$$

3.2. Some examples. For pedagogical reasons, we present here a few basic results on Besov classes. We recall in particular their interest in modelling inhomogeneous spatial regularities and justify why we allow p to be very small in our results.

Firstly, Hölder's inequality implies $\|\beta_{j,\cdot}\|_{p,\infty}\le c2^{j(1/p-1/p')}\|\beta_{j,\cdot}\|_{p',\infty}$ for all $(\alpha+1)^{-1}<p\le p'$ and all compactly supported function f on $[0,1]$. The factor c only depends on p,p' and the wavelet basis. We deduce,

$$\big\{f\in\mathcal{WB}^\alpha_{p',\infty}(R),\,\mathrm{supp}f\subset[0,1]\big\}\subset\mathcal{WB}^\alpha_{p,\infty}\big(cR\big).$$

This inclusion roughly means that the condition "f belongs to a Besov class" is milder the smaller p is when we work with compactly supported densities. This claim no longer holds in the non-compact case (see, for example, the minimax rates in Figure 1: they tend to be faster when p is small).

Secondly, a few isolated singularities have no impact on the global smoothness exponent α when p is small enough. This point is stressed by the proposition below. It works for both compactly and non-compactly supported densities and is proved in Section 5.2.

Proposition 2. *Let $p\in(0,+\infty)$, $u\ge 1$, $m_1<\cdots<m_u$, $\alpha\in(1/p-1,1/p)\cap(0,\tau)$, r be the largest integer smaller than α, and f be an integrable map on $\mathbb{R}$. Suppose that f admits r derivatives on $\mathbb{R}\setminus\{m_1,\dots,m_u\}$. Moreover, there is some $A>0$ such that*

$$\left|f^{(r)}(y)-f^{(r)}(x)\right|\ \le A|x-m_\ell|^{-1/p}|y-x|^{\alpha-r}$$

for all $\ell\in\{1,\dots,u\}$ and $x,y\in(m_\ell,m_{\ell+1})$ such that $y\ge x$ (using the convention $m_{u+1}=+\infty$). We suppose that this inequality also holds true with $\ell=1$ for all $x,y<m_1$ such that $y\le x$. Moreover, we assume $f(x)\le A|x-m_\ell|^{-(1/p-\alpha)}$ for all $x\in[m_\ell-\varepsilon,m_\ell+\varepsilon]$, $\ell\in\{1,\dots,u\}$, and some $\varepsilon\in(0,\min_{1\le\ell\le u-1}(m_{\ell+1}-m_\ell))$ (this last condition reduces to $\varepsilon>0$ when $u=1$). Then, f belongs to $\mathcal{WB}^\alpha_{p,\infty}(R)$ where $R=c(A+\|f\|_1)$ and where c only depends on α,p,u,ε and the wavelets.

An interesting point about this proposition is that it can be combined with the forthcoming Theorems 6 and 8 to guarantee a fast estimation rate for the "usual" distributions. We shall go into a little more detail on this aspect later when f is defined for $q>1$, $\theta\in(0,1)$, $\pi_1,\pi_2,\pi_3>0$, and $x\in\mathbb{R}$ by

$$f(x)=\begin{cases}0 & \text{if } x\le 0\\ \pi_1x^{-1/q} & \text{if } x\in(0,1]\\ \pi_2 & \text{if } x\in(1,2]\\ \pi_3x^{-1/\theta} & \text{if } x>2.\end{cases}\tag{14}$$

Above, π_1,π_2,π_3 must be such that f is a density. Note merely here that $f\in\mathcal{WB}^\alpha_{p,\infty}(R)$ for all $p\in(q/(1+q\tau),q)$, $\alpha=1/p-1/q$, and some R only depending on $p,q,\theta,\pi_1,\pi_2,\pi_3$ and the wavelets.

3.3. Shape constraint. We consider a positive integer s and introduce a condition of s-monotony on the tails of f. We define for $b \geq a$,

$$\mathcal{M}^s(a,b) = \{f \in \mathcal{D}, f(a-\cdot) \text{ and } f(b+\cdot) \text{ are } s\text{-monotone on } (0,+\infty)\},$$

where $\mathcal{D}$ designs the set of densities on $\mathbb{R}$. In the particular case where $s = 1$, saying that $f \in \mathcal{M}^s(a,b)$ is equivalent to saying that f is non-decreasing on $(-\infty, a)$ and non-increasing on $(b, +\infty)$. If now $s = 2$, $f \in \mathcal{M}^s(a,b)$ amounts to assuming that f is convex on $(-\infty, a)$ and on $(b, +\infty)$. The condition $f \in \mathcal{M}^s(a,b)$ becomes more stringent as s grows.

Note that this condition does not imply anything about the restriction of f to $[a,b]$. We may therefore work with densities that do not have the same properties on $[a,b]$ and on $\mathbb{R} \setminus [a,b]$. We like to say that the core of the distribution is in $[a,b]$. What is outside then corresponds to the tails of f. Saying that $f \in \mathcal{M}^s(a,b)$ amounts to saying that we put a shape constraint on its tails. Such an interpretation can nevertheless be a bit misleading because it suggests that the probability $\mathbb{P}(X \in [a,b])$ is close to 1. This will never be assumed in the sequel.

Naturally, any compactly supported density belongs to $\mathcal{M}^s(a,b)$ if a and b are far enough away. But the same holds true for most of the "usual" densities with infinite support (such as the Gaussian, Gamma, Inverse Gamma, Cauchy, Log-logistic distributions to name but a few). By way of illustration, suppose that f admits an expression of the form

$$f(x) = \frac{P(x)}{Q(x)} e^{R(x)} \tag{15}$$

when x is large enough. In this formula, P, Q and R are signomials and $Q(x) \neq 0$ when x is large enough (R may be the zero function). The first s-derivatives of f cannot change sign on $(b, +\infty)$ if b is large enough. This property implies that $f(b+\cdot)$ is s-monotone on $(0, +\infty)$ as f is a density. Note that the s-monotony of the left-hand tails of f can be guarantee if we suppose, for example, equality (15) but for $f(-x)$ in place of $f(x)$.

The choice of a, b is up to the statistician and we may take $a = b$. A function $f \in \mathcal{M}^s(a,b)$ is then unimodal. It satisfies a shape constraint of level s on each side of the mode. It is also smooth in the sense that it belongs to a Besov class as the next proposition shows. It is proved in Section 5.3.

Proposition 3. *Consider $s \geq 1$, $\tau \geq s$, $m_0 \in \mathbb{R}$ and a density $f \in \mathcal{M}^s(a,b)$ with $a = b = m_0$. Suppose that f satisfies for all $x \neq m_0$,*

$$f(x) \leq L^{1+1/q} |x - m_0|^{-1/q}, \tag{16}$$

where $L > 0$ and $q \in (1, +\infty]$. Then, f belongs to $\mathcal{B}^s_{p,\infty}(R)$ for all $p \in (1/(s+1), 1/(s+1/q))$, $R = cL^{(1+1/q)(s-1/p+1)/(1-1/q)}$ and c only depending on the wavelet basis, p, q, s. When $q = \infty$, (16) amounts to assuming that f is bounded from above by L.

In this proposition, the aim of (16) is to control the behaviour of f at the vicinity of $x = m_0$. The closer q is to 1, the faster f can tend to infinity. When $q = \infty$, there is no longer any singularity, but f may not be continuous.

Note that the assumptions we put on f are not sufficient to ensure that $f \in \mathfrak{B}^s_{p,\infty}(R)$ or $f \in \mathcal{B}^s_{p,\infty}(R)$ with $p \geq 1$. The function f may not belong to $\mathcal{B}^s_{p,\infty}(R)$ or $\mathfrak{B}^s_{p,\infty}(R)$ when $p \geq 1$ and $sp > 1$

merely because the densities in these sets are continuous. As to the case $p < 1$, f is in $\mathfrak{B}^s_{p,\infty}(R)$ only if a condition on its tails is added (see the preceding section).

3.4. Tail dominance condition. We consider $j, k \in \mathbb{Z}$, and define for any non-negative and integrable function f,

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

We then consider $b \geq a$, $\theta \in [0,1]$, $M > 0$, and the set $\mathcal{WT}_\theta(a,b,M)$ composed of non-negative and integrable functions f satisfying

$$|\{k \in \mathbb{Z}_j(a,b), F_{j,k}(f) \geq t\}| \leq M t^{-\theta} 2^{j(1-\theta)} \tag{17}$$

for all $j \geq 0$ and $t > 0$. In this inequality,

$$\mathbb{Z}_j(a,b) = \mathbb{Z} \cap \left((-\infty, -1/2 + 2^j a] \cup [1/2 + 2^j b, +\infty)\right).$$

As for $\mathcal{M}^s(a,b)$, the condition $f \in \mathcal{WT}_\theta(a,b,M)$ actually concerns the restriction of f to $\mathbb{R}\setminus[a,b]$ and therefore the "tails of f". The aim of (17) is to control the number of coefficients $F_{j,k}(f)$ that are large and to put a dominance constraint on them. Roughly speaking, the smaller θ, M, are, the smaller $F_{j,k}(f)$ must be and the smaller the tails of f should be. The extreme cases for θ are $\theta = 0$ and $\theta = 1$. In the first case, $\mathcal{WT}_0(a,b,M)$ only contains compactly supported functions. In the second case, the constraint disappears: any density belongs to $\mathcal{WT}_1(a,b,1)$.

We may also replace the set $\mathbb{Z}_j(a,b)$ in (17) by $\mathbb{Z}$. The resulting class is denoted by $\mathcal{WT}_\theta(M)$. Following [Sar24b], we say that the "weak tail dominance condition" is met when $f \in \mathcal{WT}_\theta(M)$. This terminology can, a priori, lead to confusion as the entire function is considered for $\mathcal{WT}_\theta(M)$. However, the inclusion $\mathcal{WT}_\theta(M) \subset \mathcal{WT}_\theta(a,b,M)$ is always true. Conversely, we show in Section 5.4:

Proposition 4. *Consider $\theta \in [0,1)$, $b \geq a$, $M > 0$ and a density f. If $f \in \mathcal{WT}_\theta(a,b,M)$, then $f \in \mathcal{WT}_\theta((b-a+2)^{1-\theta} + M)$.*

It is therefore the behaviour of the tails of f that determines whether f lies in $\mathcal{WT}_\theta(M)$ or not. The two sets, $\mathcal{WT}_\theta(M)$ and $\mathcal{WT}_\theta(a,b,M)$, essentially model the same thing. Nonetheless, as we also work with densities of $\mathcal{M}^s(a,b)$, it seems more natural to consider $\mathcal{WT}_\theta(a,b,M)$. From a mathematical point of view, the interest of using different values for a and b lies in the fact that this can reduce M.

We now present a version of Proposition 3.1 of [Sar24a] adapted to our tail condition. A sketch of its proof is given in Section 5.5.

Proposition 5. *Let $\theta \in (0,1)$, $b \geq a$, $M \geq 1$ and f be a density.*

1. Suppose that f satisfies for all $x \notin [a,b]$,

$$f(x) \leq M^{1/\theta}\left[(x-b)^{-1/\theta}\mathbf{1}_{x>b} + (a-x)^{-1/\theta}\mathbf{1}_{x<a}\right]. \tag{18}$$

Then, f belongs to $\mathcal{WT}_\theta(a,b,2M+2)$.

2. Suppose that $f \in \mathcal{WT}_\theta(a,b,M) \cap \mathcal{M}^1(a,b)$. Then, f satisfies

$$f(x) \leq (2M)^{1/\theta}\left[(x-b)^{-1/\theta}\mathbf{1}_{x>b} + (a-x)^{-1/\theta}\mathbf{1}_{x<a}\right] \tag{19}$$

for all $x \notin [a,b]$.

The interest of this result is that it relates the assumption $f \in \mathcal{WT}_\theta(a,b,M)$ to a more tangible condition: f is bounded by a map decreasing as $|x|^{-1/\theta}$ outside the interval $[a,b]$. These notions are equivalent under an additional shape constraint, but not otherwise. Consider indeed $\alpha, \beta > 0$ and

$$f(x) = \pi \sum_{d=1}^{\infty} \mathbf{1}_{\left[2^{\alpha d}, 2^{\alpha d} + 2^{-\beta d}\right]}(x). \tag{20}$$

Above, $\pi > 0$ is defined so that f is a density. We show in Section 5.6 that f belongs to $\mathcal{WT}_\theta(M)$ for all $\theta \in (0,1)$ and some M only depending on α, β, θ. However, (18) is never true uniformly for all $x \notin [a,b]$, whatever a, b, θ, M. This example is perhaps a little artificial. This is due to the fact that it cannot be monotone at infinity since then Proposition 5 would apply. Despite this, it is smooth in the sense that it belongs to the Besov spaces $\mathcal{B}^\alpha_{p,\infty}$ when $p < 1$ and $\alpha \in (1/p - 1, \min\{1/p, \tau\})$.

4. Main results

We study in this section the theoretical properties of our estimator $\hat{f}$. We recall that $\hat{f}$ is defined by (8) where $\widehat{K}_j$ gathers the coefficients we keep at scale 2^j. They are obtained by minimizing the criterion $\mathrm{crit}_j(\cdot)$ defined by (6) in which $\widehat{\mathcal{E}}_j(K_j)$ is defined via (9). This procedure is applied to all levels j between $\widehat{J}_0$ and $\widehat{J}_1$ where $\widehat{J}_0$ is given by (11) and $\widehat{J}_1$ by (12).

We consider the assumptions described in the preceding section. We suppose that f is smooth and its tails are dominated with a certain shape. In Section 4.1 below, the level of the shape constraint is assumed to be smaller than the smoothness index. The reverse case is dealt in Section 4.2.

4.1. When smoothness prevails. We consider $p \in (0, +\infty]$, $s \geq 1$, $\alpha \in (\max\{1/p - 1, s\}, \tau)$, $\theta \in (0,1)$, $R > 0$, $M_1, M_2 \geq 1$, $b \geq a$ and

$$\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2) = \begin{cases} \mathcal{WB}^\alpha_{p,\infty}(R) \cap \mathcal{WT}_\theta(M_1) \cap \mathcal{WT}_\theta(a,b,M_2) \cap \mathcal{M}^s(a,b) & \text{if } p \neq 1 \\ \mathcal{B}^\alpha_{p,\infty}(R) \cap \mathcal{WT}_\theta(M_1) \cap \mathcal{WT}_\theta(a,b,M_2) \cap \mathcal{M}^s(a,b) & \text{if } p = 1. \end{cases}$$

A function f in this class satisfies the three preceding constraints with $s < \alpha$.

We need the following notations:

$$\theta_0 = \frac{\alpha}{2\alpha+1-1/p}$$

$$\gamma = \begin{cases} \dfrac{\alpha}{2\alpha+1} & \text{if } \theta \le \theta_0 \\ \dfrac{\alpha - s\theta/p - \alpha\theta + s\theta + \alpha s\theta}{1-\theta/p+\alpha+s-2s\theta/p+s\theta+2\alpha s\theta} & \text{if } \theta > \theta_0 \end{cases}$$

$$\nu_n = \begin{cases} \log^{2\gamma} n & \text{if } \theta = 1/2 \text{ and } p = 1 \\ \log n & \text{if } \theta = \theta_0 \text{ and } \theta \neq 1/2 \\ 1 & \text{otherwise} \end{cases}$$

$$\beta_1 = \begin{cases} \dfrac{1}{2\alpha+1} & \text{if } \theta \le \theta_0 \\ \dfrac{(s+1)(1-\theta)}{1-\theta/p+\alpha+s-2s\theta/p+s\theta+2\alpha s\theta} & \text{if } \theta > \theta_0 \end{cases}$$

$$\beta_2 = \begin{cases} \dfrac{\alpha+1-1/p}{(1-\theta)(2\alpha+1)} & \text{if } \theta \le \theta_0 \\ \dfrac{(s+1)(1+\alpha-1/p)}{1-\theta/p+\alpha+s-2s\theta/p+s\theta+2\alpha s\theta} & \text{if } \theta > \theta_0 \end{cases}$$

$$M = \begin{cases} M_1 & \text{if } \theta \le \theta_0 \\ M_2 & \text{if } \theta > \theta_0. \end{cases}$$

The theorem below gives an upper-bound of the maximal risk of our estimator when f lies in $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$. It is proved in Section 6.

Theorem 6. *Let $s \ge 1$, $p \in (0,+\infty]$, $\alpha \in (\max\{1/p-1,s\},\tau)$, $\theta \in (0,1)$, $R>0$, $M_1,M_2 \ge 1$ and $b \ge a$. Then, there exist $\varsigma_0, \varrho_0, \kappa_0$, and n_0 such that for all $n \ge n_0$, $\kappa \ge \kappa_0$, $\varrho \ge \varrho_0$, $\varsigma \ge \varsigma_0$,*

$$\sup_{f \in \mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)} \mathbb{E}\left[d_1(f,\hat{f})\right] \le c\left[R^{\beta_1} M^{\beta_2} \nu_n n^{-\gamma} + M_2^{(s+1)/(2s\theta+1)} n^{-(s\theta+1-\theta)/(2s\theta+1)}\right]. \tag{21}$$

Moreover, $\varsigma_0, \varrho_0, \kappa_0$ only depends on the wavelet basis. The term c only depends on $p, s, \alpha, \theta, \kappa, \varrho, \varsigma$ and the wavelet basis, and n_0 only depends on $p, \alpha, \theta, s, \kappa, \varrho, \varsigma, R, M_1, M_2, b-a$ and the wavelet basis.

Before discussing the result, let us note that this theorem provides the (theoretical) values of $\varsigma, \varrho, \kappa$. An important point is that they do not depend on the parameters $\alpha, p, s, \theta, R, M_1, M_2, a, b$ that appear in the definition of the class $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$. The latter can therefore be unknown to the statistician, and our estimator is adaptive.

We would also like to point out that the conditions on p, α are very mild. The parameter p may be smaller than 1 and the condition $\alpha > 1/p-1$ is essentially minimal, see Proposition 4 of [Sar24b] for more details. In particular, we may estimate densities f that possess quite severe singularities (see Propositions 2 or 3).

For information, the proof of this theorem involves an approximation result on s-monotone functions (Lemma 2) that may be of independent interest.

We now propose to use the minimax approach to assess the relevance of (21). We define for all class $\mathcal{F}$ of functions, the minimax risk $\mathcal{R}(\mathcal{F})$ by

$$\mathcal{R}(\mathcal{F}) = \inf_{\tilde{f}} \sup_{f\in\mathcal{F}} \mathbb{E}\left[d_1(f,\tilde{f})\right],$$

where the infimum is taken over all estimators $\tilde{f}$.

The preceding theorem gives an (asymptotic) upper-bound on $\mathcal{R}(\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2))$. It may be compared to the following lower bound (to be proved in Section 7).

Proposition 7. *Consider $s \geq 1$, $p \in (0,+\infty]$, $\alpha > \max\{1/p-1, s\}$, $\theta \in (0,1)$, τ large enough, $R > 0$, $M_1, M_2 \geq 1$ and $b > a$. Then, there exist R_0, M_0 and n_0 such that: for all $R \geq R_0$, $M_1 \geq M_0$, $M_2 \geq M_0$, and $n \geq n_0$,*

$$\mathcal{R}(\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M,M)) \geq cR^{\beta_1}M^{\beta_2}n^{-\gamma}. \tag{22}$$

Moreover, c is positive and only depends on p, s, α, θ and the wavelet basis.

This proposition shows that our estimator is rate optimal except possibly when $\theta = \theta_0$ and $(s\theta+1-\theta)/(2s\theta+1) < \gamma$. If we exclude these two cases, we see that we do not lose any log factors, even when the smoothness condition is very mild. This is a very attractive property as many competing procedures are known to involve undesirable log factors in the results. For more information on this subject, we refer to [Sar24b]. The gap when $\theta = \theta_0$ and $(s\theta+1-\theta)/(2s\theta+1) < \gamma$ may be due to inequality (22) not being sufficiently accurate. For example, there is a log factor in the lower bound when $\theta = \theta_0$ if we remove the shape constraint as shown by Proposition 3 of [Sar24b]. There is also little hope of significantly improving our convergence rate when $(s\theta+1-\theta)/(2s\theta+1) < \gamma$. Indeed, the proof of the proposition actually shows the following inequality: under the conditions of Proposition 7, we have for all $R \geq R_0$, $M_1 \geq M_0$, $M_2 \geq M_0$, and $n \geq n_0$,

$$\mathcal{R}(\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M,M)) \geq Cn^{-\gamma_2}, \tag{23}$$

where

$$\gamma_2 = (\alpha - s\theta/p - \alpha\theta + s\theta + \alpha s\theta)/(1-\theta/p+\alpha+s-2s\theta/p+s\theta+2\alpha s\theta),$$

where C is positive and only depends on $p, s, \alpha, \theta, M_1, M_2, R$ and the wavelet basis. The major difference between this result and (22) lies in the fact that we may use the formula for γ when $\theta > \theta_0$ even when $\theta \leq \theta_0$. Elementary albeit cumbersome computations show that $(s\theta+1-\theta)/(2s\theta+1) < \gamma$ implies $1-\theta/p+s\theta < 0$ (see also the figure below). When this last inequality is fulfilled, the map $\alpha \mapsto \gamma_2$ is non-increasing and tends to $(s\theta+1-\theta)/(2s\theta+1)$ when $\alpha \to +\infty$. As the minimax risk does not increase when α increases, we deduce that for all $\varepsilon > 0$, and τ, n large enough,

$$\mathcal{R}(\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M,M)) \geq C_\varepsilon n^{-(s\theta+1-\theta)/(2s\theta+1)-\varepsilon},$$

where $C_\varepsilon > 0$ only depends on $\varepsilon, p, s, \alpha, \theta, M_1, M_2, R$ and the wavelet basis.

In Theorem 6 and Proposition 7, the parameters a and b can be as close or as far away as we wish. It is noteworthy that our asymptotic bounds do not depend on them, except through M_2 (roughly speaking, M_2 can always be taken not larger than M_1, and the further apart a, b are, the smaller M_2 is). The condition $b \neq a$ that appears in the proposition is important as it guarantees $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M,M) \neq \emptyset$. Think, for example, of a density f of $\mathcal{M}^2(0,0)$. It is not differentiable at 0 and does not therefore belong to $\mathcal{B}^\alpha_{\infty,\infty}(R)$ if $\alpha > 1$.

We now introduce the minimax risk $\mathcal{R}(\mathscr{F}^{\alpha}_{p,\theta}(R,M))$ on

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

The main difference between $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$ and $\mathscr{F}^{\alpha}_{p,\theta}(R,M)$ is that we have deleted the shape constraint for the second set. We can then compare the two minimax risks to measure the impact of this condition on the optimal rates of convergence.

The following result is proved in [Sar24b]: for all $p \in (0,+\infty]$, $\alpha \in (1/p-1,\tau)$, $\theta \in (0,p]$, $R>0$, $M \geq 1$, and n large enough,

$$\mathcal{R}\left(\mathscr{F}^{\alpha}_{p,\theta}(R,M)\right) \leq c\left[R^{\beta_{1,s}} M^{\beta_{2,s}} \nu_n n^{-\gamma_s} + M n^{-(1-\theta)}\right],$$

where

$$\gamma_s = \begin{cases} \alpha/(2\alpha+1) & \text{if } \theta \leq \theta_0 \\ \alpha(1-\theta)/(\alpha+1-\theta/p) & \text{if } \theta > \theta_0, \end{cases}$$

where $\beta_{1,s}, \beta_{2,s}$ only depend on α, p, θ ($\beta_1 = \beta_{1,s}$ and $\beta_2 = \beta_{2,s}$ when $\theta \leq \theta_0$). The reverse inequality also holds true, at least when M and R are large enough. Although the assumption $\theta \leq p$ appears in [Sar24b], the latter is actually not necessary and the above remains true even when $\theta \in [p,1)$ (the arguments are the same).

We present two graphs below to compare the minimax rates more easily. The graph on the left refers to $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$ and the one on the right to $\mathscr{F}^{\alpha}_{p,\theta}(R,M)$. The parameter p is placed on the x-axis and θ on the y-axis. The dotted area in the graph on the left indicates that the optimal rate is at least as fast as that shown (up to log factors at the junction point $\theta = \theta_0 = (\alpha+1)/(2\alpha-s+1)$). Everywhere else, the formula corresponds to the minimax rate (up to logs when $\theta = \theta_0$).

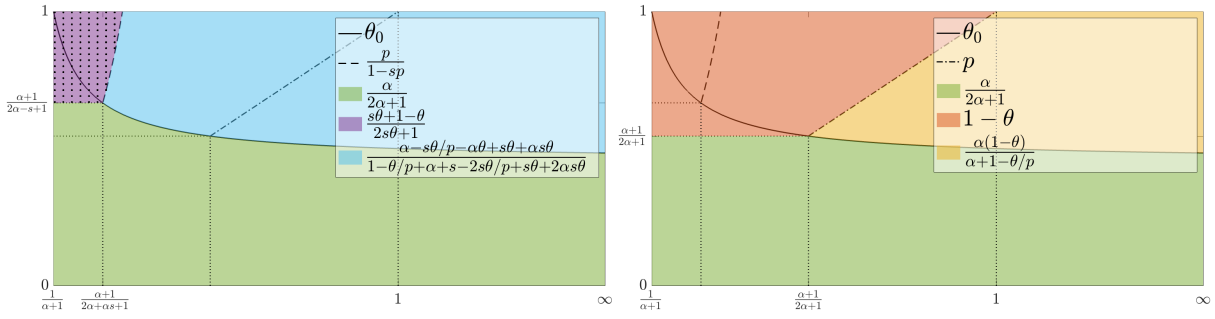

FIGURE 1. Comparison of the rates of convergence with and without shape constraint on the tails of the distribution

The first thing to note is that a large green area is common to both graphs. Inside this zone, the optimal rate is the standard one $n^{-\alpha/(2\alpha+1)}$. The shape assumption adds nothing here.

There is also a green area on the left which is red on the right. This region corresponds to the set

$$\mathcal{A} = \{(p,\theta) \in (0,+\infty] \times (0,1),\, \alpha > 1/p - 1, \theta > (\alpha+1)/(2\alpha+1), \\ \theta < \min\left\{\alpha/(2\alpha+1-1/p), (\alpha+1)/(2\alpha-s+1)\right\}\}.$$

When $(p,\theta) \in \mathcal{A}$, the shape assumption allows to recover the classical rate of convergence $n^{-\alpha/(2\alpha+1)}$. Without it, the minimax rate is much slower. This rate is recovered even when $s = 1$, i.e. when f is monotone before a and after b. The parameter s still has a role to play, since the larger it is, the wider the zone $\mathcal{A}$.

Colours other than green never match between the two graphs. When the rate is not the standard one, taking into account the shape constraint hence leads to an improvement in the estimation of f. This improvement can be very significant: for example, the rate can be made arbitrarily slow by choosing θ very close to 1 in the graph on the right. This phenomenon does not occur when there is the shape constraint: the rate gets closer to $n^{-s/(2s+1)}$ when $\theta \to 1$. It is therefore easier, in some sense, to estimate shape constrained tails than smooth tails.

When $a = b$, the entire distribution satisfies a shape and smoothness constraint. It is worth asking whether our approach might not lead to better results in this case than those we would have under a pure shape constraint that would ignore its regularity. The answer is positive: wherever we look on the graph on the left, the rate is always in the range $[n^{-\alpha/(2\alpha+1)}, n^{-s/(2s+1)})$. This exponent $s/(2s+1)$ corresponds to the expected rate of convergence of a bounded compactly supported density $f \in \mathcal{M}^s(0,0)$. More details on the estimation of a s-monotone function under the global $\mathbb{L}^1$ loss can be found in [Bir89, DL12]. Hellinger results are to be found in [Gao08, GW09].

The estimation of a density f of $\mathfrak{B}^\alpha_{p,\infty}(R)$ when $p < 1$ may also be improved by assuming a shape constraint on its tails. We recall in this respect that the weak tail dominance condition is satisfied with $\theta = p$ and M of the order of R^p when $f \in \mathfrak{B}^\alpha_{p,\infty}(R)$ and $p < 1$. This point has been shown in Proposition 1 of [Sar24b]. Consequently: for all $s \geq 1$, $p \in (0,1)$, $\alpha \in (\max\{1/p-1, s\}, \tau)$, $R \geq 1$, $b \geq a$, and n large enough,

$$\sup_{f \in \mathfrak{B}^\alpha_{p,\infty}(R) \cap \mathcal{M}^s([a,b])} \mathbb{E}\left[d_1(f,\hat{f})\right] \leq cR^\beta \nu_n n^{-\gamma}.$$

The terms ν_n and γ can here be written as

$$\gamma = \begin{cases} \dfrac{\alpha}{2\alpha+1} & \text{if } p \in (0,1/2] \text{ or } \alpha < \alpha_0 \\ 1 - p + \dfrac{sp(2p-1)\,(\alpha-\alpha_0)}{\alpha - s + sp(2\alpha+1)} & \text{if } p \in (1/2,1) \text{ and } \alpha \geq \alpha_0 \end{cases}$$

$$\nu_n = \begin{cases} \log n & \text{if } p \in (1/2,1) \text{ and } \alpha = \alpha_0 \\ 1 & \text{otherwise} \end{cases}$$

and $\alpha_0 = (1-p)/(2p-1)$. Besides,

$$\beta = \begin{cases} \dfrac{\alpha p}{(2\alpha+1)(1-p)} & \text{if } p \in (0,1/2] \text{ or } \alpha < \alpha_0 \\ \dfrac{\alpha p(s+1)}{\alpha - s + sp(2\alpha+1)} & \text{if } p \in (1/2,1) \text{ and } \alpha \geq \alpha_0. \end{cases}$$

This result can be compared with (16) in [Sar24b] where there is no shape constraints. The conclusion is that we obtain a better rate of convergence when $p \in (1/2, 1)$ and $\alpha > \alpha_0$ (and the same otherwise).

4.2. Higher degree of shape. In the previous section, we assumed that the shape degree s was smaller than the smoothness index α. We now turn to the other case $s \geq \alpha$. We take $M_1 = M_2$ and denote these two numbers by M. We then set

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

The theorem below is proved in Section 6 and gives an upper-bound on the (asymptotic) maximal risk of our estimator when f lies in $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M)$.

Theorem 8. *Let $p \in (0, +\infty]$, $\alpha \in ((1/p-1)_+, \tau)$, $s \in [\alpha, \tau)$, $\theta \in (0,1)$, $R > 0, M \geq 1$, $b \geq a$. Suppose $b \neq a$ if $s > \alpha$. Then, there exist $\varsigma_0, \varrho_0, \kappa_0$, and n_0 such that for all $n \geq n_0$, $\kappa \geq \kappa_0$, $\varrho \geq \varrho_0$, $\varsigma \geq \varsigma_0$,*

$$\sup_{f \in \mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M)} \mathbb{E}\left[d_1(f, \hat{f})\right] \leq C n^{-\alpha/(2\alpha+1)}, \tag{24}$$

where

$$C = \begin{cases} cR^{1/(2\alpha+1)}(b-a)^{(\alpha+1-1/p)/(2\alpha+1)} & \text{if } s > \alpha \\ c\left[R^{1/(2\alpha+1)}(b-a)^{(\alpha+1-1/p)/(2\alpha+1)} + \log_+(R) + \log_+(M)\right] & \text{if } s = \alpha, \end{cases}$$

and where c only depends on $p, s, \alpha, \theta, \kappa, \varrho, \varsigma$ and the wavelet basis. Moreover, $\varsigma_0, \varrho_0, \kappa_0$ only depends on the wavelet basis and n_0 only depends on $p, \alpha, \theta, s, \kappa, \varrho, \varsigma, R, M, b-a$ and the wavelet basis.

We thus recover the estimation rate of a compactly supported density of smoothness α. However, f is not supposed to be compactly supported. The parameter θ can even be arbitrarily close to 1. The whole point of this theorem lies in the fact that it applies to densities whose tails tend very slowly to 0. As Proposition 1 shows, such a result is beyond the reach of wavelet methods that only deal with non-negative resolutions as soon as $\theta > (\alpha+1)/(2\alpha+1)$.

The term $R^{1/(2\alpha+1)}(b-a)^{(\alpha+1-1/p)/(2\alpha+1)}$ in the factor in front of $n^{-\alpha/(2\alpha+1)}$ corresponds to what one would expect when estimating a compactly supported density $f \in \mathcal{B}^{\alpha}_{p,\infty}(R)$ on $[a,b]$. The sum $\log_+(R) + \log_+(M)$ when $\alpha = s$ can be seen as the price to pay for estimating its tails.

We now illustrate this theorem by applying it to the density f defined by (14). Consider $\tau \geq 2$, $\alpha = \tau - 1$, and set $1/p = \alpha + 1/q$. Note that f lies in $\mathscr{F}^{\alpha,\alpha}_{p,\theta}(0,2,R,M)$ for some R, M. We therefore get the estimation rate $n^{-\alpha/(2\alpha+1)} = n^{-(\tau-1)/(2\tau-1)}$. In particular, if we choose a large value of τ, the exponent in the rate can be made as close as $1/2$ as we wish. This estimation rate can thus be very fast, even if f has a peak at 0, and that its tails tend very slowly to 0.

It is also possible to deal with pure shape constraints. Consider indeed $q \in (1, +\infty]$, $\theta \in (0,1)$, $L, M \geq 1$, $s \in [1, \tau)$, $m_0 \in \mathbb{R}$, and

$$\mathcal{S}^s_{q,\theta}(m_0, L, M) = \Big\{ f \in \mathcal{M}^s(m_0, m_0),\ \forall x \neq m_0,\\ f(x) \leq \min\left\{ L^{1+1/q} |x - m_0|^{-1/q}, M^{1/\theta} |x - m_0|^{-1/\theta} \right\} \Big\}.$$

A function f in $\mathcal{S}^s_{q,\theta}(m_0, L, M)$ therefore satisfies the assumption of s-monotony. More precisely, $f(m_0 + \cdot)$ and $f(m_0 - \cdot)$ are s-monotone on $(0, +\infty)$. At the junction point m_0, f may admit a singularity up to the order $1/q$. Its tails decreases as $|x|^{-1/\theta}$.

We deduce from Propositions 3 and 5,

$$\mathcal{S}^s_{q,\theta}(m_0, L, M) \subset \mathscr{F}^{s,s}_{p,\theta}(m_0, m_0, R, M'),$$

where $p \in (1/(s+1), 1/(s+1/q))$, where R is of the order of $L^{(1+1/q)(s-1/p+1)/(1-1/q)}$, and where M' is of the order of M. The above theorem therefore ensures that $\hat{f}$ estimates the densities of $\mathcal{S}^s_{q,\theta}(m_0, L, M)$ at the rate $n^{-s/(2s+1)}$. More precisely: for all n large enough,

$$\sup_{f \in \mathcal{S}^s_{q,\theta}(m_0, L, M)} \mathbb{E}\left[d_1(f, \hat{f}) \right] \leq c \left[\log_+ L + \log_+ M \right] n^{-s/(2s+1)}, \tag{25}$$

where c only depends on $q, s, \theta, \kappa, \varrho, \varsigma$ and the wavelet basis.

It may be interesting to compare the performance of our estimator with that of an estimator dedicated to shape constraints, such as that of Grenander. The latter is perfectly suited to estimating unimodal densities (with known mode), which corresponds in our paper to a shape constraint of level $s = 1$. Naturally, the Grenander estimator does not have the same versatility as ours, and can be a disaster if the assumed shape constraint is not true. Risk bounds for the $\mathbb{L}^1$ loss can nevertheless be established when f is unimodal. The following result is due to [Bir89]: for all n large enough, the Grenander estimator $\hat{f}_{\mathrm{gre}}$ of a unimodal density with mode at 0 satisfies

$$\sup_{f \in \mathcal{S}^1_{\infty,0}(0, L, M)} \mathbb{E}\left[d_1(f, \hat{f}_{\mathrm{gre}}) \right] \leq c \left(\log(1 + LM) \right)^{1/3} n^{-1/3},$$

where

$$\mathcal{S}^1_{\infty,0}(0, L, M) = \left\{ f \in \mathcal{M}^1(0,0),\ \|f\|_\infty \leq L,\ \mathrm{supp}\ f \subset [-M, M] \right\}.$$

This risk bound for the Grenander estimator is therefore similar to ours. The rate is the same. In both cases, L and M contribute logarithmically to the factor in front of $n^{-1/3}$. More surprising is that the exponent on this log factor is not the same.

There are three possible explanations for this difference. The first is that our underlying calculations are not sufficiently accurate. The second is that our estimator is worse than the Grenander estimator. The third is more involved. A careful analysis of the proofs of Theorems 6 and 8 indicates that (21) and (24) remain true when we replace $\mathbb{E}[d_1(f, \hat{f})]$ by $\mathbb{E}[\|f - \hat{f}\|_{\mathcal{B}^0_{1,1}}]$. In other terms, we may replace the $\mathbb{L}^1$ loss by a suitable Besov norm (see Section 3.2 of [Sar24b] for instance for its definition). We have $\|f\|_1 \leq \|f\|_{\mathcal{B}^0_{1,1}}$ but the reverse inequality is not true in general. A difference in the minimax results is therefore possible.

5. Proofs of Sections 2 and 3

5.1. Proof of Proposition 1. The result relies on the following lemma:

Lemma 1. *Let $\widehat{K}_{0,\alpha}$ be a (possibly random) subset of $\mathbb{Z}$, and for each $j \geq 0$, $\widehat{K}_{j,\beta}$ be a (possibly random) subset of $\mathbb{Z}$. Then, the estimator $\hat{f}$ defined by (2) satisfies*

$$\mathbb{E}\left[d_1(f,\hat{f})\right] \geq (4/27) \sum_{\substack{k\in\mathbb{Z} \\ \alpha_{0,k}\leq 1/n}} \alpha_{0,k}.$$

As f satisfies (3), we have

$$c_1 k^{-1/\theta} \leq \alpha_{0,k} \leq c_2 k^{-1/\theta}$$

for all $k \geq k_0$, up to a modification of c_1, c_2. We deduce from the lemma that for all n large enough,

$$\mathbb{E}\left[d_1(f,\hat{f})\right] \geq (4c_1/27) \sum_{k\geq (c_2 n)^\theta} k^{-1/\theta} \geq c_3 n^{-(1-\theta)}.$$

□

Proof of Lemma 1. We first observe that for all $k, k' \in \mathbb{Z}$ and $j \geq 0$,

$$\int \bar{\psi}_{j,k}(x)\mathbf{1}_{[k',k'+1]}(x)\,\mathrm{d}x = 0 \tag{26}$$

$$\int \bar{\phi}_{0,k}(x)\mathbf{1}_{[k',k'+1]}(x)\,\mathrm{d}x = \mathbf{1}_{k=k'} \tag{27}$$

as the basis is bi-orthogonal with $\phi = \mathbf{1}_{[0,1]}$.

Consider now the set

$$A = \bigcup_{\substack{k\in\mathbb{Z} \\ \widehat{\alpha}_{0,k}\mathbf{1}_{k\in\widehat{K}_{0,\alpha}}\geq\alpha_{0,k}}} [k,k+1].$$

We deduce from (1) with $J_0 = 0$, and from (26), (27),

$$\begin{aligned} d_1(f,\hat{f}) &\geq \left|\int_A \hat{f} - \int_A f\right| \\ &\geq \sum_{k\in\mathbb{Z}} \left(\widehat{\alpha}_{0,k}\mathbf{1}_{k\in\widehat{K}_{0,\alpha}} - \alpha_{0,k}\right)_+. \end{aligned}$$

The same argument with

$$A = \bigcup_{\substack{k\in\mathbb{Z} \\ \widehat{\alpha}_{0,k}\mathbf{1}_{k\in\widehat{K}_{0,\alpha}}\leq\alpha_{0,k}}} [k,k+1]$$

leads to

$$d_1(f,\hat{f}) \geq \sum_{k\in\mathbb{Z}} \left(\alpha_{0,k} - \widehat{\alpha}_{0,k}\mathbf{1}_{k\in\widehat{K}_{0,\alpha}}\right)_+.$$

By putting the two inequalities together,

$$2d_1(f,\hat{f}) \geq \sum_{k\in\widehat{K}_{0,\alpha}} |\widehat{\alpha}_{0,k} - \alpha_{0,k}| + \sum_{k\notin\widehat{K}_{0,\alpha}} \alpha_{0,k}.$$

In particular,

$$2d_1(f,\hat{f}) \geq \sum_{\substack{k\in\mathbb{Z}\\ \alpha_{0,k}\leq 1/n}} \alpha_{0,k}\mathbf{1}_{\widehat{\alpha}_{0,k}=0}.$$

Therefore,

$$\begin{aligned} 2\mathbb{E}\left[d_1(f,\hat{f})\right] &\geq \sum_{\substack{k\in\mathbb{Z}\\ \alpha_{0,k}\leq 1/n}} \alpha_{0,k}\left(1-\alpha_{0,k}\right)^n \\ &\geq (8/27) \sum_{\substack{k\in\mathbb{Z}\\ \alpha_{0,k}\leq 1/n}} \alpha_{0,k} \end{aligned}$$

when $n \geq 3$ (which is assumed throughout the paper). □

5.2. Proof of Proposition 2. Consider $j \geq 0$, $k \in \mathbb{Z}$, $\ell \in \{1,\dots,u\}$, $I_{j,k} = [2^{-j}(k-L_{\mathrm{wav}}), 2^{-j}(k+L_{\mathrm{wav}})]$, $K_{j,\ell} = [L_{\mathrm{wav}} + 2^j m_\ell + 1, -L_{\mathrm{wav}} + 2^j m_{\ell+1} - 1]$ with the convention $m_{u+1} = +\infty$.

Suppose $r \geq 1$ and $k \in K_{j,\ell}$. We deduce from Taylor's theorem that there exists a polynomial expansion Q of order $r-1$ satisfying

$$f(x) - Q(x) = \frac{1}{(r-1)!}\int_{2^{-j}(k-L_{\mathrm{wav}})}^{x} (x-t)^{r-1}\ f^{(r)}(t)\ \mathrm{d}t$$

when $x \in I_{j,k}$. The polynomial function

$$P(x) = Q(x) + \frac{f^{(r)}\left(2^{-j}(k-L_{\mathrm{wav}})\right)}{(r-1)!}\int_{2^{-j}(k-L_{\mathrm{wav}})}^{x} (x-t)^{r-1} \quad \mathrm{d}t$$

is of degree at most $r \leq \tau - 1$. It satisfies for all $x \in I_{j,k}$,

$$\begin{aligned} &|f(x) - P(x)| \\ &\leq A\left[(r-1)!\right]^{-1}\left[2^{-j}(k-L_{\mathrm{wav}}) - m_\ell\right]^{-1/p}\int_{2^{-j}(k-L_{\mathrm{wav}})}^{x} (x-t)^{r-1}\left[t - 2^{-j}(k-L_{\mathrm{wav}})\right]^{\alpha-r}\ \mathrm{d}t. \end{aligned}$$

In particular,

$$\sup_{x\in I_{j,k}} |f(x) - P(x)| \leq c_1 A 2^{-j(\alpha-1/p)}\left|k - L_{\mathrm{wav}} - 2^j m_\ell\right|^{-1/p}.$$

This last inequality also holds true when $r = 0$ by choosing suitably P. Moreover,

$$\begin{aligned} |\beta_{j,k}| &\leq c_2 2^{j/2}\int_{2^{-j}(k-L_{\mathrm{wav}})}^{2^{-j}(k+L_{\mathrm{wav}})} |f(x) - P(x)|\ \mathrm{d}x \\ &\leq c_3 2^{-j/2} \sup_{x\in I_{j,k}} |f(x) - P(x)| \\ &\leq c_4 A 2^{-j(\alpha+1/2-1/p)}\left|k - L_{\mathrm{wav}} - 2^j m_\ell\right|^{-1/p}. \end{aligned}$$

We deduce from all of this: for all $\ell \in \{1, \dots, u\}$ and $k \in K_{j,\ell}$,

$$|\beta_{j,k}|^p \leq c_5 A^p 2^{-jp(\alpha+1/2-1/p)} \left|k - L_{\text{wav}} - 2^j m_\ell\right|^{-1}.$$

By doing a similar reasoning: for all $k \leq -L_{\text{wav}} + 2^j m_1 - 1$,

$$|\beta_{j,k}|^p \leq c_6 A^p 2^{-jp(\alpha+1/2-1/p)} \left|k + L_{\text{wav}} - 2^j m_1\right|^{-1}.$$

When $k \in [-L_{\text{wav}} + 2^j m_\ell - 1, L_{\text{wav}} + 2^j m_\ell + 1]$ and j large enough to ensure $2^{-j}(2L_{\text{wav}} + 1) \leq \varepsilon$,

$$\begin{aligned} |\beta_{j,k}|^p &\leq c_7 A^p 2^{jp/2} \left(\int_{-2^{-j}(2L_{\text{wav}}+1)}^{2^{-j}(2L_{\text{wav}}+1)} |x|^{-1/p+\alpha} \, \mathrm{d}x \right)^p \\ &\leq c_8 A^p 2^{-jp(\alpha+1/2-1/p)}. \end{aligned}$$

When $2^{-j}(2L_{\text{wav}} + 1) > \varepsilon$, we merely say that

$$|\beta_{j,k}|^p \leq c_9 2^{jp/2} \|f\|_1^p \leq c_{10} \|f\|_1^p 2^{-jp(\alpha+1/2-1/p)}.$$

Gathering all these inequalities yields

$$\|\beta_{j,\cdot}\|_{p,\infty}^p \leq c_{11}(1 + \|f\|_1^p) 2^{-jp(\alpha+1/2-1/p)}$$

as wished. □

5.3. Proof of Proposition 3.

5.3.1. An approximation lemma. The following lemma is an essential result that is not only used in the proof of Proposition 3, but also, later, in the proofs of Lemma 12 and Proposition 9.

Lemma 2. *Suppose that $f \in \mathcal{M}^s([a,b])$ with $b \geq a$, $s \geq 1$, $\tau \geq s$ and $m \in \mathbb{R}$. Then, for all $\ell \geq 1$ and all $k_0 \geq 3L_{wav} + (s-1)\ell + b2^j$,*

$$\sum_{k \geq k_0} 2^{-j/2} |\beta_{j,k}| \leq c\ell^{1-s} f_{j,k_0-(s-1)\ell-2L_{wav}}.$$

Moreover, for all $k_0 \leq -3L_{wav} - (s-1)\ell + a2^j$,

$$\sum_{k \leq k_0} 2^{-j/2} |\beta_{j,k}| \leq c\ell^{1-s} f_{j,k_0+(s-1)\ell+2L_{wav}},$$

Above, c only depends on s and the wavelet basis, $f_{j,k}$ is defined by (35) and we recall that $L_{wav} \in \mathbb{N}^\star$ is defined so that $[-L_{wav}, L_{wav}]$ contains the support of $\phi, \bar{\phi}, \psi, \bar{\psi}$.

Its proof is based on the elementary result below whose proof is postponed at the end of the section.

Lemma 3. *Let $a < b$, $s \geq 2$ and f be a $s-2$ times differentiable function on $[a,b]$ such that $f^{(s-2)}$ is either convex or concave. Then, there is a polynomial function of degree at most $s-1$ such that*

$$\int_a^b |f(x) - P(x)| \, \mathrm{d}x \leq c(b-a)^s \left| f_l^{(s-1)}(b) - f_r^{(s-1)}(a) \right|.$$

In this inequality, c only depends on s and $f_l^{(s-1)}$, $f_r^{(s-1)}$ denote the left and right derivatives of $f^{(s-2)}$.

Proof of Lemma 2. The proof when $s = 1$ is almost the same as the proof of Lemma 9 of [Sar24a] and is therefore skipped. We suppose from now on that $s \geq 2$. We moreover assume that s is even, the case s odd can be treated similarly.

Consider $k \geq k_0$ and apply Lemma 3. There is a polynomial function P_k of degree at most $s-1$ such that

$$\int_{2^{-j}(k-L_{\mathrm{wav}})}^{2^{-j}(k+L_{\mathrm{wav}})} |f(x) - P_k(x)| \,\mathrm{d}x \leq c_1 2^{-js} \left[f^{(s-1)}(2^{-j}(k+L_{\mathrm{wav}})) - f^{(s-1)}(2^{-j}(k-L_{\mathrm{wav}}))\right].$$

In this formula, $f^{(s-1)}$ may denote either the left or right derivative of $f^{(s-2)}$ according to the reader's wishes. In particular,

$$\begin{aligned} 2^{-j/2}|\beta_{j,k}| &= 2^{-j/2} \left| \int (f(x) - P_k(x)) \, \psi_{j,k}(x) \,\mathrm{d}x \right| \\ &\leq c_2 \int_{2^{-j}(k-L_{\mathrm{wav}})}^{2^{-j}(k+L_{\mathrm{wav}})} |f(x) - P_k(x)| \,\mathrm{d}x \\ &\leq c_3 2^{-js} \left[f^{(s-1)}(2^{-j}(k+L_{\mathrm{wav}})) - f^{(s-1)}(2^{-j}(k-L_{\mathrm{wav}}))\right]. \end{aligned}$$

Therefore,

$$\sum_{k \geq k_0} 2^{-j/2}|\beta_{j,k}| \leq c_4 2^{-js} |f^{(s-1)}(2^{-j}(k_0 - L_{\mathrm{wav}}))|. \tag{28}$$

The assumption of s-monotony implies:

$$\begin{aligned} |f^{(s-j+1)}(2^{-j}(k-L_{\mathrm{wav}}))| &\leq 2^j \ell^{-1} \left[|f^{(s-j)}(2^{-j}(k-L_{\mathrm{wav}}-\ell))| - |f^{(s-j)}(2^{-j}(k-L_{\mathrm{wav}}))|\right] \\ &\leq 2^j \ell^{-1} |f^{(s-j)}(2^{-j}(k-L_{\mathrm{wav}}-\ell))|. \end{aligned}$$

Therefore, by induction,

$$|f^{(s-1)}(2^{-j}(k-L_{\mathrm{wav}}))| \leq (2^j \ell^{-1})^{s-1} f(2^{-j}(k-L_{\mathrm{wav}}-(s-1)\ell)).$$

Let us now observe that $\operatorname{supp} \psi \supset [L_1, L_2]$ for some $L_1, L_2 \in \mathbb{R}$ such that $-L_{\mathrm{wav}} \leq L_1 < L_2 \leq L_{\mathrm{wav}}$. For all $k \in \mathbb{Z}$ such that $k \geq 2^j b + 2L_{\mathrm{wav}} - L_1$,

$$f_{j,k-2L_{\mathrm{wav}}} \geq c_5 2^{-j} f(2^{-j}(k + L_2 - 2L_{\mathrm{wav}})) \geq c_5 2^{-j} f(2^{-j}(k - L_{\mathrm{wav}})).$$

Therefore, if k satisfies $k \geq 2^j b + 3L_{\mathrm{wav}} + (s-1)\ell \geq 2^j b + 2L_{\mathrm{wav}} - L_1 + (s-1)\ell$,

$$|f^{(s-1)}(2^{-j}(k-L_{\mathrm{wav}}))| \leq c_6 2^{js} \ell^{-(s-1)} f_{j,k-(s-1)\ell-2L_{\mathrm{wav}}}.$$

We conclude using (28). □

Proof of Lemma 3. Depending on the parity of s, $f^{(s-2)}$ is either convex or concave. We suppose throughout the proof that $f^{(s-2)}$ is convex, the proof in the other case is similar. We set

$$\upsilon = \frac{f^{(s-2)}(b) - f^{(s-2)}(a)}{b-a},$$

and define for all $x \in [a, b]$,

$$\begin{aligned} g_1(x) &= f^{(s-2)}(a) + \upsilon(x-a) \\ g_2(x) &= f^{(s-2)}(a) + f_r^{(s-1)}(a)(x-a). \end{aligned}$$

Note that

$$g_2(x) \leq f^{(s-2)}(x) \ \leq g_1(x)$$

for all $x \in [a,b]$. We set $P = g_1$ when $s = 2$ and

$$P(x) = \sum_{k=0}^{s-3} \frac{f^{(k)}(a)(x-a)^k}{k!} + \int_a^x \frac{g_1(t)}{(s-3)!}(x-t)^{s-3} \ \mathrm{d}t$$

when $s \geq 3$. In both cases, P is a polynomial function of degree at most $s-1$. Moreover, $f(x) \leq P(x)$ for all $x \in [a,b]$. When $s \geq 3$, we have:

$$\begin{aligned} P(x) - f(x) &\leq \int_a^x \frac{g_1(t) - g_2(t)}{(s-3)!}(x-t)^{s-3} \ \mathrm{d}t \\ &\leq ((s-3)!)^{-1} \left[\upsilon - f_r^{(s-1)}(a)\right] \int_a^x (x-t)^{s-3} \ (t-a) \, \mathrm{d}t \\ &\leq \left((s-3)! \times (s^2 - 3s + 2)\right)^{-1} \left[\upsilon - f_r^{(s-1)}(a)\right] (x-a)^{s-1}. \\ &\leq c \left[f_l^{(s-1)}(b) - f_r^{(s-1)}(a)\right] (x-a)^{s-1}, \end{aligned}$$

where c only depends on s. This last inequality also holds true when $s = 2$. The result then follows from integration. □

5.3.2. Proof of Proposition 3. We define

$$\dot{\mathbb{Z}}_{j,r,-} = \left\{k \in \mathbb{Z}, \, f_{j,k} \ > 2^{-r}\right\},$$

where $f_{j,k}$ is given by (35). We begin by showing:

Lemma 4. *Under the assumptions of Proposition 3, we have when $q < \infty$: for all $j \in \mathbb{Z}$ and $r > 0$,*

$$|\dot{\mathbb{Z}}_{j,r,-}| \leq c \left[1 + L^{1+q} 2^{rq + j(1-q)}\right],$$

where c only depends on q and the wavelet basis.

Proof of Lemma 4. We have for all $k \geq L_{\mathrm{wav}} + 2^j m_0 + 1$,

$$\begin{aligned} f_{j,k} &\leq \int_{2^{-j}(k - L_{\mathrm{wav}})}^{2^{-j}(k + L_{\mathrm{wav}})} L^{1+1/q} |x - m_0|^{-1/q} \, \mathrm{d}x \\ &\leq c_1 L^{1+1/q} 2^{-j(1-1/q)} \left[k - L_{\mathrm{wav}} - 2^j m_0\right]^{-1/q}. \end{aligned}$$

Therefore, the number of $k \geq L_{\mathrm{wav}} + 2^j m_0 + 1$ such that $f_{j,k} > 2^{-r}$ is not larger than $c_2 L^{1+q} 2^{rq + j(1-q)}$. The same reasoning applies when $k \leq -L_{\mathrm{wav}} + 2^j m_0 - 1$. The number of $k \in \mathbb{Z}$ satisfying this inequality is therefore not larger than $c_3[1 + L^{1+q} 2^{rq + j(1-q)}]$. □

Proof of Proposition 3. We consider for $j, r \geq 1$,

$$z_{j,r} = \begin{cases} 1 + \min\left\{L^{1+q} 2^{rq + j(1-q)}, 2^r\right\} & \text{if } q < \infty \\ 1 + 2^r & \text{if } q = \infty. \end{cases}$$

We define $c_s \geq s + 2L_{\mathrm{wav}}$ large enough so that $|\dot{\mathbb{Z}}_{j,r,-}| \leq c_s z_{j,r}$. The existence of c_s is guaranteed by the above lemma and by the fact that f is a density. We define $r_j \geq 1$ when $q < \infty$

and $L^{1+q}2^{j(1-q)} < 1$ as the smallest number satisfying $L^{1+q}2^{r_j q + j(1-q)} \geq 1$. When $q = \infty$ or $L^{1+q}2^{j(1-q)} \geq 1$, $r_j = 1$. Note that $z_{j,r_j} = 3$ when $q = \infty$ and $z_{j,r_j} \in [1, 1+2^q]$ otherwise.

We define for all $r \geq r_j$ the smallest integer $\bar{k}_{j,r}$ satisfying

$$\bar{k}_{j,r} \geq L_{\mathrm{wav}} + 2^j m_0 + c_s \sum_{r'=1}^{r} z_{j,r'}$$

and the largest integer $\underline{k}_{j,r}$ such that

$$\underline{k}_{j,r} \leq -L_{\mathrm{wav}} + 2^j m_0 - c_s \sum_{r'=1}^{r} z_{j,r'}.$$

As f is unimodal with mode at m_0, $f_{j,k} \leq 2^{-r}$ for all $k \geq \bar{k}_{j,r}$ or $k \leq \underline{k}_{j,r}$. Moreover, as f satisfies (16), we have for all $k \in \mathbb{Z}$,

$$\begin{aligned} f_{j,k} &\leq L^{1+1/q} \sup_{t\in\mathbb{R}} \int_t^{t+2^{1-j}L_{\mathrm{wav}}} |x|^{-1/q} \, \mathrm{d}x \\ &\leq c_1 L^{1+1/q} 2^{-j(1-1/q)}. \end{aligned}$$

Note that for all $r \geq r_j$,

$$\begin{aligned} \bar{k}_{j,r+1} - \bar{k}_{j,r} &\geq 2L_{\mathrm{wav}} + (s-1) z_{j,r+1} \\ \underline{k}_{j,r} - \underline{k}_{j,r+1} &\geq 2L_{\mathrm{wav}} + (s-1) z_{j,r+1} \end{aligned}$$

as $c_s \geq s + 2L_{\mathrm{wav}}$ and $z_{j,r+1} \geq 1$. We apply Lemma 2 with ℓ defined as the integer part of $z_{j,r+1}$. As $\ell \geq z_{j,r+1}/2$, we have

$$2^{-j/2} \sum_{\substack{k \geq \bar{k}_{j,r+1} \\ k \leq \underline{k}_{j,r+1}}} |\beta_{j,k}| \leq c_2 z_{j,r+1}^{1-s} \min\left\{ 2^{-r}, L^{1+1/q} 2^{-j(1-1/q)} \right\}.$$

Suppose $q < \infty$ and apply Hölder's inequality:

$$\begin{aligned} &\sum_{r=r_j+1}^{\infty} \sum_{\substack{k \in [\bar{k}_{j,r}, \bar{k}_{j,r+1}] \\ \text{or } k \in [\underline{k}_{j,r+1}, \underline{k}_{j,r}]}} |\beta_{j,k}|^p \\ &\qquad \leq c_3 \sum_{r=r_j+1}^{\infty} z_{j,r+1}^{1-ps} 2^{jp/2} 2^{-rp} \\ &\qquad \leq c_4 \sum_{r=r_j+1}^{\infty} 2^{jp/2} \min\left\{ 2^{-rp(1+s-1/p)}, \left(L^{1+q}2^{j(1-q)}\right)^{1-ps} 2^{rqp(1/p-s-1/q)} \right\} \\ &\qquad \leq c_5 L^{p(1+1/q)(s-1/p+1)/(1-1/q)} 2^{-jp(s+1/2-1/p)}. \end{aligned} \tag{29}$$

When $q=\infty$, we rather have

$$\sum_{r=r_j+1}^{\infty} \sum_{\substack{k\in[\bar{k}_{j,r},\bar{k}_{j,r+1}]\\ \text{or } k\in[\underline{k}_{j,r+1},\underline{k}_{j,r}]}} |\beta_{j,k}|^p \le c_6 \sum_{r=r_j+1}^{\infty} 2^{r(1-ps)} 2^{jp/2} \min\left\{2^{-rp}, L^{p(1+1/q)} 2^{-jp(1-1/q)}\right\}$$

$$\le c_7 L^{p(1+1/q)(s-1/p+1)/(1-1/q)} 2^{-jp(s+1/2-1/p)}. \tag{30}$$

Observe now that the number of $k\in\mathbb{Z}$ in the interval $[\underline{k}_{j,r_j+1},\bar{k}_{j,r_j+1}]$ can be bounded from above by a term only depending on the wavelets and q,s. Therefore,

$$\begin{aligned}\sum_{k\in[\underline{k}_{j,r_j+1},\bar{k}_{j,r_j+1}]} |\beta_{j,k}|^p &\le c_8 \sup_{k\in[\underline{k}_{j,r_j+1},\bar{k}_{j,r_j+1}]} |\beta_{j,k}|^p \\ &\le c_9 2^{jp/2} \sup_{k\in[\underline{k}_{j,r_j+1},\bar{k}_{j,r_j+1}]} f_{j,k}^p. \\ &\le c_{10} \min\left\{2^{jp/2}, L^{p(1+1/q)} 2^{-jp(1/2-1/q)}\right\} \\ &\le c_{11} L^{p(1+1/q)(s-1/p+1)/(1-1/q)} 2^{-jp(s+1/2-1/p)}. \qquad (31)\end{aligned}$$

Inequalities (29), (30) and (31) ensure that f belongs to the Besov class. □

5.4. Proof of Proposition 4. For all $t>0$,

$$|\{k\in\mathbb{Z}, F_{j,k}(f)\ge t\}| \le t^{-\theta} \sum_{k\in(2^j a-1/2,2^j b+1/2)} F_{j,k}^{\theta}(f) + |\{k\in\mathbb{Z}_j(a,b), F_{j,k}(f)\ge t\}|.$$

Now,

$$\begin{aligned}\sum_{k\in(2^j a-1/2,2^j b+1/2)} F_{j,k}^{\theta}(f) &\le \left(\sum_{k\in(2^j a-1/2,2^j b+1/2)} F_{j,k}(f)\right)^{\theta} \left((b-a)2^j+2\right)^{1-\theta} \\ &\le (b-a+2)^{1-\theta} 2^{j(1-\theta)}.\end{aligned}$$

This ends the proof. □

5.5. Sketch of the proof of Proposition 5. For all $k\ge 1/2+2^j b$,

$$\begin{aligned}F_{j,k}(f) &\le M^{1/\theta} \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} (x-b)^{-1/\theta}\,\mathrm{d}x \\ &\le M^{1/\theta} 2^{-j} \left(2^{-j}(k-1/2)-b\right)^{-1/\theta}.\end{aligned}$$

As $F_{j,k}(f)\le 1$, we only need to control the number of k such that $F_{j,k}(f)\ge t$ where $t\in(0,1]$. Besides, the number of $k\ge 1/2+2^j b$ such that $F_{j,k}(f)\ge t$ is bounded by $M2^{j(1-\theta)}t^{-\theta}+1\le (1+M)2^{j(1-\theta)}t^{-\theta}$. We do the same reasoning when $k\le -1/2+2^j a$ to get $f\in\mathcal{WT}_\theta(a,b,2M+2)$.

As to (19), we do as in the proof of Proposition 3.1 of [Sar24a]. Consider $t>0$ and $x>b$ such that $f(x)\ge t$. The proof of [Sar24a] ensures that x must satisfy $x<b+2Mt^{-\theta}$. In particular, $f(b+2Mt^{-\theta})\le t$ for all $t>0$. This gives (19) when $x>b$. The proof when $x<a$ is similar. □

5.6. Proof of the counter-example in Section 3.4. We show here the assertion below Proposition 5.

We have for all $j, k \in \mathbb{Z}$,

$$F_{j,k}\left(\mathbf{1}_{\left[2^{\alpha d}, 2^{\alpha d}+2^{-\beta d}\right]}\right) \leq \min\left\{2^{-\beta d}, 2^{-j}\right\}.$$

The left-hand side is even zero when $2^{\alpha d} \geq 2^{-j}(k+1/2)$ or $2^{\alpha d}+2^{-\beta d} \leq 2^{-j}(k-1/2)$. We now define

$$A = \sum_{k\neq 0} F^{\theta}_{j,k}(f)$$

where f is given by (20). As f is a density,

$$\sum_{k\in\mathbb{Z}} F^{\theta}_{j,k}(f) \leq 1 + A.$$

Moreover,

$$\begin{aligned}
A &\leq \pi^{\theta} \sum_{k\neq 0}\sum_{d\geq 1} F^{\theta}_{j,k}\left(\mathbf{1}_{\left[2^{\alpha d}, 2^{\alpha d}+2^{-\beta d}\right]}\right) \\
&\leq \pi^{\theta} \sum_{k\neq 0}\sum_{d\geq 1} \min\left\{2^{-\beta d\theta}, 2^{-j\theta}\right\} \mathbf{1}_{2^{\alpha d+j}-1/2\leq k\leq 2^{\alpha d+j}+2^{-\beta d+j}+1/2} \\
&\leq \pi^{\theta} \sum_{d\geq 1} \min\left\{2^{-\beta d\theta}, 2^{-j\theta}\right\}\left(2^{-\beta d+j}+2\right) \\
&\leq c 2^{j(1-\theta)}.
\end{aligned}$$

□

6. Proofs of Theorems 6 and 8

We suppose throughout this section that f lies in the set $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$. The proof of Theorem 6 corresponds to the case $s<\alpha$. The proof of Theorem 8 corresponds to the case $s\geq\alpha$, $M_1 = M_2 = M$.

6.1. Preliminary results. We gather here preliminary results that are needed to show the theorems. Their proofs are deferred to Section 6.3. The first lemma can be deduced from Lemma 12 of [Sar24b] plus a union bound.

Lemma 5. *For all $\xi > 0$ and probability $1-e^{-\xi}$: for all $j\in\mathbb{Z}$,*

$$|\widehat{\mathbb{Z}}_j(\alpha)| \leq c\left[\mathbb{E}[|\widehat{\mathbb{Z}}_j(\alpha)|] + \log(1+|j|) + \xi\right], \tag{32}$$

where c is universal and where we recall that $\widehat{\mathbb{Z}}_j(\alpha)$ is defined by (10).

We now need:

Lemma 6. *For all $j \in \mathbb{Z}$, $t > 0$*

$$\left|\left\{k \in \mathbb{Z}, \alpha_{j,k} \geq t2^{j/2}\right\}\right| \leq c_1 \left[1 + (b-a)2^j + M_2 2^{j(1-\theta)} t^{-\theta}\right]. \tag{33}$$

Likewise,

$$|\{k \in \mathbb{Z},\, f_{j,k} \geq t\}| \leq c_2 \left[1 + (b-a)2^j + M_2 2^{j(1-\theta)} t^{-\theta}\right], \tag{34}$$

where

$$f_{j,k} = \int f(x) \mathbf{1}_{supp\ \psi_{j,k}}(x)\, \mathrm{d}x. \tag{35}$$

Moreover, for all $q \geq 1$,

$$|\{k \in \mathbb{Z},\, f_{j,k} \geq t\}| \leq c_3 \|f\|_q^q t^{-q} 2^{-j(q-1)}. \tag{36}$$

Above, c_1, c_2 only depend on θ and the wavelet basis, and c_3 only depends on q and the wavelet basis. Moreover, c_3 increases when q grows up.

The lemma below follows from the arguments developed in the proofs of Lemmas 22 and 23 of [Sar24b].

Lemma 7. *For all $j \in \mathbb{Z}$,*

$$\mathbb{E}\left[|\widehat{\mathbb{Z}}_j(\alpha)|\right] \leq c \left[1 + (b-a)2^j + M_2 n^\theta 2^{j(1-\theta)}\right], \tag{37}$$

where c only depends on θ.

We now carry out a more involved result that gives the values of ς_0 and κ_0 that appear in the theorems. In this lemma,

$$\sigma_{j,k}^2 = \int f(x)\psi^2(2^j x - k)\, \mathrm{d}x$$

for all $j, k \in \mathbb{Z}$.

Lemma 8. *There exists $\varsigma_0 > 0$ only depending on the wavelets such that: for all $\varsigma \geq \varsigma_0$ there exists an event $\mathcal{A}$ of probability $1 - n^{-4}$ on which the following holds true: for all $j \in \mathbb{Z}$, $r \leq \bar{r}_{n,j}$, and $k \in \widehat{\mathbb{Z}}_{j,r,-}$,*

$$\frac{\sigma_{j,k}^2}{2} \leq \widehat{\sigma}_{j,k}^2 \leq 2\sigma_{j,k}^2. \tag{38}$$

Besides, for all $K_j \subset \mathbb{Z}$,

$$\mathcal{E}_j(\alpha, K_j) \leq c \left[\sqrt{\frac{|K_j|}{n}} + \sqrt{\frac{\log((1+|j|)n)}{n}}\right], \tag{39}$$

where

$$\mathcal{E}_j(\alpha, K_j) = 2^{-j/2} \sum_{k \in K_j} |\widehat{\alpha}_{j,k} - \alpha_{j,k}|, \tag{40}$$

and where c only depends on the wavelet basis. Moreover, for all $K_j \subset \widehat{\mathbb{Z}}_j$,

$$\mathcal{E}_j(K_j) \leq (\kappa_0/2)\widehat{\mathcal{E}}_j(K_j), \tag{41}$$

where

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

and where κ_0 only depends on the wavelet basis.

We are now interested in the thinnest and coarsest resolutions $\widehat{J}_1$ and $\widehat{J}_0$. Lemmas 9 and 10 below roughly say that $\widehat{J}_1$ and $|\widehat{J}_0|$ are of the order of $\log n$. Lemma 10 also provides the value of ϱ_0 that appears in Theorems 6 and 8. The proof of Lemma 9 is omitted as it follows from Lemmas 17 and 25 of [Sar24b].

Lemma 9. *We have $\mathbb{E}\big[\widehat{J}_1^2\big] \leq c\log^2 n$ where c only depends on α, p, R and the wavelets.*

Lemma 10. *There exists ϱ_0 such that if $\varrho \geq \varrho_0$,*

$$|\widehat{J}_0| \leq c_1 \left[\log_+(b-a) + \log_+\left(M_2 n^\theta\right)\right] \tag{42}$$

with probability $1 - n^{-4}$. We also have

$$\mathbb{E}\big[\widehat{J}_0^2\big] \leq c_2 \left[\log_+^2(b-a) + \log_+^2\left(M_2 n^\theta\right)\right] \tag{43}$$

Above, c_1, c_2 only depend on θ, ϱ and the wavelets.

The lemma below deals with the estimation errors associated with the father wavelet.

Lemma 11. *We have,*

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z})\right] \leq c \Bigg[&\sqrt{\frac{\log\left[n\left(\log_+\left(b-a\right) + \log_+\left(M_2 n^\theta\right)\right)\right]}{n}} \\ &+ \frac{\log^2\left[n\left(\log_+\left(b-a\right) + \log_+\left(M_2 n^\theta\right)\right)\right]}{n}\Bigg],\end{aligned} \tag{44}$$

where c only depends on ϱ.

The quantities

$$\widehat{T} = \sum_{j=0}^{\widehat{J}_1} 2^{-j/2} \sum_{k\in\mathbb{Z}\setminus\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}} |\beta_{j,k}| + \sum_{j=\widehat{J}_1+1}^{\infty} 2^{-j/2} \sum_{k\in\mathbb{Z}} |\beta_{j,k}| \tag{45}$$

$$\widehat{\bar{T}} = \sum_{j=\widehat{J}_0}^{-1} 2^{-j/2} \sum_{k\in\mathbb{Z}\setminus\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}} |\beta_{j,k}| \tag{46}$$

are involved in our analysis. We present below a lemma to bound them.

Lemma 12. *On the event $\mathcal{A}$ on which the results of Lemmas 8 and 10 hold true,*

$$\widehat{T} \le c\Big[R^{1/(1+\alpha)}(b-a+1)^{(1+\alpha-1/p)/(\alpha+1)}\left(\log n+\log_+\log_+ R\right)^{\alpha/(1+\alpha)} n^{-\alpha/(1+\alpha)} + (b-a+1)(\log n+\log_+\log_+ R)n^{-1} + (\log^2 n+\log_+^2 R)n^{-1}\Big] \tag{47}$$

$$\widehat{\widehat{T}} \le c\frac{\big(b-a+\log_+(M_2n^\theta)\big)\log\big[n\big(\log_+(b-a)+\log_+(M_2n^\theta)\big)\big]}{n} \tag{48}$$

where c only depends on $\alpha, p, \theta, \varsigma, \varrho$ and the wavelets.

We define for $j \in \mathbb{Z}$, and $K_j \subset \widehat{\mathbb{Z}}_j$,

$$U_j(K_j) = B_j(K_j) + \kappa\widehat{\mathcal{E}}_j(K_j)$$
$$B_j(K_j) = 2^{-j/2}\sum_{k\in\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}\setminus K_j}|\beta_{j,k}|.$$

We then put

$$U_j = \inf_{K_j\subset\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}} U_j(K_j). \tag{49}$$

The last proposition of this section is designed to provide a bound on U_j when f is in $\mathscr{F}^{\alpha,s}_{p,\theta}(a,b,R,M_1,M_2)$. It is lengthy because many cases have to be considered.

Proposition 9. *The following points hold true on the event $\mathcal{A}$ on which the results of Lemmas 8 and 10 are valid. In these inequalities,*

$$u_j = (\log n)\frac{\log((|j|+1)n)}{n^{1/2}},$$

and v_j, w_j are non-negative numbers satisfying

$$\sum_{j=0}^{\infty} v_j \le R^{1/(2\alpha+1)}M_1^{(1+\alpha-1/p)/((2\alpha+1)(1-\theta))}n^{-\alpha/(2\alpha+1)}$$
$$\sum_{j=0}^{\infty} w_j \le R^{1/(2\alpha+1)}(b-a)^{(1+\alpha-1/p)/(2\alpha+1))}n^{-\alpha/(2\alpha+1)}.$$

Besides, c only depends on $\alpha, p, \theta, s, \varrho, \varsigma, \kappa$ and the wavelets.

1. If $\theta \le 1/2$ and $j \le -1$,

$$U_j \le c\left[\left\{1+(b-a)2^j + M_2 2^{j(1-\theta)}\left(1+(\log n)\mathbf{1}_{\theta=1/2}\right)\right\}n^{-1/2}+u_j\right]. \tag{50}$$

2. If $\theta > 1/2$ and $j \le -1$ such that $2^{2js(1-\theta)} \le nM_2^{-2s}$,

$$U_j \le c\Big[M_2^{(s+1)/(2s\theta+1)}n^{(\theta-s\theta-1)/(2s\theta+1)}2^{j(1-\theta)(s+1)/(2s\theta+1)} + \left(1+(b-a)2^j\right)n^{-1/2}+u_j\Big]. \tag{51}$$

3. If $\theta \le 1/2$, and $j \ge 0$,

$$U_j \le c\left[M_1^{1/(2(1-\theta))}\left(1+(\log n)\mathbf{1}_{\theta=1/2}\right)2^{j/2}n^{-1/2}+u_j\right]. \tag{52}$$

4. If $\theta = 1/2$, $p = 1$, and $j \geq 0$,

$$U_j \leq c\left[R2^{-j\alpha} + u_j\right]. \tag{53}$$

5a. If $\theta \geq 1/2$, $p < 1$, and $j \geq 0$ such that

$$M_1^{(1-2/p)/(1-\theta)} R^2 n \geq 2^{j(2\alpha+1)},$$

then

$$\begin{aligned} U_j \leq c\Big\{ & R^{-(2\theta-1)/(1-2\theta/p)} M_1^{(1-1/p)/(1-2\theta/p)} n^{-\theta(1/p-1)/(2\theta/p-1)} 2^{-j(\theta-\alpha-\theta/p+2\alpha\theta)/(2\theta/p-1)} \\ & \times \left(1 + \log_+\left(M_1^{1-2/p} R n^{1/2} 2^{-j(\alpha+1/2)}\right) \mathbf{1}_{\theta=1/2}\right) + u_j + v_j \Big\}. \end{aligned} \tag{54}$$

The same inequality is true when $\theta < 1/2$, $p \geq 1$ and $j \geq 0$ such that

$$2^{j(2\alpha+1)} \geq R^2 M_1^{(1-2/p)/(1-\theta)} n.$$

5b. If $\theta > 1/2$, $p < 1$, and $j \geq 0$ such that

$$M_1^{(1-2\theta/p)/(1-\theta)} M_2^{-2/p} R^2 n \geq 2^{j(2\alpha+1)},$$

then

$$\begin{aligned} U_j \leq c\Big\{ & R^{-(2\theta-1)/(1-2\theta/p)} M_2^{(1-1/p)/(1-2\theta/p)} n^{-\theta(1/p-1)/(2\theta/p-1)} 2^{-j(\theta-\alpha-\theta/p+2\alpha\theta)/(2\theta/p-1)} \\ & \times \log_+\left((b-a)^{(2\theta-p)/(p\theta)} R^{-2} M_2^{1/\theta} n^{-1} 2^{j(2\alpha+1)}\right) \\ & + M_1^{-1/(2(1-\theta))} (b-a) 2^{j/2} n^{-1/2} + u_j + v_j \Big\}. \end{aligned} \tag{55}$$

5c. If $\theta < 1/2$, $p \geq 1$, and $j \geq 0$ such that

$$2^{j(2\alpha+1)} \geq R^2 M_1^{(1-2\theta/p)/(1-\theta)} M_2^{-2/p} n,$$

then

$$\begin{aligned} U_j \leq c\Big[& R^{-(2\theta-1)/(1-2\theta/p)} M_2^{(1-1/p)/(1-2\theta/p)} n^{-\theta(1/p-1)/(2\theta/p-1)} 2^{-j(\theta-\alpha-\theta/p+2\alpha\theta)/(2\theta/p-1)} \\ & + R(b-a)^{1-1/p} 2^{-j\alpha} + (b-a) M_2^{-1/(p-2\theta)} R^{p/(p-2\theta)} n^{\theta/(p-2\theta)} 2^{-j(\theta+\alpha p)/(p-2\theta)} \\ & + u_j + v_j \Big]. \end{aligned} \tag{56}$$

6. If $p < 1$ and $j \geq 0$,

$$\begin{aligned} U_j \leq c\Big[& R^p M_1^{-(1-p)/(2(1-\theta))} n^{-(1-p)/2} 2^{-jp(\alpha+1/2-1/(2p))} \\ & \times \log_+\left(M_1^{-(p-2)/(2(1-\theta))} R^{-p} n^{-p/2} 2^{jp(\alpha+1/2)}\right) + u_j + v_j \Big]. \end{aligned} \tag{57}$$

7a. If $\theta > 1/2$ and $j \geq 0$ such that $2^{2j(1-\theta)s} \leq n M_2^{-2s}$,

$$\begin{aligned} U_j \leq c\Big[& M_2^{(s+1)/(2s\theta+1)} n^{(\theta-s\theta-1)/(2s\theta+1)} 2^{j(1-\theta)(s+1)/(2s\theta+1)} \\ & + \left(2^{-j} + (b-a)\right) M_1^{-1/(2(1-\theta))} \log_+\left(M_1^{1/(1-\theta)} 2^j\right) \log_+\left(\frac{M^{1/(1-\theta)} 2^j}{1+(b-a)2^j}\right) 2^{j/2} n^{-1/2} \\ & + u_j + v_j + w_j \Big]. \end{aligned} \tag{58}$$

7b. If $\theta > 1/2$ and $j \geq 0$ such that $2^{2j(1-\theta)s} \leq nM_1^{-2s}$,

$$U_j \leq c\left[M_1^{(s+1)/(2s\theta+1)} n^{(\theta-s\theta-1)/(2s\theta+1)} 2^{j(1-\theta)(s+1)/(2s\theta+1)} + u_j + w_j\right]. \tag{59}$$

7c. If $j \geq 0$ is such that $2^{-2j(q-1)s} \leq nA_q^{-2qs}$, where $q \geq 1$ and A_q are such that $A_q \geq \|f\|_q$, then

$$U_j \leq c_q\left[A_q^{q(s+1)/(2sq+1)} 2^{-j(s+1)(q-1)/(2qs+1)} n^{-(1+sq-q)/(2sq+1)} + u_j + w_j\right]. \tag{60}$$

Here, c_q only depends on $\alpha, p, q, s, \varrho, \varsigma$ and the wavelets. It does not depend on q if $q \in [1,2]$.

7d. If $\theta = 1/2$, $p > 1$, and $j \geq 0$ such that $2^{js} \leq nM_2^{-2s}$ and

$$2^{2j(\alpha-1/(2p)+1)} \geq R^2 M_2^{-2/p} n,$$

then

$$\begin{aligned} U_j \leq c\Big[& M_2 2^{j/2} n^{-1/2} \log_+\left(R^{2(s+1)} M_2^{-2(s+1/p)} n^{2+s-1/p} 2^{-j(2-2s/p+2\alpha+3s+2\alpha s-1/p)}\right) \\ & + R\log_+\left(R^{-2} M_2^{2/p} n^{-1} 2^{2j(\alpha-1/(2p)+1)}\right)\left(2^{-j(\alpha+1-1/p)} + (b-a)^{1-1/p} 2^{-j\alpha}\right) \\ & + \left(2^{-j} + (b-a)\right) 2^{-j(2\alpha p+1)/(2(p-1))} R^{p/(p-1)} M_2^{-1/(p-1)} n^{1/(2(p-1))} \\ & \qquad \times \log_+\left(\frac{R^{-2p/(p-1)} M_2^{2/(p-1)} 2^{2jp(\alpha-1/(2p)+1)/(p-1)} n^{-p/(p-1)}}{1+(b-a)2^j}\right) \\ & + u_j + v_j + w_j\Big]. \end{aligned} \tag{61}$$

8. If $p \geq 1$ and $j \geq 0$ such that $2^{j(1-\theta/p+\alpha-s+s\theta)} \geq RM_2^{s-1/p}$,

$$\begin{aligned} U_j \leq c\Big[& R^{(s\theta-\theta+1)/(1-\theta/p+s\theta)} M_2^{(1-1/p)/(1-\theta/p+s\theta)} 2^{-j(\alpha-s\theta/p-\alpha\theta+s\theta+\alpha s\theta)/(1-\theta/p+s\theta)} \\ & + R\left((b-a)^{1-1/p} 2^{-j\alpha} + 2^{-j(\alpha+1-1/p)}\right) \log_+\left(2^{j(1-\theta/p+\alpha-s+s\theta)} R^{-1} M_2^{1/p-s}\right) \\ & + u_j + v_j + w_j\Big]. \end{aligned} \tag{62}$$

6.2. Proofs of Theorems 6 and 8. The proposition below establishes a general risk bound for our estimator. It involves the notations introduced in Section 6.1. It is proved at the end of the section.

Proposition 10. *Suppose that $f \in \mathscr{F}_{p,\theta}^{\alpha,s}(a,b,R,M_1,M_2)$ for some $p \in (0,+\infty]$, $s \in [1,\tau)$, $\alpha \in ((1/p-1)_+,\tau)$, $\theta \in (0,1)$, $R > 0$, $M_1, M_2 \geq 1$, $b \geq a$. Let $\hat{f}$ be the estimator defined by (8) for some $\varsigma \geq \varsigma_0$, $\kappa \geq \kappa_0$, where ς_0, κ_0 are given by Lemma 8 and where ϱ_0 is given by Lemma 10. Then,*

$$\mathbb{E}\left[d_1\big(f,\hat{f}\big)\right] \leq c\left\{\mathbb{E}\left[\sum_{j=\widehat{J}_0}^{\widehat{J}_1} U_j \mathbf{1}_{\mathcal{A}}\right] + \mathbb{E}\left[\left(\widehat{T}+\widehat{\widehat{T}}\right)\mathbf{1}_{\mathcal{A}}\right] + \mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha,\mathbb{Z})\mathbf{1}_{\mathcal{A}}\right] + C(\log n)/n^2\right\}, \tag{63}$$

where $\mathcal{A}$ is an event on which the results of Lemmas 8 and 10 are true, where $\widehat{T}$ and $\widehat{\widehat{T}}$ are given by (45) and (46), where U_j is defined by (49), and where the error term $\mathcal{E}_{\widehat{J}_0}(\alpha,\mathbb{Z})$ is defined by (40). Moreover, c only depends on the wavelets. As to C, it only depends on $\alpha, p, \theta, \rho, R, M_1, M_2, b-a$ and the wavelets.

Thereby, we only need to control the right-hand side of (63) to show the theorem. We apply Lemmas 11 and 12 to get a suitable bound on

$$\mathbb{E}\left[\left(\widehat{T}+\widehat{\widehat{T}}\right)\mathbf{1}_{\mathcal{A}}\right]+\mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha,\mathbb{Z})\mathbf{1}_{\mathcal{A}}\right].$$

We now use Proposition 9 to deal with

$$\mathbb{E}\left[\sum_{j=\widehat{J}_0}^{\widehat{J}_1} U_j\mathbf{1}_{\mathcal{A}}\right].$$

Observe that the sum $\sum_{j=0}^{\infty} v_j$ that appears in this proposition is not larger than the right-hand side of (21). We also have the rough upper-bound,

$$\mathbb{E}\left[\sum_{j=\widehat{J}_0}^{\widehat{J}_1} u_j\mathbf{1}_{\mathcal{A}}\right]\leq c'\mathbb{E}\left[\left(\widehat{J}_1+|\widehat{J}_0|+1\right)(\log n)\frac{\log((|\widehat{J}_0|+\widehat{J}_1+1)n)}{n^{1/2}}\right],$$

which, combined with Lemmas 9 and 10, ensures that the expectation tends to 0 faster than the right-hand sides of (21) and (24).

In the following, we will always suppose that n is large enough in the sense that $n\geq n_0$ for some n_0 depending only on $p,\alpha,\theta,s,\kappa,\varrho,\varsigma,R,M_1,M_2,b-a$, the wavelet basis, and possibly varying line by line. When $\theta\leq\theta_0$, $j_0\geq 0$ denotes the smallest integer satisfying

$$2^{j_0(2\alpha+1)}\ \geq R^2M_1^{(1-2/p)/(1-\theta)}n. \tag{64}$$

When $\theta>\theta_0$, $j_0\geq 0$ is rather the smallest integer satisfying

$$2^{j_0(2\alpha+1)}\ \geq R^2M_1^{(1-2\theta/p)/(1-\theta)}M_2^{-2/p}n. \tag{65}$$

When, moreover, $1-\theta/p+s\theta>0$ (and $\alpha>s$), we define the smallest integer $j_1\geq 0$ such that

$$2^{j_1(1-\theta/p+\alpha+s-2s\theta/p+s\theta+2\alpha s\theta)}\ \geq M_2^{-s-1/p}R^{2s\theta+1}n^{1-\theta/p+s\theta}. \tag{66}$$

This number is also defined if $1-\theta/p+s\theta=0$, $\alpha>s$ and $M_2^{-s-1/p}R^{2s\theta+1}\geq 1$. We have $j_0=j_1$ when $p=2\theta$. We have $j_0<j_1$ when $p>2\theta$ and $\alpha>s$, and $j_0>j_1$ when $p<2\theta$ and $\alpha>s$. Note also that $\theta_0<1/2$ when $p>1$, $\theta_0>1/2$ when $p<1$ and $\theta_0=1/2$ when $p=1$.

Below, we explain which formulas we choose in Proposition 9 to bound the terms U_j, $j\in\{\widehat{J}_0,\dots,\widehat{J}_1\}$. By summing up all these terms, and possibly applying Lemma 30 in [Sar24b], or the bounds on $|\widehat{J}_0|$, $\widehat{J}_1$, we get the theorems.

Case 1: $\alpha>s$, $p\geq 1$ and $\theta\leq\theta_0$, $\theta\neq 1/2$. We use (50) when $j\leq -1$, (52) when $j\in[0,j_0]$ and (54) when $j>j_0$.

Case 2: $\alpha>s$, $p=1$ and $\theta=\theta_0=1/2$. We define j_2 as the smallest integer such that

$$2^{j_2(2\alpha+1)}\ \geq R^2M_1^{2(1-2/p)}(n/\log^2 n).$$

We use (50) when $j\leq -1$, (52) when $j\in[0,j_2]$ and (53) when $j>j_2$.

Case 3: $\alpha>s$, $p\geq 1$ and $\theta\in(\theta_0,1/2)$. We use (50) when $j\leq -1$, (52) when $j\in[0,j_0]$, (56) when $j\in(j_0,j_1)$, and (62) when $j\geq j_1$

Case 4: $\alpha>s$, $p\geq 1$, $\theta>1/2$. We use (51) when $j\leq -1$, (58) when $j\in[0,j_1)$, and (62) when $j\geq j_1$.

Case 5: $\alpha > s$, $p > 1$ and $\theta = 1/2$. We use (50) when $j \le -1$, (52) when $j \in [0, j_0]$, (61) when $j \in (j_0, j_1)$, and (62) when $j \ge j_1$

Case 6: $\alpha > s$, $p < 1$ and $\theta < 1/2$. We use (50) when $j \le -1$, (52) when $j \in [0, j_0]$ and (57) when $j > j_0$.

Case 7: $\alpha > s$, $p < 1$ and $\theta = 1/2$. We use (50) when $j \le -1$, (54) when $j \in [0, j_0)$ and (57) when $j \ge j_0$.

Case 8: $\alpha > s$, $p < 1$ and $\theta \in (1/2, \theta_0]$. We use (51) when $j \le -1$, (54) when $j \in [0, j_0)$ and (57) when $j \ge j_0$.

Case 9: $\alpha > s$, $p < 1$, $\theta > \theta_0$ and either $1-\theta/p+s\theta > 0$ or $1-\theta/p+s\theta = 0$ with $M_2^{-s-1/p} R^{2s\theta+1} \ge 1$. Note that $p < 2\theta$ here and hence $j_1 < j_0$. We then use (51) when $j \le -1$, (58) when $j \le j_1$ and (55) when $j \in (j_1, j_0)$ and (57) if $j \ge j_0$.

Case 10: $\alpha > s$, $p < 1$, $\theta > \theta_0$ and $1 - \theta/p + s\theta < 0$. We use (51) when $j \le -1$, (55) when $j \in [0, j_0)$ and (57) if $j \ge j_0$. Note that the sum of (55) leads to the rate $n^{-\theta(1/p-1)/(2\theta/p-1)}$ that is faster than $n^{-(s\theta+1-\theta)/(2s\theta+1)}$.

Case 11: $\alpha > s$, $p < 1$, $\theta > \theta_0$ and $1 - \theta/p + s\theta = 0$ with $M_2^{-s-1/p} R^{2s\theta+1} < 1$. We use (51) when $j \le -1$, (55) when $j \in [0, j_0)$ and (57) if $j \ge j_0$. The sum of (55) leads to the rate $n^{-\theta(1/p-1)/(2\theta/p-1)} = n^{-\gamma} = n^{-(s\theta+1-\theta)/(2s\theta+1)}$. The factor in front of n is not larger than $M_2^{(s+1)/(2s\theta+1)}$.

Case 12: $\alpha \le s$. We consider an arbitrary number $\theta' \in (\theta, 1)$ and observe that $f \in \mathcal{WT}_{\theta'}(M')$ with $M' = M_1^{(1-\theta')/(1-\theta)}$. We may therefore pretend from now on that $\theta > 1/2$. The proof of Lemma 25 of [Sar24b] ensures the existence of $q_1 > 1$ and $q_2 > 0$ such that $\|f\|_{q_1}^{q_1} \le c_1 R'^{q_2}$ where $R' = 1+R$. Besides, c_1, q_1, q_2 only depend on α, p and the wavelets. An interpolation inequality using that f is a density gives: for all $q \in (1, q_1]$,

$$\|f\|_q^q \le c_2 R'^{c_3(q-1)},$$

where c_2 only depends on α, p, q and the wavelets, and where c_3 only depends on α, p and the wavelets. We define $j_3 \ge 0$ as the largest integer such that

$$2^{j_3} \ \le M_1^{-1/(1-\theta)} n^{1/(2s+1)}.$$

We use (51) when $j \le -1$, (59) when $j \in [0, j_3)$ and (60) when $j \ge j_3$ with q suitably chosen.

To understand why this yields the result, let us denote by

$$\mu_j = c_q R'^{c_3(q-1)q(s+1)/(2sq+1)} 2^{-j(s+1)(q-1)/(2qs+1)} n^{-(1+sq-q)/(2sq+1)} \tag{67}$$

the leading term in the right-hand side of (60). If we take $q \in (1, \min\{q_1, 2\}]$, the factor c_q may be chosen independently of q. We have,

$$\begin{aligned}
\sum_{j=j_3}^{\infty} \mu_j &\le c' \sum_{j=j_3}^{\infty} R'^{c_3(q-1)q(s+1)/(2sq+1)} 2^{-j(s+1)(q-1)/(2qs+1)} n^{-(1+sq-q)/(2sq+1)} \\
&\le c'' \ (q-1)^{-1} R'^{c_3(q-1)q(s+1)/(2sq+1)} M_1^{(s+1)(q-1)/((1-\theta)(2qs+1))} n^{-s/(2s+1)} \\
&\le c''' \ (q-1)^{-1} R'^{c_4(q-1)} M_1^{c_4(q-1)} n^{-s/(2s+1)},
\end{aligned}$$

where c'', c''' do not depend on q, and where

$$c_4 = \max\left\{2c_3(s+1)/(4s+1), (s+1)/((1-\theta)(2s+1))\right\}.$$

We then choose

$$q = 1 + \frac{1}{c_4 \log(R' M_1)},$$

when $R'M_1$ is large enough and $q = \min\{q_1, 2\}$ otherwise.

□

Proof of Proposition 10. The triangle inequality leads to

$$d_1(f, \hat{f}) \leq c_1 \left[A + \mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z}) \right] + \sum_{k=\widehat{J}_1+1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}|$$

where

$$A = \sum_{j=\widehat{J}_0}^{\widehat{J}_1} \left\{ 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| - 2^{-j/2} \sum_{k \in \widehat{K}_j} |\beta_{j,k}| + \mathcal{E}_j(\widehat{K}_j) \right\}$$

and where c_1 only depends on the wavelet basis.

The reverse triangle inequality gives

$$A \leq \sum_{j=\widehat{J}_0}^{\widehat{J}_1} \left\{ 2^{-j/2} \sum_{k \in \mathbb{Z}} |\beta_{j,k}| - 2^{-j/2} \sum_{k \in \widehat{K}_j} |\widehat{\beta}_{j,k}| + 2\mathcal{E}_j(\widehat{K}_j) \right\}.$$

We deduce from (41) and from the condition $\kappa \geq \kappa_0$,

$$\begin{aligned} -2^{-j/2} \sum_{k \in \widehat{K}_j} |\widehat{\beta}_{j,k}| + 2\mathcal{E}_j(\widehat{K}_j) &\leq -2^{-j/2} \sum_{k \in \widehat{K}_j} |\widehat{\beta}_{j,k}| + \kappa \widehat{\mathcal{E}}_j(\widehat{K}_j) \\ &\leq \mathrm{crit}_j(\widehat{K}_j). \end{aligned}$$

By using (7): for all $K_j \subset \widehat{\mathbb{Z}}_j$,

$$\begin{aligned} \mathrm{crit}_j(\widehat{K}_j) &\leq \mathrm{crit}_j(K_j) \\ &\leq -2^{-j/2} \sum_{k \in K_j} |\widehat{\beta}_{j,k}| + \kappa \widehat{\mathcal{E}}_j(K_j). \end{aligned}$$

We put everything together and take the expectation of the result to get

$$\mathbb{E}\left[d_1(f, \hat{f}) \mathbf{1}_{\mathcal{A}} \right] \leq c_2 \left\{ \mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z}) \mathbf{1}_{\mathcal{A}} \right] + \mathbb{E}\left[\sum_{j=\widehat{J}_0}^{\widehat{J}_1} U_j \mathbf{1}_{\mathcal{A}} \right] + \mathbb{E}\left[\left(\widehat{T} + \widehat{\bar{T}} \right) \mathbf{1}_{\mathcal{A}} \right] \right\}.$$

Now,

$$\begin{aligned} \|\hat{f}\|_1 &\leq \frac{c_3}{n} \sum_{i=1}^{n} \left[\sum_{k \in \mathbb{Z}} \left| \phi(2^{\widehat{J}_0} X_i - k) \right| + \sum_{j=\widehat{J}_0}^{\widehat{J}_1} \sum_{k \in \mathbb{Z}} |\psi(2^j X_i - k)| \right] \\ &\leq c_4 (\widehat{J}_1 + |\widehat{J}_0| + 1). \end{aligned}$$

We deduce from Cauchy-Schwarz inequality,

$$\mathbb{E}\left[d_1\big(f,\hat{f}\big)\mathbf{1}_{\mathcal{A}^c}\right] \leq c_5\left(\mathbb{E}[\widehat{J}_1^2]+\mathbb{E}[\widehat{J}_0^2]+1\right)^{1/2}\left[\mathbb{P}\left(\mathcal{A}^c\right)\right]^{1/2}.$$

We conclude using (43), Lemma 9, and $\mathbb{P}\left(\mathcal{A}^c\right) \leq c_6 n^{-4}$. □

6.3. Proofs of Section 6.1

6.3.1. Proof of Lemma 6. The proof of (33) and (34) is similar and we may therefore only focus on (34). We put

$$\overline{\mathbb{Z}}_{j,\mathrm{ref}} = \mathbb{Z}\cap\left[-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b\right].$$

We have,

$$\left|\{k\in\mathbb{Z},\, f_{j,k}\geq t\}\right| \leq \left|\left\{k\in\overline{\mathbb{Z}}_{j,\mathrm{ref}}, f_{j,k}\geq t\right\}\right| + \left|\left\{k\notin\overline{\mathbb{Z}}_{j,\mathrm{ref}}, f_{j,k}\geq t\right\}\right|.$$

The cardinal of the first set can be bounded from above by $2^j(b-a)+2L_{\mathrm{wav}}+1$. By using the monotonic properties of f, we deduce for all $k\geq L_{\mathrm{wav}}+2^j b$,

$$f_{j,k}\leq L_{\mathrm{wav}}2^{1-j}f(2^{-j}(k-L_{\mathrm{wav}})).$$

We now use (19) to get

$$f_{j,k}\leq L_{\mathrm{wav}}2^{1-j}(2M_2)^{1/\theta}\left(2^{-j}(k-L_{\mathrm{wav}})-b\right)^{-1/\theta}.$$

The number of $k\geq L_{\mathrm{wav}}+2^j b$ such that $f_{j,k}\geq t$ is hence upper-bounded by $2^\theta L_{\mathrm{wav}}^\theta(2M_2)2^{j(1-\theta)}t^{-\theta}+1$. We do a similar reasoning when $k\leq -L_{\mathrm{wav}}+2^j a$ hence the result.

We now turn to the proof of (36). We have,

$$\left|\{k\in\mathbb{Z},\, f_{j,k}\geq t\}\right| \leq t^{-q}\sum_{k\in\mathbb{Z}} f_{j,k}^q,$$

and we apply Hölder's inequality:

$$\begin{aligned}\sum_{k\in\mathbb{Z}} f_{j,k}^q &\leq \sum_{k\in\mathbb{Z}}\left(\int_{2^{-j}(k-L_{\mathrm{wav}})}^{2^{-j}(k+L_{\mathrm{wav}})}(f(x))^q\,\mathrm{d}x\right)\left(2^{1-j}L_{\mathrm{wav}}\right)^{q-1}\\ &\leq c\|f\|_q^q 2^{-j(q-1)}.\end{aligned}$$

□

6.3.2. Sketch of the proof of Lemma 8. The lemma mainly follows from Proposition 6, Lemmas 1 and 14 of [Sar24b]. We have,

$$\widehat{\sigma}_{j,k}^2-\sigma_{j,k}^2=\frac{1}{n}\sum_{i=1}^n\left[\psi^2(2^jX_i-k)-\mathbb{E}\left(\psi^2(2^jX_i-k)\right)\right].$$

The set $\mathcal{F}_j=\left\{\psi^2(2^j\cdot-k),\, k\in\mathbb{Z}\right\}$ is composed of piecewise constants functions on at most one interval. It is therefore VC subgraph. As ψ is bounded, we may apply Proposition 6 of [Sar24b] with some $\xi>\log n$. We deduce an event of probability $1-e^{-\xi}$ on which: for all $k\in\mathbb{Z}$,

$$\begin{aligned}\left|\widehat{\sigma}_{j,k}^2-\sigma_{j,k}^2\right| &\leq c_1\left[\sqrt{(\sigma_{j,k}^2/n)\log_+(1/\sigma_{j,k}^2)}+\sqrt{\sigma_{j,k}^2(\xi/n)}+\xi/n\right]\\ &\leq (1/3)\sigma_{j,k}^2+c_2\xi/n.\end{aligned}$$

If now ς and ξ are suitably chosen, we deduce from a union bound that with probability $1-n^{-4}$

$$\left|\widehat{\sigma}_{j,k}^2-\sigma_{j,k}^2\right|\leq (1/3)\sigma_{j,k}^2+(1/3)\widehat{\sigma}_{j,k}^2$$

holds true for all $j\in\mathbb{Z}$ and $k\in\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}$. This leads to (38).

The reasoning behind (39) is the same than that the one used to prove Lemma 2 of [Sar24b]. Only the end changes (just adapt the union bound argument and note that $\sigma_j^2(K_j)$ can be bounded independently of K_j).

We can mimic the reasoning used to prove Lemma 14 of [Sar24b] to show (41). We do not rewrite it here because it is very long and that there are only minor differences. For example, we deal here with possibly negative values of j whereas he only controls $\mathcal{E}_j(K_j)$ for non-negative j. We therefore need to change the union bounds slightly, which explains why the absolute value of j appears in the definition of $\widehat{\mathcal{E}}_j(K_j)$. The sets $\widehat{\mathbb{Z}}_{j,r}$ do not have exactly the same definition than in [Sar24b] (the supremum norm of ψ is involved in our definition). The reason is that we are relying here on the partition

$$\widehat{\mathbb{Z}}_j=\widehat{\mathbb{Z}}_{j,r_j,-}\bigcup\left(\bigcup_{r=r_j+1}^{\infty}\widehat{\mathbb{Z}}_{j,r}\right)$$

to define $\widehat{\mathcal{E}}_j(K_j)$ by (9). □

6.3.3. Proof of Lemma 10. We consider $\xi\geq 4$ and deduce from (32) and (37) that on an event of probability $1-e^{-\xi}$: for all $j\leq -1$,

$$|\widehat{\mathbb{Z}}_j(\alpha)|\leq c_1\left[1+(b-a)2^j+M_2n^\theta 2^{j(1-\theta)}+\log(1+|j|)+\xi\right].$$

For all $\xi>0$ and $j\leq 0$ such that $\log((1+|j|)n)\geq\xi/4$,

$$|\widehat{\mathbb{Z}}_j(\alpha)|\leq c_2\left[1+(b-a)2^j+M_2n^\theta 2^{j(1-\theta)}+\log((1+|j|)n)\right].$$

In particular, when

$$|j|\ \geq\max\left\{(1-\theta)^{-1}\log_2(M_2n^\theta),\log_2(b-a),e^{\xi/4}n^{-1}-1\right\},$$

we have

$$|\widehat{\mathbb{Z}}_j(\alpha)|\leq\varrho\log\left((1+|j|)n\right)$$

for any $\varrho\geq 4c_2$. Therefore,

$$|\widehat{J_0}|\ \leq c_3\max\left\{\log_+(M_2n^\theta),\log_+(b-a),e^{\xi/4}n^{-1}-1\right\},$$

with probability $1-e^{-\xi}$. We use this result with $\xi=4\log n$ to get (42). Besides, (43) follows by integration. □

6.3.4. Proof of Lemma 11. We have,

$$|\widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)| \leq \varrho \log\left[\left(1+|\widehat{J}_0|\right) n\right]. \tag{68}$$

We deduce from (39) and (42) that with probability $1-2n^{-4}$,

$$\mathcal{E}_{\widehat{J}_0}(\alpha, \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) \leq c_1 \sqrt{\frac{\log\left(n \max\left\{\log_+(M_2 n^\theta), \log_+(b-a)\right\}\right)}{n}}.$$

Note that

$$\begin{aligned}\mathcal{E}_{\widehat{J}_0}(\alpha, \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) &\leq \sup_{j\in\mathbb{Z}} \sup_{i\in\{1,\dots,n\}} \left\{\sum_{k\in\mathbb{Z}} \phi(2^j X_i - k) + \mathbb{E}\left[\sum_{k\in\mathbb{Z}} \phi(2^j X_i - k)\right]\right\} \\ &\leq 2.\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha, \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha))\right] \leq c_2 \sqrt{\frac{\log\left(n \max\left\{\log_+(M_2 n^\theta), \log_+(b-a)\right\}\right)}{n}}. \tag{69}$$

We now focus on $\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha))$. We sort the set $\widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)$: we write $\widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha) = \{k_1, \dots, k_r\}$, where $k_1 < k_2 < \cdots < k_r$. We have,

$$\begin{aligned}\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) &= \sum_{k\in\mathbb{Z}\setminus\widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)} \int_{2^{-\widehat{J}_0}k}^{2^{-\widehat{J}_0}(k+1)} f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{2^{-\widehat{J}_0}k_1} f(x)\,\mathrm{d}x + \sum_{\ell=1}^{r-1} \int_{2^{-\widehat{J}_0}(k_\ell+1)}^{2^{-\widehat{J}_0}k_{\ell+1}} f(x)\,\mathrm{d}x + \int_{2^{-\widehat{J}_0}(k_r+1)}^{\infty} f(x)\,\mathrm{d}x.\end{aligned}$$

We introduce the intervals $K_\ell = [k_\ell, k_\ell + 1]$ and note

$$\left\{2^{\widehat{J}_0} X_1, \dots, 2^{\widehat{J}_0} X_n\right\} \subset \bigcup_{\ell=1}^{r} K_\ell.$$

Moreover, each interval K_ℓ contains at least one element of $\{2^{\widehat{J}_0} X_1, \dots, 2^{\widehat{J}_0} X_n\}$. We may hence define the smallest integer $n_\ell \in \{1, \dots, n\}$ satisfying $2^{\widehat{J}_0} X_{(n_\ell)} \in K_\ell$. We recall that $X_{(n_\ell)}$ is the n_ℓ^{th} smallest value of $X_1, \dots, X_n$. The numbers n_ℓ are increasing when ℓ grows up. Besides $n_1 = 1$ and we set $n_{r+1} = 1 + n$ so that

$$\left\{2^{\widehat{J}_0} X_{(n_\ell)}, \dots, 2^{\widehat{J}_0} X_{(n_{\ell+1}-1)}\right\} \subset K_\ell$$

holds true for all $\ell \in \{1, \dots, r\}$. In particular, $k_\ell \leq 2^{\widehat{J}_0} X_{(n_\ell)}$ and $2^{\widehat{J}_0} X_{(n_{\ell+1}-1)} \leq k_\ell + 1$. We deduce,

$$\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) \leq \int_{-\infty}^{X_{(1)}} f(x)\,\mathrm{d}x + \sum_{\ell=1}^{r-1} \int_{X_{(n_\ell+1-1)}}^{X_{(n_{\ell+1})}} f(x)\,\mathrm{d}x + \int_{X_{(n)}}^{\infty} f(x)\,\mathrm{d}x.$$

Let F be the cumulative distribution function of X and $U_i = F(X_i)$. The above ensures,

$$\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) \leq U_{(1)} + \sum_{\ell=1}^{r-1} \left(U_{(n_{\ell+1})} - U_{(n_{\ell+1}-1)}\right) + 1 - U_{(n)}.$$

We set $U_{(0)} = 0$ and $U_{(n+1)} = 1$ to get

$$\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) \leq (r+1) \max_{0\leq i\leq n} \left\{U_{(i+1)} - U_{(i)}\right\}.$$

We deduce from (68) and (42) that on an event $\mathcal{A}$ of probability $1 - n^{-4}$,

$$\begin{aligned}\mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha))\mathbf{1}_{\mathcal{A}}\right] \leq c_3 \log\left(n \max\left\{\log_+(M_2 n^\theta), \log_+(b-a)\right\}\right) \\ \times \mathbb{E}\left[\max_{0\leq i\leq n} \left\{U_{(i+1)} - U_{(i)}\right\}\right].\end{aligned}$$

We apply Lemma 13 below with the inequality $\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha)) \leq 2$ to get

$$\mathbb{E}\left[\mathcal{E}_{\widehat{J}_0}(\alpha, \mathbb{Z} \setminus \widehat{\mathbb{Z}}_{\widehat{J}_0}(\alpha))\right] \leq c_4 \log\left(n \max\left\{\log_+(M_2 n^\theta), \log_+(b-a)\right\}\right) \frac{\log n}{n}.$$

We put this result with (69) to conclude. □

Lemma 13. *Let $n \geq 2$ and $U_1, \dots, U_n$ be n independent random variables that are uniformly distributed on $[0,1]$. Set $U_{(0)} = 0$ and $U_{(n+1)} = 1$. Then,*

$$\mathbb{E}\left[\max_{0\leq i\leq n} \left\{U_{(i+1)} - U_{(i)}\right\}\right] \leq \frac{\log\left(n(n+1)\right)}{n-1}.$$

Proof of Lemma 13. The density φ of $U_{(i+1)} - U_{(i)}$ is given for $x \in [0,1]$ by

$$\varphi(x) = n(1-x)^{n-1}.$$

We then have,

$$\begin{aligned}\mathbb{E}\left[\max_{0\leq i\leq n} \left\{U_{(i+1)} - U_{(i)}\right\}\right] &\leq \frac{1}{n-1} \log\left(\sum_{i=0}^{n} \mathbb{E}\left[e^{(n-1)\left(U_{(i+1)} - U_{(i)}\right)}\right]\right) \\ &\leq \frac{1}{n-1} \log\left(n(n+1) \int_0^1 e^{(n-1)t}(1-t)^{n-1}\,\mathrm{d}t\right) \\ &\leq \frac{\log(n(n+1))}{n-1}.\end{aligned}$$

□

6.3.5. Proof of Lemma 12. Lemma 8 implies $f_{j,k} \leq c_1 \log((1+|j|)n)/n$ on $\mathcal{A}$ for all $k \notin \widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}$ and even for all $k \in \mathbb{Z}$ if $j \geq \widehat{J}_1 + 1$ (as ψ is bounded from below by a positive constant on its support). We define $r'_{n,j} \geq 1$ as the largest integer such that

$$2^{-r'_{n,j}+1} \geq \min\left\{1, c_1 \log((1+|j|)n)/n\right\}.$$

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

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

and

$$T_{j,r} = \sum_{k \in \overline{\mathbb{Z}}_{j,r}} 2^{-j/2} \, |\beta_{j,k}|.$$

We deduce from the above,

$$\widehat{T} \leq \sum_{j=0}^{\infty} \sum_{r=r'_{n,j}}^{\infty} T_{j,r}$$

$$\widehat{\widehat{T}} \leq \sum_{j=\widehat{J}_0}^{-1} \sum_{r=r'_{n,j}}^{\infty} T_{j,r}.$$

Consider now the smallest integer $\bar{k}_{j,r} \geq L_{\mathrm{wav}} + 2^j b$ such that $f_{j,\bar{k}_{j,r}} \leq 2^{-r+1}$, and the largest integer $\underline{k}_{j,r} \leq -L_{\mathrm{wav}} + 2^j a$ such that $f_{j,\underline{k}_{j,r}} \leq 2^{-r+1}$. We have,

$$T_{j,r} \leq 2^{-j/2} \left[\sum_{k=\bar{k}_{j,r}}^{\bar{k}_{j,r}+2L_{\mathrm{wav}}-1} |\beta_{j,k}| + \sum_{k=\underline{k}_{j,r}-2L_{\mathrm{wav}}+1}^{\underline{k}_{j,r}} |\beta_{j,k}| + \sum_{\substack{k \leq \underline{k}_{j,r}-2L_{\mathrm{wav}} \\ \text{or } k \geq \bar{k}_{j,r}+2L_{\mathrm{wav}}}} |\beta_{j,k}| \right.$$

$$\left. + \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \right].$$

We deduce from Lemma 2 with $s = 1$, and from the inequality $2^{-j/2}|\beta_{j,k}| \leq c_2 f_{j,k}$,

$$T_{j,r} \leq c_3 \left[2^{-r} + 2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \right]. \tag{70}$$

We always have

$$2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \leq c_4 2^{-r} \left[1 + 2^j (b-a) \right].$$

Therefore,

$$\widehat{\widehat{T}} \leq c_5 \sum_{j=\widehat{J}_0}^{-1} \sum_{r=r'_{n,j}}^{\infty} 2^{-r} \left[1 + 2^j (b-a) \right]$$

$$\leq c_6 \frac{\log((1+|\widehat{J}_0|)n)}{n} \left[|\widehat{J}_0| + 1 + b - a \right].$$

Using (42) then leads to (48).

We now turn to the proof of (47). We derive from Lemma 25 of [Sar24b] that $f_{j,k} \leq c_7 R^{q_1} 2^{-j(1-1/q_2)}$ for some $q_1, q_2 > 0$. Therefore, when $k \in \overline{\mathbb{Z}}_{j,r}$ with $r \geq \bar{r}'_{n,j}$,

$$f_{j,k} \ \leq c_8 \min\left\{ \frac{\log((1+j)n)}{n}, R^{q_1} 2^{-j(1-1/q_2)} \right\}, \tag{71}$$

and hence

$$f_{j,k} \ \leq c_9 \frac{\log n + \log_+ \log_+ R}{n}.$$

We define $r''_{n,j} \geq 1$ as the largest integer such that

$$2^{-r''_{n,j}+1} \ \geq \min\left\{ 1, c_8 \frac{\log((1+j)n)}{n}, c_8 R^{q_1} 2^{-j(1-1/q_2)} \right\},$$

and $r''_n \geq 1$ as the largest integer such that

$$2^{-r''_n+1} \ \geq \min\left\{ 1, c_9 \frac{\log n + \log_+ \log_+ R}{n} \right\}.$$

We deduce,

$$\widehat{T} \leq \sum_{j=0}^{\infty} \sum_{r=\max\{r''_{n,j}, r''_n\}}^{\infty} T_{j,r}.$$

When $p \geq 1$, the smoothness of f and a suitable version of Hölder's inequality lead to

$$2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \leq c_{10} R 2^{-j(\alpha+1-1/p)} \left(1 + 2^j (b-a)\right)^{1-1/p}.$$

Without loss of generality, we may assume that $b - a \geq 1$. We then have,

$$2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \leq c_{11} \min\left\{ 2^{-r+j}(b-a), R(b-a)^{1-1/p} 2^{-j\alpha} \right\}.$$

When $p < 1$, we use (38) in [Sar24b] to get

$$2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \leq c_{12} R^p 2^{-jp(\alpha+1-1/p)} 2^{-r(1-p)}.$$

Therefore,

$$2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}| \leq c_{13} \min\left\{ 2^{-r+j}(b-a), R^p 2^{-jp(\alpha+1-1/p)} 2^{-r(1-p)} \right\}.$$

In both cases,

$$\sum_{j=0}^{\infty} 2^{-j/2} \sum_{\substack{k \in [-L_{\mathrm{wav}}+2^j a, L_{\mathrm{wav}}+2^j b] \\ k \in \overline{\mathbb{Z}}_{j,r}}} |\beta_{j,k}|$$
$$\leq c_{14} \left\{ R^{1/(1+\alpha)} (b-a)^{(1+\alpha-1/p)/(\alpha+1)} 2^{-r\alpha/(\alpha+1)} + (b-a) 2^{-r} \right\}.$$

We deduce from (70) and (71),

$$\sum_{j=0}^{\infty}\sum_{r=\max\{r''_{n,j},r''_n\}}^{\infty} T_{j,r} \le c_{15}\left\{R^{1/(1+\alpha)}(b-a)^{(1+\alpha-1/p)/(\alpha+1)}\left(\frac{\log n+\log_+\log_+ R}{n}\right)^{\alpha/(\alpha+1)}\right.$$
$$\left.+(b-a)\frac{\log n+\log_+\log_+ R}{n}\right\}$$
$$+c_{16}\sum_{j=0}^{\infty}\min\left\{\frac{\log((1+j)n)}{n}, R^{q_1}2^{-j(1-1/q_2)}\right\}.$$

It then remains to compute the last sum to end the proof. □

6.3.6. Proof of Proposition 9. On the event $\mathcal{A}$, we deduce from (38) that $f_{j,k} \ge c_0 2^{-r}$ for all $k \in \widehat{\mathbb{Z}}_{j,r,-}$, $j \in \mathbb{Z}$, $r \in \{0,\dots,\bar{r}_{n,j}\}$ and some $c_0 > 0$. All the results stated in this section are valid on this event $\mathcal{A}$. We define

$$\overline{\mathbb{Z}}_{j,r,-} = \left\{k \in \mathbb{Z},\ f_{j,k} \ge c_0 2^{-r}\right\}.$$

The condition on the tails of f entails when $j \ge 0$,

$$|\overline{\mathbb{Z}}_{j,r,-}| \le c_1 M_1 2^{j(1-\theta)}2^{r\theta}.$$

This inequality may be improved in some cases as shown by Lemma 6. We actually have for all $j \ge 0$,

$$|\widehat{\mathbb{Z}}_{j,r,-}| + |\widehat{\mathbb{Z}}'_{j,r}| + |\overline{\mathbb{Z}}_{j,r,-}| < c_w \min\Big\{M_1 2^{j(1-\theta)}2^{r\theta}, (b-a)2^j + M_2 2^{j(1-\theta)}2^{r\theta},$$
$$2^r, \|f\|_q^q 2^{rq-j(q-1)}\Big\}. \tag{72}$$

When $j \le -1$, the inequality

$$|\widehat{\mathbb{Z}}_{j,r,-}| + |\widehat{\mathbb{Z}}'_{j,r}| + |\overline{\mathbb{Z}}_{j,r,-}| < c_w\left[1 + (b-a)2^j + M_2 2^{j(1-\theta)}2^{r\theta}\right] \tag{73}$$

is also true. We increase c_w if necessary so that $c_w \ge s + 2L_{\mathrm{wav}}$. Note that c_w may be chosen independently of q when $q \ge 1$ is small enough, say $q \le 2$.

We define $\bar{k}_{j,0} \in \mathbb{Z}$ as the smallest integer satisfying $\bar{k}_{j,0} \ge L_{\mathrm{wav}} + 2^j b$ and $\underline{k}_{j,0} \in \mathbb{Z}$ is the largest integer such that $\underline{k}_{j,0} \le -L_{\mathrm{wav}} + 2^j a$. We then set

$$\overline{\mathbb{Z}}_{j,\mathrm{ref}} = \mathbb{Z} \cap \left[\underline{k}_{j,0}, \bar{k}_{j,0}\right]. \tag{74}$$

We consider $r_j \ge 0$ to be specified later on. We then set for $x \ge 0$ and $r \ge 0$,

$$\widetilde{\mathcal{E}}_{j,r}(x) = x2^{-r/2}\log_+\left(\min\left\{2^r, |\widehat{\mathbb{Z}}'_{j,r}|\right\}/x\right)n^{-1/2},$$
$$\widetilde{\mathcal{E}}_{j,r,-}(x) = x^{1/2}n^{-1/2}.$$

Above, the convention $0 \times \log_+(\cdot/0) = 0$ is used when necessary. We recall that $2^r \le c_3 n/\log n \le c_3 n$ when $r \le \bar{r}_{n,j}$, and $|\widehat{\mathbb{Z}}_{j,r}| \le |\widehat{\mathbb{Z}}_{j,r,-}| \le c_4 2^r$. Elementary maths therefore lead to: for all $K_j \subset \widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}$,

$$U_j(K_j) \le c_5\left[B_j(K_j) + \widetilde{\mathcal{E}}_j(K_j) + u_j\right] \tag{75}$$

where c_5 only depends on κ, ς and the wavelets, where $u_j = (\log n)(\log((|j|+1)n))n^{-1/2}$ and where

$$\widetilde{\mathcal{E}}_j(K_j) = \widetilde{\mathcal{E}}_{j,r_j,-}(K_j \cap \widehat{\mathbb{Z}}_{j,r_j,-}) + \sum_{r=r_j+1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r}\left(|K_j \cap \widehat{\mathbb{Z}}_{j,r}|\right).$$

We decompose the bias and variance as

$$B_j(K_j) \leq B_j^{(1)}(K_j) + B_j^{(2)}(K_j) + B_j^{(3)}(K_j),$$

where

$$\begin{aligned}
B_j^{(1)}(K_j) &= 2^{-j/2} \sum_{k \in (\widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-} \cap \overline{\mathbb{Z}}_{j,\mathrm{ref}}) \setminus (K_j \cup \widehat{\mathbb{Z}}_{j,r_j,-})} |\beta_{j,k}| \\
B_j^{(2)}(K_j) &= 2^{-j/2} \sum_{k \in \widehat{\mathbb{Z}}_{j,r_j,-} \setminus K_j} |\beta_{j,k}| \\
B_j^{(3)}(K_j) &= 2^{-j/2} \sum_{k \in \widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-} \setminus (\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cup \widehat{\mathbb{Z}}_{j,r_j,-} \cup K_j)} |\beta_{j,k}|.
\end{aligned}$$

Likewise, we have,

$$\widetilde{\mathcal{E}}_j(K_j) \leq \widetilde{\mathcal{E}}_j^{(1)}(K_j) + \widetilde{\mathcal{E}}_j^{(2)}(K_j) + \widetilde{\mathcal{E}}_j^{(3)}(K_j),$$

where

$$\begin{aligned}
\widetilde{\mathcal{E}}_j^{(1)}(K_j) &= \widetilde{\mathcal{E}}_j((K_j \cap \overline{\mathbb{Z}}_{j,\mathrm{ref}}) \setminus \widehat{\mathbb{Z}}_{j,r_j,-}) \\
\widetilde{\mathcal{E}}_j^{(2)}(K_j) &= \widetilde{\mathcal{E}}_j(K_j \cap \widehat{\mathbb{Z}}_{j,r_j,-}) \\
\widetilde{\mathcal{E}}_j^{(3)}(K_j) &= \widetilde{\mathcal{E}}_j(K_j \setminus (\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cup \widehat{\mathbb{Z}}_{j,r_j,-})).
\end{aligned}$$

The lemmas below show how the first two bias and error terms can be controlled when $j \geq 0$. They are proved in Sections 6.3.19 and 6.3.20.

Lemma 14. *There exist $n_0 \geq 1$ and for each $j \geq 0$, a subset $K_{j,ref}^{(1)}$ of $\overline{\mathbb{Z}}_{j,ref}$ such that: for all $n \geq n_0$,*

$$\sum_{j=0}^{\infty} \left\{ B_j^{(1)}(K_{j,ref}^{(1)}) + \widetilde{\mathcal{E}}_j^{(1)}(K_{j,ref}^{(1)}) \right\} \leq cR^{1/(2\alpha+1)}(b-a)^{(1+\alpha-1/p)/(2\alpha+1)} n^{-\alpha/(2\alpha+1)}. \tag{76}$$

Above, c only depends on α, p, θ and the wavelets. Moreover, n_0 only depends on $\alpha, p, b-a, R$.

Lemma 15. *Suppose that $r_j \geq 0$ is defined for all $j \geq 0$ as the smallest integer such that*

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

Then, there exist $n_0 \geq 1$ and for each $j \geq 0$, a subset $K_{j,ref}^{(2)}$ of $\widehat{\mathbb{Z}}_{j,r_j,-}$ such that

$$\sum_{j=0}^{\infty} \left\{ B_2^{(2)}(K_{j,ref}^{(2)}) + \widetilde{\mathcal{E}}_j^{(2)}(K_{j,ref}^{(2)}) \right\} \leq c_1 R^{1/(2\alpha+1)} M_1^{(1+\alpha-1/p)/((2\alpha+1)(1-\theta))} n^{-\alpha/(2\alpha+1)}.$$

Above, c only depends on α, p, θ and the wavelets. Moreover, n_0 only depends on $\alpha, p, b-a, M_1$.

The two following lemmas are designed to deal with the first and third bias and variance terms. They are proved in Section 6.3.21.

Lemma 16. *Consider $j \geq 0$, $\underline{r} \geq r_j + 1$, and*

$$K_j^{(1,3)} = \bigcup_{r=\underline{r}}^{\bar{r}_{n,j}} \widehat{\mathbb{Z}}_{j,r}.$$

We have $B_j^{(3)}\big(K_j^{(1,3)}\big) = 0$ if $\underline{r} = r_j + 1$. Moreover, the two following inequalities hold true when $p \geq 1$:

$$B_j^{(1)}\big(K_j^{(1,3)}\big) + B_j^{(3)}\big(K_j^{(1,3)}\big) \leq c_1 R M_1^{1-1/p} 2^{\underline{r}(1-1/p)\theta} 2^{-j(\alpha-\theta/p+\theta)} \tag{77}$$

$$B_j^{(1)}\big(K_j^{(1,3)}\big) + B_j^{(3)}\big(K_j^{(1,3)}\big) \leq c_2 R \left\{ M_2^{1-1/p} 2^{\underline{r}(1-1/p)\theta} 2^{-j(\alpha-\theta/p+\theta)} + (b-a)^{1-1/p} 2^{-j\alpha} \right\}. \tag{78}$$

We also have when $\theta < 1/2$,

$$\widetilde{\mathcal{E}}_j^{(1)}\big(K_j^{(1,3)}\big) + \widetilde{\mathcal{E}}_j^{(3)}\big(K_j^{(1,3)}\big) \leq c_3 M_1 2^{j(1-\theta)} 2^{\underline{r}(\theta-1/2)} n^{-1/2} \tag{79}$$

$$\widetilde{\mathcal{E}}_j^{(1)}\big(K_j^{(1,3)}\big) + \widetilde{\mathcal{E}}_j^{(3)}\big(K_j^{(1,3)}\big) \leq c_4 \left[M_2 2^{j(1-\theta)} 2^{\underline{r}(\theta-1/2)} n^{-1/2} + \left[1 + (b-a)2^j\right] n^{-1/2} 2^{-\underline{r}/2} \right]. \tag{80}$$

Lemma 17. *Consider $j \geq 0$, $p < 1$, $\bar{r} \geq r_j$ and*

$$K_j^{(1,3)} = \left(\bigcup_{r=r_j+1}^{\bar{r}} \widehat{\mathbb{Z}}_{j,r} \right) \bigcup \left(\bigcup_{r=\bar{r}+1}^{\bar{r}_{n,j}} K_{j,r} \right),$$

where

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

Then,

$$B_j^{(1)}(K_j^{(1,3)}) + B_j^{(3)}(K_j^{(1,3)}) \leq c_1 R^p 2^{-jp(\alpha+1-1/p)} 2^{-\bar{r}(1-p)/2} n^{-(1-p)/2}.$$

If $\bar{r} = r_j$,

$$\widetilde{\mathcal{E}}_j^{(1)}\big(K_j^{(1,3)}\big) + \widetilde{\mathcal{E}}_j^{(3)}\big(K_j^{(1,3)}\big) \leq c_2 R^p 2^{-jp(\alpha+1-1/p)} 2^{-\bar{r}(1-p)/2} n^{-(1-p)/2} \times \log_+ \left(M_1 R^{-p} n^{-p/2} 2^{jp(\alpha+1-\theta/p)} 2^{\bar{r}(\theta-p/2)} \right).$$

If $\bar{r} \geq r_j + 1$ and $\theta \geq 1/2$,

$$\begin{aligned} \widetilde{\mathcal{E}}_j^{(1)}\big(K_j^{(1,3)}\big) + \widetilde{\mathcal{E}}_j^{(3)}\big(K_j^{(1,3)}\big) \leq{}& c_3 M_1 2^{j(1-\theta)} \left(2^{\bar{r}(\theta-1/2)} + (\bar{r} - r_j) \mathbf{1}_{\theta=1/2} \right) n^{-1/2} \\ &+ c_4 R^p 2^{-jp(\alpha+1-1/p)} 2^{-\bar{r}(1-p)/2} n^{-(1-p)/2} \\ &\times \log_+ \left(M_1 R^{-p} n^{-p/2} 2^{jp(\alpha+1-\theta/p)} 2^{\bar{r}(\theta-p/2)} \right). \end{aligned} \tag{81}$$

We also have if $\bar r \geq r_j + 1$ *and* $\theta > 1/2$,

$$\begin{aligned}\widetilde{\mathcal{E}}_j^{(1)}\big(K_j^{(1,3)}\big) + \widetilde{\mathcal{E}}_j^{(3)}\big(K_j^{(1,3)}\big) &\leq c_5 M_2 2^{j(1-\theta)} 2^{\bar r(\theta-1/2)} n^{-1/2} + c_6 (b-a) 2^{-r_j/2} 2^j n^{-1/2} \\ &\quad + c_7 R^p 2^{-jp(\alpha+1-1/p)} 2^{-\bar r(1-p)/2} n^{-(1-p)/2} \\ &\quad \times \Big[\log_+\Big(M_2 R^{-p} n^{-p/2} 2^{jp(\alpha+1-\theta/p)} 2^{\bar r(\theta-p/2)}\Big) \\ &\quad + \log_+\Big((b-a) R^{-p} n^{-p/2} 2^{jp(\alpha+1)} 2^{-\bar r p/2}\Big)\Big]. \end{aligned} \tag{82}$$

The shape constraint has not been used in the previous four lemmas. We present two results to take it into account.

In the sequel, $\dot\ell_{j,r}$ stands for one of these three elements:

$$\dot\ell_{j,r} = c_w \Big[1 + (b-a)2^j + M_2 2^{j(1-\theta)} 2^{r\theta}\Big] \tag{83}$$

$$\dot\ell_{j,r} = c_w M_1 2^{j(1-\theta)} 2^{r\theta} \tag{84}$$

$$\dot\ell_{j,r} = c_w \Big[1 + A_q^q 2^{rq - j(q-1)}\Big]. \tag{85}$$

In these equalities, $q \geq 1$ and $A_q \geq \|f\|_q$ are to be specified and $c_w \geq s + 2L_{\mathrm{wav}}$ appears in (72) and (73).

Note that $\bar k_{j,0}$ and $\underline{k}_{j,0}$ have already been defined. We now define for $r \geq 1$ the smallest integer $\bar k_{j,r}$ satisfying

$$\bar k_{j,r} \geq L_{\mathrm{wav}} + 2^j b + \sum_{r'=1}^{r} \dot\ell_{j,r'}$$

and the largest integer $\underline{k}_{j,r}$ such that

$$\underline{k}_{j,r} \leq -L_{\mathrm{wav}} + 2^j a - \sum_{r'=1}^{r} \dot\ell_{j,r'}.$$

We put for all $r \geq 0$,

$$\mathbb{K}_{j,r} = \widehat{\mathbb{Z}}_{j,\bar r_{n,j},-} \cap \big([\underline{k}_{j,r+1}, \underline{k}_{j,r}) \cup (\bar k_{j,r}, \bar k_{j,r+1}]\big). \tag{86}$$

We show in Section 6.3.22:

Lemma 18. *Consider* $j \in \mathbb{Z}$, $\underline r \geq 0$, $\bar r \geq \underline r$, $\dot\ell_{j,r}$ *given by (83) and*

$$K_j^{(3)} = \bigcup_{r=\underline r}^{\bar r} \mathbb{K}_{j,r}.$$

We suppose that $\underline r = 0$ *if* $j \leq -1$ *and that* $p \geq 1$ *if* $\underline r \geq 1$. *Then:*

$$\begin{aligned} B_j^{(3)}(K_j^{(3)}) &\leq c_1 M_2^{1-s} 2^{-\bar r(s\theta-\theta+1)} 2^{-j(1-\theta)(s-1)} \\ &\quad + c_1 R \Big[\underline r 2^{-j(\alpha+1-1/p)} + (b-a)^{1-1/p} 2^{-j\alpha} \underline r + M_2^{1-1/p} 2^{\underline r(1-1/p)\theta} 2^{-j(\alpha-\theta/p+\theta)}\Big] \mathbf{1}_{\underline r \geq 1}. \end{aligned}$$

Moreover, if $\theta \geq 1/2$,

$$\begin{aligned}
&\widetilde{\mathcal{E}}_j^{(3)}(K_j^{(3)}) \\
&\leq c_2 \left[\left(1 + 2^j(b-a)\right) \min \left\{ (r_j+1)2^{-r_j/2} \log_+ \left(2^{r_j}/(1+(b-a)2^j)\right), 2^{-\underline{r}/2} \log_+ \left(2^{\underline{r}}/(1+(b-a)2^j)\right) \right\} \right. \\
&\qquad \left. + M_2 2^{j(1-\theta)} \left\{ 2^{\bar{r}(\theta-1/2)} + (\bar{r}-\underline{r}+1)\mathbf{1}_{\theta=1/2} \right\} \right] n^{-1/2}.
\end{aligned}$$

In these inequalities, c_1, c_2 *only depend on* s, θ *and the wavelets. All the above remains true when* $j \geq 0$ *up to mild modifications when the formulas (84) and (85) are used for* $\dot{\ell}_{j,r}$ *in place of (83). In the first case, we need to formally replace in the inequalities* M_2 *by* M_1 *and* $b-a$ *by* 0. *In the second case, we replace* M_2 *by* A_q^q *and* $b-a$ *by* 0 *(and then,* c_1, c_2 *only depend on* s, q *and the wavelets. We also have the following property:* c_1 *and* c_2 *do not depend on* q *if* $q \in [1,2]$*).*

Lemma 19. *Consider* $j \geq 0$ *such that* $2^{j(1-\theta/p+\alpha-s+s\theta)} \geq R M_2^{s-1/p}$ *and* $p \geq 1$. *Then,*

$$\begin{aligned}
B_j^{(3)}(\emptyset) \leq c \Big[& R^{(s\theta-\theta+1)/(1-\theta/p+s\theta)} M_2^{(1-1/p)/(1-\theta/p+s\theta)} 2^{-j(\alpha-s\theta/p-\alpha\theta+s\theta+\alpha s\theta)/(1-\theta/p+s\theta)} \\
& + R \left\{ (b-a)^{1-1/p} 2^{-j\alpha} + 2^{-j(\alpha+1-1/p)} \right\} \log_+ \left(2^{j(1-\theta/p+\alpha-s+s\theta)} R^{-1} M_2^{1/p-s}\right) \Big],
\end{aligned}$$

where c *only depends on* α, p, s, θ *and the wavelets. This result also holds true if we replace formally* M_2 *by* M_1 *and* $b-a$ *by* 0 *(in the assumption and inequality).*

The following remains to be done to prove (50)–(62).

- Choose $r_j \geq 0$.
- Define a set K_j for which the bias term $B_j(K_j)$ and the variance term $\widetilde{\mathcal{E}}_j(K_j)$ are of the right-order of magnitude. The preceding lemmas are particularly well suited to solving this problem.
- Conclude by applying (75).

6.3.7. Proof of (50). We take $K_j = \widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}$ and $r_j = 0$. Therefore,

$$\begin{aligned}
B_j(K_j) + \widetilde{\mathcal{E}}_j(K_j) &\leq \sum_{r=1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r}\left(|\widehat{\mathbb{Z}}_{j,r}|\right) \\
&\leq c_1 \sum_{r=1}^{\bar{r}_{n,j}} \left[1 + (b-a)2^j + M_2 2^{j(1-\theta)} 2^{r\theta}\right] 2^{-r/2} n^{-1/2} \\
&\leq c_2 \left[1 + (b-a)2^j + M_2 2^{j(1-\theta)} n^{-1/2} + M_2 2^{j/2} \, (\log n) n^{-1/2} \, \mathbf{1}_{\theta=1/2}\right].
\end{aligned}$$

□

6.3.8. Proof of (51). We define $r_j = 0$ so that $B_j^{(2)}(K_j^{(2)}) + \widetilde{\mathcal{E}}_j^{(2)}(K_j^{(2)}) = 0$ for all $K_j^{(2)}$. Moreover,

$$\begin{aligned} B_j^{(1)}(\overline{\mathbb{Z}}_{j,\mathrm{ref}}) + \widetilde{\mathcal{E}}_j^{(1)}(\overline{\mathbb{Z}}_{j,\mathrm{ref}}) &\leq \sum_{r=1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r}\left(|\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r}|\right) \\ &\leq c_1 \sum_{r=1}^{\bar{r}_{n,j}} \left[1 + (b-a)2^j\right] 2^{-r/2} \log_+\left(2^r/(1+(b-a)2^j)\right) n^{-1/2} \\ &\leq c_2 \left[1 + (b-a)2^j\right] n^{-1/2} \end{aligned}$$

thanks to Lemma 30 of [Sar24b]. We now define the smallest integer $\bar{r} \geq 0$ such that

$$2^{\bar{r}(2s\theta+1)} \geq n M_2^{-2s} 2^{-2j(1-\theta)s}$$

and set

$$K_j^{(3)} = \bigcup_{r=0}^{\bar{r}} \mathbb{K}_{j,r}.$$

We use Lemma 18 to get a bound on

$$B_j^{(3)}(K_j^{(3)}) + \widetilde{\mathcal{E}}_j^{(3)}(K_j^{(3)}).$$

We put everything together to conclude. □.

6.3.9. Proof of (52). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

We take $K_j = \widehat{\mathbb{Z}}_{j,\bar{r}_{n,j},-}$ so that

$$\begin{aligned} B_j(K_j) + \widetilde{\mathcal{E}}_j(K_j) &\leq \widetilde{\mathcal{E}}_{j,r_j,-}\left(|\widehat{\mathbb{Z}}_{j,r_j,-}|\right) + \sum_{r=r_j+1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r}\left(|\widehat{\mathbb{Z}}_{j,r}|\right) \\ &\leq c\left[1 + M_1 2^{j(1-\theta)} 2^{r_j\theta}\right]^{1/2} n^{-1/2} \\ &\quad + c \sum_{r=r_j+1}^{\bar{r}_{n,j}} \left[1 + M_1 2^{j(1-\theta)} 2^{r\theta}\right] 2^{-r/2} n^{-1/2}. \end{aligned}$$

We use the definition of r_j to conclude. □

6.3.10. Proof of (53). We choose $K_j = \emptyset$ so that

$$\begin{aligned} B_j(K_j) + \widetilde{\mathcal{E}}_j(K_j) &\leq 2^{-j/2} \sum_{k\in\mathbb{Z}} |\beta_{j,k}| \\ &\leq R 2^{-j\alpha}. \end{aligned}$$

□

6.3.11. Proof of (54). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

When $\theta \geq 1/2$ and $p < 1$, we define the smallest integer $\bar{r} \geq r_j$ satisfying

$$2^{(\theta - p/2)(\bar{r} - r_j - 1)} \geq M_1^{(p-2)/(2(1-\theta))} R^p n^{p/2} 2^{-jp(\alpha+1/2)},$$

set

$$K_j^{(1,3)} = \left(\bigcup_{r=r_j+1}^{\bar{r}} \widehat{\mathbb{Z}}_{j,r} \right) \bigcup \left(\bigcup_{r=\bar{r}+1}^{\bar{r}_{n,j}} K_{j,r} \right),$$

apply Lemma 17 (with the bound given by (81)) and conclude thanks to Lemma 15.

When $\theta < 1/2$ and $p \geq 1$, we define $\underline{r} \geq r_j + 1$ as the smallest integer such that

$$2^{(1-2\theta/p)(\underline{r}-1)} \geq M_1^{2/p} R^{-2} 2^{2j(1+\alpha-\theta/p)} n^{-1}.$$

We then consider the set

$$K_j^{(1,3)} = \bigcup_{r=\underline{r}}^{\bar{r}_{n,j}} \widehat{\mathbb{Z}}_{j,r},$$

apply Lemma 16 (with (77) and (79)) and conclude thanks to Lemma 15. □

6.3.12. Proof of (55). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

We define $\bar{r} \geq r_j + 1$ as the smallest integer satisfying

$$2^{(\theta - p/2)(\bar{r} - r_j - 1)} \geq M_1^{(p-2\theta)/(2(1-\theta))} M_2^{-1} R^p n^{p/2} 2^{-jp(\alpha+1/2)},$$

set

$$K_j^{(1,3)} = \left(\bigcup_{r=r_j+1}^{\bar{r}} \widehat{\mathbb{Z}}_{j,r} \right) \bigcup \left(\bigcup_{r=\bar{r}+1}^{\bar{r}_{n,j}} K_{j,r} \right),$$

apply Lemma 17 (with the bound given by (82)) and Lemma 15. □

6.3.13. Proof of (56). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

We define $\underline{r} \geq r_j + 1$ as the smallest integer satisfying

$$2^{(\underline{r}-1)(1-2\theta/p)} \geq M_2^{2/p} R^{-2} 2^{2j(1+\alpha-\theta/p)} n^{-1}.$$

and set

$$K_j^{(1,3)} = \bigcup_{r=\underline{r}}^{\bar{r}_{n,j}} \widehat{\mathbb{Z}}_{j,r}.$$

We then consider the set and apply Lemma 16 (with (78) and (80)). □

6.3.14. Proof of (57). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

We set

$$K_j^{(1,3)} = \bigcup_{r=r_j+1}^{\bar{r}_{n,j}} K_{j,r}.$$

and apply Lemma 17 with $\bar{r} = r_j$. □

6.3.15. Proof of (58). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j.$$

We define $\underline{r} = 0$ and $\bar{r} \geq 0$ as the smallest integer such that

$$2^{\bar{r}(2s\theta+1)} \geq n M_2^{-2s} 2^{-2j(1-\theta)s}.$$

We define $\dot{\ell}_{j,r}$ given by (83) and set

$$K_j^{(3)} = \bigcup_{r=0}^{\bar{r}} \mathbb{K}_{j,r}.$$

We then apply Lemmas 18, 14 and 15. □

6.3.16. Proofs of (59) and (60). The proof is the same than in the preceding section but by defining $r_j = 0$, and $\dot{\ell}_{j,r}$ either by (84) or (85). □

6.3.17. Proof of (61). We define $r_j \geq 0$ as the smallest integer such that

$$2^{r_j} \geq M_1^{1/(1-\theta)} 2^j,$$

and $\underline{r}, \bar{r} \geq 0$, as the smallest integers such that

$$\begin{aligned} 2^{\bar{r}(s+1)} &\geq n M_2^{-2s} 2^{-js} \\ 2^{\underline{r}(1-1/p)} &\geq R^{-2} M_2^{2/p} 2^{2j(\alpha-1/(2p)+1)} n^{-1}. \end{aligned}$$

We define $\dot{\ell}_{j,r}$ by (83) and set

$$K_j^{(3)} = \bigcup_{r=\underline{r}}^{\bar{r}} \mathbb{K}_{j,r}.$$

We conclude by applying Lemmas 18, 14 and 15. □

6.3.18. Proofs of (62). We apply Lemma 19. □

6.3.19. Proof of Lemma 14. We consider the largest $j_0 \geq 0$ such that

$$2^{j_0(1+2\alpha)} \leq R^2 n (b-a)^{1-2/p}. \tag{87}$$

We set $K_{j,\mathrm{ref}} = \overline{\mathbb{Z}}_{j,\mathrm{ref}}$ when $j \in \{0, \dots, j_0\}$. When $j \geq j_0 + 1$, the definition of $K_{j,\mathrm{ref}}$ depends on whether $p \geq 1$ or $p < 1$. When $p \geq 1$, $K_{j,\mathrm{ref}} = \emptyset$. When $p < 1$,

$$K_{j,\mathrm{ref}} = \left\{ k \in \overline{\mathbb{Z}}_{j,\mathrm{ref}},\, |\beta_{j,k}| \ \geq \eta \right\},$$

with

$$\eta = R^{p/(p-2)} n^{1/(p-2)} 2^{-jp(\alpha+1/2)/(p-2)}. \tag{88}$$

For all $j \geq 0$,

$$|\overline{\mathbb{Z}}_{j,\mathrm{ref}}| \leq c_1 \left[1 + (b-a)2^j \right].$$

Therefore, when $j \in \{0, \dots, j_0\}$,

$$\begin{aligned} B_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) + \widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) &\leq \widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \\ &\leq \sum_{r=r_j+1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r} \left(|\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r}| \right) \\ &\leq n^{-1/2} \sum_{r=r_j+1}^{\bar{r}_{n,j}} |\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}|\, 2^{-r/2} \log_+ \left(2^r / |\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}| \right). \end{aligned}$$

We define the smallest integer r'_j such that $2^{r'_j} \ \geq 1 + (b-a)2^j$. We deduce from Lemma 30 of [Sar24b],

$$\sum_{r=r'_j+1}^{\bar{r}_{n,j}} |\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}| 2^{-r/2} \log_+ \left(2^r / |\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}| \right) \leq c_2 \left(1 + (b-a)2^j \right)^{1/2}.$$

We deal with smaller values of r by noticing that $|\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}| \leq c_3 2^r$ and hence

$$\begin{aligned} \sum_{r=1}^{r'_j} |\overline{\mathbb{Z}}_{j,\mathrm{ref}} \cap \widehat{\mathbb{Z}}_{j,r,-}| &\leq c_4 \sum_{r=1}^{r'_j} 2^{r/2} \\ &\leq c_5 \left(1 + (b-a)2^j \right)^{1/2}. \end{aligned}$$

Thereby, by putting everything together, we get for all $j \in \{0, \dots, j_0\}$,

$$B_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) + \widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \leq c_6 \left(1 + (b-a)2^j \right)^{1/2} n^{-1/2}.$$

We now deal with larger value of j. When $j > j_0$ and $p \geq 1$, the smoothness condition plus Lemma 18 of [Sar24b] lead to

$$\begin{aligned} B_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) + \widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) &= B_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \\ &\leq c_7 R 2^{-j(\alpha+1-1/p)} |\overline{\mathbb{Z}}_{j,\mathrm{ref}}|^{1-1/p} \\ &\leq c_8 R \left[2^{-j(\alpha+1-1/p)} + (b-a)^{1-1/p} 2^{-j\alpha} \right]. \end{aligned}$$

We reapply this lemma when $j > j_0$ and $p < 1$ to bound the bias term:

$$\begin{aligned} B_j^{(1)}(K_{j,\mathrm{ref}}) &\leq c_9 2^{-j/2}\eta^{1-p}R^p 2^{-jp(\alpha+1/2-1/p)} \\ &\leq c_{10}R^{p/(2-p)}n^{-(1/p-1)/(2/p-1)}2^{-j(1+\alpha-1/p)/(2/p-1)}. \end{aligned}$$

Moreover, the assumption $f \in \mathcal{WB}^{\alpha}_{p,\infty}(R)$ and (88) lead to

$$\begin{aligned} |K_{j,\mathrm{ref}}^{(1)}| &\leq \eta^{-p}R^p 2^{-jp(\alpha+1/2-1/p)} \\ &\leq R^{2p/(2-p)}n^{1/(2/p-1)}2^{-2j(1+\alpha-1/p)/(2/p-1)}. \end{aligned} \tag{89}$$

As $|\widehat{\mathbb{Z}}_{j,r}| \leq c_{11}2^r$,

$$\widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \leq c_{12}\sum_{r=r_j+1}^{\bar{r}_{n,j}} \min\{|K_{j,\mathrm{ref}}^{(1)}|, 2^r\}\, 2^{-r/2}\log_+\left(2^r/\min\{|K_{j,\mathrm{ref}}^{(1)}|, 2^r\}|\right).$$

Define the smallest integer r_j'' such that $2^{r_j''} \geq |K_{j,\mathrm{ref}}^{(1)}|$. The sum can be split into two parts: the first part runs from 1 to r_j'' and the second from $r_j''+1$ to ∞. By mimicking what was done above when $j \leq j_0$, we deduce

$$\widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \leq c_{13}n^{-1/2}\ |K_{j,\mathrm{ref}}^{(1)}|^{1/2},$$

and using (89),

$$\widetilde{\mathcal{E}}_j^{(1)}(K_{j,\mathrm{ref}}^{(1)}) \leq c_{14}R^{p/(2-p)}n^{-(1/p-1)/(2/p-1)}2^{-j(1+\alpha-1/p)/(2/p-1)}.$$

In the end, all this leads to (76). □

6.3.20. Proof of Lemma 15. We consider the largest $j_0 \geq 0$ such that

$$2^{j_0(1+2\alpha)} \leq R^2 n M_1^{(1-2/p)/(1-\theta)}.$$

We set $K_{j,\mathrm{ref}}^{(2)} = \widehat{\mathbb{Z}}_{j,r_j,-}$ when $j \in \{0,\dots,j_0\}$. When $j \geq j_0+1$, and $p \geq 1$, $K_{j,\mathrm{ref}}^{(2)} = \emptyset$. When $j \geq j_0+1$ and $p < 1$,

$$K_{j,\mathrm{ref}}^{(2)} = \left\{k \in \widehat{\mathbb{Z}}_{j,r_j,-},\ |\beta_{j,k}|\ \geq \eta\right\},$$

with η defined by (88).

When $j \in \{0,\dots,j_0\}$,

$$\begin{aligned} B_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) + \widetilde{\mathcal{E}}_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) &\leq \widetilde{\mathcal{E}}_{j,r_j,-}(\widehat{\mathbb{Z}}_{j,r_j,-}) \\ &\leq n^{-1/2}\ \left|\widehat{\mathbb{Z}}_{j,r_j,-}\right|^{1/2} \\ &\leq c_1 n^{-1/2}M_1^{1/2}2^{r_j\theta/2}2^{j(1-\theta)/2} \\ &\leq c_2 n^{-1/2}M_1^{1/(2(1-\theta))}2^{j/2} \end{aligned}$$

thanks to (72). When $j > j_0$ and $p \geq 1$, we do as in the proof of Lemma 14 to get

$$\begin{aligned} B_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) + \widetilde{\mathcal{E}}_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) &= B_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) \\ &\leq c_3 R 2^{-j(\alpha+1-1/p)} |\widehat{\mathbb{Z}}_{j,r_{j,-}}|^{1-1/p} \\ &\leq c_4 R M_1^{(1-1/p)/(1-\theta)} 2^{-j\alpha}. \end{aligned}$$

Likewise, when $j > j_0$ and $p < 1$

$$B_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) \ \leq c_5 R^{p/(2-p)} n^{-(1/p-1)/(2/p-1)} 2^{-j(1+\alpha-1/p)/(2/p-1)}.$$

We also have in this case,

$$\begin{aligned} \mathcal{E}_j^{(2)}(K_{j,\mathrm{ref}}^{(2)}) \ &\leq n^{-1/2} |K_{j,\mathrm{ref}}^{(2)}|^{1/2} \\ &\leq R^{p/(2-p)} n^{-(1/p-1)/(2/p-1)} 2^{-j(1+\alpha-1/p)/(2/p-1)}. \end{aligned}$$

thanks to (89). The lemma can be deduced from all these results. □

6.3.21. Proofs of Lemmas 16 and 17.

Proof of Lemma 16. We omit the exponent (1,3) on K_j to lighten the notations. We have,

$$B_j^{(1)}(K_j) + B_j^{(3)}(K_j) \leq 2^{-j/2} \sum_{k \in \widehat{\mathbb{Z}}_{j,\underline{r}-1,-}} |\beta_{j,k}|.$$

When $p > 1$ we apply a suitable version of Hölder's inequality (see Lemma 18 of [Sar24b] if needed) to get

$$2^{-j/2} \sum_{k \in \widehat{\mathbb{Z}}_{j,\underline{r}-1,-}} |\beta_{j,k}| \leq c_1 2^{-j/2} \|\beta_{j,\cdot}\|_{p,\infty} \left|\widehat{\mathbb{Z}}_{j,\underline{r}-1,-}\right|^{1-1/p}.$$

This result also holds true when $p = 1$ if we replace the weak norm by its strong version. We then use (72) and the smoothness assumption to get the bound on the bias.

As to the variance terms, we merely remark

$$\widetilde{\mathcal{E}}_j^{(1)}(K_j) + \widetilde{\mathcal{E}}_j^{(3)}(K_j) \leq \sum_{r=\underline{r}}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r} \left(|\widehat{\mathbb{Z}}_{j,r}| \right),$$

and conclude by using (72). □

Proof of Lemma 17. We omit in this proof the exponent (1,3) on K_j. The smoothness assumption ensures that $|K_{j,r}|$ is no larger than

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

We deduce from a classical inequality in weak spaces (stated in (38) in [Sar24b]),

$$\sum_{k \in \widehat{\mathbb{Z}}_{j,r} \setminus K_{j,r}} 2^{-j/2} |\beta_{j,k}| \ \leq c_1 k_{j,r} 2^{-r/2} n^{-1/2}.$$

Therefore,

$$\begin{aligned} B_j^{(1)}(K_j) + B_j^{(3)}(K_j) &\leq \sum_{r=\bar{r}+1}^{\bar{r}_{n,j}} \sum_{k \in \widehat{\mathbb{Z}}_{j,r} \setminus K_{j,r}} 2^{-j/2} \, |\beta_{j,k}| \\ &\leq c_2 \sum_{r=\bar{r}+1}^{\bar{r}_{n,j}} R^p n^{-(1-p)/2} 2^{-jp(\alpha+1-1/p)} 2^{-r(1-p)/2} \\ &\leq c_3 R^p n^{-(1-p)/2} 2^{-jp(\alpha+1-1/p)} 2^{-\bar{r}(1-p)/2}. \end{aligned}$$

Moreover,

$$\widetilde{\mathcal{E}}_j^{(1)}(K_j) + \widetilde{\mathcal{E}}_j^{(3)}(K_j) \leq \sum_{r=r_j+1}^{\bar{r}} \widetilde{\mathcal{E}}_{j,r}(|\widehat{\mathbb{Z}}_{j,r}|) + \sum_{r=\bar{r}+1}^{\bar{r}_{n,j}} \widetilde{\mathcal{E}}_{j,r}(|K_{j,r}|).$$

The first term is zero if $\bar{r} = r_j$. It can be bounded thanks to (72) and (73) in the contrary case. The second term can be controlled by using the bounds on $|\widehat{\mathbb{Z}}'_{j,r}|$ and $|K_{j,r}|$ and by applying Lemma 30 of [Sar24b]. □

6.3.22. Proofs of Lemmas 18 and 19. As f is monotone on $(-\infty, a)$ and on $(b, +\infty)$, we have $f_{j,k} \leq c_0' 2^{-r}$ when $k \in \mathbb{K}_{j,r}$ where $c_0' = \max\{c_0, 1\}$ (the case $r = 0$ comes from the inequality $f_{j,k} \leq 1$ as f is a density and $c_0' \geq 1$).

We apply Lemma 2 to get

$$\sum_{\substack{k \geq \overline{k}_{j,r+1} \\ \text{or } k \leq \underline{k}_{j,r+1}}} 2^{-j/2} |\beta_{j,k}| \leq c_1 \dot{\ell}_{j,r}^{1-s} 2^{-r}. \tag{90}$$

Throughout this proof, we omit the exponent on K_j to lighten the notations. We begin by showing the following lemma.

Lemma 20. *Suppose $\theta \geq 1/2$, consider $j \in \mathbb{Z}$, $\underline{r}, \bar{r} \geq 0$, $\dot{\ell}_{j,r}$ given by (83) and*

$$K_j = \bigcup_{r=\underline{r}}^{\bar{r}} \mathbb{K}_{j,r}.$$

Then,

$$\begin{aligned} \widetilde{\mathcal{E}}_j^{(3)}(K_j) \leq c \left[1 + 2^j(b-a)\right] \min \Big\{ & (r_j+1) 2^{-r_j/2} \log_+ \left(2^{r_j}/(1+(b-a)2^j)\right), \\ & 2^{-\underline{r}/2} \ \log_+ \left(2^{\underline{r}}/(1+(b-a)2^j)\right) \Big\} n^{-1/2} \\ & + c M_2 2^{j(1-\theta)} \left\{ 2^{\bar{r}(\theta-1/2)} + (\bar{r} - \underline{r} + 1) \mathbf{1}_{\theta=1/2} \right\} n^{-1/2}, \end{aligned} \tag{91}$$

where c only depends on θ and the wavelets. The same inequality holds true when $j \geq 0$ and the definition of $\mathbb{K}_{j,r}$ rather uses (84) or (85), up to the following modifications: we replace M_2 by M_1 and $b-a$ by 0 in the first case, and M_2 by A_q^q, θ by q, and $b-a$ by 0 in the second case (c then only depends on q and the wavelets. It can be bounded independently of q when $q \in [1,2]$).

Proof of Lemma 20. We recall that $f_{j,k} \leq c_0' 2^{-r'}$ when $k \in \mathbb{K}_{j,r'}$, and $f_{j,k} \geq c_0 2^{-r}$ when $k \in \widehat{\mathbb{Z}}_{j,r}$ with $r \leq \bar{r}_{n,j}$. We deduce $\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r} = \emptyset$ when $r' > r + c_1$. Hence,

$$K_j \cap \widehat{\mathbb{Z}}_{j,r} \subset \bigcup_{r'=\underline{r}}^{\min\{r+c_1,\bar{r}\}} \left(\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r} \right).$$

Therefore,

$$\widetilde{\mathcal{E}}_j^{(3)}(K_j) \leq \sum_{r=r_j+1}^{\bar{r}_{n,j}} \sum_{r'=\underline{r}}^{\min\{r+c_1,\bar{r}\}} \widetilde{\mathcal{E}}_{j,r}(|\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r}|).$$

We consider the smallest integer $\underline{r}' \geq 0$ such that

$$M_2 2^{j(1-\theta)} 2^{\underline{r}'\theta} \geq 1 + (b-a)2^j.$$

We deduce from (72),

$$\begin{aligned}
&\sum_{r=\max\{\underline{r}',r_j+1\}}^{\bar{r}} \sum_{r'=\max\{\underline{r},\underline{r}'\}}^{\min\{r+c_1,\bar{r}\}} \widetilde{\mathcal{E}}_{j,r}(|\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r}|) \\
&\quad \leq c_2 \sum_{r=\max\{\underline{r},\underline{r}',r_j+1\}}^{\bar{r}} \sum_{r'=\max\{\underline{r},\underline{r}'\}}^{r+c_1} M_2 2^{j(1-\theta)} 2^{r'\theta} 2^{-r/2} \log_+ \left(2^{r\theta}/2^{r'\theta} \right) n^{-1/2} \\
&\quad \leq c_3 \sum_{r=\underline{r}}^{\bar{r}} M_2 2^{j(1-\theta)} 2^{r(\theta-1/2)} n^{-1/2} \\
&\quad \leq c_4 M_2 2^{j(1-\theta)} \left[2^{\bar{r}(\theta-1/2)} + (\bar{r} - \underline{r} + 1)\mathbf{1}_{\theta=1/2} \right] n^{-1/2}.
\end{aligned}$$

To get the third line, we relied on Lemma 30 of [Sar24b]. Moreover,

$$\begin{aligned}
&\sum_{r=\max\{\bar{r},\underline{r}'\}}^{\bar{r}_{n,j}} \sum_{r'=\max\{\underline{r},\underline{r}'\}}^{\min\{r+c_1,\bar{r}\}} \widetilde{\mathcal{E}}_{j,r}(|\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r}|) \\
&\quad \leq c_5 \sum_{r=\max\{\bar{r},\underline{r}'\}}^{\bar{r}_{n,j}} \sum_{r'=\max\{\underline{r},\underline{r}'\}}^{\bar{r}} M_2 2^{j(1-\theta)} 2^{r'\theta} 2^{-r/2} \log_+ \left(2^{r\theta}/2^{r'\theta} \right) n^{-1/2} \\
&\quad \leq c_6 \sum_{r=\max\{\bar{r},\underline{r}'\}}^{\bar{r}_{n,j}} M_2 2^{j(1-\theta)} 2^{\bar{r}\theta} 2^{-r/2} \log_+ \left(2^{r\theta}/2^{\bar{r}\theta} \right) n^{-1/2} \\
&\quad \leq c_7 M_2 2^{j(1-\theta)} 2^{\bar{r}(\theta-1/2)} n^{-1/2}.
\end{aligned} \tag{92}$$

When either r or r' is smaller than $\underline{r}'$, $|\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r}| \leq c_8(1+(b-a)2^j)$. Hence,

$$\begin{aligned}
&\sum_{r=r_j+1}^{\bar{r}_{n,j}} \sum_{r'=\underline{r}}^{\min\{r+c_1,\bar{r}\}} \widetilde{\mathcal{E}}_{j,r}(|\mathbb{K}_{j,r'} \cap \widehat{\mathbb{Z}}_{j,r}|)\mathbf{1}_{\min\{r,r'\}\leq \underline{r}'} \\
&\quad \leq c_9 \sum_{r=\max\{r_j+1,\underline{r}-c_1\}}^{\bar{r}_{n,j}} \sum_{r'=\underline{r}}^{r+c_1} \left(1+(b-a)2^j\right) 2^{-r/2} n^{-1/2} \log_+\left(2^r/(1+(b-a)2^j)\right) \\
&\quad \leq c_{10}\left[1+2^j(b-a)\right] \min\Big\{(r_j+1)2^{-r_j/2}\ \log_+\left(2^{r_j}/(1+(b-a)2^j)\right), \\
&\qquad\qquad 2^{-\underline{r}/2}\ \log_+\left(2^{\underline{r}}/(1+(b-a)2^j)\right)\Big\} n^{-1/2}.
\end{aligned}$$

In the last inequality, we use the elementary result

$$\sum_{r=r_0}^{\infty} r 2^{-r/2} \log_+(2^r/x) \leq c\, r_0 2^{-r_0/2} \log_+(2^{r_0}/x),$$

that holds true for all $r_0 \geq 1$ and $x \geq 1$.

When $j \geq 0$ and $\mathbb{K}_{j,r}$ is defined via (85), the term c in (91) can be bounded independently of q because c_w does not depend on q when $q \in [1,2]$ and because the term c in the first equation of Lemma 30 of [Sar24b] can be taken as a bounded function of a_1 if a_1 lies in a a compact set such as $[1,2]$ (and $k=1$). □

Proof of Lemma 18. We have,

$$\begin{aligned}
B_j^{(3)}(K_j) &\leq \sum_{r=0}^{\underline{r}-1} 2^{-j/2} \sum_{k\in\mathbb{K}_{j,r}} |\beta_{j,k}| + \sum_{\substack{k\geq \bar{k}_{j,\bar{r}+1}+1 \\ \text{or } k\leq \underline{k}_{j,\bar{r}+1}-1}} 2^{-j/2}|\beta_{j,k}| \\
&\leq \sum_{r=0}^{\underline{r}-1} 2^{-j/2} \sum_{k\in\mathbb{K}_{j,r}} |\beta_{j,k}| \ + c_2 M_2^{1-s} 2^{-\bar{r}(s\theta-\theta+1)} 2^{-j(1-\theta)(s-1)}.
\end{aligned}$$

The formula for the second term comes from (90) by noticing that $\dot{\ell}_{j,r} \ \geq c_w M_2 2^{j(1-\theta)} 2^{r\theta}$. Note that the first term of this inequality is 0 if $\underline{r}=0$. Otherwise, we reproduce some of the arguments in the proof of Lemma 16. When $p>1$,

$$2^{-j/2} \sum_{k\in\mathbb{K}_{j,r}} |\beta_{j,k}| \leq c_3 2^{-j/2} \|\beta_{j,\cdot}\|_{p,\infty} |\mathbb{K}_{j,r}|^{1-1/p}.$$

We use $f \in \mathcal{WB}^{\alpha}_{p,\infty}(R)$ to bound the weak norm. As to $|\mathbb{K}_{j,r}|$, we have

$$|\mathbb{K}_{j,r}| \leq c_4\left[1+(b-a)2^j + M_2 2^{j(1-\theta)} 2^{r\theta}\right].$$

This leads to

$$2^{-j/2} \sum_{k\in\mathbb{K}_{j,r}} |\beta_{j,k}| \ \leq c_5 R\left[2^{-j(\alpha+1-1/p)} + (b-a)^{1-1/p} 2^{-j\alpha} + M_2^{1-1/p} 2^{r(1-1/p)\theta} 2^{-j(\alpha-\theta/p+\theta)}\right].$$

It then remains to sum this inequality to get the bound on the bias term. The proof when $p=1$ is similar. □

Proof of Lemma 19. Consider some $\bar{r} \geq 0$ to be specified later on. We have,

$$B_j^{(3)}(\emptyset) \leq \sum_{r=0}^{\bar{r}} 2^{-j/2} \sum_{k \in \mathbb{K}_{j,r}} |\beta_{j,k}| + \sum_{r=\bar{r}+1}^{\infty} 2^{-j/2} \sum_{k \in \mathbb{K}_{j,r}} |\beta_{j,k}|,$$

where $\mathbb{K}_{j,r}$ is defined via (83). By doing as in the proof of Lemma 18,

$$\begin{aligned} B_j^{(3)}(\emptyset) \leq c_1 R \left[\bar{r} 2^{-j(\alpha+1-1/p)} + \bar{r}(b-a)^{1-1/p} 2^{-j\alpha} + M_2^{1-1/p} 2^{\bar{r}(1-1/p)\theta} 2^{-j(\alpha-\theta/p+\theta)} \right] \\ + c_2 M_2^{1-s} 2^{-\bar{r}(s\theta-\theta+1)} 2^{-j(1-\theta)(s-1)} \end{aligned}$$

We then choose $\bar{r}$ suitably. The remaining part of the lemma is obtained by choosing a different formula for $\dot{\ell}_{j,r}$. □

7. Proof of Proposition 7

Since the result is stated for R, M_1 and M_2 large enough, we only need to show

$$\inf_{\tilde{f}} \sup_{f \in \mathscr{F}_{p,\theta}^{\alpha,s}(a,b,c'R,c'M,c'M)} \mathbb{E}\left[d_1(f, \hat{f}) \right] \geq c R^{\beta_1} M^{\beta_2} n^{-\gamma}.$$

for some $c' > 0$ only depending on p, s, θ, α and the wavelet basis.

In the sequel, we need to build functions with special properties. For this, we rely on the two elementary results below. They are proved in Sections 7.1 and 7.2.

Lemma 21. *Consider $q_1 \geq 0$ and $q_2 \geq 0$. There exist two polynomial functions ς_0 and ς_1 such that:*

- *ς_0, ς_1 are non-negative and non-increasing on $[0,1]$.*
- *$\varsigma_0(0) = \varsigma_1(0) = 1$ and $\varsigma_0(1) = \varsigma_1(1) = 0$.*
- *For all $k \in \{1, \ldots, q_1 + 1\}$, $\varsigma_0^{(k)}(0) = \varsigma_1^{(k)}(0) = \varsigma_0^{(k)}(1) = \varsigma_1^{(k)}(1) = 0$.*
- *For all $k \in \{0, \ldots, q_2\}$,*

$$\int_0^1 x^k \varsigma_0(x)\, \mathrm{d}x = \int_0^1 x^k \varsigma_1(x)\, \mathrm{d}x.$$

- *ς_0 and ς_1 do not coincide almost everywhere on $[0,1]$.*

Lemma 22. *Consider $q \geq 0$, $\ell \geq 1$, and $\varepsilon_1, \ldots, \varepsilon_\ell > 0$ small enough. There exists a polynomial function ζ on $[0,1]$ such that:*

- *For all $x \in [0,1]$, $\zeta(x) \geq 0$. Moreover, $\zeta(0) = 0$ and $\zeta(1) = 1$.*
- *For all $k \in \{1, \ldots, \ell + q + 1\}$, $\zeta^{(k)}(0) = 0$.*
- *For all $k \in \{1, \ldots, \ell\}$, $\zeta^{(k)}(1) = (-1)^k \varepsilon_k$.*
- *For all $k \in \{\ell + 1, \ldots, \ell + q + 1\}$, $\zeta^{(k)}(1) = 0$.*
- *For all $k \in \{1, \ldots, \ell + q + 1\}$, $\sup_{x \in [0,1]} |\zeta^{(k)}(x)| \leq c$ where c only depends on ℓ, k and q.*

The following lemma is useful for showing the regularity of the forthcoming functions. It is proved in Section 7.3.

Lemma 23. *Consider $p \in (0,+\infty]$, $\alpha > \max\{1/p-1,1\}$, $q \in [\alpha,\alpha+1]\cap\mathbb{Z}$ and τ large enough. We suppose that f is even and admits bounded derivatives up to order $q+1$. We consider $x_2 > x_1 > 0$ and we denote for $k \in \{1,\dots,q+1\}$,*

$$\begin{aligned} A_{1,k} &= \sup_{x\in[0,x_1]} |f^{(k)}(x)| \\ A_{2,k} &= \sup_{x\in[x_1,x_2]} |f^{(k)}(x)| \\ A_k &= \sup_{x\in\mathbb{R}} |f^{(k)}(x)|. \end{aligned}$$

We also assume that $f(x) = ae^{-bx}$ for some $a \geq 0$, $b \in (0,1]$ when $x \geq x_2$. We then have $f \in \mathcal{B}^\alpha_{p,\infty}(cR)$ where

$$\begin{aligned} R = 1 &+ A_{1,q}^{q+1-\alpha} A_{1,q+1}^{-q+\alpha} (x_1+1)^{1/p} + A_{2,q}^{q+1-\alpha} A_{2,q+1}^{-q+\alpha} (x_2-x_1+1)^{1/p} \\ &+ A_q^{q+1-\alpha} A_{q+1}^{-q+\alpha} + ab^{\alpha-1/p} e^{-bx_2}, \end{aligned}$$

and where c only depends on α, p and the wavelet basis.

Without loss of generality, we assume in the sequel that $b > 0$ and $a < 0$. We then denote $x_0 = \min\{b,|a|\}$. We also consider $r \geq 1$ and denote the elements of $\{0,1\}^{r+1}$ by $(\delta_k)_{0\leq k\leq r}$. We introduce the Hamming distance Δ defined for $\delta,\delta' \in \{0,1\}^{r+1}$ by

$$\Delta(\delta,\delta') = \sum_{k=0}^{r} |\delta_k - \delta'_k|.$$

We consider positive numbers b_0, ℓ_0, and set for all $k \geq 0$,

$$\begin{aligned} b_k &= b_0(1+1/r)^{-k} \\ \ell_k &= \ell_0(1+1/r)^{2k} \\ x_{k+1} &= x_k + \ell_k. \end{aligned}$$

We define the smallest integer $q_{\alpha,s} \geq 2$ larger than $\alpha - s + 1$. Let then ς_0 and ς_1 be the maps given by Lemma 21 with $q_1 = q_{\alpha,s}$ and $q_2 = s$.

We define for $x \geq 0$ and $\delta \in \{0,1\}^{r+1}$,

$$\begin{aligned} g_{1,\delta}(x) = b_0 &\mathbf{1}_{[0,x_0)}(x) \\ &+ \frac{r}{r+1} \sum_{k=0}^{r} b_k \left[1 + r^{-1}\varsigma_{\delta_k}((x-x_k)/\ell_k)\right] \mathbf{1}_{[x_k,x_{k+1})}(x) \\ &+ b_{r+1}\,\varsigma_0\left(\frac{x-x_{r+1}}{r\ell_0}\right)\mathbf{1}_{[x_{r+1},x_{r+1}+r\ell_0]}(x). \end{aligned} \tag{93}$$

We extend it to make it even on $\mathbb{R}$.

The lemma below describes some of its properties. Its proof is elementary and is therefore skipped. It simply follows from Lemma 21.

Lemma 24. *For all $\delta \in \{0,1\}^{r+1}$, $g_{1,\delta}$ belongs to $\mathcal{M}^1(a,b)$, is non-negative and compactly supported on the interval $[-x_{r+1} - r\ell_0, x_{r+1} + r\ell_0]$. It is $q_{\alpha,s}+1$ times differentiable. Moreover, for all $x \in \mathbb{R}$ and $k \in \{1,\dots,q_{\alpha,s}+1\}$,*

$$\left|g_{1,\delta}^{(k)}(x)\right| \leq c_1 r^{-1} b_0 \ell_0^{-k},$$

where c_1 only depends on ς_0, ς_1. For all $k \in \{0,\dots,r\}$, $\delta, \delta' \in \{0,1\}^{r+1}$, and $x \in [x_k, x_{k+1}]$,

$$g_{1,\delta}(x) - g_{1,\delta'}(x) = \frac{b_k}{1+r}\big(\delta_k - \delta_k'\big)\xi_1((x-x_k)/\ell_k)$$

where $\xi_1 = \varsigma_1 - \varsigma_0$. Moreover, for all $q \in \{0,\dots,s\}$,

$$\int_{x_0}^{x_k} x^q g_{1,\delta}(x)\,\mathrm{d}x = \int_{x_0}^{x_k} x^q g_{1,\delta'}(x)\,\mathrm{d}x.$$

We also have for all $\delta' \in \{0,1\}^{r+1}$,

$$d_1(g_{1,\delta}, g_{1,\delta'}) \geq c_2 b_0 (\ell_0/r)\Delta(\delta,\delta')$$
$$d_2^2(g_{1,\delta}, g_{1,\delta'}) \leq c_3 b_0^2 (\ell_0/r^2)\Delta(\delta,\delta')$$

where c_2, c_3 are positive and only depend on ς_0 and ς_1.

As we will see below, this result is sufficient to prove the lower bound when $s = 1$. We now suppose $s \geq 2$ and define functions whose tails satisfy stronger shape constraints. We consider $u \in \{1,\dots,s\}$, $x \geq 0$, $\delta \in \{0,1\}^{r+1}$, and set

$$g_{u+1,\delta}(x) = \frac{J_u}{\ell_0 r}\zeta_{u+1}\left(2(x - x_0/2)/x_0\right)\mathbf{1}_{[x_0/2,x_0)}(x) + \left(\frac{1}{\ell_0 r}\int_x^{x_{r+1}+r\ell_0} g_{u,\delta}(t)\,\mathrm{d}t\right)\mathbf{1}_{[x_0,x_{r+1}+r\ell_0]}(x)$$

where

$$J_u = \int_{x_0}^{x_{r+1}+r\ell_0} g_{u,\delta}(t)\,\mathrm{d}t.$$

In this equality, $\zeta_{u+1}(\cdot)$ comes from Lemma 22 with $\ell = u$, $q = q_{\alpha,s}$, and

$$\varepsilon_k = \begin{cases} (x_0/2)^k (\ell_0 r)^{-k} J_{u-k}/J_u, & \text{if } k \in \{1,\dots,u-1\} \\ (x_0/2)^u (\ell_0 r)^{-u+1} b_0/J_u & \text{if } k = u. \end{cases} \tag{94}$$

This definition is possible if the ε_k are small enough, which we suppose for the moment. The map $g_{u+1,\delta}$ is then extended to be even on $\mathbb{R}$. We show in Section 7.4:

Lemma 25. *Consider $u \in \{2,\dots,s\}$ and $\delta \in \{0,1\}^{r+1}$. Then:*

- *$g_{u,\delta} \in \mathcal{M}^u(a,b)$ and is compactly supported on the interval $[-x_{r+1} - r\ell_0, x_{r+1} + r\ell_0]$.*
- *$g_{u,\delta}$ admits derivatives up to order $q_{\alpha,s}+u$. Moreover,*

$$g_{u,\delta}^{(k)}(x_0) = \begin{cases} (-1)^k (\ell_0 r)^{-k-1} J_{u-k-1} & \text{if } k \in \{1,\dots,u-2\} \\ (-1)^{u-1}(\ell_0 r)^{-u+1} b_0 & \text{if } k = u-1 \\ 0 & \text{if } k \in \{u,\dots,q_{\alpha,s}+u\}. \end{cases}$$

For all $x \notin [-x_0, x_0]$ and $k \in \{u,\dots,q_{\alpha,s}+u\}$,

$$|g_{u,\delta}^{(k)}(x)| \leq c_1 r^{-u} b_0 \ell_0^{-k},$$

where c_1 only depends on ς_0, ς_1.

- *For all $k \in \{0, \dots, r+1\}$, $\delta' \in \{0,1\}^{r+1}$, $q \in \{0, \dots, s-u+1\}$,*
$$\int_{x_0}^{x_k} x^q g_{u,\delta}(x)\,\mathrm{d}x = \int_{x_0}^{x_k} x^q g_{u,\delta'}(x)\,\mathrm{d}x.$$
- *We have $g_{u,\delta}(x_0) \le c_2 b_0$ where c_2 only depends on u.*
- *We have*
$$\int_{x_{r+1}}^{x_{r+1}+r\ell_0} g_{u,\delta}(x)\,\mathrm{d}x = \frac{b_{r+1} r \ell_0}{(u-1)!}\int_0^1 t^{u-1}\varsigma_0(t)\,\mathrm{d}t.$$
In particular, J_u does not depend on δ.
- *For all $k \in \{0, \dots, r\}$, $\delta' \in \{0,1\}^{r+1}$, $x \in [x_k, x_{k+1}]$,*
$$g_{u,\delta}(x) - g_{u,\delta'}(x) = \frac{b_k(\ell_k/\ell_0)^{u-1}}{(r+1)r^{u-1}}(\delta_k - \delta'_k)\xi_u((x-x_k)/\ell_k)$$
where $\xi_u(x) = \int_x^1 \xi_{u-1}(t)\,\mathrm{d}t$ and $\xi_1 = \varsigma_1 - \varsigma_0$.
- *We have for all $\delta' \in \{0,1\}^{r+1}$,*
$$d_1(g_{u,\delta}, g_{u,\delta'}) \ge c_3 \ell_0 b_0 r^{-u}\Delta(\delta,\delta')$$
$$d_2^2(g_{u,\delta}, g_{u,\delta'}) \le c_4 \ell_0 b_0^2 r^{-2u}\Delta(\delta,\delta')$$
where c_3, c_4 are positive and only depend on $\varsigma_0, \varsigma_1, u$.

We consider $\upsilon \in (0,1)$, a density $\omega \in \mathscr{F}_{p,\theta}^{\alpha,s}(a,b,R,M,M)$, and

$$\varphi(x) = (1/2)(M/(r\ell_0))^{1/\theta}\left[\zeta_\varphi\big(2(x-x_0/2)/x_0\big)\mathbf{1}_{[x_0/2,x_0]}(x) + e^{-(x-x_0)/(r\ell_0)}\mathbf{1}_{x>x_0}\right],$$

where $\zeta_\varphi(\cdot)$ is defined by Lemma 22 with $q=0$, $\ell = q_{\alpha,s}+s$ and $\varepsilon_k = (x_0/2)^k(\ell_0 r)^{-k}$. The map φ is extended to an even function on $\mathbb{R}$. We then set for all $\delta \in \{0,1\}^{r+1}$, $x \in \mathbb{R}$,

$$f_\delta(x) = \upsilon g_{s,\delta}(x) + \varphi(x) + p\omega(x),$$

where

$$p = 1 - \int_{\mathbb{R}} \left(\upsilon g_{s,\delta}(x) + \varphi(x)\right)\,\mathrm{d}x.$$

Note that p does not depend on δ because neither does J_s (see Lemma 24 or 25).

We now specify the parameters r, ℓ_0, b_0 that are involved in our analysis. When $\theta > \theta_0$, we define r as the smallest integer such that

$$r^{1+s} \ge R^{\beta_1} M^{\beta_2} n^{1-\gamma}. \tag{95}$$

When $\theta \le \theta_0$, r is the smallest integer such that

$$r \ge R^{2\beta_1} M^{2\beta_2} n^{1/(2\alpha+1)}. \tag{96}$$

We define in the two cases ℓ_0 by

$$\ell_0^{1+\alpha-1/p} = r^{1/p-1} R^{\beta_1 - 1} M^{\beta_2} n^{-\gamma}. \tag{97}$$

and

$$b_0 = R\ell_0^{\alpha-1/p} r^{s-1/p}. \tag{98}$$

Let us observe that $r \to \infty$, $b_0 \to 0$, $b_0 \ell_0 r \to 0$ when $n \to +\infty$. In particular, $J_s \to 0$. We also have $\ell_0 r \ \to +\infty$ when $\theta > \theta_0$ and $n \to +\infty$. When $\theta \leq \theta_0$, $\ell_0 r$ is rather of the order of $M^{1/(1-\theta)}$ (and is not smaller than $M^{1/(1-\theta)}$). In both cases, $\ell_0 r$ can be taken as large as we wish by choosing either n large enough or M large enough.

The fourth point of Lemma 25 ensures that $J_{u-k} \ \leq c_1 b_0 (r\ell_0)$. The fifth point, $J_u \geq c_2 b_0 r \ell_0$. In particular, the term ε_k in (94) is no larger than $c_3 x_0^k (\ell_0 r)^{-k}$, whatever $k \in \{1, \ldots, u\}$. They can therefore be assumed as small as we wish. The maps $g_{u,\delta}$ are then well defined. Likewise, the map φ above is well defined. We also deduce that p may be taken between 0 and 1. The f_δ are therefore densities.

We know from the above that f_δ lies in $\mathcal{M}^s(a,b)$. Its smoothness is ensured by Lemma 23. More precisely, we apply Lemma 23 to deal with $g_{s,\delta}$. We choose $x_1 = x_0$ and $x_2 = x_{r+1} + r\ell_0$. The last point of Lemma 22 ensures that

$$A_{1,k} \ \leq c_3 (\ell_0 r)^{-1} J_u \leq c_4 b_0,$$

which can be as small as wished. Moreover, Lemma 25 ensures

$$A_{2,k} \ \leq c_5 r^{-s} b_0 \ell_0^{-k}.$$

We deduce that $g_{s,\delta}$ belongs to $\mathcal{B}^\alpha_{p,\infty}(R')$ with

$$\begin{aligned} R' &\leq c_6 \left[1 + b_0 (1 + x_0^{1/p}) + r^{-s} b_0 \ell_0^{-\alpha} (r\ell_0)^{1/p} \right] \\ &\leq c_7 \left[1 + b_0 \ell_0^{-\alpha + 1/p} r^{1/p - s} \right]. \end{aligned}$$

We deduce from (95)–(98) that $R' \leq c_8 R$. The smoothness of φ is also guaranteed by Lemma 23. It ensures that $\varphi \in \mathcal{B}^\alpha_{p,\infty}(R'')$ with

$$\begin{aligned} R'' &\leq c_9 \left[1 + M^{1/\theta} (r\ell_0)^{1/p - \alpha - 1/\theta} \right] \\ &\leq c_{10} R \end{aligned}$$

as n or M is large enough. In conclusion, f_δ does belong to $\mathcal{B}^\alpha_{p,\infty}(c_{11} R)$.

Since $g_{s,\delta}$ is bounded from above by $c_{12} b_0$ when $|x| \geq x_0$ and is compactly supported, $g_{s,\delta}$ belongs to $\mathcal{WT}_\theta(M')$ with $M' = c_{13} \left[1 + b_0^\theta (r\ell_0) \right] \leq c_{14} M$ (see Propositions 4 and 5). Clearly, $\varphi \in \mathcal{WT}_\theta(c_{15} M)$. All this ensures $f_\delta \ \in \mathscr{F}^{\alpha,s}_{p,\theta}(a, b, c_{16} R, c_{17} M, c_{18} M)$.

We deduce from the last point of Lemmas 24 and 25, and from the formulas (95)–(98), that for all $\delta, \delta' \in \{0,1\}^{r+1}$,

$$\begin{aligned} d_1(f_\delta, f_{\delta'}) &= \upsilon d_1(g_{s,\delta}, g_{s,\delta'}) \\ &\geq c_{19} \upsilon \ell_0 b_0 r^{-s} \Delta(\delta, \delta') \\ &\geq c_{20} \upsilon r^{-1} R^{\beta_1} M^{\beta_2} n^{-\gamma} \Delta(\delta, \delta'). \end{aligned}$$

Moreover,

$$d_2^2(f_\delta, f_{\delta'}) \leq c_{21} \upsilon^2 \ell_0 b_0^2 r^{-2s} \Delta(\delta, \delta'). \tag{99}$$

The Hellinger distance $h(f_\delta, f_{\delta'})$ between f_δ and $f_{\delta'}$ is defined by

$$h^2(f_\delta, f_{\delta'}) = \frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{f_\delta(x)} - \sqrt{f_{\delta'}(x)} \right)^2 \mathrm{d}x.$$

As f_δ and $f_{\delta'}$ coincide on $[-x_0, x_0]$ and on $\mathbb{R} \setminus [-x_{r+1}, x_{r+1}]$,

$$h^2(f_\delta, f_{\delta'}) \leq \sup_{x \in [x_0, x_{r+1}]} \frac{d_2^2(f_\delta, f_{\delta'})}{8\varphi(x)}. \tag{100}$$

Note that $\inf_{x\in[x_0,x_{r+1}]} \varphi(x) \geq c_{22}(M/(r\ell_0))^{1/\theta}$. This inequality with (99) and (100) leads to

$$h^2(f_\delta, f_{\delta'}) \leq c_{23}\upsilon^2 n^{-1}\Delta(\delta, \delta').$$

We conclude by applying Assouad's lemma, and by taking υ small enough, see Lemma 31 of [Sar24b] for instance. □

Note: we may get (23) by using the same reasoning but by applying formula (95) all the time, even when $\theta \leq \theta_0$.

7.1. Proof of Lemma 21. We set $\varsigma_0(x) = (1 - x^{q_1+2})^{q_1+2}$ and turn to the definition of ς_1. We introduce the linear map L defined for any polynomial R by

$$L(R) = \Big(R(0), R'(0), \dots, R^{(q_1+2)}(0), R(1), R'(1), \dots, R^{(q_1+2)}(1), \\ \int_0^1 R(x)\,\mathrm{d}x, \int_0^1 xR(x)\,\mathrm{d}x, \dots, \int_0^1 x^{q_2}R(x)\,\mathrm{d}x\Big).$$

This map cannot be injective. There exist therefore a non-zero polynomial R such that $L(R) = 0$. Since 0 and 1 are roots of R of order at least $q_1 + 2$, we may find another polynomials R_1, R_2 such that

$$R(x) = (1 - x)^{q_1+2}R_1(x)$$
$$R'(x) = x^{q_1+1}(1 - x)^{q_1+1}R_2(x).$$

We set

$$\varsigma_1(x) = \varsigma_0(x) + \varepsilon R(x),$$

where

$$\varepsilon = \min\left\{\left[\sup_{x\in[0,1]} |R_1(x)|\right]^{-1}, (q_1 + 2)^2\left[\sup_{x\in[0,1]} |R_2(x)|\right]^{-1}\right\}.$$

Note that

$$\varsigma_1(x) = (1 - x^{q_1+2})^{q_1+2} + \varepsilon(1 - x)^{q_1+2}R_1(x)$$

is non-negative when $x \in [0, 1]$. Besides,

$$\varsigma_1'(x) = -(q_1 + 2)^2x^{q_1+1}(1 - x^{q_1+2})^{q_1+1} + \varepsilon x^{q_1+1}(1 - x)^{q_1+1}R_2(x)$$

is non-positive when $x \in [0, 1]$. This ends the proof. □

7.2. Proof of Lemma 22. We define by induction maps $\varsigma_{\ell,\varepsilon_1,\dots,\varepsilon_\ell}$ where $\ell \geq 1$ and $\varepsilon_1,\dots,\varepsilon_\ell \in \mathbb{R}$. More precisely, we set when $\ell = 1$, $\varepsilon_1 \in \mathbb{R}$, and $x \in [0,1]$,

$$\varsigma_{1,\varepsilon_1}(x) = 1 + \varepsilon_1 - \varepsilon_1 x - \varepsilon_1 \int_x^1 (1-t^{q+2})^{q+2}\,\mathrm{d}t.$$

We then set when $\ell \geq 1$, $\varepsilon_1,\dots,\varepsilon_{\ell+1} \in \mathbb{R}$ and $x \in [0,1]$,

$$\varsigma_{\ell+1,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_{\ell+1}}(x) = 1 - \int_x^1 (\varsigma_{\ell,-\varepsilon_2,-\varepsilon_3,\dots,-\varepsilon_{\ell+1}}(t) - 1 - \varepsilon_1)\,\mathrm{d}t.$$

The following properties hold true whatever $\ell \geq 1$ and $\varepsilon_1,\dots,\varepsilon_\ell \in \mathbb{R}$. They are proved by induction.

- $\varsigma_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(1) = 1$.
- For all $k \in \{1,\dots,\ell\}$, $\varsigma^{(k)}_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(1) = (-1)^k \varepsilon_k$.
- For all $x \in [0,1]$,
$$|\varsigma_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(x) - 1| \leq c_1 \left[|\varepsilon_1| + \cdots + |\varepsilon_\ell|\right],$$
where c_1 only depends on q and ℓ.
- For all $x \in [0,1]$, and $k \in \{1,\dots,\ell-1\}$,
$$\left|\varsigma^{(k)}_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(x)\right| \leq c_2 \left[|\varepsilon_1| + \cdots + |\varepsilon_\ell|\right],$$
where c_2 only depends on q and ℓ.
- For all $x \in [0,1]$,
$$\varsigma^{(\ell)}_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(x) = (-1)^\ell \varepsilon_\ell \left[1 - \left(1 - x^{q+2}\right)^{q+2}\right].$$
In particular, $\varsigma^{(k)}_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(1) = 0$ for all $k \in \{\ell+1,\dots,\ell+q+1\}$.

We consider $\ell \geq 1$, $\varepsilon_1,\dots,\varepsilon_\ell > 0$ small enough and set for all $x \in [0,1]$,

$$\zeta(x) = Q(x) \times \varsigma_{\ell,\varepsilon_1,\varepsilon_2,\dots,\varepsilon_\ell}(x)$$

where $Q(x) = 1 - (1 - x^{\ell+q+2})^{\ell+q+2}$. This function suits. □

7.3. Proof of Lemma 23. We show the result when $p < \infty$. The case $p = \infty$ is similar. The map $\psi_{j,k}$ is compactly supported on the interval $[2^{-j}(k - L_{\mathrm{wav}}), 2^{-j}(k + L_{\mathrm{wav}})]$. We consider $r \in \{0,\dots,q\}$ and apply the Taylor–Lagrange formula to build a polynomial map P of degree r satisfying

$$\sup_{x \in [2^{-j}(L_{\mathrm{wav}}-k), 2^{-j}(L_{\mathrm{wav}}+k)]} |f(x) - P(x)| \leq c_1 \left[\sup_{x \in [2^{-j}(k-L_{\mathrm{wav}}), 2^{-j}(L_{\mathrm{wav}}+k)]} |f^{(r+1)}(x)|\right] 2^{-j(r+1)}, \tag{101}$$

where c_1 only depends on the wavelets and r. Define

$$K_1 = \mathbb{Z} \bigcap \left[-2^j x_1 + L_{\mathrm{wav}}, 2^j x_1 - L_{\mathrm{wav}}\right]$$
$$K_2 = \mathbb{Z} \bigcap \left(\left[-2^j x_2 + L_{\mathrm{wav}}, -2^j x_1 - L_{\mathrm{wav}}\right] \bigcup \left[2^j x_1 + L_{\mathrm{wav}}, 2^j x_2 - L_{\mathrm{wav}}\right]\right)$$
$$K_3 = \mathbb{Z} \bigcap \left(\left(-\infty, -2^j x_2 - L_{\mathrm{wav}}\right] \bigcup \left[2^j x_2 + L_{\mathrm{wav}}, +\infty\right)\right).$$

Since $\psi_{j,k}$ is orthogonal to polynomials of degree no larger than $\tau-1$,

$$\beta_{j,k} = 2^{j/2} \int_{\mathbb{R}} (f(x) - P(x))\, \psi(2^j x - k)\, \mathrm{d}x.$$

We deduce from (101),

$$|\beta_{j,k}| \leq c_2 A_{r+1} 2^{-j(r+3/2)}$$

for all $k \in \mathbb{Z}$. Moreover, when $k \in K_u$ for some $u \in \{1,2\}$,

$$|\beta_{j,k}| \leq c_3 A_{u,r+1} 2^{-j(r+3/2)}.$$

By putting these two results together:

$$\begin{aligned}\sum_{k \in \mathbb{Z}\setminus(K_1\cup K_2\cup K_3)} |\beta_{j,k}|^p &\leq c_4 A_{r+1}^p 2^{-jp(r+3/2)} \\ &\leq c_4 A_{r+1}^p 2^{-jp(r+3/2-1/p)}\end{aligned}$$

and

$$\sum_{k\in K_u} |\beta_{j,k}|^p \leq c_5 A_{u,r+1}^p L_u 2^{-jp(r+3/2-1/p)},$$

where $L_1 = x_1 + 1$ and $L_2 = x_2 - x_1 + 1$.

We turn to K_3 and observe that if $k \in K_3$ such that $k \geq 2^j x_2 + L_{\mathrm{wav}}$,

$$|\beta_{j,k}| \leq c_6 a b^{r+1} e^{-b2^{-j}(k - L_{\mathrm{wav}})} 2^{-j(r+3/2)}.$$

A similar result holds true when $k \leq -2^j x_2 - L_{\mathrm{wav}}$. Therefore:

$$\sum_{k\in K_3} |\beta_{j,k}|^p \leq c_7 a^p b^{p(r+1-1/p)} e^{-bpx_2}\ 2^{-jp(r+3/2-1/p)}.$$

All these inequalities are valid for $r \in \{0,\dots,q\}$. We apply them with $r = q-1$ and $r = q$. We deduce that for all $\theta \in [0,1]$, $u \in \{1,2\}$,

$$\sum_{k\in K_u} |\beta_{j,k}|^p \leq c_8 \left(L_u A_{u,q}^p 2^{-jp(q+1/2-1/p)}\right)^{\theta} \left(L_u A_{u,q+1}^p 2^{-jp(q+3/2-1/p)}\right)^{1-\theta}$$

$$\sum_{k\in K_3} |\beta_{j,k}|^p \leq c_9 a^p e^{-bpx_2} \left(b^{p(q-1/p)} 2^{-jp(q+1/2-1/p)}\right)^{\theta} \left(b^{p(q+1-1/p)} 2^{-jp(q+3/2-1/p)}\right)^{1-\theta}$$

$$\sum_{k\in\mathbb{Z}\setminus(K_1\cup K_2\cup K_3)} |\beta_{j,k}|^p \leq c_{10} \left(A_q^p 2^{-jp(q+1/2-1/p)}\right)^{\theta} \left(A_{q+1}^p 2^{-jp(q+3/2-1/p)}\right)^{1-\theta}.$$

We then choose $\theta = q+1-\alpha$. □

7.4. Proof of Lemma 25. We first observe that the conclusion of the lemma also applies when $u = 1$ (see Lemma 24). We may therefore prove the lemma by induction. We suppose that it is true for $u \in \{1,\dots,s-1\}$ and shows that it also holds true for $u+1$.

- For all $x > x_0$, $g'_{u+1,\delta}(x) = -(1/(\ell_0 r)) g_{u,\delta}(x)$. A similar result holds true when $x < x_0$. As $g_{u,\delta} \in \mathcal{M}^u(a,b)$, we deduce that $g_{u+1,\delta} \in \mathcal{M}^{u+1}(a,b)$.
- The second point of the lemma comes from the equality $g^{(k)}_{u+1,\delta}(x) = -(1/(\ell_0 r)) g^{(k-1)}_{u,\delta}(x)$ that is valid for all $x > x_0$, and $k \in \{1,\dots,q_{\alpha,s}+u+1\}$, and from the definition of ζ_{u+1}.

- For the third point, we consider $q \in \{0, \dots, s-u\}$ and observe that

$$\begin{aligned}\int_{x_0}^{x_k} x^q \left(g_{u+1,\delta}(x) - g_{u+1,\delta'}(x)\right) \mathrm{d}x \\ = \frac{1}{(q+1)\ell_0 r} \int_{x_0}^{x_k} \left(x^{q+1} - x_0^{q+1}\right) \left(g_{u,\delta}(x) - g_{u,\delta'}(x)\right) \mathrm{d}x.\end{aligned}$$

- We remark that $x_{r+1} - x_0 \le c_1 \ell_0 r$. Besides, $g_{u,\delta}$ is non-increasing on (x_0, ∞). Therefore,

$$\begin{aligned} g_{u+1,\delta}(x_0) &\le (\ell_0 r)^{-1} g_{u,\delta}(x_0)(x_{r+1} - x_0 + \ell_0 r) \\ &\le c_2 g_{u,\delta}(x_0) \end{aligned}$$

  hence the fourth point.
- We have,

$$\begin{aligned}\int_{x_{r+1}}^{x_{r+1}+r\ell_0} g_{u,\delta}(x)\,\mathrm{d}x &= (\ell_0 r)^{-1} \int_{x_{r+1}}^{x_{r+1}+r\ell_0} (t - x_{r+1}) g_{u-1,\delta}(t)\,\mathrm{d}t \\ &= (\ell_0 r)^{-2}\, \frac{1}{2} \int_{x_{r+1}}^{x_{r+1}+r\ell_0} (t - x_{r+1})^2 g_{u-2,\delta}(t)\,\mathrm{d}t \\ &= \dots \\ &= (\ell_0 r)^{-u+1}\, \frac{1}{(u-1)!} \int_{x_{r+1}}^{x_{r+1}+r\ell_0} (t - x_{r+1})^{u-1} g_{1,\delta}(t)\,\mathrm{d}t \\ &= b_{r+1} (\ell_0 r)^{-u+1}\, \frac{1}{(u-1)!} \int_{x_{r+1}}^{x_{r+1}+r\ell_0} (t - x_{r+1})^{u-1} \varsigma_0((t - x_{r+1})/(r\ell_0))\,\mathrm{d}t \\ &= \frac{b_{r+1} r \ell_0}{(u-1)!} \int_0^1 t^{u-1} \varsigma_0(t)\,\mathrm{d}t.\end{aligned}$$

- We have for all $x \in [x_k, x_{k+1}]$,

$$\begin{aligned} g_{u+1,\delta}(x) - g_{u+1,\delta'}(x) &= \frac{1}{\ell_0 r} \int_x^{x_{r+1}+r\ell_0} \left(g_{u,\delta}(t) - g_{u,\delta'}(t)\right) \mathrm{d}t \\ &= \frac{1}{\ell_0 r} \int_x^{x_{k+1}} \left(g_{u,\delta}(t) - g_{u,\delta'}(t)\right) \mathrm{d}t \end{aligned}$$

  because of the third point (and because $g_{u+1,\delta}(x)$ does not depend on δ when $x \ge x_{r+1}$). Therefore,

$$\begin{aligned} g_{u+1,\delta}(x) - g_{u+1,\delta'}(x) &= \frac{b_k (\ell_k/\ell_0)^{u-1}}{(1+r)\ell_0 r^u} (\delta_k - \delta_k') \int_x^{x_{k+1}} \xi_u((t - x_k)/\ell_k)\,\mathrm{d}t \\ &= \frac{b_k (\ell_k/\ell_0)^u}{(1+r) r^u} (\delta_k - \delta_k') \int_{(x-x_k)/\ell_k}^1 \xi_u(t)\,\mathrm{d}t. \end{aligned}$$

- We deduce from the last point,

$$\begin{aligned} d_1(g_{u+1,\delta}, g_{u+1,\delta'}) &= \frac{2}{(1+r)r^u} \sum_{k=0}^{r} b_k (\ell_k/\ell_0)^u |\delta_k - \delta'_k| \int_{x_k}^{x_{k+1}} |\xi_{u+1}((x-x_k)/\ell_k)| \,\mathrm{d}x \\ &= \frac{2}{(1+r)r^u} \sum_{k=0}^{r} b_k \ell_k (\ell_k/\ell_0)^u |\delta_k - \delta'_k| \int_0^1 |\xi_{u+1}(x)| \,\mathrm{d}x \\ &\geq c_3 r^{-u-1} b_0 \ell_0 \left(\int_0^1 |\xi_{u+1}(x)| \,\mathrm{d}x \right) \Delta(\delta, \delta'). \end{aligned}$$

We get the lower bound by noticing that ξ_{u+1} cannot be zero almost everywhere. The proof for the $\mathbb{L}^2$ distance follows the same principle.

□

References

[Bal04] Fadoua Balabdaoui. *Nonparametric estimation of a k-monotone density: A new asymptotic distribution theory*. PhD thesis, 2004.

[BF96] Peter J Bickel and Jianqing Fan. Some problems on the estimation of unimodal densities. *Statistica Sinica*, 6(1):23–45, 1996.

[BFW13] Fadoua Balabdaoui, Simon Foucart, and Jon A Wellner. On the hermite spline conjecture and its connection to k-monotone densities. *arXiv:1301.3190*, 2013.

[Bir89] Lucien Birgé. The grenander estimator: A nonasymptotic approach. *The Annals of Statistics*, 17(4):1532–1549, 1989.

[Bir97] Lucien Birgé. Estimation of unimodal densities without smoothness assumptions. *The Annals of Statistics*, 25(3):970–981, 1997.

[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.

[Che20] Xin Chen. *Penalized unimodal spline density estimate with application to M-estimation*. PhD thesis, Colorado State University, 2020.

[DJ96] Bernard Delyon and Anatoli Juditsky. On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, 3(3):215–228, 1996.

[DJKP96] David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2):508–539, 1996.

[DL93] Ronald A DeVore and George G Lorentz. *Constructive approximation*. Springer Berlin, Heidelberg, 1993.

[DL12] Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer Science & Business Media, 2012.

[DL18] Cécile Durot and Hendrik P. Lopuhaä. Limit Theory in Monotone Function Estimation. *Statistical Science*, 33(4):547–567, 2018.

[Gao08] Fuchang Gao. Entropy Estimate for k-Monotone Functions via Small Ball Probability of Integrated Brownian Motions. *Electronic Communications in Probability*, 13:121–130, 2008.

[GJ14] Piet Groeneboom and Geurt Jongbloed. *Nonparametric estimation under shape constraints*. Cambridge University Press, 2014.

[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.

[GW09] FuChang Gao and Jon A Wellner. On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Science in China Series A: Mathematics*, 52(7):1525–1538, 2009.

[HH02] Peter Hall and Li-Shan Huang. Unimodal density estimation using kernel methods. *Statistica Sinica*, 12(4):965–990, 2002.

[HK05] Peter Hall and Kee-Hoon Kang. Unimodal kernel density estimation by data sharpening. *Statistica Sinica*, 15(1):73–98, 2005.

[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*. Springer Science & Business Media, 2012.

[JLL04] Anatoli Juditsky and Sophie Lambert-Lacroix. On minimax density estimation on $\mathbb{R}$. *Bernoulli*, 10(2):187–220, 2004.

[KPT96] Gérard Kerkyacharian, Dominique Picard, and Karine Tribouley. Lp adaptive density estimation. *Bernoulli*, 2(3):229–247, 1996.

[Lep15] Oleg Lepski. Adaptive estimation over anisotropic functional classes via oracle approach. *The Annals of Statistics*, 43(3):1178–1242, 2015.

[LM17] Hendrik P Lopuhaä and Eni Musta. Smooth estimation of a monotone hazard and a monotone density under random censoring. *Statistica Neerlandica*, 71(1):58–82, 2017.

[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.

[Mey12] Mary C Meyer. Nonparametric estimation of a smooth density with shape restrictions. *Statistica Sinica*, 22(2):681–701, 2012.

[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.

[Sar21] Mathieu Sart. Minimax bounds for besov classes in density estimation. *Electronic Journal of Statistics*, 15(1):3184–3216, 2021.

[Sar24a] Mathieu Sart. About the optimal estimation of a density with infinite support under Hellinger loss. *Electronic Journal of Statistics*, 18(2):4526–4577, 2024.

[Sar24b] Mathieu Sart. Non-linear wavelet density estimation on the real line. Hal, 2024.

[VDVVDL03] Aad Van Der Vaart and Mark Van Der Laan. Smooth estimation of a monotone density. *Statistics: A Journal of Theoretical and Applied Statistics*, 37(3):189–203, 2003.

[Wan95] Yazhen Wang. The l_1 theory of estimation of monotone and unimodal densities. *Journal of Nonparametric Statistics*, 4(3):249–261, 1995.

[Weg70] Edward J Wegman. Maximum likelihood estimation of a unimodal density function. *The Annals of Mathematical Statistics*, 41(2):457–471, 1970.

Université Jean Monnet, CNRS, Ecole Centrale de Lyon, INSA Lyon, Universite Claude Bernard Lyon 1, ICJ UMR5208, 42023 Saint-Etienne, France

Email address: mathieu.sart@univ-st-etienne.fr