Electronic Journal of Statistics
Vol. 18 (2024) 4526–4577
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2306

About the optimal estimation of a density with infinite support under Hellinger loss

Mathieu Sart

Universite Jean Monnet, CNRS, Ecole Centrale de Lyon, INSA Lyon, Universite Claude Bernard Lyon 1, ICJ UMR5208, 42023 Saint-Etienne, France,
e-mail: mathieu.sart@univ-st-etienne.fr

Abstract: The aim of this paper is to give a complete description of the optimal estimation rates for the Hellinger loss when the square root of the density belongs to a Besov ball $\mathfrak{B}^{\alpha}_{p,\infty}(R)$. We make them explicit without further conditions when $p < 2$, and under a tail dominance condition when p is larger. We also show that these rates can be improved when the density is assumed to be unimodal.

MSC2020 subject classifications: 62G05, 62G07.
Keywords and phrases: Besov classes, minimax rates, shape constraints, non-parametric estimation.

Received April 2023.

Contents

1 Introduction 4527
2 Minimax rates under smoothness assumptions 4529
  2.1 Wavelet basis 4530
  2.2 Besov classes 4530
  2.3 Tail dominance condition 4531
  2.4 Minimax risk 4532
3 Mixing shape and smoothness constraints 4535
4 Proofs 4538
  4.1 Proof of Proposition 2.1 4538
  4.2 Proof of Theorem 2.2 4540
  4.3 Proof of Theorem 2.3 4546
  4.4 Sketch of the proof of Proposition 2.4 4548
  4.5 Proof of Proposition 3.1 4549
  4.6 Proof of Theorem 3.2 4550
    4.6.1 Proofs of Lemmas 6 and 7: preliminary results 4559
    4.6.2 Proof of Lemma 6 4562
    4.6.3 Proof of Lemma 7 4563
  4.7 Proof of Theorem 3.3 4565
  4.8 Proof of the claims of Section 2.4 4571
References 4575

1. Introduction

We consider n independent and identically distributed random variables $X_1, \dots,$ X_n with values in $\mathbb{R}$. We suppose that their distribution is absolutely continuous with respect to the Lebesgue measure and denote their density by f.

An important challenge in the density estimation problem is to determine as accurately as possible the minimax risk. The latter can be defined as follows. Let $\mathcal{D}(\mathbb{R})$ be the set of densities on $\mathbb{R}$, $\mathscr{F}$ be a subset of $\mathcal{D}(\mathbb{R})$, and $\mathscr{L}$ be a loss function. The minimax risk is

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

where the infimum is taken over all estimators $\hat{f}$. Different choices are possible for $\mathscr{L}$. Among them are the q^{th} powers of the $\mathbb{L}^q$ distances $\mathscr{L} = d_q^q$, or the square of the Hellinger distance $\mathscr{L} = h^2$. We recall that h is defined for all $f_1, f_2 \in \mathcal{D}(\mathbb{R})$ by

$$h^2(f_1, f_2) = \frac{1}{2} \int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 \mathrm{d}x.$$

The role of the minimax risk is to give a baseline against which to compare when proposing a statistical estimation procedure. We are more precisely interested here in the optimal estimation rate, that is in the sequence $(\varepsilon_n)_{n \geq 1}$ satisfying

$$0 < \liminf_{n \to +\infty} \varepsilon_n^{-1} \mathcal{R}(\mathscr{F}, \mathscr{L}) \leq \limsup_{n \to +\infty} \varepsilon_n^{-1} \mathcal{R}(\mathscr{F}, \mathscr{L}) < +\infty.$$

An optimal estimation procedure $\hat{f}$ is therefore a procedure whose risk $\mathbb{E}[\mathscr{L}(f, \hat{f})]$ converges at the rate ε_n under the sole condition that f lies in $\mathscr{F}$. This minimax point of view thus makes it possible to discard certain procedures that are not rate optimal, even in the a priori simple case where f is a smooth density on $\mathbb{R}$.

To formalize things a little more, we state that f is smooth if f belongs to a ball $\mathfrak{B}^{\alpha}_{p,\infty}(R)$ of a Besov space. In a nutshell, the parameter R is an upper-bound of the (quasi) Besov norm of the elements f of $\mathfrak{B}^{\alpha}_{p,\infty}(R)$. This (quasi) norm measures the variations of f by means of a (quasi) $\mathbb{L}^p$ norm and according to the smoothness exponent α. The larger p is, the more uniformly the regularity of f is measured. The latter is therefore likely to have much smaller local variations if p is large than if p is small. Note also that R induces a constraint on the (quasi) $\mathbb{L}^p$ norm of f and hence on its tails when $p < 1$ (the smaller p is, the lighter they should be). There are several possible equivalent definitions of R, and we choose one in Section 2. For the sake of rigour, we assume throughout this introduction that R is large enough ($\mathfrak{B}^{\alpha}_{p,\infty}(R)$ does not contain densities with compact support in $[0,1]$ if R is too small when $\alpha > (1/p - 1)_+$).

The minimax rates have been studied by many authors when $\mathscr{L} = d_q^q$. They are now fully known, up to log factors, when the density is also compactly supported, that is when it belongs to

$$\mathscr{F} = \left\{f \in \mathfrak{B}^{\alpha}_{p,\infty}(R),\ \operatorname{supp} f \subset [0,1]\right\}.$$

A summary of these rates can be found in [Sar21]. Let us just mention that the case $p \geq q$ can be easily solved with linear estimators. This is no longer true when $p < q$, see [DJKP96]. To be optimal, an estimator must, in some sense, adapt to local variations of the density. When, moreover, α is allowed to be smaller than $1/p$, the statistical estimation procedure must be able to cope with singularities to be optimal.

In recent years, a special endeavour has been made by statisticians to remove the assumption of compact support. For the $\mathbb{L}^q$ loss, results can be found in [JLL04, RBRTM11, GL11, Lep13, GL14, LW19, Sar23]. Other statistical frameworks have also been involved in this effort. We may cite the regression model, the problem of estimating the conditional density, the hazard rate, the intensity of a Poisson process, or the density in the convolution structure model. For more details, we refer to [RBR10, LW19, BC21, CGC21, CL23].

The aim of the present manuscript is to deal with the Hellinger loss $\mathscr{L} = h^2$. The latter naturally appears in the study of maximum likelihood estimators, see [BM98, DW16, KS16] for some references. This is also true for the T- and ρ-estimators, the founding references being [Bir06a] and [BBS17]. In the case of the Hellinger loss, the assumption of regularity is traditionally put on $\sqrt{f}$, and we will also adopt this point of view here. Note that the minimax risk has already been investigated in [Bir06a] when $\sqrt{f}$ is compactly supported and belongs to a Besov ball. The whole point of this paper is to understand how the minimax risk evolves when f is no longer assumed to be compactly supported.

For the $\mathbb{L}^q$ losses, the estimation rates remain noticeably the same as in the compact case (within possible log factors) when the tails of f are light enough, say when $f(x) \leq |x|^{-b}$ for some large b and all $|x| \geq 1$. This point has been revealed by [GL14]. Actually, there are not even logarithmic losses when $q = 1$, see [Sar23]. The situation turns out to be completely different for the Hellinger loss.

First, the minimax risk for the Hellinger loss does not tend to 0 if the only assumption made on the density is $\sqrt{f} \in \mathfrak{B}^{\alpha}_{p,\infty}(R)$ with $p \geq 2$. A supplementary condition on the tails of f is required to ensure the convergence of the minimax risk. We propose here to use the one of [Sar23]. This phenomenon can be explained by the importance that the Hellinger distance gives to the estimation errors in the tails of f. A similar result is true for the $\mathbb{L}^1$ loss when $f \in \mathfrak{B}^{\alpha}_{p,\infty}(R)$ but not for the other $\mathbb{L}^q$ losses [GL14, Sar23]. We prove that the minimax risk achieves the rate $n^{-\gamma}$ where $\gamma \in (0, 2\alpha/(2\alpha+1)]$ depends on the tails of f. But contrary to the $\mathbb{L}^q$ losses (including $q = 1$), we never have $\gamma = 2\alpha/(2\alpha+1)$ if the tail dominance condition allows $f(x) \leq |x|^{-b}$, and this, whatever the value of $b > 1$.

Second, the optimal rate of convergence is $n^{1-p/2}$ when $p < 2$ and no additional assumption is made. This result is valid for all $\alpha > 1/p - 1/2$. This rate contrasts with the classical rate $n^{-2\alpha/(2\alpha+1)}$ associated with compactly supported densities. A faster rate can be obtained under the tail dominance condition of [Sar23]. But, as above, it is not possible to recover the rate $n^{-2\alpha/(2\alpha+1)}$ if the density is allowed to be slightly fat tailed. In the remaining case $\alpha \leq 1/p - 1/2$, the minimax risk does not tend to 0 even when the density is compactly supported on $[0,1]$.

In the results mentioned above, the tails of f may not tend monotonically to 0. In other words, the density can be alternately increasing and decreasing, and this, an infinite number of times over an interval of infinite length. The fact that the density can oscillate as many times as we like is exploited in the proof of our lower bound. It is therefore natural to wonder whether banning this possibility might not improve the results. This leads us to study the minimax risk under the following three conditions: 1.) $\sqrt{f} \in \mathfrak{B}^{\alpha}_{p,\infty}(R)$ with $\alpha > \max\{1, 1/p - 1/2\}$ and $p > 0$ 2.) f is unimodal on $\mathbb{R}$ 3.) the tail dominance condition of [Sar23].

The idea of mixing two types of constraints in the density model – one of regularity and one of unimodality – also appears in [EL00, VDVVDL03, HK05, DL14, LM17, LM19] to cite a few papers. An overview of what is being done in the literature may be found in [DL18]. Unfortunately, adding the constraint "f is unimodal" to the assumption "f is smooth" generally has no impact on the convergence rates. This phenomenon may be due to the losses functions that are used or to the supplementary assumptions that are made. For instance, no improvement is to be expected for h^2 when f is compactly supported (see Theorem 3.3).

The situation is quite different in the non-compact case. The optimal estimation rates under the three above points depend on α, p and the tails of f. They are always faster than the classical rate $n^{-2/3}$ corresponding to the estimation of a bounded unimodal density with compact support. They are also always faster than the ones that can be obtained without the unimodality assumption, i.e. under points 1.) and 3.) only. Mixing a shape and smoothness constraint can therefore lead to better rates than would have been possible under these constraints taken separately.

We present our results in Sections 2 and 3. The proofs are postponed to Section 4. Throughout this paper, we suppose $n \geq 2$. Moreover, $c, c_1, c_2, \dots$ are terms that may vary from line to line. To lighten the notations, we define for all class $\mathscr{F}$ of functions,

$$\mathcal{R}(\mathscr{F}) = \mathcal{R}\left(\mathscr{F}, h^2\right).$$

We denote for $p > 0$ and $x = (x_k)_{k \in \mathbb{Z}}$ the weak (quasi) ℓ^p norm of x by

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

When $p = \infty$, we set $\|x\|_{\infty,\infty} = \|x\|_\infty$.

2. Minimax rates under smoothness assumptions

We present in this section the classes of functions we use to model the smoothness of f and the size of its tails. We then carry out the associated minimax rates.

2.1. Wavelet basis

A classical way to measure the regularity of a function is to decompose it in a wavelet basis, and to put conditions on its wavelet coefficients. We deal here with the special bi-orthogonal basis of [CDF92] where the father wavelet is $\phi = \mathbb{1}_{[0,1]}$, where the mother wavelet ψ is piecewise constant and where their duals $\bar{\phi}$ and $\bar{\psi}$ are compactly supported and Hölder continuous with exponent $\tau \in \mathbb{N}^\star$. The wavelet ψ is also orthogonal to polynomials of degree $\tau - 1$.

In this basis, any square integrable function f can be written as

$$f = \sum_{k \in \mathbb{Z}} \alpha_{J_0,k}(f) \bar{\phi}_{J_0,k} + \sum_{j=J_0}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k}(f) \bar{\psi}_{j,k}, \tag{2.1}$$

where $J_0 \in \mathbb{Z}$ is an arbitrary number to be chosen, where

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

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

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

2.2. Besov classes

We consider $p \in (0, +\infty]$, $\alpha \in ((1/p - 1)_+, \tau)$ and introduce the standard Besov space $\mathcal{B}^\alpha_{p,\infty}$. By definition, it is composed of functions f of $\mathbb{L}^{\max\{p,1\}}(\mathbb{R})$ satisfying $\|f\|_{\mathcal{B}^\alpha_{p,\infty}} < \infty$ where

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

see [DJ97]. The quantity $\|f\|_{\mathcal{B}^\alpha_{p,\infty}}$ refers to the (quasi) Besov norm of f. The Besov ball $\mathfrak{B}^\alpha_{p,\infty}(R)$ is thus defined for $R > 0$ by

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

In the present paper, we pay particular attention to the strong and weak Besov classes $\mathcal{B}^\alpha_{p,\infty}(R)$ and $\mathcal{W}\mathcal{B}^\alpha_{p,\infty}(R)$. They are defined as follows:

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

We can classify the above conditions on the wavelet coefficients by order of importance: they are the weakest for the weak Besov classes, then the strong Besov classes, and finally the Besov balls.

2.3. Tail dominance condition

We describe here a supplementary assumption that is intended to control the tails of the density.

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

$$F_{j,k}(f) = \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} f(x)\, \mathrm{d}x. \tag{2.2}$$

We set for $M > 0$, and $\theta \in (0,1)$,

$$\mathcal{T}_\theta(M) = \left\{ f \in \mathbb{L}^1(\mathbb{R}),\ f \geq 0,\ \forall j \geq 0, \|F_{j,\cdot}(f)\|_\theta^\theta \leq M 2^{j(1-\theta)} \right\}$$
$$\mathcal{WT}_\theta(M) = \left\{ f \in \mathbb{L}^1(\mathbb{R}),\ f \geq 0,\ \forall j \geq 0, \|F_{j,\cdot}(f)\|_{\theta,\infty}^\theta \leq M 2^{j(1-\theta)} \right\}.$$

The case $\theta = 0$ corresponds to compactly supported functions:

$$\begin{aligned}\mathcal{T}_\theta(M) &= \mathcal{WT}_\theta(M) \\ &= \left\{ f \in \mathbb{L}^1(\mathbb{R}),\ f \geq 0,\ \forall j \geq 0, |\{k \in \mathbb{Z},\ F_{j,k}(f) > 0\}| \leq M 2^j \right\}.\end{aligned}$$

In this formula, $|\cdot|$ denotes the size of the set between the two bars. A density belonging to one of these classes is therefore a density whose tails are sufficiently light. The smaller θ is, the lighter they are.

In line with [Sar23], we say that the "weak tail dominance condition" is fulfilled if $f \in \mathcal{WT}_\theta(M)$. The "strong tail dominance condition" is met if $f \in \mathcal{T}_\theta(M)$. This terminology "tail dominance condition" has been initially proposed by Alexander Goldenshluger and Oleg Lepski in [GL14]. Their condition do not exactly match with ours though (our conditions are always implied by theirs, see [CL20] where the condition $f \in \mathcal{T}_\theta(M)$ also appears).

We recall – see Proposition 1 of [Sar23] – that a compactly supported density on $[-L, L]$ satisfies our tail dominance condition with $\theta = 0$ and $M = 2L + 2$. This bound on M can be a bit pessimistic though. Think for example about the density f defined for $a > 0$ and $x \in \mathbb{R}$ by

$$f(x) = \frac{1}{2}\mathbb{1}_{[-a-1,-a]}(x) + \frac{1}{2}\mathbb{1}_{[a,a+1]}(x).$$

It belongs to $\mathcal{T}_0(6)$ whereas $L = a + 1$ may be taken arbitrarily large. In the non-compact case, a density f satisfying $f(x) \leq A^b |x|^{-b}$ for all $|x| \geq 1$ and some $A > 0$, $b > 1$, lies in $\mathcal{WT}_\theta(M)$ with $\theta = 1/b$ and M only depending on b, A. The (strong) tail dominance condition is automatically fulfilled with $\theta = p$ when f belongs to a Besov ball $\mathfrak{B}_{p,\infty}^\alpha(R)$ with $p < 1$. A variant of this last claim, that is useful when dealing with a smoothness assumption on $\sqrt{f}$, is the following.

Proposition 2.1. *Let $p \in (0,2)$, $R > 0$, $\alpha \in (1/p - 1/2, \tau)$ and $f \in \mathcal{D}(\mathbb{R})$. Then, if $\sqrt{f}$ belongs to $\mathfrak{B}_{p,\infty}^\alpha(R)$, f belongs to $\mathcal{T}_{p/2}(c_1 R^p)$. Conversely, if $\sqrt{f} \in \mathcal{B}_{p,\infty}^\alpha(R)$ and $f \in \mathcal{T}_{p/2}(R^p)$, then $\sqrt{f} \in \mathfrak{B}_{p,\infty}^\alpha(c_2 R)$. The terms c_1, c_2 only depend on the wavelet basis and α, p.*

2.4. Minimax risk

We now investigate the minimax risk under the preceding conditions. We consider $p \in (0, +\infty]$, $\alpha \in ((1/p - 1/2)_+, \tau)$, $\theta \in [0, p/2] \cap [0, 1)$, $R > 0$, $M \geq 1$. We define when $p \neq 2$,

$$\mathscr{S}^{\alpha}_{p,\theta}(R, M) = \left\{ f \in \mathcal{D}(\mathbb{R}),\ \sqrt{f} \in \mathcal{WB}^{\alpha}_{p,\infty}(R),\ f \in \mathcal{WT}_{\theta}(M) \right\}.$$

When $p = 2$, we rather set

$$\mathscr{S}^{\alpha}_{p,\theta}(R, M) = \left\{ f \in \mathcal{D}(\mathbb{R}),\ \sqrt{f} \in \mathcal{B}^{\alpha}_{p,\infty}(R),\ f \in \mathcal{WT}_{\theta}(M) \right\}.$$

We recall that τ is defined in Section 2.1 and exclusively depends on the wavelets. As there is a wavelet basis for each value of $\tau \in \mathbb{N}^{\star}$, we can take it arbitrarily large.

The theorem below gives a non-asymptotic upper-bound of the minimax risk when f belongs to $\mathscr{S}^{\alpha}_{p,\theta}(R, M)$.

Theorem 2.2. *For all $p \in (0, +\infty]$, $\alpha \in ((1/p - 1/2)_+, \tau)$, $\theta \in [0, p/2] \cap [0, 1)$, $R > 0$, $M \geq 1$,*

$$\mathcal{R}(\mathscr{S}^{\alpha}_{p,\theta}(R, M)) \leq c_1 \left[\varepsilon_n + (\log n) n^{-1} \right], \tag{2.3}$$

where

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

and where c_1 is a positive number only depending on p, α, θ and the wavelet basis.

This result can be compared with the following lower-bound:

Theorem 2.3. *For all $p \in (0, +\infty]$, $\alpha \in ((1/p - 1/2)_+, \tau)$, $\theta \in [0, p/2] \cap [0, 1)$, there are R_0, M_0 such that for all $R \geq R_0$, $M \geq M_0$ and n large enough,*

$$\mathcal{R}(\mathscr{S}'^{\alpha}_{p,\theta}(R, M)) \geq c_2\, \varepsilon_n,$$

where ε_n is given in the preceding theorem, and where $\mathscr{S}'^{\alpha}_{p,\theta}(R, M)$ is a subset of $\mathscr{S}^{\alpha}_{p,\theta}(R, M)$. Moreover,

$$\mathscr{S}'^{\alpha}_{p,\theta}(R, M) \subset \left\{ f \in \mathcal{D}(\mathbb{R}), \sqrt{f} \in \mathcal{B}^{\alpha}_{p,\infty}(R), f \in \mathcal{T}_{\theta}(M), \sup_{x \in \mathbb{R}} |x| f^{\theta}(x) \leq M, \|f\|_{\infty} \leq 1 \right\}.$$

When $\theta < p/2$ or when $\theta = p/2$ with $M \leq R^p$, we also have

$$\mathscr{S}'^{\alpha}_{p,\theta}(R, M) \subset \left\{ f \in \mathcal{D}(\mathbb{R}), \sqrt{f} \in \mathfrak{B}^{\alpha}_{p,\infty}(R) \right\}.$$

Above, c_2, M_0, R_0 are positive numbers only depending on p, α, θ and the wavelet basis.

When $\theta = 0$, we recover the usual estimation rate, and this, for all possible values of α and p satisfying $\alpha \in ((1/p - 1/2)_+, \tau)$. The case $\alpha \leq (1/p - 1/2)_+$ is treated below.

We observe that the optimal estimation rate is strongly affected by the parameter θ, i.e, the tails of f. The larger θ is, the slower the rate is. However, the choice of the dominance condition (whether weak or strong) has no influence on the rate. We can also assume, without changing the results, that the density is fat tailed, i.e. its tails are smaller than the inverse of a power of $|x|$. As explained in the introduction, this deterioration of rates when the density is slightly fat tailed does not occur for the $\mathbb{L}^q$ losses (whatever $q \geq 1$, and up to possible log factors).

When $p \geq 2$, the minimax rate can be made arbitrarily slow by letting θ tend to 1. Actually, it is not possible to estimate the density under the sole assumption that $\sqrt{f}$ belongs to a Besov ball $\mathfrak{B}^\alpha_{p,\infty}(R)$ with R large enough (see the proof of Theorem 2.3). The situation appears to be quite different when $p < 2$. The tail dominance condition is indeed always satisfied in this case with $\theta = p/2$. More precisely, we derive from the above: for all $p \in (0, 2)$, $\alpha \in ((1/p - 1/2)_+, \tau)$, $R \geq R_0$, and n large enough,

$$c_2 \, R^p n^{-(1-p/2)} \leq \mathcal{R}\left(\left\{f \in \mathcal{D}(\mathbb{R}),\ \sqrt{f}\ \in \mathfrak{B}^\alpha_{p,\infty}(R)\right\}\right) \leq c_1 \, R^p n^{-(1-p/2)}.$$

The rate is much slower than the standard rate $n^{-2\alpha/(2\alpha+1)}$ we would have had if the density was compactly supported though.

We will not insist on this point but the preceding rates can be reached by an adaptive estimator (that is by an estimator whose construction does not involve p, α, θ, R, M).

The proof of Theorem 2.2 is based on an oracle inequality (or model selection inequality) for the Hellinger loss. A crude version of this one is as follows: for all suitable collection $(V_m)_{m\in\mathcal{M}}$ of finite dimensional linear spaces, not containing too many spaces per dimension, we may build an estimator $\hat{f}_1$ satisfying

$$\mathbb{E}\left[h^2(f, \hat{f}_1)\right] \leq c \inf_{m\in\mathcal{M}} \left\{d_2^2\big(\sqrt{f}, V_m\big) + \frac{\dim V_m}{n}\right\}, \tag{2.4}$$

where c is a constant and where d_2 denotes the $\mathbb{L}^2$ distance.

Such an inequality can be derived from the T-estimation theory of [Bir06a]. Interestingly, the latter also leads to results for the $\mathbb{L}^2$ and $\mathbb{L}^1$ losses. More precisely, we may define estimators $\hat{f}_2$ and $\hat{f}_3$ such that

$$\mathbb{E}\left[d_2^2(f, \hat{f}_2)\right] \leq c \inf_{m\in\mathcal{M}} \left\{d_2^2\big(f, V_m\big) + \|f\|_\infty \frac{\dim V_m}{n}\right\} \tag{2.5}$$

$$\mathbb{E}\left[d_1(f, \hat{f}_3)\right] \leq c \inf_{m\in\mathcal{M}} \left\{d_1\big(f, V_m\big) + \sqrt{\frac{\dim V_m}{n}}\right\}, \tag{2.6}$$

where c is another constant. We refer to [Bir06a] and [Bir14] for more details.

We show in the proof of Theorem 2.2 that it is possible to choose the collection $(V_m)_{m\in\mathcal{M}}$ so that the infimum in (2.4) tends to 0 at the rate indicated by (2.3). The conclusion to be drawn is that an oracle inequality of this form is sufficiently precise to obtain the optimal rates of convergence under (weak) smoothness assumptions in the non compact setting.

Observe now that (2.5) and (2.6) are very similar to (2.4). More precisely, the only thing to do to go from h^2 to d_2^2 is to replace $\sqrt{f}$ by f and add a sup norm to the variance term. The bounds we get for one should therefore also work for the other, up to minor modifications. In particular, the rates should correspond when f is bounded and when we suppose "$\int f^{1/2}(x)\,\mathrm{d}x \leq M$" for the Hellinger loss to make the comparison fair as $\int f(x)\,\mathrm{d}x = 1$ for the $\mathbb{L}^2$ loss. This is not true: the optimal estimation rate of a bounded function in $\mathfrak{B}^{\alpha}_{p,\infty}(R)$ for the $\mathbb{L}^2$ loss is the standard one $n^{-2\alpha/(2\alpha+1)}$ when $p \in [1,2]$ and $\alpha > 1/p - 1/2$, even in the non-compact case (at least up to log factors, see [RBRTM11]). This rate is faster than the rate $n^{-\alpha/(2\alpha+1-1/p)}$ given by Theorem 2.2 when $\sqrt{f} \in \mathfrak{B}^{\alpha}_{p,\infty}(R)$ is bounded and in $\mathcal{T}_{1/2}(M)$ (this last condition implies $\int (f(x))^{1/2}\,\mathrm{d}x \leq M$).

This paradox is in fact due to the variance term in (2.5) that may be too large. A term of the order of "model dimension over n" is of the right order of magnitude when we deal with h^2. But not necessarily for d_2^2.

Another way of looking at it, which is perhaps a little more revealing, is to apply these inequalities to a single model. Introduce for all $j \geq 0$,

$$K_j = \left\{k \in \mathbb{Z},\, \mathbb{P}\left(X \in \operatorname{supp}\, \psi_{j,k}\right) \;\geq 1/n\right\}$$
$$K_{-1} = \left\{k \in \mathbb{Z},\, \mathbb{P}\left(X \in \operatorname{supp}\, \phi_{0,k}\right) \;\geq 1/n\right\}$$

and define for $J \geq 0$,

$$V = \left\{ \sum_{k\in K_{-1}} \gamma_{-1,k}\bar{\phi}_{0,k} + \sum_{j=0}^{J}\sum_{k\in K_j} \gamma_{j,k}\bar{\psi}_{j,k},\ \forall j \geq -1,\, k \in K_j,\ \gamma_{j,k}\in\mathbb{R} \right\}.$$

In the unrealistic but very favourable situation where the sets K_j are known, it is possible to build estimators that satisfy (2.4), (2.5), (2.6) with $(V_m)_{m\in\mathcal{M}}$ reduced to V.

The size of each set K_j can be bounded by $c' n^{\theta} 2^{j(1-\theta)}$ when $f \in \mathcal{WT}_{\theta}(M)$ (as ϕ and ψ are compactly supported). In particular, the dimension of V is no larger than

$$\dim V \;\leq c'' M 2^{J(1-\theta)} n^{\theta}.$$

We show in Section 4.8:

$$\begin{aligned}
d_2^2\big(\sqrt{f}, V\big) &\leq c'''\left[R^2 2^{-2J\alpha} + \varepsilon_n\right] && \text{if } \sqrt{f} \in \mathcal{B}^{\alpha}_{2,\infty}(R) \text{ and } f \in \mathcal{T}_{1/2}(M)\\
d_2^2\big(f, V\big) &\leq c'''\left[R^2 2^{-2J\alpha} + \varepsilon_n'\right] && \text{if } f \in \mathcal{B}^{\alpha}_{2,\infty}(R)\\
d_1(f, V) &\leq c'''\left[R 2^{-J\alpha} + \varepsilon_n''\right] && \text{if } f \in \mathcal{B}^{\alpha}_{1,\infty}(R) \cap \mathcal{T}_{1/2}(M).
\end{aligned}$$

Above,

$$\varepsilon'_n = R^{2/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} + n^{-1}$$
$$\varepsilon''_n = R^{1/(2\alpha+1)} M^{2\alpha/(2\alpha+1)} n^{-\alpha/(2\alpha+1)} + M n^{-1/2}.$$

We now choose 2^J to optimize the right-hand sides of (2.4), (2.5) and (2.6). We deduce the rate $n^{-2\alpha/(4\alpha+1)}$ for $\hat{f}_1$, $n^{-2(1-\theta)\alpha/(1-\theta+2\alpha)}$ for $\hat{f}_2$, and $n^{-\alpha/(4\alpha+1)}$ for $\hat{f}_3$. The first is the right rate of convergence. Not the other two.

Yet, estimating such a function for the $\mathbb{L}^1$ or $\mathbb{L}^2$ loss is not a difficult problem. By way of example, a suitable linear estimator $\tilde{f}$ in V or in

$$V' = \left\{ \sum_{k\in\mathbb{Z}} \gamma_{-1,k} \bar{\phi}_{0,k} + \sum_{j=0}^{J} \sum_{k\in\mathbb{Z}} \gamma_{j,k} \bar{\psi}_{j,k},\ \forall j \geq -1,\ k \in K_j,\ \gamma_{j,k} \in \mathbb{R} \right\},$$

suits (if J is correctly chosen): it satisfies for n large enough,

$$\sup_{f\in\mathcal{B}^{\alpha}_{2,\infty}(R)} \mathbb{E}\left[d_2^2(f,\tilde{f})\right] \leq C n^{-2\alpha/(2\alpha+1)}$$
$$\sup_{f\in\mathcal{B}^{\alpha}_{1,\infty}(R)\cap\mathcal{T}_{1/2}(M)} \mathbb{E}\left[d_1(f,\tilde{f})\right] \leq C n^{-\alpha/(2\alpha+1)},$$

and converges therefore faster than the rates given by (2.5) and (2.6). The proof of these inequalities is given in Section 4.8. More general results for the $\mathbb{L}^1$ loss are to be found in [Sar23].

In Theorems 2.2 and 2.3, we assumed $\alpha > 1/p - 1/2$ when $p < 2$. This condition is necessary to ensure the convergence of the minimax risk, even when the density is compactly supported. We may indeed show:

Proposition 2.4. *For all $p \in (0,2)$, $R > 0$, $\tau > 1/p - 1/2$ and $\alpha = 1/p - 1/2$,*

$$\mathcal{R}\left(\left\{f \in \mathcal{D}(\mathbb{R}),\ \sqrt{f}\ \in \mathfrak{B}^{\alpha}_{p,\infty}(R),\ \textit{supp}\ f\ \subset [0,1]\right\}\right) \geq 1/16.$$

It is interesting to note that the exponent in the optimal rate does not tend to 0 when $\alpha \to 1/p - 1/2$. There is thus a kind of discontinuity at the boundary $\alpha = 1/p - 1/2$. A similar phenomenon occurs for the $\mathbb{L}^1$ distance but not for the other $\mathbb{L}^q$ distances, see [Sar21, Sar23].

3. Mixing shape and smoothness constraints

As explained in the previous section, the assumption "f is compactly supported" cannot be weakened to include densities whose tails are bounded by $|x|^{-1/\theta}$ without this having a substantial impact on the results. Such a minimax approach is always a little pessimistic though. The target function may well have properties other than regularity. Many densities, for example, have tails that tend monotonically to 0. This leads us to wonder whether adding the constraint "f is unimodal on $\mathbb{R}$" might improve the results. If this is true when $\alpha > 1$ and $\theta \neq 0$, this will indicate that the estimation errors can be better controlled when the tails are not allowed to oscillate.

Throughout this section, we suppose that f is unimodal with unknown mode, that is $f \in \mathscr{U}$ where

$$\mathscr{U} = \{f \in \mathcal{D}(\mathbb{R}), \text{ there exists } m \in \mathbb{R} \text{ such that } f \text{ is non-decreasing on } (-\infty, m] \text{ and non-increasing on } [m, +\infty)\} .$$

It turns out that the weak tail dominance condition can be written more simply when f is unimodal. Consider $\theta \in (0,1)$, $M \geq 1$ and

$$\mathbb{L}^{\theta,\infty}(M) = \left\{f \in \mathcal{D}(\mathbb{R}),\ \forall t > 0,\ |\{x \in \mathbb{R},\ f(x) \geq t\}| < M t^{-\theta}\right\}.$$

When f is unimodal with mode at $m \in \mathbb{R}$, f belongs to $\mathbb{L}^{\theta,\infty}(M)$ implies

$$f(x) \leq \frac{M^{1/\theta}}{|x-m|^{1/\theta}}, \tag{3.1}$$

for all $x \neq m$. Conversely, if f satisfies (3.1), f lies in $\mathbb{L}^{\theta,\infty}(cM)$ for all $c > 2$. This inequality is most informative when x moves significantly away from m ($f(m)$ may be infinite). It can therefore be seen as a condition on the tails of f. In any way, it is equivalent to the weak tail dominance condition:

Proposition 3.1. *Consider $M \geq 1$, $\theta \in (0,1)$ and suppose that $f \in \mathscr{U}$. If f belongs to $\mathbb{L}^{\theta,\infty}(M)$, then*

$$|\{k \in \mathbb{Z},\ F_{j,k}(f) \geq t\}| \leq c_1 \left[1 + M 2^{j(1-\theta)} t^{-\theta}\right] \tag{3.2}$$

for all $j \in \mathbb{Z}$ and $t > 0$. In particular, f belongs to $\mathcal{WT}_\theta(2c_1 M)$. Conversely, if f belongs to $\mathcal{WT}_\theta(M)$, then f lies in $\mathbb{L}^{\theta,\infty}(c_2 M)$. The terms c_1 and c_2 are constants.

We now consider $p \in (0, +\infty]$, $\alpha \in (\max(1/p - 1/2, 1), \tau)$, $\theta \in (0, p/2] \cap [0, 1)$, $R > 0$ and $M \geq 1$. We study here the minimax risk on

$$\mathscr{U}\mathscr{S}^{\alpha}_{p,\theta}(R, M) = \mathscr{U} \cap \mathscr{S}^{\alpha}_{p,\theta}(R, M).$$

It involves the following parameters:

$$\begin{aligned} t &= \frac{(2\alpha+1)(1-\theta)}{1 + 2\alpha + 2\theta + 4\alpha\theta - 6\theta/p} \\ \gamma &= 2t\alpha/(2\alpha+1) + 2(1-t)/3 \\ \beta_1 &= \frac{2(1-\theta)}{1 + 2\alpha + 2\theta + 4\alpha\theta - 6\theta/p} \\ \beta_2 &= \frac{1 + 2\alpha - 2/p}{1 + 2\alpha + 2\theta + 4\alpha\theta - 6\theta/p}. \end{aligned} \tag{3.3}$$

We now state:

Theorem 3.2. *For all $p \in (0, +\infty]$, $\alpha \in (\max(1/p - 1/2, 1), \tau)$, $\theta \in (0, p/2] \cap (0, 1)$, $R > 0$, $M \geq 1$,*

$$\mathcal{R}(\mathscr{U}\mathscr{S}^{\alpha}_{p,\theta}(R, M)) \leq c\left[R^{\beta_1} M^{\beta_2} n^{-\gamma} + \upsilon_n\right], \tag{3.4}$$

where c only depends on p, α, θ and the wavelet basis. The term υ_n only depends on $p, \alpha, \theta, M, R, n$, and tends to 0 faster than $n^{-\gamma}$.

In this inequality, υ_n is smaller than $R^{\beta_1} M^{\beta_2} n^{-\gamma}$ when $n \geq n_0$ for some n_0 only depending on p, α, θ, M, R. It does, however, appear in (3.4) as the result is non-asymptotic. Its expression is not displayed here as it is a little cumbersome. It can be found in the proof in the theorem, see (4.21) and (4.27).

In (3.3), t is between 0 and 1. The smaller θ is, the larger t is, and the closer the rate is to the standard estimation rate $n^{-2\alpha/(2\alpha+1)}$ of a compactly supported density. However, and this is a major improvement on the previous section, the exponent γ gets closer to $2/3$ when θ becomes very close to 1. Without the additional shape constraint, the rate became arbitrarily slow.

We can check that

$$\gamma > \max\left\{2\alpha(1-\theta)/(2\alpha + 1 - 2\theta/p), 2/3\right\}.$$

We recall that the first term in the maximum refers to the exponent we have when we estimate a fat tailed density f whose square root is α-smooth. The second is the usual exponent corresponding to the estimation of a unimodal density. Here, γ is larger than these two exponents. Thereby, associating a smoothness assumption with a shape constraint may lead to faster rates of convergence than those achievable under these assumptions taken separately.

This phenomenon occurs even with a very mild smoothness assumption: ours does not even guarantee the continuity of f when $p \in (1/\tau, 1)$ and $\alpha \in (\max(1/p - 1/2, 1), 1/p]$. Consider indeed two sequences $(a_k)_{k\geq 1}$ and $(b_k)_{k \geq 1}$ of non-negative numbers. The first sequence $(a_k)_k$ is assumed to be non-decreasing, and the second $(b_k)_k$ is non-increasing. We also suppose:

$$\sum_{k=1}^{\infty} a_k (k(k+1))^{-1} + \sum_{k=1}^{\infty} b_k = 1/2 \tag{3.5}$$

$$\sum_{k=2}^{\infty} b_k^{\theta} \leq M \tag{3.6}$$

$$\sum_{k=1}^{\infty} a_k^{p/2} (k(k+1))^{\alpha p - 1} + \sum_{k=1}^{\infty} b_k^{p/2} \leq R^p. \tag{3.7}$$

We then define the function f for $x \geq 0$ by

$$f(x) = \sum_{k=1}^{\infty} a_k \mathbb{1}_{[1/(k+1), 1/k)}(x) + \sum_{k=1}^{\infty} b_k \mathbb{1}_{[k-1,k)}(x),$$

and extend it to an even function on $\mathbb{R}$. We may check that f is unimodal. Equality (3.5) ensures that f is a density, (3.6) gives an upper-bound on the $\mathbb{L}^{\theta}$

(quasi) norm of f. In particular, $f \in \mathcal{WT}_\theta(c_1 M)$ for some c_1, see Proposition 3.1. Elementary maths based on (3.7) lead to $\sqrt{f} \in \mathfrak{B}^\alpha_{p,\infty}(c_2 R)$ for some $c_2 > 0$ only depending on the wavelet basis, p, α when $p \in (1/\tau, 1)$ and $\alpha \in (\max(1/p - 1/2, 1), 1/p]$. In conclusion, $f \in \mathscr{U}\mathscr{S}^\alpha_{p,\theta}(c_1 R, c_2 M)$. This function f can admit an infinite number of discontinuities and not be bounded (and of course not be compactly supported).

Although Theorem 3.2 only establishes a minimax risk bound, the proof gives an estimator that achieves this rate adaptively, i.e., without a priori knowledge of p, α, θ, R, M.

We still need to verify that (3.4) is sharp. This is the aim of the following theorem.

Theorem 3.3. *For all $p \in (0, +\infty]$, $\alpha \in (\max(1/p - 1/2, 1), \tau)$, $\theta \in (0, p/2) \cap (0, 1)$, there are R_0, M_0 such that for all $R \geq R_0$, $M \geq M_0$ and n large enough,*

$$\mathcal{R}(\mathscr{U}\mathscr{S}'^\alpha_{p,\theta}(R, M)) \geq c R^{\beta_1} M^{\beta_2} n^{-\gamma},$$

where $\mathscr{U}\mathscr{S}'^\alpha_{p,\theta}(R, M)$ is a subset of $\mathscr{U}\mathscr{S}^\alpha_{p,\theta}(R, M)$ and where c is a positive term only depending on α, p, θ and the wavelet basis. Moreover,

$$\mathscr{U}\mathscr{S}'^\alpha_{p,\theta}(R, M) \subset \left\{ f \in \mathscr{U},\ \sqrt{f} \in \mathfrak{B}^\alpha_{p,\infty}(R),\ \int_{\mathbb{R}} f^\theta(x)\, \mathrm{d}x \leq M \right\}.$$

All the above remains true when $\theta = p/2$ under the additional condition $M \leq R^p$. The lower bound is also true when $\theta = 0$. In this case, the densities of $\mathscr{U}\mathscr{S}'^\alpha_{p,\theta}(R, M)$ are compactly supported in $[-M, M]$, $t = 1$, $\gamma = 2\alpha/(2\alpha + 1)$, $\beta_1 = 2/(1 + 2\alpha)$, $\beta_2 = (1 + 2\alpha - 2/p)/(1 + 2\alpha)$.

4. Proofs

4.1. Proof of Proposition 2.1

We only show that if $\sqrt{f}$ belongs to $\mathfrak{B}^\alpha_{p,\infty}(R)$, then $f \in \mathcal{T}_{p/2}(c_1 R^p)$. The proof of the converse is straightforward (just apply Cauchy-Schwarz inequality). To simplify the notations, we omit the square root of f in the wavelets coefficients.

We define

$$F_{k,j_1,j_2} = \sum_{\substack{k_1 \in \mathbb{Z} \\ k_2 \in \mathbb{Z}}} |\beta_{j_1,k_1}| |\beta_{j_2,k_2}| I_{j,k,j_1,k_1,j_2,k_2},$$

where

$$I_{j,k,j_1,k_1,j_2,k_2} = \int_{2^{-j}(k-1/2)}^{2^{-j}(k+1/2)} \left| \bar{\psi}_{j_1,k_1} \bar{\psi}_{j_2,k_2} \right|.$$

Since $p < 2$, $\|\cdot\|_1 \leq \|\cdot\|_{p/2}$, and hence

$$(F_k(f))^{p/2} \leq 2 \sum_{\substack{j_1 \geq -1 \\ j_2 \geq j_1}} (F_{k,j_1,j_2})^{p/2}. \tag{4.1}$$

We consider a real number $\bar{L} > 0$ large enough to ensure that supp $\bar{\psi}_{j',k'} \subset [2^{-j'_+}(-\bar{L}+k'), 2^{-j'_+}(\bar{L}+k')]$, where $j'_+ = \max\{j,0\}$, and set

$$K_{j,j',k'} = \left[-\bar{L} + 2^{j-j'_+}\left(k'-\bar{L}\right), \bar{L} + 2^{j-j'_+}\left(k'+\bar{L}\right)\right].$$

Note that $K_{j,j',k}$ contains at most

$$|K_{j,j',k'}| \le c_1\left[1+2^{j-j'}\right] \tag{4.2}$$

integers. Moreover, $I_{j,k,j_1,k_1,j_2,k_2} = 0$ if $k_1 \notin K_{j_1,j_2,k_2}$ or if $k_2 \notin K_{j_2,j_1,k_1}$. If $\bar{L}$ is large enough, the integral is also zero if $k \notin K_{j,j_1,k_1}$ or if $k \notin K_{j,j_2,k_2}$. The same thing is true if $k_1 \notin K_{j_1,j,k}$ or $k_2 \notin K_{j_2,j,k}$. In any case, we have $I_{j,k,j_1,k_1,j_2,k_2} \le c_2 r_{j_1,j_2}$ where

$$r_{j_1,j_2} = 2^{\min\{j_1/2+j_2/2-j,-(j_2-j_1)/2\}}.$$

We deduce from Cauchy-Schwarz inequality,

$$\begin{aligned} F_{k,j_1,j_2} \le c_2 r_{j_1,j_2} &\left(\sum_{k_1\in\mathbb{Z}} \beta_{j_1,k_1}^2 \mathbb{1}_{k\in K_{j,j_1,k_1}} \sum_{k_2\in\mathbb{Z}} \mathbb{1}_{k_2\in K_{j_2,j_1,k_1}\cap K_{j_2,j,k}}\right)^{1/2} \\ &\times\left(\sum_{k_2\in\mathbb{Z}} \beta_{j_2,k_2}^2 \mathbb{1}_{k\in K_{j,j_2,k_2}} \sum_{k_1\in\mathbb{Z}} \mathbb{1}_{k_1\in K_{j_1,j_2,k_2}\cap K_{j_1,j,k}}\right)^{1/2}. \end{aligned}$$

By using the inequality $\|\cdot\|_1 \le \|\cdot\|_{p/2}$ again,

$$\begin{aligned} &(F_{k,j_1,j_2})^{p/2} \\ &\le c_3 r_{j_1,j_2}^{p/2}\left(\sum_{k_1\in\mathbb{Z}} \beta_{j_1,k_1}^p \mathbb{1}_{k\in K_{j,j_1,k_1}}\left(\sum_{k_2\in\mathbb{Z}} \mathbb{1}_{k_2\in K_{j_2,j_1,k_1}\cap K_{j_2,j,k}}\right)^{p/2}\right)^{1/2} \\ &\quad\times\left(\sum_{k_2\in\mathbb{Z}} \beta_{j_2,k_2}^p \mathbb{1}_{k\in K_{j,j_2,k_2}}\left(\sum_{k_1\in\mathbb{Z}} \mathbb{1}_{k_1\in K_{j_1,j_2,k_2}\cap K_{j_1,j,k}}\right)^{p/2}\right)^{1/2}. \end{aligned}$$

A new application of Cauchy-Schwarz leads to

$$\begin{aligned} &\sum_{k\in\mathbb{Z}}(F_{k,j_1,j_2})^{p/2} \\ &\le c_3 r_{j_1,j_2}^{p/2}\left(\sum_{k\in\mathbb{Z}}\sum_{k_1\in\mathbb{Z}} \beta_{j_1,k_1}^p \mathbb{1}_{k\in K_{j,j_1,k_1}}\left(\sum_{k_2\in\mathbb{Z}} \mathbb{1}_{k_2\in K_{j_2,j_1,k_1}\cap K_{j_2,j,k}}\right)^{p/2}\right)^{1/2} \\ &\quad\times\left(\sum_{k\in\mathbb{Z}}\sum_{k_2\in\mathbb{Z}} \beta_{j_2,k_2}^p \mathbb{1}_{k\in K_{j,j_2,k_2}}\left(\sum_{k_1\in\mathbb{Z}} \mathbb{1}_{k_1\in K_{j_1,j_2,k_2}\cap K_{j_1,j,k}}\right)^{p/2}\right)^{1/2} \end{aligned}$$

$$\leq c_3 r_{j_1,j_2}^{p/2} R^p 2^{-(j_1+j_2)(p/2)(\alpha+1/2-1/p)} \left(\sup_{k,k_1 \in \mathbb{Z}} |K_{j,j_1,k_1}||K_{j_2,j_1,k_1} \cap K_{j_2,j,k}|^{p/2} \right)^{1/2}$$
$$\times \left(\sup_{k,k_2 \in \mathbb{Z}} |K_{j,j_2,k_2}||K_{j_1,j_2,k_2} \cap K_{j_1,j,k}|^{p/2} \right)^{1/2}.$$

We now use (4.2) to get if $j_2 \geq j$ and $j_1 \geq j$

$$\sum_{k \in \mathbb{Z}} (F_{k,j_1,j_2})^{p/2} \leq c_4 R^p 2^{-(p/2)(j_1+j_2)(\alpha-1/p+1/2)}.$$

If $j_2 \geq j$ and $j_1 < j$,

$$\sum_{k \in \mathbb{Z}} (F_{k,j_1,j_2})^{p/2} \leq c_4 R^p 2^{j(1-p/2)/2} 2^{-(p/2)j_2(\alpha-1/p+1/2)} 2^{-\alpha(p/2)j_1}.$$

If $j_2 \leq j$, and $j_1 \leq j$,

$$\sum_{k \in \mathbb{Z}} (F_{k,j_1,j_2})^{p/2} \leq c_4 R^p 2^{j(1-p/2)} 2^{-(p/2)j_1\alpha} 2^{-(p/2)j_2\alpha}.$$

We conclude thanks to (4.1). □

4.2. Proof of Theorem 2.2

Our proof relies on the result below that is due to [Bir06a] (see his Theorem 6 and Proposition 8).

Proposition 4.1. *Let $(V_m)_{m \in \mathcal{M}}$ be an at most countable collection of linear spaces of $\mathbb{L}^2(\mathbb{R})$ with finite dimension. Let $(\Delta_m)_{m \in \mathcal{M}}$ be a family of non-negative weights such that*

$$\sum_{m \in \mathcal{M}} e^{-\Delta_m} \leq 1.$$

Then, there is a density estimator $\hat{f}$ such that

$$\mathbb{E}\left[h^2(f, \hat{f})\right] \leq c \inf_{m \in \mathcal{M}} \left\{ d_2^2(\sqrt{f}, V_m) + \frac{\dim V_m + \Delta_m}{n} \right\}.$$

In the above inequality, c is a universal constant.

Without loss of generality, we may assume in the sequel that we have another independent sample $X_1', \dots, X_n'$ of X. We set for $j \geq 0$ and $k \in \mathbb{Z}$,

$$I_{j,k} = \{x \in \mathbb{R},\, \psi_{j,k}(x) \neq 0\}.$$

When $j = -1$, we rather set

$$I_{-1,k} = \{x \in \mathbb{R},\, \phi_{0,k}(x) \neq 0\}.$$

We then consider the (random) set

$$\widetilde{\mathbb{Z}}_j = \{k \in \mathbb{Z},\ \exists i \in \{1, \ldots, n\},\ X'_i \in I_{j,k}\}.$$

We now order the sample $X'_{(1)} < X'_{(2)} < \cdots < X'_{(n)}$ and define the smallest integer $\widetilde{J} \geq 0$ satisfying

$$\min_{1 \leq i \leq n-1} \left(X'_{(i+1)} - X'_{(i)} \right) > 2^{1-\widetilde{J}} L_\psi.$$

In this inequality, $L_\psi \geq 1$ stands for a real number such that supp $\psi \subset [-L_\psi, L_\psi]$.

Let $\widetilde{\mathcal{K}}$ be the collection of all sets of the form $\mathbf{K} = (K_j)_{j \in \{-1,\ldots,\widetilde{J}\}}$ where K_j denotes a finite subset of $\widetilde{\mathbb{Z}}_j$. We define for all such $\mathbf{K}$ the linear space

$$V_{\mathbf{K}} = \left\{ \sum_{k \in K_{-1}} \gamma_{-1,k} \bar{\phi}_{0,k} + \sum_{j=0}^{\widetilde{J}} \sum_{k \in K_j} \gamma_{j,k} \bar{\psi}_{j,k},\ \forall j \geq -1,\ k \in K_j,\ \gamma_{j,k} \in \mathbb{R} \right\}.$$

The dimension of this linear space is not larger than

$$\dim V_{\mathbf{K}} \leq \sum_{j=-1}^{\widetilde{J}} |K_j|.$$

For all $\mathbf{K} \in \widetilde{\mathcal{K}}$, we set

$$\Delta_{\mathbf{K}} = \sum_{j=-1}^{\widetilde{J}} \left\{ |K_j| + |K_j| \log\left(e|\widetilde{\mathbb{Z}}_j|/|K_j| \right) - \log\left(1 - e^{-1}\right) \right\},$$

where we use the convention $0 \times \log(e|\widetilde{\mathbb{Z}}_j|/0) = 0$. It follows from Proposition 2.5 of [Mas07] that

$$\sum_{\mathbf{K} \in \widetilde{\mathcal{K}}} e^{-\Delta_{\mathbf{K}}} \leq 1.$$

We apply Proposition 4.1 conditionally to the independent sample $X'_1, \ldots, X'_n$ and take the expectation of the result. By cleaning it a little, we get

$$\mathbb{E}\left[h^2(f, \hat{f}) \right] \leq c_1 \mathbb{E}\left[\inf_{\mathbf{K} \in \widetilde{\mathcal{K}}} \left\{ d_2^2\big(\sqrt{f}, V_{\mathbf{K}}\big) + \frac{1}{n} \sum_{j=-1}^{\widetilde{J}} |K_j| \log_+ \left(|\widetilde{\mathbb{Z}}_j|/|K_j| \right) + \frac{\widetilde{J}+1}{n} \right\} \right],$$

where c_1 is universal, where $\log_+(x) = \log(e+x)$, and where $0 \times \log_+(|\widetilde{\mathbb{Z}}_j|/0) = 0$. To simplify the notations, we set in the sequel

$$\beta^\star_{j,k} = \begin{cases} \beta_{j,k}\big(\sqrt{f}\big) & \text{if } j \geq 0 \\ \alpha_{0,k}\big(\sqrt{f}\big) & \text{if } j = -1. \end{cases}$$

We deduce from the above inequality, and from (2.1) with $J_0 = 0$,

$$\begin{aligned}\mathbb{E}\left[h^2(f,\hat{f})\right] &\leq c_2 \left\{ \mathbb{E}\left[\sum_{j=-1}^{\widetilde{J}} \inf_{K_j \subset \widetilde{\mathbb{Z}}_j} \left\{ \sum_{k \in \mathbb{Z} \setminus K_j} (\beta^\star_{j,k})^2 + \frac{|K_j| \log_+(|\widetilde{\mathbb{Z}}_j|/|K_j|)}{n} \right\}\right] \right. \\ &\qquad \left. + \frac{\mathbb{E}\big[\widetilde{J}\big]+1}{n} + \mathbb{E}\left[\sum_{j=\widetilde{J}+1}^{\infty} \sum_{k\in\mathbb{Z}} (\beta^\star_{j,k})^2\right]\right\} \\ &\leq c_2 \left[A + R_1 + R_2 + T\right],\end{aligned}$$

where

$$\begin{aligned} A &= \frac{\mathbb{E}\big[\widetilde{J}\big]+1}{n}, \\ R_1 &= \frac{\mathbb{E}[|\widetilde{\mathbb{Z}}_{-1}|]}{n}, \\ R_2 &= \sum_{j=0}^{\infty} \mathbb{E}\left[\inf_{K_j \subset \widetilde{\mathbb{Z}}_j} \left\{ \sum_{k \in \widetilde{\mathbb{Z}}_j \setminus K_j} (\beta^\star_{j,k})^2 + \frac{|K_j| \log_+(|\widetilde{\mathbb{Z}}_j|/|K_j|)}{n} \right\}\right], \\ T &= \mathbb{E}\left[\sum_{k \notin \widetilde{\mathbb{Z}}_{-1}} \alpha_{0,k}^2 + \sum_{j=0}^{\widetilde{J}} \sum_{k \notin \widetilde{\mathbb{Z}}_j} (\beta^\star_{j,k})^2 + \sum_{j=\widetilde{J}+1}^{\infty} \sum_{k\in\mathbb{Z}} (\beta^\star_{j,k})^2\right]. \end{aligned}$$

This oracle inequality has the same flavour as that obtained by [Sar23] for the $\mathbb{L}^1$ loss (see his inequality (14)). We can hence use some of its results to reduce the size of this proof. First, note that an upper-bound on R_1 is given by his Lemma 23: $R_1 \leq c_3 M n^{-(1-\theta)}$. For T, A and R_2, we show:

Lemma 1. *There exists $c_4 > 0$ only depending on p, α and the wavelet basis such that*

$$T \leq c_4 \varepsilon_n.$$

Lemma 2. *There exist $c_5, c_6 > 0$ only depending on p, α and the wavelet basis such that*

$$\begin{aligned} A &\leq c_5 \frac{\log n + \log(1+R)}{n} \\ &\leq c_6 \left[\varepsilon_n + \frac{\log n}{n}\right]. \end{aligned}$$

Lemma 3. *There exists $c_7 > 0$ only depending on p, α and the wavelet basis such that*

$$R_2 \leq c_7 \varepsilon_n.$$

It then remains to put all these bounds together to conclude. □

Proof of Lemma 1. Define the number $\tilde{n}_{j,k}$ of $i \in \{1, \dots, n\}$ such that $X'_i \in I_{j,k}$. We have $\tilde{n}_{j,k} \le 1$ if $k \notin \widetilde{\mathbb{Z}}_j$ or if $j \ge \widetilde{J} + 1$. Hence,

$$T \le 2 \sum_{j=-1}^{\infty} \sum_{k \in \mathbb{Z}} (\beta^\star_{j,k})^2 \mathbb{P}\left[\tilde{n}_{j,k} \le 1\right].$$

Set

$$\begin{aligned} f_{j,k} &= \int f(x) \mathbb{1}_{I_{j,k}}(x)\, \mathrm{d}x \\ \overline{\mathbb{Z}}_j &= \{k \in \mathbb{Z},\ f_{j,k} \ge 1/n\}. \end{aligned} \tag{4.3}$$

We have,

$$\begin{aligned} T &\le 2 \sum_{j=-1}^{\infty} \sum_{k \in \mathbb{Z}} (\beta^\star_{j,k})^2 \left[(1 - f_{j,k})^n + n f_{j,k}(1 - f_{j,k})^{n-1}\right] \\ &\le 4T_1 + 4T_2, \end{aligned}$$

where

$$\begin{aligned} T_1 &= \sum_{j=-1}^{\infty} \sum_{k \notin \overline{\mathbb{Z}}_j} (\beta^\star_{j,k})^2 \\ T_2 &= n \sum_{j=-1}^{\infty} \sum_{k \in \overline{\mathbb{Z}}_j} (\beta^\star_{j,k})^2 f_{j,k}(1 - f_{j,k})^{n-1}. \end{aligned}$$

Define $b_{j,k}$ such that $(\beta^\star_{j,k})^2 = 2^{-j/2}|b_{j,k}|$. For all $j \ge 0$,

$$\|b_{j,\cdot}\|_{p/2} \le R^2 2^{-j(2\alpha + 1/2 - 1/(p/2))}$$

if $\sqrt{f} \in \mathcal{B}^\alpha_{p,\infty}(R)$. This inequality also holds true for the weak $\ell^{p/2}$ (quasi) norm if $\sqrt{f}$ belongs to the weak Besov class. We conclude by using Lemma 21 of [Sar23] with his $\beta_{j,k}$ replaced by $b_{j,k}$, p by $p/2$, α by 2α and R by R^2. □

Proof of Lemma 2. Let $\xi > 0$, $q > 1$ and suppose that the $\mathbb{L}^q$ norm of f is finite: $\|f\|_q < \infty$. Lemma 17 of [Sar23] ensures that $\widetilde{J} \le c_1 [1 + \log(1 + \xi)]$, with probability $1 - n^2\|f\|_q/\xi$. In this inequality, c_1 is a term only depending on q and ψ. We deduce,

$$\begin{aligned} \mathbb{E}[\widetilde{J}] &\le c_2 \left[1 + \int_0^\infty \mathbb{P}\left(\widetilde{J} \ge c_1 [1 + \log(1 + \xi)]\right) (1 + \xi)^{-1}\, \mathrm{d}\xi\right] \\ &\le c_3 \left[1 + c_1 n^2 \|f\|_q \int_{\max\{n^2\|f\|_q, 1\}}^\infty \xi^{-1}(1 + \xi)^{-1} d\xi \right. \\ &\qquad \left. + c_1 \int_0^{\max\{n^2\|f\|_q, 1\}} (1 + \xi)^{-1} d\xi \right] \end{aligned}$$

$$\leq c_4 \left[1 + \log(1 + n^2 \|f\|_q)\right].$$

It then remains to bound $\|f\|_q$ for some $q > 1$.

We consider $q \in (\max\{1, p/2\}, p(\alpha + 1/2))$ if p is finite and $q > 1$ if p is infinite. When $\sum_{j=-1}^{\infty} 2^{(j/2)(1-1/q)} \|\beta^\star_{j,\cdot}\|_{2q}$ is finite, $\sqrt{f} \in \mathbb{L}^{2q}(\mathbb{R})$, and

$$\begin{aligned} \|f\|_q^{1/2} &= \left\|\sqrt{f}\right\|_{2q} \\ &\leq c_5 \sum_{j=-1}^{\infty} 2^{(j/2)(1-1/q)} \|\beta^\star_{j,\cdot}\|_{2q}. \end{aligned} \tag{4.4}$$

Note that $|\beta^\star_{j,k}| \leq c_6 \sqrt{f_{j,k}} \leq c_6$ as f is a density and hence

$$\|\beta^\star_{j,\cdot}\|_{2q}^{2q} \leq c_7 \sum_{k \in \mathbb{Z}} f_{j,k} \leq c_8.$$

We moreover have when p is finite and $j \geq 0$,

$$\left\|\beta^\star_{j,\cdot}\right\|_{2q}^{2q} \leq c_9 \left\|\beta^\star_{j,\cdot}\right\|_{p,\infty}^{p} \leq c_9 R^p 2^{-jp(\alpha + 1/2 - 1/p)}.$$

When $p = \infty$ and $j \geq 0$, we rather have,

$$\begin{aligned} \left\|\beta^\star_{j,\cdot}\right\|_{2q}^{2q} &\leq c_{10} \left(\sum_{k \in \mathbb{Z}} f_{j,k}\right) \left\|\beta^\star_{j,\cdot}\right\|_{\infty}^{2(q-1)} \\ &\leq c_{11} R^{2(q-1)} 2^{-j(q-1)(2\alpha+1)}. \end{aligned}$$

In both cases,

$$\sum_{j=0}^{\infty} 2^{(j/2)(1-1/q)} \left\|\beta^\star_{j,\cdot}\right\|_{2q} < c_{12} R^r, \tag{4.5}$$

where $r = p/(2q)$ if p is finite, and $r = 1 - 1/q$ if $p = \infty$. We conclude by (4.4). □

Proof of Lemma 3 when $p \geq 2$. By choosing $K_j = \emptyset$ or $K_j = \widetilde{\mathbb{Z}}_j$,

$$R_2 \leq \sum_{j=0}^{\infty} \min\left\{\mathbb{E}\left[\sum_{k \in \widetilde{\mathbb{Z}}_j} (\beta^\star_{j,k})^2\right], \frac{\mathbb{E}[|\widetilde{\mathbb{Z}}_j|]}{n}\right\}.$$

It follows from Lemma 23 of [Sar23] that

$$\mathbb{E}\left[|\widetilde{\mathbb{Z}}_j|\right] \leq c_1 M n^\theta 2^{j(1-\theta)}. \tag{4.6}$$

By using a suitable version of Hölder's inequality – see [CVNRF15] – we get

$$\sum_{k \in \widetilde{\mathbb{Z}}_j} (\beta^\star_{j,k})^2 \leq c_2 \|\beta^\star_{j,\cdot}\|_{p,\infty}^2 |\widetilde{\mathbb{Z}}_j|^{1-2/p}.$$

When $p \neq 2$, we deduce from $\sqrt{f} \in \mathcal{WB}^{\alpha}_{p,\infty}(R)$,

$$\sum_{k \in \widetilde{\mathbb{Z}}_j} (\beta^{\star}_{j,k})^2 \leq c_2 R^2 2^{-2j(\alpha+1/2-1/p)} |\widetilde{\mathbb{Z}}_j|^{1-2/p}.$$

This last inequality is also true when $p = 2$ and $\sqrt{f} \in \mathcal{B}^{\alpha}_{p,\infty}(R)$. We deduce from (4.6) and Jensen's inequality,

$$R_2 \leq c_3 \sum_{j=0}^{\infty} \min \left\{ M n^{-(1-\theta)} 2^{j(1-\theta)}, R^2 M^{1-2/p} n^{\theta(1-2/p)} 2^{-j(2\alpha+\theta(1-2/p))} \right\}.$$

It remains to compute the right-hand side of this inequality to prove the result. ☐

Proof of Lemma 3 when $p < 2$. We set for $j \geq 0$,

$$\widetilde{K}_j = \left\{ k \in \widetilde{\mathbb{Z}}_j,\ (\beta^{\star}_{j,k})^2 \geq 1/n \right\}$$

and observe as $\sqrt{f} \in \mathcal{WB}^{\alpha}_{p,\infty}(R)$,

$$|\widetilde{K}_j| \leq n^{p/2} R^p 2^{-jp(\alpha+1/2-1/p)}.$$

By using a classical inequality in weak spaces, see (35) of [Sar23],

$$\sum_{k \in \widetilde{\mathbb{Z}}_j \setminus \widetilde{K}_j} (\beta^{\star}_{j,k})^2 \leq c_1 n^{-(1-p/2)} R^p 2^{-jp(\alpha+1/2-1/p)}.$$

Therefore,

$$R_2 \leq c_2 \sum_{j=0}^{\infty} \min \left\{ \frac{\mathbb{E}[|\widetilde{\mathbb{Z}}_j|]}{n}, \mathbb{E}\left[\sum_{k \in \widetilde{\mathbb{Z}}_j \setminus \widetilde{K}_j} (\beta^{\star}_{j,k})^2 + \frac{|\widetilde{K}_j| \log_+(|\widetilde{\mathbb{Z}}_j|/|\widetilde{K}_j|)}{n} \right] \right\}.$$

By doing as in the preceding proof for the first term, and by using Jensen's inequality,

$$R_2 \leq c_3 \sum_{j=0}^{\infty} \min \Big\{ M n^{-(1-\theta)} 2^{j(1-\theta)}, n^{-(1-p/2)} R^p 2^{-jp(\alpha+1/2-1/p)} \times \log_+ \left(M R^{-p} n^{\theta-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right) \Big\}.$$

Elementary computations allows to bound the right-hand side of this inequality from above (see Lemma 30 of [Sar23]). ☐

4.3. Proof of Theorem 2.3

Since the theorem is stated for R and M large enough, we may without loss of generality prove the theorem with R replaced by $c_1 R$ and M by $c_2 M$ where c_1, c_2 only depend on α, p, θ and the wavelet basis.

Let $\ell \geq 1$ be the smallest integer such that $(-2^\ell, 2^\ell)$ contains the supports of $\bar{\phi}$ and $\bar{\psi}$. We consider two integers $j \geq -1$, $j_0 \geq 0$ such that $2^{j_0+j-\ell} \geq 12$. We define $\underline{k} \geq 1$ as the smallest integer satisfying $1+2\underline{k} \geq 2^{j_0+j-\ell-1}$, and $\bar{k} \geq 1$ as the largest integer satisfying $4\bar{k}+2\underline{k}+1 \leq 2^{j_0+j-\ell}$. We endow $\mathcal{D} = \{0,1\}^{\bar{k}}$ with the Hamming distance Δ defined for all $\delta, \delta' \in \mathcal{D}$ by

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

We consider $b > 0$ and set for $\delta \in \mathcal{D}$,

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

Let $g_0 \in \mathfrak{B}^\alpha_{p,\infty}(R_{g_0})$ be a compactly supported density on $[0,2]$ satisfying $\inf_{x \in [1/2,1]} g_0(x) \geq 1/4$ and $\|g_0\|_\infty \leq 1$. We then consider

$$\kappa = 4 \max \ \{2^{1/2} \|\bar{\phi}\|_\infty, \|\bar{\psi}\|_\infty\}$$

and set for $x \in \mathbb{R}$,

$$g(x) = \kappa b 2^{j/2} g_0(2^{-j_0} x).$$

Let ζ be a density, compactly supported on $(-3,0)$, bounded by $1/2$, and such that $\sqrt{\zeta} \in \mathfrak{B}^\alpha_{p,\infty}(R/\max\{2^{1/p}, 2\})$. Such a density does exist (recall that R is large enough in the proof). We put

$$q = \int (g(x) + h_\delta(x))^2 \, \mathrm{d}x,$$

and define for $x \in \mathbb{R}$,

$$f_\delta(x) = (1-q)\zeta(x) + (g(x) + h_\delta(x))^2.$$

We now state:

Lemma 4. *There are a_1, a_2, a_3, a_4 such that if*

$$\begin{aligned} b^2 2^{j_0+j} &\leq a_1 \\ b 2^{j/2} 2^{j_0(1/p-\alpha)} &\leq a_2 R \\ b 2^{j_0/p} 2^{j(\alpha+1/2)} \mathbb{1}_{j \geq 0} &\leq a_3 R \\ b^{2\theta} 2^{j\theta} 2^{j_0} &\leq a_4 M \end{aligned}$$

then, f_δ is a density belonging to $\mathcal{T}_\theta(M)$ such that $\sqrt{f_\delta} \in \mathcal{B}^\alpha_{p,\infty}(R)$. If $b 2^{j/2} 2^{j_0/p} \leq a_2 R$, $\sqrt{f_\delta} \in \mathfrak{B}^\alpha_{p,\infty}(R)$. Moreover $\|f_\delta\|_\infty \leq 1/2 + a_5 b^2 2^j$. For all $x \in \mathbb{R}$,

$$|x| f_\delta^\theta(x) \leq a_6 M,$$

and for all $\delta, \delta' \in \mathcal{D}$,

$$h^2(f_\delta, f_{\delta'}) = a_7 b^2 \Delta(\delta, \delta').$$

The terms $a_1, a_2, a_3, a_4, a_5, a_6, a_7$ above are positive and only depend on g_0, p, θ and the wavelet basis.

The proof of this lemma is given after the present proof. We define

$$\mathscr{S}'^{\alpha}_{p,\theta}(R, M) = \{f_\delta,\, \delta \in \mathcal{D}\}.$$

It follows from Assouad's lemma – see [Bir06b] – that if $b^2 = 1/(2a_6 n)$,

$$\mathcal{R}(\mathscr{S}'^{\alpha}_{p,\theta}(R, M)) \geq c_0 n^{-1} 2^{j_0 + j},$$

where c_0 only depends on the wavelet basis, g_0, p, θ. It then remains to choose j and j_0.

We first suppose either $\theta < p/2$ or $\theta = p/2$ and $R \geq M^{1/p}$. We then define $j \geq 0$ as the largest integer such that

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

We then consider c_1 small enough and the largest integer $j_0 \geq 0$ such that

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

We may check that the conditions of the lemma are satisfied.

We now suppose $\theta = p/2$ and $R < M^{1/p}$. We set $j = -1$, consider c_2 small enough and define $j_0 \geq 0$ as the largest integer such that $2^{j_0} \leq c_2 M n^\theta$. All the conditions of the lemma are met, hence the result.

Remark. We can see from this proof why the minimax risk does not tend to 0 when the tail dominance condition is not fulfilled and $p \geq 2$. Formally, this means choosing $\theta = 1$ and $M = 1$. The proof of Lemma 4 with $j = -1$ ensures that f_δ is a density such that $\sqrt{f_\delta} \in \mathfrak{B}^\alpha_{p,\infty}(R)$ if $b^2 2^{j_0+j} \leq a_1$ and $b 2^{j/2} 2^{j_0/p} \leq a_2 R$. We then choose b^2 as above and $2^{j_0} \leq c_2 n$ for some c_2. □

Proof of Lemma 4. First, observe that

$$q \leq 2 \left[\int g^2(x)\,\mathrm{d}x + \int h_\delta^2(x)\,\mathrm{d}x \right]$$

is not larger than 1 if we choose a_1 appropriately. This entails that f_δ is a density.

We have supp $h_\delta \subset [2^{\ell-j}(2\underline{k}+1), 2^{\ell-j}(4\bar{k}+2\underline{k}+1)] \subset [2^{j_0-1}, 2^{j_0}]$, supp $g \subset [0, 2^{j_0+1}]$ and $g(x) \geq \|h_\delta\|_\infty$ for all $x \in [2^{j_0-1}, 2^{j_0}]$. We deduce $g + h_\delta \geq 0$ and

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

We also have

$$h_\delta \in \mathcal{B}^\alpha_{p,\infty}\left(b\bar{k}^{1/p}2^{j(\alpha+1/2-1/p)}\mathbb{1}_{j\geq 0}\right)\cap \mathfrak{B}^\alpha_{p,\infty}\left(b\bar{k}^{1/p}2^{j(\alpha+1/2-1/p)}\right),$$
$$g \in \mathcal{B}^\alpha_{p,\infty}\left(c_1 b 2^{j/2}2^{j_0(1/p-\alpha)}\right)\cap \mathfrak{B}^\alpha_{p,\infty}\left(c_1 b 2^{j/2}2^{j_0/p}\right),$$

where c_1 only depends on g_0 and the wavelet basis. Therefore, we may consider a_2 and a_3 so that $\sqrt{f_\delta} \in \mathcal{B}^\alpha_{p,\infty}(R)$. If the supplementary conditions are fulfilled, $\sqrt{f_\delta} \in \mathfrak{B}^\alpha_{p,\infty}(R)$.

Note that ζ belongs to $\mathcal{T}_\theta(M/2)$ if M is large enough, which is assumed throughout the proof of Theorem 2.3. Besides,

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

see Lemma 2.1 of [CL20]. A similar result holds true for g^2 and hence $f_\delta \in \mathcal{T}_\theta(M)$ if a_4 is small enough.

As to the Hellinger distance, we have

$$h^2(f_\delta, f_{\delta'}) = \frac{1}{2}\int (h_\delta(x) - h_{\delta'}(x))^2 \, \mathrm{d}x,$$

and we conclude using that the supports of $\bar{\psi}_{j,2^{\ell+1}k}$ are disjoint.

Finally, for all $x \geq 0$,

$$f_\delta(x) \leq 2\left(\|g\|_\infty^2 + \|h_\delta\|_\infty^2\right)$$
$$\leq c_2 b^2 2^j,$$

where c_2 only depends on g_0 and the wavelet basis. Since f_δ is compactly supported on $[-3, 2^{j_0+1}]$, we get for all $x \geq 0$,

$$|x| f_\delta^\theta(x) \leq 2^{j_0+1}\left[c_2 b^2 2^j\right]^\theta$$
$$\leq 2a_4 c_2^\theta M.$$

Moreover, as $|f_\delta(x)| \leq 1/2$ when $x \in [-3, 0]$, and M is at least 1, this inequality is actually true for all $x \in \mathbb{R}$, up to a multiplicative factor. □

4.4. Sketch of the proof of Proposition 2.4

Let φ_δ be the map defined in the proof of Proposition 4 of [Sar23] with his α replaced by 2α and his p replaced by $p/2$. In other words,

$$\varphi_\delta(x) = \frac{1}{D}\sum_{j=j_0}^{j_1} 2^j \sum_{k\in K_j} \delta_{j,k}\mathbb{1}_{I_{j,k}}(x),$$

where the $I_{j,k} \subset [0, 1/2)$ are disjoint intervals of size 2^{-j}, where $\delta_{j,k} \in \{0,1\}$, where $|K_j| = n^{p/2}+1$, where j_0 is the smallest integer such that $2^{j_0} \geq 4(n^{p/2}+1)$, where $j_1 \geq j_0$ is to be specified, and where $D = (j_1 - j_0 + 1)(n^{p/2} + 1)$. We define for $x \in \mathbb{R}$,

$$f_\delta(x) = \varphi_\delta(x) + \varphi_{1-\delta}(x - 1/2).$$

Note that f_δ is a compactly supported density on $[0, 1]$ such that $\sqrt{f_\delta(x)} = \sqrt{\varphi_\delta(x)} + \sqrt{\varphi_{1-\delta}(x - 1/2)}$ and

$$\sqrt{\varphi_\delta(x)} = \frac{1}{\sqrt{D}} \sum_{j=j_0}^{j_1} 2^{j/2} \sum_{k \in K_j} \delta_{j,k} \mathbb{1}_{I_{j,k}}(x).$$

The lemma below is proved as Lemma 36 of [Sar23] (just replace the ℓ^1–ℓ^p inequality by Hölder's inequality in the first line of his proof when $p \in (1, 2)$).

Lemma 5. *For all $\varepsilon > 0$, j_1 large enough, and $\delta = (\delta_{j,k})_{j,k}$, $\sqrt{\varphi_\delta}$ belongs to $\mathfrak{B}^\alpha_{p,\infty}(\varepsilon)$.*

We deduce that $\sqrt{f_\delta}$ lies in $\mathfrak{B}^\alpha_{p,\infty}(R)$ if j_1 is large enough. Now, for all δ, δ' of the form $\delta = (\delta_{j,k})_{j,k}$, $\delta' = (\delta'_{j,k})_{j,k}$,

$$h^2(f_\delta, f_{\delta'}) = \frac{1}{D} \sum_{j=j_0}^{j_1} \sum_{k \in K_j} |\delta_{j,k} - \delta'_{j,k}|.$$

We conclude by using Assouad's Lemma (see [Bir06b], Lemma 2) and by taking j_1 large enough. □

4.5. Proof of Proposition 3.1

We consider $m \in \mathbb{R}$ such that f is non-increasing on $[m, +\infty)$ and non-decreasing on $(-\infty, m]$.

We first assume that $f \in \mathbb{L}^{\theta,\infty}(M)$. Then, for all $j, k \in \mathbb{Z}$ such that $k \geq 1/2 + 2^j m$,

$$F_{j,k}(f) \ \leq 2^{-j} f(2^{-j}(k - 1/2)).$$

If $k \geq 1/2 + 2^j m$ is such that $f(2^{-j}(k - 1/2)) \geq t2^j$, then $2^{-j}(k - 1/2)$ belongs to an interval of length at most $M2^{-j\theta}t^{-\theta}$. There are therefore at most $c[1 + M2^{j(1-\theta)}t^{-\theta}]$ integers $k \geq 1/2 + 2^j m$ such that $F_{j,k}(f) \ \geq t$. We can follow a similar line of reasoning to deal with k not larger than $-1/2 + 2^j m$. This leads to (3.2).

We now suppose that $f \in \mathcal{WT}_\theta(M)$. We consider $t > 0$ and $x > m$ such that $f(x) \geq t$. Let $j \geq 0$ large enough to ensure that $2^{1-j} \ < Mt^{-\theta}$ and $x \geq m + 2^{-j}$. Suppose that there exists $k \geq 3/2 + 2^j m$ such that $x \geq 2^{-j}(k - 1/2)$. We then have

$$f(x) \leq f(2^{-j}(k - 1/2)) \leq 2^j F_{j,k-1}(f).$$

Therefore, any $k \in [3/2 + 2^j m, 1/2 + 2^j x]$ is such that $F_{j,k-1}(f) \geq 2^{-j}t$. We deduce from $f \in \mathcal{WT}_\theta(M)$, that

$$\left|[3/2 + 2^j m, 1/2 + 2^j x]\right| \leq Mt^{-\theta}2^j,$$

and hence $2^j(x-m)-2 \leq Mt^{-\theta}2^j$. Therefore, x belongs to an interval of length at most $2^{1-j} + Mt^{-\theta} < 2Mt^{-\theta}$. We do the same reasoning when $x < m$ to complete the proof. □

4.6. Proof of Theorem 3.2

As in the proof of Theorem 2.2, we suppose that we have another independent sample $X_1', \dots, X_n'$ of X. We consider an integer $L_{\text{wav}} \geq 1$ such that supp $\psi \subset [-L_{\text{wav}}, L_{\text{wav}}]$ and the interval

$$I_{j,k} = \left[2^{-j}(k - L_{\text{wav}}), 2^{-j}(k + L_{\text{wav}})\right]. \tag{4.7}$$

It contains the supports of $\phi_{j,k}$ and $\psi_{j,k}$. We consider some $\rho > 0$ whose value is to be specified later on, set $u_n = n(2\rho \log n)^{-1}$, and introduce the smallest integer $\bar{r}_n \geq 0$ such that $2^{-\bar{r}_n} \leq u_n^{-1}$. Note that the integer r_n is well defined when $u_n \geq 1$, that is when n is large enough, and more precisely when $n/\log n \geq 2\rho$. As the Hellinger distance between two densities is no larger than 1, we may throughout this proof assume that n fulfils this property (otherwise, the risk bound still holds up to an increase of the risk of $2\rho(\log n)/n$). We define for all $j \in \mathbb{Z}$ and $r \geq 0$,

$$\widetilde{\mathbb{Z}}_{j,r} = \left\{ k \in \mathbb{Z},\ 2^{-r} < \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i' \in I_{j,k}} \leq 2^{-r+1} \right\}$$
$$\widetilde{\mathbb{Z}}_{j,r,-} = \left\{ k \in \mathbb{Z},\ \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i' \in I_{j,k}} > 2^{-r} \right\}.$$

Let $J_0 \leq 0$ and $J_1 \geq 0$. Let, for each $j \in \{J_0, \dots, J_1\}$, a subset K_j of $\widetilde{\mathbb{Z}}_{j,\bar{r}_n,-}$. We put $\mathbf{K} = (K_j)_{J_0 \leq j \leq J_1}$ and group together all the possible sets $\mathbf{K}$ into a collection $\widetilde{\mathcal{K}}_{J_0,J_1}$. We also set

$$\widetilde{\mathcal{K}} = \bigcup_{\substack{J_0 \leq 0 \\ J_1 \geq 0}} \widetilde{\mathcal{K}}_{J_0,J_1}.$$

Consider $\mathbf{K} \in \widetilde{\mathcal{K}}$ and J_0, J_1 such that $\mathbf{K} \in \widetilde{\mathcal{K}}_{J_0,J_1}$. We associate this set with the linear space

$$V_{\mathbf{K}} = \left\{ \sum_{k \in \widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}} \gamma_{J_0-1,k}\bar{\phi}_{J_0,k} + \sum_{j=J_0}^{J_1} \sum_{k \in K_j} \gamma_{j,k}\bar{\psi}_{j,k},\ \forall j,k,\ \gamma_{j,k} \in \mathbb{R} \right\}.$$

Its dimension is not larger than

$$\dim V_{\mathbf{K}} \leq |\widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}| + \sum_{j=J_0}^{J_1} |K_j|.$$

The set K_j also writes

$$K_j = \left(K_j \cap \widetilde{\mathbb{Z}}_{j,r_j,-}\right) \bigcup \bigcup_{r=r_j+1}^{\bar{r}_n} \left(K_j \cap \widetilde{\mathbb{Z}}_{j,r}\right),$$

and this, for any choice of $r_j \in \{0,\dots,\bar{r}_n\}$ (the set to the right of the union sign is empty if $r_j = \bar{r}_n$). We put

$$\begin{aligned}\Delta_{\mathbf{K}} = \sum_{j=J_0}^{J_1} \inf_{r_j \le \bar{r}_n} \Bigg[\sum_{r=r_j+1}^{\bar{r}_n} \Big\{ & |K_j \cap \widetilde{\mathbb{Z}}_{j,r}| + |K_j \cap \widetilde{\mathbb{Z}}_{j,r}| \log\left(e|\widetilde{\mathbb{Z}}_{j,r}|/|K_j \cap \widetilde{\mathbb{Z}}_{j,r}|\right) \\ & - \log\left(1-e^{-1}\right)\Big\} \\ + \Big\{ & |K_j \cap \widetilde{\mathbb{Z}}_{j,r_j,-}| + |K_j \cap \widetilde{\mathbb{Z}}_{j,r_j,-}| \log\left(e|\widetilde{\mathbb{Z}}_{j,r_j,-}|/|K_j \cap \widetilde{\mathbb{Z}}_{j,r_j,-}|\right) \\ & - \log\left(1-e^{-1}\right)\Big\}\Bigg] \\ + |J_0| + J_1 + (|J_0| + J_1 + 2)\log(1+\bar{r}_n) + \log(2(1-1/e)^{-1}).\end{aligned}$$

Above, we use the convention $0 \times \log(e|\cdot|/0) = 0$ when necessary. Elementary, albeit cumbersome, computations using Proposition 2.5 of [Mas07] yield

$$\sum_{\mathbf{K} \in \widetilde{\mathcal{K}}} e^{-\Delta_{\mathbf{K}}} \le 1.$$

We consider an event $\mathcal{A}$ of probability $1-1/n$ not depending on $X_1,\dots,X_n$ but rather on $X'_1,\dots,X'_n$ and to be specified later on. We apply Proposition 4.1 conditionally to $X'_1,\dots,X'_n$ when $\mathcal{A}$ occurs to define an estimator $\hat{f}$. When $\mathcal{A}$ does not happen, $\hat{f}$ is any density estimator (as $\mathcal{A}$ is not known by the statistician, we should take the same, but the whole point is that we do not need any of its properties). We take the expectation of the result and simplify it a bit. This leads to: for all $J_0 \le 0$ and $J_1 \ge 0$:

$$\mathbb{E}\left[h^2(f,\hat{f})\mathbb{1}_{\mathcal{A}}\right] \le c\mathbb{E}\left[\left\{\sum_{j=J_0}^{J_1} U_j + T + A + \frac{|\widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}| + (|J_0| + J_1 + 1)\log(1+\bar{r}_n)}{n}\right\}\mathbb{1}_{\mathcal{A}}\right], \quad (4.8)$$

where

$$U_j = \inf_{0 \le r_j \le \bar{r}_n} \left\{ U_{j,r_j,-} + \sum_{r=r_j+1}^{\bar{r}_n} U_{j,r} \right\}$$

$$U_{j,r_j,-} = \inf_{K_j \subset \widetilde{\mathbb{Z}}_{j,r_j,-}} \sum_{k \in \widetilde{\mathbb{Z}}_{j,r_j,-} \setminus K_j} \beta_{j,k}^2 + \frac{|K_j|}{n} \log_+\left(|\widetilde{\mathbb{Z}}_{j,r_j,-}|/|K_j|\right)$$

$$U_{j,r} = \inf_{K_j \subset \widetilde{\mathbb{Z}}_{j,r}} \sum_{k \in \widetilde{\mathbb{Z}}_{j,r} \setminus K_j} \beta_{j,k}^2 + \frac{|K_j|}{n} \log_+ \left(|\widetilde{\mathbb{Z}}_{j,r}| / |K_j| \right)$$

$$T = \sum_{j=J_0}^{J_1} \sum_{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{j,\bar{r}_n,-}} \beta_{j,k}^2 + \sum_{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}} \alpha_{J_0,k}^2$$

$$A = \sum_{j=J_1+1}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k}^2,$$

where $\alpha_{J_0,k} = \alpha_{J_0,k}(\sqrt{f})$, $\beta_{j,k} = \beta_{j,k}(\sqrt{f})$, and where $\log_+(x) = \log(e+x)$.

Since $\hat{f}$ is always a density, $h^2(f,\hat{f}) \leq 1$. The triangle inequality then ensures,

$$\begin{aligned} \mathbb{E}\left[h^2(f,\hat{f})\right] &\leq \mathbb{E}\left[h^2(f,\hat{f})\mathbb{1}_{\mathcal{A}}\right] + \mathbb{P}(A^c) \\ &\leq \mathbb{E}\left[h^2(f,\hat{f})\mathbb{1}_{\mathcal{A}}\right] + 1/n. \end{aligned} \tag{4.9}$$

Within one residual term, inequality (4.8) hence gives an upper-bound on the Hellinger risk $\mathbb{E}[h^2(f,\hat{f})]$ of $\hat{f}$. We now need the two following lemmas, to be proven after the present proof.

Lemma 6. *There exist $\rho > 0$ and an event $\mathcal{A}$ meeting the above conditions on which the three inequalities below hold true. First,*

$$T \leq c \left[(J_1 - J_0 + 1) \frac{\log n}{n} + M 2^{J_0(1-\theta)} \left(\frac{\log n}{n} \right)^{1-\theta} \right],$$

Second,

$$A \leq c R^{q_1} 2^{-J_1(1-1/q_2)}.$$

The terms $q_1 > 0, q_2 > 1$ only depend on α, p and the wavelet basis. Third, the cardinality of $\widetilde{\mathbb{Z}}_{j,r,-}$ can be bounded from above for all $j \in \mathbb{Z}$ and $r \in \{0,\dots,\bar{r}_n\}$ by

$$|\widetilde{\mathbb{Z}}_{j,r,-}| \leq c \left[1 + M 2^{r\theta} 2^{j(1-\theta)} \right]. \tag{4.10}$$

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

Lemma 7. *The following results hold true on the event $\mathcal{A}$ and are written for the value of ρ given by Lemma 6. For all $j \in \mathbb{Z}$ and $r \leq \bar{r}_n$,*

$$U_{j,r} \leq c \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(2^{r/2} \frac{1 + M 2^{r\theta} 2^{j(1-\theta)}}{\sqrt{n}} \right) \tag{4.11}$$

$$U_{j,r} \leq c \frac{1 + M 2^{r\theta + j(1-\theta)}}{n}. \tag{4.12}$$

We also have when $p \geq 2$ and $j \geq 0$,

$$U_{j,r} \leq c R^2 M^{1-2/p} 2^{-j(2\alpha - 2\theta/p + \theta)} 2^{r\theta(1-2/p)}. \tag{4.13}$$

When $p < 2$ and $j \geq 0$,

$$U_{j,r,-} \leq cR^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \times \log_+ \left(MR^{-p} 2^{r\theta} n^{-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right). \tag{4.14}$$

In these inequalities, c only depends on α, p, θ and the wavelet basis.

We are now in position to prove Theorem 3.2. We define J_1 as the smallest integer such that $R^{q_1} 2^{-J_1(1-1/q_2)} \leq 1/n$ where q_1, q_2 appear in Lemma 6. We define J_0 as the largest negative integer such that

$$M 2^{J_0(1-\theta)} \leq \left(\frac{\log n}{n} \right)^\theta .$$

We deduce from (4.8), (4.9) some c_1, c_2 such that

$$\mathbb{E}\left[h^2(f, \hat{f}) \right] \leq c_1 \mathbb{E}\left[\sum_{j=J_0}^{J_1} U_j \mathbb{1}_{\mathcal{A}} \right] + c_2 \frac{\log n + \log(1+R) + \log M}{n} \log(1 + \bar{r}_n).$$

The factor c_1 only depends on the wavelet basis, whereas c_2 only depends on α, p, θ and the wavelet basis. It then remains to bound

$$U = \sum_{j=J_0}^{J_1} U_j$$

on $\mathcal{A}$ from above. We treat cases $p \geq 2$ and $p < 2$ separately.

Proof of Theorem 3.2 when $p \geq 2$. Note that

$$U_j \leq \sum_{r=1}^{\bar{r}_n} U_{j,r}.$$

For all $j \geq 0$ and $r \leq \bar{r}_n$, Lemma 7 implies:

$$U_{j,r} \leq c_1 \min \left\{ \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(\frac{M 2^{r(\theta+1/2)} 2^{j(1-\theta)}}{\sqrt{n}} \right), R^2 M^{1-2/p} 2^{-j(2\alpha - 2\theta/p + \theta)} 2^{r\theta(1-2/p)}, \frac{M 2^{r\theta} 2^{j(1-\theta)}}{n} \right\}.$$

We first compute

$$U^{(1)} = \sum_{j=0}^{\infty} \sum_{r=j}^{\bar{r}_n} U_{j,r}$$

$$\leq c_2 \sum_{r=0}^{\bar{r}_n} \sum_{j=0}^{\infty} \min \left\{ \frac{2^{-r/2-j/2}}{\sqrt{n}} \log_+ \left(M \frac{2^{r(\theta+1/2)} 2^{3j/2}}{\sqrt{n}} \right), \right.$$
$$\left. R^2 M^{1-2/p} 2^{-2j\alpha} 2^{r\theta(1-2/p)}, \frac{M 2^{r\theta} 2^j}{n} \right\}. \qquad (4.15)$$

Let us denote by $U_r^{(1)}$ the sum in j. We may bound it from above by proceeding as follows. First, let us assume that

$$n M^{-2} 2^{-r(2\theta+1)} \leq 1.$$

Lemma 30 of [Sar23] entails:

$$U_r^{(1)} \leq c_3 \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(M \frac{2^{r(\theta+1/2)}}{\sqrt{n}} \right). \qquad (4.16)$$

In the contrary case, we may consider the largest integer $j_r \geq 0$ such that

$$2^{3j_r} \leq n M^{-2} 2^{-r(2\theta+1)}.$$

We then have,

$$U_r^{(1)} \leq \sum_{j=0}^{j_r} \frac{M 2^{r\theta} 2^j}{n} + \sum_{j=j_r+1}^{\infty} \frac{2^{-r/2-j/2}}{\sqrt{n}} \log_+ \left(M \frac{2^{r(\theta+1/2)} 2^{3j/2}}{\sqrt{n}} \right)$$
$$\leq c_4 M^{1/3} n^{-2/3} 2^{-r(1-\theta)/3}. \qquad (4.17)$$

To get the last inequality, Lemma 30 of [Sar23] is used once again. By grouping (4.16) and (4.17), we thus have for all $r \geq 0$,

$$U_r^{(1)} \leq c_5 \left\{ M^{1/3} n^{-2/3} 2^{-r(1-\theta)/3} \right.$$
$$\left. + 2^{-r/2} n^{-1/2} \log_+ \left(M 2^{r(\theta+1/2)} n^{-1/2} \right) \mathbb{1}_{nM^{-2} \leq 2^{r(2\theta+1)}} \right\}. \qquad (4.18)$$

This bound is obtained by using only the minimum between terms 1 and 3 in (4.15). We can also make a similar reasoning with terms 2 and 3 only. This leads to

$$U_r^{(1)} \leq c_6 \left[R^{2/(1+2\alpha)} M^{(1+2\alpha-2/p)/(1+2\alpha)} n^{-2\alpha/(2\alpha+1)} 2^{r\theta(1-2/((1+2\alpha)p))} \right.$$
$$\left. + R^2 M^{1-2/p} 2^{r\theta(1-2/p)} \mathbb{1}_{nR^2 M^{-2/p} \leq 2^{2r\theta/p}} \right]. \qquad (4.19)$$

If

$$M^{1/3-(1+2\alpha-2/p)/(1+2\alpha)} R^{-2/(1+2\alpha)} n^{2\alpha/(2\alpha+1)-2/3} \leq 1,$$

we sum (4.18) for all r. Thereby,

$$U^{(1)} \leq c_7 \left[M^{1/3} n^{-2/3} + M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} \right]$$

$$\leq c_7 \left[M^{(1+2\alpha-2/p)/(1+2\alpha)} R^{2/(1+2\alpha)} n^{-2\alpha/(2\alpha+1)} \right. \\ \left. + M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} \right].$$

Otherwise, we define $\mathfrak{r}_0 \geq 0$ as the largest integer such that

$$2^{\mathfrak{r}_0(1+2\alpha+2\theta+4\alpha\theta-6\theta/p)/(3(2\alpha+1))} \\ \leq M^{1/3-(1+2\alpha-2/p)/(1+2\alpha)} R^{-2/(1+2\alpha)} n^{2\alpha/(2\alpha+1)-2/3}. \tag{4.20}$$

We sum (4.18) when $r \geq \mathfrak{r}_0 + 1$ and (4.19) when $r \leq \mathfrak{r}_0$ to get

$$U^{(1)} \leq c_8 \left[R^{\beta_1} M^{\beta_2} n^{-\gamma} + M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} \right. \\ \left. + R^2 M^{1-2/p} 2^{\mathfrak{r}_0 \theta(1-2/p)} (1 + \mathfrak{r}_0 \mathbb{1}_{p=2}) \mathbb{1}_{nR^2 M^{-2/p} \leq 2^{2\mathfrak{r}_0\theta/p}} \right].$$

Observe that

$$nR^2 M^{-2/p} \leq 2^{2\mathfrak{r}_0 \theta/p}$$

is possible only if

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

that is if n is small enough. We now study

$$\begin{aligned} U^{(2)} &= \sum_{j=0}^{\infty} \sum_{r<j} U_{j,r} \\ &\leq c_9 \sum_{j=0}^{\infty} \min \left\{ R^2 M^{1-2/p} 2^{-2j\alpha}, M \frac{2^j}{n} \right\} \\ &\leq c_{10} \left[R^{2/(1+2\alpha)} M^{(1+2\alpha-2/p)/(1+2\alpha)} n^{-2\alpha/(2\alpha+1)} + Mn^{-1} \right]. \end{aligned}$$

We finally deal with

$$\begin{aligned} U^{(3)} &= \sum_{j=J_0}^{0} \sum_{r=0}^{\bar{r}_n} U_{j,r} \\ &\leq c_{11} \sum_{j=0}^{|J_0|} \sum_{r=0}^{\bar{r}_n} \min \left\{ \frac{2^{-r/2}}{\sqrt{n}} \log \left(2^{r/2} \frac{1 + M 2^{r\theta} 2^{-j(1-\theta)}}{\sqrt{n}} \right), \frac{1 + M 2^{r\theta} 2^{-j(1-\theta)}}{n} \right\} \\ &\leq c_{12} \frac{|J_0|^2}{n} + c_{12} \sum_{j=0}^{|J_0|} \sum_{r=0}^{\bar{r}_n} U_{j,r}^{(3)} \end{aligned}$$

where

$$U_{j,r}^{(3)} = \min \left\{ \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(\frac{M 2^{r(\theta+1/2)} 2^{-j(1-\theta)}}{\sqrt{n}} \right), \frac{M 2^{r\theta} 2^{-j(1-\theta)}}{n} \right\}.$$

Note that

$$|J_0| \leq c_{13} \log(M(n/\log n)^\theta).$$

When $2^{2j(1-\theta)} \leq M^2/n$ (and $j \geq 0$), which is possible only if $n \leq M^2$, we do:

$$\begin{aligned}
&\sum_{2^{2j(1-\theta)} \leq M^2/n} \sum_{r=0}^{\bar{r}_n} U_{j,r}^{(3)} \\
&\leq \sum_{2^{2j(1-\theta)} \leq M^2/n} \sum_{r=0}^{\infty} \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(\frac{M 2^{r(\theta+1/2)} 2^{-j(1-\theta)}}{\sqrt{n}} \right) \\
&\leq c_{14} \sum_{2^{2j(1-\theta)} \leq M^2/n} \frac{1}{\sqrt{n}} \log_+ \left(\frac{M 2^{-j(1-\theta)}}{\sqrt{n}} \right) \\
&\leq c_{15} \frac{\log_+^2 (M^2/n)}{\sqrt{n}}.
\end{aligned}$$

When $2^{2j(1-\theta)} \geq M^2/n$ (and $j \geq 0$), we may consider the largest integer $r_j \geq 0$ such that

$$2^{r_j(2\theta+1)} \leq n M^{-2} 2^{2j(1-\theta)}.$$

We then have,

$$\begin{aligned}
&\sum_{2^{2j(1-\theta)} \leq M^2/n} \sum_{r=0}^{\bar{r}_n} U_{j,r}^{(3)} \\
&\leq c_{16} \sum_{j=0}^{|J_0|} \left\{ \sum_{r \geq r_j} \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(\frac{M 2^{r(\theta+1/2)} 2^{-j(1-\theta)}}{\sqrt{n}} \right) + \sum_{r < r_j} \frac{M 2^{r\theta} 2^{-j(1-\theta)}}{n} \right\} \\
&\leq c_{17} \sum_{j=0}^{|J_0|} M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} 2^{-j(1-\theta)/(2\theta+1)} \\
&\leq c_{18} M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)}.
\end{aligned}$$

In conclusion,

$$\begin{aligned}
U^{(3)} \leq c_{19} \Big[& M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} + \log^2(M(n/\log n)^\theta) n^{-1} \\
& + \log_+^2(M^2/n) n^{-1/2} \mathbb{1}_{n \leq M^2} \Big].
\end{aligned}$$

It then remains to sum up the different results to get a bound on U and hence (3.4) where

$$\begin{aligned}
v_n = {} & R^{2/(1+2\alpha)} M^{(1+2\alpha-2/p)/(1+2\alpha)} n^{-2\alpha/(2\alpha+1)} + M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} \\
& + \log^2(M(n/\log n)^\theta) n^{-1} + (\log n)(\log(1+\log n)) n^{-1} \\
& + (\log M) \log(1+\log n) n^{-1} + \log(1+R) \log(1+\log n) n^{-1} + M n^{-1}
\end{aligned}$$

$$+ R^2 M^{1-2/p} 2^{\mathfrak{r}_0\theta(1-2/p)}$$
$$\times (1 + \mathfrak{r}_0 \mathbb{1}_{p=2}) \mathbb{1}_{M^{-2(1+2\alpha)/p} R^{2(2\theta+1)(2\alpha+1)} n^{(2\alpha+1)(1-2\theta/p+2\theta)} \leq 1}$$
$$+ \log_+^2(M^2/n) n^{-1/2} \mathbb{1}_{n \leq M^2} \tag{4.21}$$

and where $\mathfrak{r}_0$ is given by (4.20). □

Proof of Theorem 3.2 when $p < 2$. We apply Lemma 7: U_j satisfies for all $j \in \{0, \ldots, J_1\}$, and $r_j \in \{0, \ldots, \bar{r}_n\}$,

$$U_j \leq c_1 \Big\{ R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \log_+ \left(M R^{-p} 2^{r_j\theta} n^{-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right) + \sum_{r=r_j+1}^{\bar{r}_n} 2^{-r/2} n^{-1/2} \log_+ \left(M 2^{r(\theta+1/2)} 2^{j(1-\theta)} n^{-1/2} \right) \Big\}.$$

Note that this inequality also holds true for $r_j > \bar{r}_n$. If we choose $r_j = \bar{r}_n$, we get (using $2^{\bar{r}_n} \leq c_2 n$),

$$U_j \leq c_3 R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \log_+ \left(M R^{-p} n^{\theta-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right). \tag{4.22}$$

We may refine this result when $n^{1-p} R^{-2p} 2^{2jp(\alpha+1/2-1/p)} \geq 1$. In this case, we may define the largest number $r_j \geq 0$ satisfying

$$2^{r_j} \leq n^{1-p} R^{-2p} 2^{2jp(\alpha+1/2-1/p)}.$$

We then derive from Lemma 30 of [Sar23],

$$\begin{aligned} U_j &\leq c_4 \Big\{ R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \log_+ \left(M R^{-p} 2^{r_j\theta} n^{-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right) \\ &\qquad + 2^{-r_j/2} n^{-1/2} \log_+ \left(M 2^{r_j(\theta+1/2)} 2^{j(1-\theta)} n^{-1/2} \right) \Big\} \\ &\leq c_5 R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \\ &\qquad \times \log_+ \left(M R^{-p(1+2\theta)} n^{\theta-p/2-p\theta} 2^{j(p/2-3\theta+\alpha p+p\theta+2\alpha p\theta)} \right). \end{aligned} \tag{4.23}$$

We now have when $j \geq 0$,

$$U_j \leq U_{j,+} + U_{j,-}, \tag{4.24}$$

with

$$U_{j,+} = \sum_{r=j}^{\infty} U_{j,r} \quad \text{and} \quad U_{j,-} = \sum_{r=0}^{j-1} U_{j,r}.$$

We deduce from Lemma 7,

$$U_{j,+} \leq c_6 \sum_{r \geq 0} \min \left\{ \frac{2^{-r/2-j/2}}{\sqrt{n}} \log_+ \left(M \frac{2^{r(\theta+1/2)} 2^{3j/2}}{\sqrt{n}} \right), M \frac{2^j 2^{r\theta}}{n} \right\}.$$

Elementary computations lead to

$$U_{j,+} \le \begin{cases} c_7 2^{-j/2} n^{-1/2} \log_+(M 2^{3j/2} n^{-1/2}) & \text{if } 2^{3j} \ \ge nM^{-2} \\ c_7 M^{1/(2\theta+1)} 2^{j(1-\theta)/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} & \text{otherwise.} \end{cases}$$

We also deduce from (4.12), $U_{j,-} \le c_8 M 2^j n^{-1}$. Therefore, (4.24) gives

$$U_j \le \begin{cases} c_9 M 2^j n^{-1} & \text{if } 2^{3j} \ \ge nM^{-2} \\ c_9 M^{1/(2\theta+1)} 2^{j(1-\theta)/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} & \text{otherwise.} \end{cases}$$

Suppose that $R^2 M^{-2/p} n \ge 1$ and consider the smallest integer $j_0 \ge 0$ such that

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

Then,

$$\begin{aligned} &\sum_{2^{3j} \ge nM^{-2}} U_j \\ &\le c_{10} \left\{ \sum_{j=0}^{j_0} M 2^j n^{-1} + \sum_{j=j_0+1}^{\infty} R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)} \right. \\ &\qquad \left. \times \log_+ \left(M R^{-p} n^{\theta-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right) \right\} \\ &\le c_{11} M^{(1+2\alpha-2/p)/(2\alpha+1)} R^{2/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} \log_+ \left(n^\alpha M^{1/p} R^{-1} \right). \end{aligned}$$

When $R^2 M^{-2/p} n < 1$, we only use (4.22):

$$\sum_{2^{3j} \ge nM^{-2}} U_j \le c_{12} R^p n^{-(1-p/2)} \log_+ \left(M R^{-p} n^{\theta-p/2} \right).$$

We now deal with smaller values of j. We first suppose

$$R^p n^{-(1-p/2)} \ \ge M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)}, \tag{4.25}$$

and

$$n^{2\alpha-2\theta/p-1+2\theta} R^{-4(1-\theta)} M^{-2(1+2\alpha-2/p)} \ \ge 1. \tag{4.26}$$

The first inequality allows us to consider the smallest integer $j_1 \ge 0$ such that

$$2^{j_1[(1-\theta)/(2\theta+1)+p(\alpha+1/2-1/p)]} \ge R^p M^{-1/(2\theta+1)} n^{(\theta+1)/(2\theta+1)-1+p/2}.$$

The second ensures

$$n^{1-p} R^{-2p} 2^{2jp(\alpha+1/2-1/p)} \ge 1$$

for all $j \ge j_1$. We may hence use (4.23):

$$\sum_{\substack{j \ge 0 \\ 2^{3j} < nM^{-2}}} U_j \le c_{13} \left\{ \sum_{j=0}^{j_1} M^{1/(2\theta+1)} 2^{j(1-\theta)/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} \right.$$

$$+ \sum_{j=j_1+1}^{\infty} R^p 2^{-jp(\alpha+1/2-1/p)} n^{-(1-p/2)}$$
$$\times \log_+ \big(MR^{-p(1+2\theta)} n^{\theta-p/2-p\theta} 2^{j(p/2-3\theta+\alpha p+p\theta+2\alpha p\theta)} \big) \Big\}$$
$$\leq c_{14} R^{\beta_1} M^{\beta_2} n^{-\gamma}.$$

When either (4.25) or (4.26) is not true, we merely apply (4.22):

$$\sum_{\substack{j\geq 0 \\ 2^{3j}<nM^{-2}}} U_j \leq c_{15} R^p n^{-(1-p/2)} \log_+ \left(MR^{-p} n^{\theta-p/2} \right).$$

The above provides an upper-bound on

$$\sum_{j=0}^{\infty} U_j.$$

We still have to work with negative values of j. For this, we use (4.11) and (4.12):

$$\sum_{j=J_0}^{0} U_j$$
$$\leq c_{16} \sum_{j=0}^{|J_0|} \sum_{r=0}^{r_n} \min \left\{ \frac{2^{-r/2}}{\sqrt{n}} \log_+ \left(2^{r/2} \frac{1+M2^{r\theta}2^{-j(1-\theta)}}{\sqrt{n}} \right), \frac{1+M2^{r\theta}2^{-j(1-\theta)}}{n} \right\}.$$

We have already found an upper-bound of this term when $p \geq 2$. The calculations are the same here. We now put all these results together to get (3.4) with

$$\begin{aligned} \upsilon_n = {} & R^{2/(1+2\alpha)} M^{(1+2\alpha-2/p)/(1+2\alpha)} n^{-2\alpha/(2\alpha+1)} \log_+ \left(n^{\alpha} M^{1/p} R^{-1} \right) \\ & + M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)} + \log^2 (M(n/\log n)^{\theta}) n^{-1} \\ & + R^p n^{-(1-p/2)} \log_+ \left(MR^{-p} n^{\theta-p/2} \right) [\mathbb{1}_{M^{-2/p}R^2 n<1} \\ & \qquad + \mathbb{1}_{R^p n^{-(1-p/2)} < M^{1/(2\theta+1)} n^{-(\theta+1)/(2\theta+1)}} \\ & \qquad + \mathbb{1}_{n^{2\alpha-2\theta/p-1+2\theta} R^{-4(1-\theta)} M^{-2(1+2\alpha-2/p)} <1}] \\ & + \log_+^2 (M^2/n) n^{-1/2} \mathbb{1}_{n \leq M^2} + \log(1+R) \log(1+\log n) n^{-1} \\ & + (\log M) \log(1+\log n) n^{-1} + (\log n) \log(1+\log n) n^{-1}. \end{aligned} \tag{4.27}$$

□

4.6.1. Proofs of Lemmas 6 and 7: preliminary results

Lemma 8. *There is an event $\mathcal{A}$ of probability $1 - 1/n$ on which: for all $r > 0$ such that $2^r \leq n$, and for all interval $I \subset \mathbb{R}$ such that $2^{-r} < \mathbb{P}(X \in I) \leq 2^{-r+1}$,*

$$\left| \frac{1}{n} \sum_{i=1}^{n} \{ \mathbb{1}_I(X_i') - \mathbb{E}[\mathbb{1}_I(X_i')] \} \right| \leq c \left[\sqrt{\frac{2^{-r}\log n}{n}} + \frac{\log n}{n} \right],$$

where c is numerical value. In particular, there is some $\rho > 0$ such that any interval I satisfying $n^{-1}\sum_{i=1}^n \mathbb{1}_I(X_i') \geq \rho(\log n)/n$ is such that:

$$\frac{1}{2}\mathbb{P}\left(X \in I\right) \leq \frac{1}{n}\sum_{i=1}^n \mathbb{1}_I(X_i') \leq 2\mathbb{P}\left(X \in I\right).$$

Sketch of the proof of Lemma 8. A short way to prove the lemma is to remark that the collection of functions of the form $\mathbb{1}_I$ where I is an interval is VC subgraph with finite dimension. We then apply Proposition 6 of [Sar23]. □

Lemma 9. *Let $m \in \mathbb{R}$ such that f is non-increasing on $[m, +\infty)$ and non-decreasing on $(-\infty, m]$. Then, for all $j \in \mathbb{Z}$, $k_0 \geq L_{wav} + 1 + m2^j$,*

$$\sum_{k \geq k_0} |\beta_{j,k}| \leq c f_{j,k_0-1}^{1/2}, \tag{4.28}$$

where

$$f_{j,k} = \int_{I_{j,k}} f(x)\,\mathrm{d}x = \int_{2^{-j}(k-L_{wav})}^{2^{-j}(k+L_{wav})} f(x)\,\mathrm{d}x,$$

and where we recall that $I_{j,k}$ is given by (4.7). Moreover, for all $j \in \mathbb{Z}$, $k_0 \leq -L_{wav} - 1 + m2^j$,

$$\sum_{k \leq k_0} |\beta_{j,k}| \leq c f_{j,k_0+1}^{1/2}. \tag{4.29}$$

Above, c only depends on the wavelet basis.

Proof of Lemma 9. We only show (4.28), the proof of (4.29) is similar. Since $\int \psi_{j,k}(x)\,\mathrm{d}x = 0$ for all $j, k \in \mathbb{Z}$, we deduce for all $k \geq L_{\mathrm{wav}} + 1 + m2^j$,

$$\begin{aligned}
|\beta_{j,k}| &= \left| \int \left(\sqrt{f(x)} - \sqrt{f(2^{-j}(k + L_{\mathrm{wav}}))} \right) \psi_{j,k}(x)\,\mathrm{d}x \right| \\
&\leq 2^{j/2}\|\psi\|_\infty \int_{2^{-j}(k-L_{\mathrm{wav}})}^{2^{-j}(k+L_{\mathrm{wav}})} \left(\sqrt{f(x)} - \sqrt{f(2^{-j}(k + L_{\mathrm{wav}}))} \right) \mathrm{d}x \\
&\leq 2^{1-j/2} L_{\mathrm{wav}} \|\psi\|_\infty \left(\sqrt{f(2^{-j}(k - L_{\mathrm{wav}}))} - \sqrt{f(2^{-j}(k + L_{\mathrm{wav}}))} \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{k \geq k_0} |\beta_{j,k}| &\leq c_1 2^{-j/2} \sqrt{f(2^{-j}(k_0 - L_{\mathrm{wav}}))} \\
&\leq c_2 2^{j/2} \int_{2^{-j}(k_0 - L_{\mathrm{wav}} - 1)}^{2^{-j}(k_0 - L_{\mathrm{wav}})} \sqrt{f(x)}\,\mathrm{d}x \\
&\leq c_3 \sqrt{\int_{2^{-j}(k_0 - L_{\mathrm{wav}} - 1)}^{2^{-j}(k_0 - L_{\mathrm{wav}})} f(x)\,\mathrm{d}x} \\
&\leq c_3 f_{j,k_0-1}^{1/2}.
\end{aligned}$$

□

Lemma 10. *We consider $m \in \mathbb{R}$ such that f is non-increasing on $[m, +\infty)$ and non-decreasing on $(-\infty, m]$. We set for all $r \leq \bar{r}_n$,*

$$\widetilde{\mathbb{Z}}_{j,r,+,right} = \left\{ k \in \mathbb{Z} \cap \left[L_{wav} + m2^j, +\infty \right), \ \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X'_i \in I_{j,k}} \leq 2^{-r+1} \right\}$$
$$\widetilde{\mathbb{Z}}_{j,r,+,left} = \left\{ k \in \mathbb{Z} \cap \left(-\infty, -L_{wav} + m2^j \right], \ \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X'_i \in I_{j,k}} \leq 2^{-r+1} \right\}$$

and

$$\tilde{k}_{j,r,right} = \min \widetilde{\mathbb{Z}}_{j,r,+,right} + 1,$$
$$\tilde{k}_{j,r,left} = \max \widetilde{\mathbb{Z}}_{j,r,+,left} - 1.$$

Let $\rho > 0$ and $\mathcal{A}$ be given by Lemma 8. On this event, the above sets are non-empty. Moreover, $f_{j,k} \leq c2^{-r}$ for all $j \in \mathbb{Z}$, $r \leq \bar{r}_n$, $k \geq \tilde{k}_{j,r,right} - 1$ or $k \leq \tilde{k}_{j,r,left} + 1$. Moreover,

$$\sum_{k \geq \tilde{k}_{j,r,right}} |\beta_{j,k}| \leq c2^{-r/2} \quad \text{and} \quad \sum_{k \geq \tilde{k}_{j,r,right}} \beta_{j,k}^2 \leq c2^{-r} \tag{4.30}$$
$$\sum_{k \leq \tilde{k}_{j,r,left}} |\beta_{j,k}| \leq c2^{-r/2} \quad \text{and} \quad \sum_{k \leq \tilde{k}_{j,r,left}} \beta_{j,k}^2 \leq c2^{-r}. \tag{4.31}$$

We also have when $f \in \mathcal{WT}_\theta(M)$,

$$\left| \widetilde{\mathbb{Z}}_{j,r,-} \right| \leq c \left[1 + M2^{r\theta} 2^{j(1-\theta)} \right]. \tag{4.32}$$

In these results, c only depends on θ and the wavelet basis.

Proof of Lemma 10. We deduce from Lemma 8 that

$$|\widetilde{\mathbb{Z}}_{j,r,-}| \leq \left| \left\{ k \in \mathbb{Z}, \ f_{j,k} > 2^{-r-1} \right\} \right|$$

holds true on $\mathcal{A}$. Since

$$f_{j,k} \ \leq \sum_{k'=-L_{\text{wav}}}^{L_{\text{wav}}} F_{j,k+k'}(f),$$

we may use Proposition 3.1 to get (4.32). Since $\widetilde{\mathbb{Z}}_{j,r,-}$ is finite, $\widetilde{\mathbb{Z}}_{j,r,+,\text{left}}$ and $\widetilde{\mathbb{Z}}_{j,r,+,\text{right}}$ are non-empty.

We now prove (4.30). The proof of (4.31) is similar. Lemma 8 ensures that

$$f_{j,\min \widetilde{\mathbb{Z}}_{j,r,+,\text{right}}} \ \leq c_1 \left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X'_i \in I_{j,\min \widetilde{\mathbb{Z}}_{j,r,+,\text{right}}}} + \frac{\log n}{n} \right] \leq c_2 2^{-r}.$$

Moreover, as f is non-increasing on $(m, +\infty)$, we have for all $k \geq \tilde{k}_{j,r,\mathrm{right}} - 1$, $f_{j,k} \leq c_2 2^{-r}$. Lemma 9 yields

$$\begin{aligned}\sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} |\beta_{j,k}| &\leq c_3 f^{1/2}_{j,\tilde{k}_{j,r,\mathrm{right}}-1} \\ &\leq c_4 2^{-r/2}.\end{aligned}$$

For the right part of (4.30), we merely use Cauchy-Schwarz inequality:

$$\begin{aligned}\sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} \beta^2_{j,k} &\leq \sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} (f_{j,k})^{1/2} |\beta_{j,k}| \\ &\leq c_5 2^{-r/2} \sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} |\beta_{j,k}|. \qquad \square\end{aligned}$$

4.6.2. Proof of Lemma 6

We deduce from Lemma 10,

$$\sum_{\substack{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{j,\bar{r}_n,-} \\ k \geq L_{\mathrm{wav}} + m2^j}} \beta^2_{j,k} \leq \sum_{k \geq \tilde{k}_{j,\bar{r}_n,\mathrm{right}} - 1} \beta^2_{j,k} \leq \beta^2_{j,\tilde{k}_{j,\bar{r}_n,\mathrm{right}}-1} + c_1 \frac{\log n}{n}.$$

Likewise,

$$\sum_{\substack{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{j,\bar{r}_n,-} \\ k \leq -L_{\mathrm{wav}} + m2^j}} \beta^2_{j,k} \leq \beta^2_{j,\tilde{k}_{j,\bar{r}_n,\mathrm{left}}+1} + c_2 \frac{\log n}{n}.$$

Cauchy-Schwarz inequality and Lemma 8 lead to $\beta^2_{j,k} \leq f_{j,k} \leq c_3 (\log n)/n$ when $k \notin \widetilde{\mathbb{Z}}_{j,\bar{r}_n,-}$. By putting it all together,

$$\sum_{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{j,\bar{r}_n,-}} \beta^2_{j,k} \leq c_4 \frac{\log n}{n}. \tag{4.33}$$

We also have $\alpha^2_{J_0,k} \leq f_{J_0,k} \leq c_5 (\log n)/n$ when $k \notin \widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}$. Hence,

$$\sum_{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}} \alpha^2_{J_0,k} \leq \sum_{2^r \geq c_6 n/(\log n)} 2^{-r} \left| \left\{ k \in \mathbb{Z},\ f_{J_0,k} \geq 2^{-r} \right\} \right|.$$

By doing as at the beginning of the proof of Lemma 10,

$$\left| \left\{ k \in \mathbb{Z},\ f_{J_0,k} \geq 2^{-r} \right\} \right| \leq c_7 \left[1 + M 2^{r\theta} 2^{J_0(1-\theta)} \right].$$

We deduce,

$$\sum_{k \in \mathbb{Z} \setminus \widetilde{\mathbb{Z}}_{J_0,\bar{r}_n,-}} \alpha^2_{J_0,k} \leq c_8 \left[\frac{\log n}{n} + M \left(\frac{\log n}{n} \right)^{1-\theta} 2^{J_0(1-\theta)} \right].$$

The upper-bound on T follows from this inequality and (4.33).

It then remains to bound A from above. The proof of Lemma 2 ensures that there are some $q_1 > 0$, $q_2 > 1$, such that $\|f\|_{q_2} \leq c_9 R^{q_1}$ for some c_9 only depending on the wavelet basis, α, p. Hence,

$$f_{j,k} \leq c_{10} R^{q_1} 2^{-j(1-1/q_2)}.$$

Let

$$\begin{aligned}
\overline{\mathbb{Z}}_{j,r} &= \left\{k \in \mathbb{Z},\, 2^{-r} < f_{j,k} \leq 2^{-r+1}\right\}, \\
\overline{\mathbb{Z}}_{j,r,+} &= \left\{k \in \mathbb{Z},\, f_{j,k} \leq 2^{-r+1}\right\}, \\
\overline{\mathbb{Z}}_{j,r,+,\text{right}} &= \left\{k \in \mathbb{Z} \cap \left[L_{\text{wav}} + m2^j, +\infty\right),\ f_{j,k} \leq 2^{-r+1}\right\}, \\
\overline{\mathbb{Z}}_{j,r,+,\text{left}} &= \left\{k \in \mathbb{Z} \cap \left(-\infty, -L_{\text{wav}} + m2^j\right],\ f_{j,k} \leq 2^{-r+1}\right\},
\end{aligned}$$

and

$$\begin{aligned}
\bar{k}_{j,r,\text{right}} &= \min \overline{\mathbb{Z}}_{j,r,+,\text{right}} + 1 \\
\bar{k}_{j,r,\text{left}} &= \max \overline{\mathbb{Z}}_{j,r,+,\text{left}} - 1.
\end{aligned}$$

We have,

$$\begin{aligned}
A &\leq \sum_{j=J_1+1}^{\infty} \sum_{2^r \geq c_{11} R^{-q_1} 2^{j(1-1/q_2)}} \sum_{k \in \overline{\mathbb{Z}}_{j,r}} \beta_{j,k}^2 \\
&\leq \sum_{j=J_1+1}^{\infty} \sum_{2^r \geq c_{11} R^{-q_1} 2^{j(1-1/q_2)}} \sum_{k \in [-L_{\text{wav}}+m2^j, L_{\text{wav}}+m2^j] \cap \overline{\mathbb{Z}}_{j,r}} f_{j,k} \\
&\quad + \sum_{j=J_1+1}^{\infty} \sum_{2^r \geq c_{11} R^{-q_1} 2^{j(1-1/q_2)}} \left[f_{j,\bar{k}_{j,r,\text{right}}-1} + f_{j,\bar{k}_{j,r,\text{left}}+1} + \sum_{\substack{k \geq \bar{k}_{j,r,\text{right}} \\ \text{or } k \leq \bar{k}_{j,r,\text{left}}}} \beta_{j,k}^2 \right] \\
&\leq c_{12} \sum_{j=J_1+1}^{\infty} \sum_{2^r \geq c_{11} R^{-q_1} 2^{j(1-1/q_2)}} 2^{-r} \\
&\leq c_{13} R^{q_1} 2^{-J_1(1-1/q_2)}.
\end{aligned}$$

□

4.6.3. Proof of Lemma 7

The proof of (4.12) simply ensues from (4.32). As to (4.13), we do the same reasoning as in the proof of Lemma 3. In short, for all $p \geq 2$,

$$\sum_{k \in \widetilde{\mathbb{Z}}_{j,r}} \beta_{j,k}^2 \leq c_2 R^2 2^{-2j(\alpha+1/2-1/p)} |\widetilde{\mathbb{Z}}_{j,r}|^{1-2/p}.$$

We then use (4.32). The same goes for (4.14): we set

$$K_{j,r} = \left\{ k \in \widetilde{\mathbb{Z}}_{j,r,-},\ |\beta_{j,k}| \ \geq 1/\sqrt{n} \right\}.$$

As $\sqrt{f}$ is smooth, we have:

$$|K_{j,r}| \leq c_3 n^{p/2} R^p 2^{-jp(\alpha+1/2-1/p)}.$$

A suitable inequality in weak spaces – see (35) of [Sar23] – leads to

$$\sum_{k \in \widetilde{\mathbb{Z}}_{j,r,-} \setminus K_{j,r}} \beta_{j,k}^2 \leq c_4 n^{-(1-p/2)} R^p 2^{-jp(\alpha+1/2-1/p)}.$$

We deduce,

$$\begin{aligned} U_{j,r,-} \ &\leq c_5 n^{p/2-1} R^p 2^{-jp(\alpha+1/2-1/p)} \log_+ \left(|\widetilde{\mathbb{Z}}_{j,r,-}| n^{-p/2} R^{-p} 2^{jp(\alpha+1/2-1/p)} \right) \\ &\leq c_6 n^{p/2-1} R^p 2^{-jp(\alpha+1/2-1/p)} \log_+ \left(M R^{-p} 2^{r\theta} n^{-p/2} 2^{jp(\alpha+1/2-\theta/p)} \right), \end{aligned}$$

thanks to (4.32).

We now show (4.11). We introduce the integers $\tilde{k}_{j,r,\mathrm{right}}$ and $\tilde{k}_{j,r,\mathrm{left}}$ appearing in Lemma 10. We then set

$$K_{j,r}^{(1)} = \left\{ k \in \widetilde{\mathbb{Z}}_{j,r},\ k < \tilde{k}_{j,r,\mathrm{right}} \text{ or } |\beta_{j,k}| \ \geq 1/\sqrt{n} \right\}$$
$$K_{j,r}^{(2)} = \left\{ k \in \widetilde{\mathbb{Z}}_{j,r},\ k > \tilde{k}_{j,r,\mathrm{left}} \text{ or } |\beta_{j,k}| \ \geq 1/\sqrt{n} \right\}$$

and

$$K_{j,r} \ = K_{j,r}^{(1)} \ \cap K_{j,r}^{(2)}.$$

Now,

$$\begin{aligned} \sum_{k \in \widetilde{\mathbb{Z}}_{j,r} \setminus K_{j,r}} \beta_{j,k}^2 &\leq \frac{1}{\sqrt{n}} \left[\sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} |\beta_{j,k}| + \sum_{k \leq \tilde{k}_{j,r,\mathrm{left}}} |\beta_{j,k}| \right] \\ &\leq \frac{c_7}{\sqrt{n}} 2^{-r/2}. \end{aligned}$$

Moreover,

$$\begin{aligned} |K_{j,r}| \leq & \left| \left\{ k \in \widetilde{\mathbb{Z}}_{j,r},\ k \in [\tilde{k}_{j,r,\mathrm{left}}, \tilde{k}_{j,r,\mathrm{right}}] \right\} \right| \\ & + \sqrt{n} \ \left[\sum_{k \geq \tilde{k}_{j,r,\mathrm{right}}} |\beta_{j,k}| + \sum_{k \leq \tilde{k}_{j,r,\mathrm{left}}} |\beta_{j,k}| \right]. \end{aligned}$$

The elements in the first set are either equal to $\tilde{k}_{j,r,\mathrm{left}}, \tilde{k}_{j,r,\mathrm{right}}, \tilde{k}_{j,r,\mathrm{left}}+1, \tilde{k}_{j,r,\mathrm{right}} - 1$ or in $[-L_{\mathrm{wav}} + m2^j, L_{\mathrm{wav}} + m2^j]$. Hence, using Lemma 10,

$$\begin{aligned} |K_{j,r}| &\leq c_8 \left[1 + \sqrt{n} 2^{-r/2} \right] \\ &\leq c_9 \sqrt{n} 2^{-r/2}. \end{aligned}$$

We then deduce (4.11) from (4.32). □

4.7. Proof of Theorem 3.3

Since the theorem is stated for R and M large enough, we may without loss of generality replace R by c_1R and M by c_2M, where c_1, c_2 only depend on α, p, θ and the wavelet basis. In other words, we only need to build a subset $\mathscr{U}\mathscr{S}'^{\alpha}_{p,\theta}(R, M)$ of $\mathscr{U}\mathscr{S}^{\alpha}_{p,\theta}(c_1R, c_2M)$. The densities f in this set may satisfy $\sqrt{f} \in \mathfrak{B}^{\alpha}_{p,\infty}(c_1R)$ and $\int f^{\theta}(x)\,\mathrm{d}x \leq c_2M$.

Throughout the proof, we consider $r \geq 1$ and denote the elements of $\{0,1\}^{r+1}$ by $(\delta_k)_{0\leq k\leq r}$. We define the Hamming distance Δ for $\delta, \delta' \in \{0,1\}^{r+1}$ by

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

We denote the Kullback Leibler divergence between two densities f and g by

$$K(f, g) = \int f(x) \log\left(f(x)/g(x)\right)\,\mathrm{d}x,$$

whenever it exists. We use it here for densities f and g that vanish only simultaneously (in which case the convention $0 \times \log(0/0) = 0$ is applied).

It is convenient to draw on the Varshamov-Gilbert bound to prove the lower bound. The lemma below is ready to use. It follows from Theorem 2.5 and Lemma 2.9 of [Tsy08] (see also Section E in the third preprint version of [Sar23] for the constants).

Lemma 11. *Consider $\bar{r} \geq 14$ and suppose $r = 2\bar{r}$. Then, there is a subset $\mathscr{D}$ of $\{0,1\}^{r+1}$ such that*

$$\sum_{k=0}^{r} \delta_k = \bar{r}$$

for all $\delta \in \mathscr{D}$. Moreover,

$$\Delta(\delta, \delta') \geq \bar{r}/4,$$

for all pair $(\delta, \delta') \in \mathscr{D}$ composed of distinct elements.

We now assume that there exists a family of densities $\mathscr{F} = \{f_\delta,\ \delta \in \mathscr{D}\}$ indexed by this set and satisfying

$$\begin{aligned} h^2(f_\delta, f_{\delta'}) &\geq \eta\Delta(\delta, \delta') \\ K(f_\delta, f_{\delta'}) &\leq \frac{5r}{1000n} \end{aligned} \tag{4.34}$$

for some $\eta > 0$ and all $\delta \neq \delta' \in \mathscr{D}$.

Then, there exists a numerical value $c > 0$ such that

$$\mathcal{R}(\mathscr{F}) \geq c\eta r.$$

We need to construct a suitable family $\mathscr{F}$ of densities satisfying the conditions of the previous lemma. The two results below are tailored to solve this problem. They are proven after the current proof.

Lemma 12. *We consider $q \geq 2$ and two non-negative maps ς_0, ς_1 on $[0, +\infty)$. We suppose that these functions are non-increasing, compactly supported on $[0,1]$ and with continuous derivatives up to order q. Moreover, $\varsigma_\delta(0) = 1$, $\varsigma_\delta(1) = 0$, and $\varsigma_\delta^{(s)}(0) = \varsigma_\delta^{(s)}(1) = 0$ for all $\delta \in \{0,1\}$ and $s \in \{1, \dots, q-1\}$, where $\varsigma_\delta^{(s)}$ denotes the s^{th} derivative of ς_δ. We consider a positive integer r, three positive numbers σ, b_0, ℓ_0, $x_0 \in (0, r\ell_0]$, $\varepsilon = \sigma/r$, and set for all $k \geq 0$,*

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

We then define for all $x \geq 0$ and $\delta = (\delta_k)_{0 \leq k \leq r} \in \{0,1\}^{r+1}$,

$$\begin{aligned} g_\delta(x) = b_0 \mathbb{1}_{[0,x_0)} & \\ &+ \frac{1}{1+\varepsilon} \sum_{k=0}^{r} b_k \left[1 + \varepsilon \varsigma_{\delta_k}((x - x_k)/\ell_k)\right] \mathbb{1}_{[x_k, x_{k+1})}(x) \\ &+ b_{r+1} \varsigma_0\left((x - x_{r+1})/(r\ell_0)\right) \mathbb{1}_{[x_{r+1}, +\infty)}(x). \end{aligned} \tag{4.35}$$

We extend g_δ to an even function on $\mathbb{R}$.

This function g_δ has the following properties: it is unimodal and such that $g_\delta(x_k) = b_k$ for all $k \in \{0, \dots, r\}$, $\delta \in \{0,1\}^{r+1}$. Moreover, g_δ lies in $\mathfrak{B}^{\alpha}_{p,\infty}(cR)$ for all $p \in (0, +\infty]$, $\alpha \in (\max\{1, 1/p - 1\}, q)$ and $R > 0$ satisfying

$$b_0 r^{1/p-1} \leq R \ell_0^{\alpha - 1/p} \tag{4.36}$$

$$b_0 \ell_0^{1/p} r^{1/p} \leq R. \tag{4.37}$$

Above, c only depends on the wavelet basis, ς_0, ς_1, p, q, α and σ.

Lemma 13. *Consider some $q \geq 2$. There exist two functions ς_0, ς_1 fulfilling the assumptions of Lemma 12. They satisfy*

$$\int_0^1 \varsigma_0(x)\,\mathrm{d}x = \int_0^1 \varsigma_1(x)\,\mathrm{d}x \tag{4.38}$$

and do not coincide almost everywhere on $[0,1]$.

Let $\bar{r}$ be the largest integer such that

$$\bar{r} \leq R^{\beta_1} M^{\beta_2} n^{1-\gamma}.$$

We consider the smallest integer q larger than α, and the set $\mathscr{D}$ given in the first part of Lemma 11. Let then ς_0, ς_1 be the maps given by Lemma 13. We consider $a > 0$ and set

$$\begin{aligned} \varepsilon &= 1/\bar{r} \\ \ell_0^{2(\alpha - 1/p + 1/2)} &= a R^{-2} n^{-1} r^{2/p} \end{aligned}$$

$$b_0^2 = ar^2(n\ell_0)^{-1}.$$

Note that (4.36) holds true. Moreover, $b_0\ell_0^{1/p}\varepsilon^{-1/p}$ tends to 0 when n goes to infinity when $\theta < p/2$. Hence (4.37) is true. When $\theta = p/2$,

$$b_0\ell_0^{1/p}\varepsilon^{-1/p} \le 2^{(2\alpha p-3+p+2/p)/(p+2\alpha p-2)}a^{\alpha/(1+2\alpha-2/p)}M^{1/p}.$$

We thus also have (4.37) when $M \le R^p$ and a small enough.

We consider $x_0 = r\ell_0$, $\delta \in \mathscr{D}$, and the map g_δ defined by (4.35). Let then I be the value of the integral in (4.38), and

$$I_{0,2} = \int_0^1 \varsigma_0^2(x)\,\mathrm{d}x$$
$$I_{1,2} = \int_0^1 \varsigma_1^2(x)\,\mathrm{d}x.$$

Elementary maths lead to:

$$\int (g_\delta(x))^2\,\mathrm{d}x = \frac{2ar^3}{n} + \frac{8a(1+2\varepsilon I)\,(2+\varepsilon)}{n(1+\varepsilon)^2\varepsilon^3} + \frac{8a\,(\bar{r}I_{1,2} + (\bar{r}+1)I_{0,2})}{n(1+\varepsilon)^2} + \frac{ar^3}{n(1+\varepsilon)^{2(r+1)}}I_{0,2}.$$

This integral does not depend on δ and tends to 0 when n goes to infinity. Besides,

$$\int (g_\delta(x))^{2\theta}\,\mathrm{d}x \le c_1 b_0^{2\theta}\ell_0 r,$$

for all n large enough and some $c_1 > 0$ only depending on ς_0, ς_1 and θ. In particular,

$$\int (g_\delta(x))^{2\theta}\,\mathrm{d}x \le c_1 a^{(1-2\theta/p+2\alpha\theta)/(1+2\alpha-2/p)}M \le c_1 M,$$

when $a \le 1$. We also have $\operatorname{supp} g_\delta \subset [-c_1M, c_1M]$ when $\theta = 0$.

We now apply Lemma 12 with suitable values of parameters to get a unimodal non-negative function $\zeta \in \mathfrak{B}^\alpha_{p,\infty}(c_2R)$, compactly supported on $[-1,1]$, and such that $\int(\zeta(x))^2\,\mathrm{d}x > 1$, $\int(\zeta(x))^{2\theta}\,\mathrm{d}x \le M$ (up to an increase of R, M). We then consider $s \in (0,1)$ and set

$$f_\delta = (g_\delta + s\zeta)^2.$$

As $x_0 > 1$ for n or M large enough,

$$\int f_\delta(x)\,\mathrm{d}x = \int (g_\delta(x))^2\,\mathrm{d}x + 2sb_0\int \zeta(x)\,\mathrm{d}x + s^2\int(\zeta(x))^2\,\mathrm{d}x.$$

Note that b_0 tends to 0 as n goes to infinity. We may hence find $s \in (0,1)$, not depending on δ, such that this integral is 1.

By putting all these results together, by using Lemma 12 and Proposition 3.1, we get that f_δ is a density lying in $\mathscr{U}\mathscr{S}^{\alpha}_{p,\theta}(c_3 R, c_4 M)$ and such that $\sqrt{f_\delta} \in \mathfrak{B}^{\alpha}_{p,\infty}(c_3 R)$. Its support is included in $[-c_1 M, c_1 M]$ when $\theta = 0$.

Since g_δ is unimodal, we have $g_\delta(x) \in [b_{r+1}, b_0]$ for all $x \in [-x_{r+1}, x_{r+1}]$. When $x \notin [-x_{r+1}, x_{r+1}]$, $g_\delta(x) = g_{\delta'}(x)$ for all δ, δ'. An elementary inequality yields $b_0 \leq e^3 b_{r+1}$. We then deduce from Lemma 2.7 of [Tsy08], that for all $\delta \neq \delta' \in \mathscr{D}$,

$$\begin{aligned}
K(f_\delta, f_{\delta'}) &\leq b_{r+1}^{-2} \int \left(f_\delta(x) - f_{\delta'}(x)\right)^2 \,\mathrm{d}x \\
&\leq b_{r+1}^{-2} \int \left(g_\delta(x) - g_{\delta'}(x)\right)^2 \left(g_\delta(x) + g_{\delta'}(x) + 2s\zeta(x)\right)^2 \,\mathrm{d}x \\
&\leq b_{r+1}^{-2} \int \left(g_\delta(x) - g_{\delta'}(x)\right)^2 \left(g_\delta(x) + g_{\delta'}(x)\right)^2 \,\mathrm{d}x \\
&\leq 4e^6 \int \left(g_\delta(x) - g_{\delta'}(x)\right)^2 \,\mathrm{d}x \\
&\leq \frac{8e^6}{(1+\varepsilon)^2} \varepsilon^2 \sum_{k=0}^{r} b_k^2 \ell_k |\delta_k - \delta_{k'}| \int \left(\varsigma_0(x) - \varsigma_1(x)\right)^2 \,\mathrm{d}x \\
&\leq c_5 b_0^2 \ell_0 \varepsilon^2 r \\
&\leq c_6 a r / n,
\end{aligned}$$

where c_5, c_6 only depend on ς_0, ς_1. We now choose a small enough to ensure that (4.34) holds true.

Moreover,

$$\begin{aligned}
h^2(f_\delta, f_{\delta'}) &= \frac{\varepsilon^2}{(1+\varepsilon)^2} \sum_{k=0}^{r} b_k^2 \ell_k |\delta_k - \delta_{k'}| \int_0^1 \left(\varsigma_0(x) - \varsigma_1(x)\right)^2 \,\mathrm{d}x \\
&\geq c_7 b_0^2 \ell_0 \varepsilon^2 \Delta(\delta, \delta'), \\
&\geq c_8 (a/n) \Delta(\delta, \delta'),
\end{aligned}$$

for some $c_7, c_8 > 0$.

We may hence apply Lemma 11 with η of the order of n^{-1}. This leads to

$$\mathcal{R}(\mathscr{F}) \geq c_9 r / n,$$

where $\mathscr{F} = \{f_\delta, \delta \in \mathscr{D}\}$. We conclude using the definition of r. □

Proof of Lemma 12. The only delicate point is $g_\delta \in \mathfrak{B}^{\alpha}_{p,\infty}(cR)$ for some c. We prove this result when $p < \infty$. The proof when $p = \infty$ is obtained by making slight modifications. We suppose without loss of generality that $q \geq 2$ is the smallest integer larger than α.

The q^{th} order difference operator evaluated in g_δ is defined for $h > 0$ and $x \in \mathbb{R}$ by

$$\Delta_h^q g_\delta(x) = \sum_{j=0}^{q} \binom{q}{j} (-1)^{q-j} g_\delta(x + jh).$$

Section 7 in Chapter 2 of [DL93] gives

$$\Delta_h^q g_\delta(x) = h^{q-1} \int \left(\Delta_h^1 g_\delta\right)^{(q-1)} (x + th) M(t)\,\mathrm{d}t,$$

where M is a compactly supported density function on $[0, q-1]$ and bounded by 1. Therefore,

$$\Delta_h^q g_\delta(x) = h^{q-1} \int \left[g_\delta^{(q-1)}(x + (t+1)h) - g_\delta^{(q-1)}(x + th) \right] M(t)\,\mathrm{d}t. \tag{4.39}$$

We consider $k \in \{0, \ldots, r\}$ and the map ς_{k,δ_k} defined for $x \geq 0$ by

$$\varsigma_{k,\delta_k}(x) = \left[1 + \varepsilon \varsigma_{\delta_k}((x - x_k)/\ell_k)\right] \mathbb{1}_{[x_k, x_{k+1})}(x).$$

When $k = r + 1$, we set

$$\varsigma_{r+1}(x) = \varsigma_0\left((x - x_{r+1})/(r\ell_0)\right) \mathbb{1}_{[x_{r+1}, \infty)}(x).$$

When $x < 0$, we put $\varsigma_{k,\delta_k}(x) = \varsigma_{k,\delta_k}(|x|)$ and $\varsigma_{r+1}(x) = \varsigma_{r+1}(|x|)$. These maps are $q-1$ times differentiable at all points except at $-x_{r+1}, -x_{k+1}, -x_k, x_k, x_{k+1}, x_{r+1}$. We nevertheless set

$$\begin{aligned} \varsigma_{k,\delta_k}^{(q-1)}(-x_{k+1}) &= \varsigma_{k,\delta_k}^{(q-1)}(-x_k) \\ &= \varsigma_{k,\delta_k}^{(q-1)}(x_k) \\ &= \varsigma_{k,\delta_k}^{(q-1)}(x_{k+1}) \\ &= 0 \end{aligned}$$

and $\varsigma_{r+1}^{(q-1)}(x_{r+1}) = \varsigma_{r+1}^{(q-1)}(-x_{r+1}) = 0$ so that

$$g_\delta^{(q-1)}(x) = \frac{1}{1+\varepsilon} \sum_{k=0}^{r} b_k \varsigma_{k,\delta_k}^{(q-1)}(x) + b_{r+1} \varsigma_{r+1}^{(q-1)}(x) \tag{4.40}$$

holds true on $\mathbb{R}$.

Define now

$$\bar{\alpha} = \alpha - (q-1) \in [0, 1).$$

Since $\varsigma_{\delta_k}^{(q-1)}$ is compactly supported with a continuous derivative on $\mathbb{R}$,

$$\left| \varsigma_{\delta_k}^{(q-1)}(b) - \varsigma_{\delta_k}^{(q-1)}(a) \right| \leq c_1 |b - a|^{\bar{\alpha}}$$

for all $a, b \in \mathbb{R}$, $\delta_k \in \{0, 1\}$, and some c_1 only depending on $\varsigma_0, \varsigma_1, \alpha, q$. Therefore, for all $a, b \geq 0$,

$$\left|\varsigma_{k,\delta_k}^{(q-1)}(b) - \varsigma_{k,\delta_k}^{(q-1)}(a)\right| \leq c_1 \varepsilon \ell_k^{-\alpha} |b-a|^{\bar{\alpha}}. \tag{4.41}$$

The same result holds true when $a, b \leq 0$. Suppose now that a, b have opposite signs, say $a \geq 0$ and $b \leq 0$. Then, $\varsigma_{k,\delta_k}^{(q-1)}(b) = 0$ if $|b| \leq x_k$. Otherwise, we use $\varsigma_{k,\delta_k}^{(q-1)}(x_k) = 0$ to get

$$\begin{aligned}\left|\varsigma_{k,\delta_k}^{(q-1)}(b)\right| &= \left|\varsigma_{k,\delta_k}^{(q-1)}(|b|) - \varsigma_{k,\delta_k}^{(q-1)}(x_k)\right| \\ &\leq c_1 \varepsilon \ell_k^{-\alpha} \left(|b| - x_k\right)^{\bar{\alpha}} \\ &\leq c_1 \varepsilon \ell_k^{-\alpha} \left(|b| + a\right)^{\bar{\alpha}} \\ &\leq c_1 \varepsilon \ell_k^{-\alpha} \left|b - a\right|^{\bar{\alpha}}.\end{aligned}$$

A similar reasoning applies to $|\varsigma_{k,\delta_k}^{(q-1)}(a)|$. To sum up, (4.41) holds true for all $a, b \in \mathbb{R}$ (to within a multiplication of c_1 by 2). Likewise,

$$\begin{aligned}\left|\varsigma_{r+1}^{(q-1)}(b) - \varsigma_{r+1}^{(q-1)}(a)\right| &\leq c_2 (r\ell_0)^{-\alpha} |b-a|^{\bar{\alpha}} \\ &\leq c_3 \varepsilon \ell_0^{-\alpha} \; |b-a|^{\bar{\alpha}}\end{aligned}$$

as $\alpha > 1$.

Consider now $x \in \mathbb{R}$, $h > 0$ and $t \in [0, q-1]$. There can only be one non-zero term in the sum of (4.40). Therefore,

$$\left|g_\delta^{(q-1)}(x + (t+1)h) - g_\delta^{(q-1)}(x + th)\right| \leq c_4 b_0 \varepsilon \ell_0^{-\alpha} h^{\bar{\alpha}}.$$

We deduce from (4.39) that $|\Delta_h^q g_\delta(x)|$ satisfies

$$|\Delta_h^q g_\delta(x)| \leq c_5 b_0 \varepsilon \ell_0^{-\alpha} h^{\alpha}.$$

Moreover, $\Delta_h^q g_\delta(\cdot)$ is compactly supported on $[-x_{r+1} - r\ell_0 - qh, x_{r+1} + r\ell_0]$. Observe that

$$\begin{aligned}x_{r+1} &\leq x_0 + \sum_{j=0}^{r} \ell_j \\ &\leq r\ell_0 + \ell_0 \sum_{j=0}^{r} (1+\varepsilon)^{2j} \\ &\leq c_6 r \ell_0.\end{aligned}$$

We hence get when $h \leq r\ell_0$,

$$\begin{aligned}\int_{-\infty}^{\infty} |\Delta_h^q g_\delta(x)|^p \, \mathrm{d}x &\leq c_7 r \ell_0 \left(b_0 \varepsilon \ell_0^{-\alpha} h^{\alpha}\right)^p \\ &\leq c_8 R^p h^{\alpha p},\end{aligned}$$

where the last inequality relies on (4.36). When $h \geq r\ell_0$, we bound the $\mathbb{L}^p$ (quasi) norm of $|\Delta_h^q g_\delta(\cdot)|$ as follows:

$$\begin{aligned}\int_{-\infty}^{\infty} |\Delta_h^q g_\delta(x)|^p \,\mathrm{d}x &\leq c_9 \int_{-\infty}^{\infty} |g_\delta(x)|^p \,\mathrm{d}x \\ &\leq c_{10}\left[b_0^p x_0 + \sum_{k=0}^{r} b_k^p \ell_k + b_{r+1}^p \ell_0 r\right] \\ &\leq c_{10}\left[b_0^p r\ell_0 + b_0^p \ell_0 \sum_{k=0}^{r} (1+\varepsilon)^{(2-p)k} + b_0^p \ell_0 r\right] \\ &\leq c_{11} b_0^p r \ell_0.\end{aligned}$$

We apply (4.36):

$$\int_{-\infty}^{\infty} |\Delta_h^q g_\delta(x)|^p \,\mathrm{d}x \leq c_{12} R^p r^p \ell_0^{\alpha p}.$$

Since $\alpha > 1$, $r^p \leq r^{\alpha p}$ and the right-hand side of the last inequality is not larger than $c_{12} R^p h^{\alpha p}$.

We may also use (4.37) in place of (4.36) to get $\|g_\delta\|_p \leq c_{13} R$. The conclusion $g_\delta \in \mathfrak{B}_{p,\infty}^{\alpha}(c_{14}R)$ then stems from another (equivalent) definition of Besov balls. See Section 10 in Chapter 2 of [DL93] for more details. □

Proof of Lemma 13. We introduce for all $p_1, p_2 \geq 1$ and $x \in [0, +\infty)$,

$$f_{p_1,p_2}(x) = (1 - x^{p_1})^{p_2} \mathbb{1}_{[0,1]}(x).$$

This map is non-increasing on $[0,1]$, $p-1$ times differentiable where $p = \min\{p_1, p_2\}$, and such that $f_{p_1,p_2}(0) = 1$, $f_{p_1,p_2}(1) = 0$, $f_{p_1,p_2}^{(s)}(0) = f_{p_1,p_2}^{(s)}(1) = 0$ for all $s \in \{1, \dots, p-1\}$. It fulfils the assumptions of Lemma 12 when $p \geq q+2$. We define the first function ς_0 by $\varsigma_0 = f_{q+3,q+3}$.

We then consider $t \in [0,1]$ and set

$$\varsigma_1 = t f_{q+3,q+2} + (1-t) f_{q+2,q+3}.$$

Since $f_{q+2,q+3} \leq f_{q+3,q+3} \leq f_{q+3,q+2}$, the integral of ς_1 evolves continuously from $\int f_{q+2,q+3}$ to $\int f_{q+3,q+2}$ as t varies from 0 to 1. There is therefore some $t \in [0,1]$ such that (4.38) holds true.

We conclude by noticing that ς_1 satisfies the assumptions of Lemma 12 and cannot coincide almost everywhere with ς_0 because they are polynomials of different degrees on $[0,1]$. □

4.8. Proof of the claims of Section 2.4

We introduce the linear estimators

$$\tilde{f} = \sum_{k \in K_{-1}} \widehat{\alpha}_{0,k} \bar{\phi}_{0,k} + \sum_{j=0}^{J} \sum_{k \in K_j} \widehat{\beta}_{j,k} \bar{\psi}_{j,k},$$

$$\tilde{f}' = \sum_{k \in \mathbb{Z}} \widehat{\alpha}_{0,k} \bar{\phi}_{0,k} + \sum_{j=0}^{J} \sum_{k \in \mathbb{Z}} \widehat{\beta}_{j,k} \bar{\psi}_{j,k}$$

where

$$\widehat{\alpha}_{0,k} = \frac{1}{n}\sum_{i=1}^{n}\phi_{0,k}(X_i) \quad \text{and} \quad \widehat{\beta}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}\psi_{j,k}(X_i).$$

We define $\bar{f}(x)$ and $\bar{f}'(x)$ for $x \in \mathbb{R}$ by $\bar{f}(x) = \mathbb{E}[\tilde{f}(x)]$, $\bar{f}'(x) = \mathbb{E}[\tilde{f}'(x)]$. All the statements concerning the results for the $\mathbb{L}^1$ loss can be deduced from the lemma below:

Lemma 14. *If $f \in \mathcal{B}^{\alpha}_{1,\infty}(R) \cap \mathcal{T}_{1/2}(M)$ with $\alpha \in (0,\tau)$, $R > 0$, $M \geq 1$,*

$$\max\left\{d_1(f,\bar{f}), d_1(f,\bar{f}')\right\} \leq c_1\left[R2^{-J\alpha} + R^{1/(2\alpha+1)}M^{2\alpha/(2\alpha+1)}n^{-\alpha/(2\alpha+1)} + Mn^{-1/2}\right] \tag{4.42}$$

$$\max\left\{\mathbb{E}\left[d_1(\tilde{f},\bar{f})\right], \mathbb{E}\left[d_1(\tilde{f}',\bar{f}')\right]\right\} \leq c_2 M 2^{J/2} n^{-1/2}, \tag{4.43}$$

where c_1, c_2 only depend on α and the wavelets.

Proof of Lemma 14. We have,

$$\max\left\{d_1(f,\bar{f}), d_1(f,\bar{f}')\right\} \leq c_1\left[\sum_{k\not\in K_{-1}}|\alpha_{0,k}(f)| + \sum_{j=0}^{J}2^{-j/2}\sum_{k\not\in K_j}|\beta_{j,k}(f)| + \sum_{j=J+1}^{\infty}2^{-j/2}\sum_{k\in\mathbb{Z}}|\beta_{j,k}(f)|\right].$$

The condition $f \in \mathcal{B}^{\alpha}_{1,\infty}(R)$ ensures

$$\sum_{j=J+1}^{\infty}2^{-j/2}\sum_{k\in\mathbb{Z}}|\beta_{j,k}(f)| \leq c_2 R 2^{-J\alpha}.$$

Now, $2^{-j/2}|\beta_{j,k}(f)| \leq c_3 f_{j,k}$ where $f_{j,k}$ is defined by (4.3). Set

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

and observe that $|\overline{\mathbb{Z}}_{j,r}| \leq c_4 M 2^{r/2} 2^{j/2}$. We deduce,

$$2^{-j/2}\sum_{k\not\in K_j}|\beta_{j,k}(f)| \leq c_5\sum_{2^r\geq n} M2^{-r/2}2^{j/2} \leq c_6 M 2^{j/2} n^{-1/2}.$$

A similar reasoning applies for the father wavelet coefficients. Now,

$$\sum_{j=0}^{J}2^{-j/2}\sum_{k\not\in K_j}|\beta_{j,k}(f)| \leq c_7\sum_{j=0}^{J}\min\left\{R2^{-j\alpha}, M2^{j/2}n^{-1/2}\right\} \leq c_8 R^{1/(2\alpha+1)}M^{2\alpha/(2\alpha+1)}n^{-\alpha/(2\alpha+1)}.$$

This shows (4.42). As to (4.43), we merely note

$$\begin{aligned}\max\left\{\mathbb{E}\left[d_1(\tilde{f},\bar{f})\right],\mathbb{E}\left[d_1(\tilde{f}',\bar{f}')\right]\right\} &\leq c_9\left[\sum_{k\in\mathbb{Z}} f_{-1,k}^{1/2}+\sum_{j=0}^{J}\sum_{k\in\mathbb{Z}} f_{j,k}^{1/2}\right]n^{-1/2}\\ &\leq c_{10}M\left[1+\sum_{j=0}^{J}2^{j/2}\right]n^{-1/2}\\ &\leq c_{11}M2^{J/2}n^{-1/2}.\end{aligned}$$

□

The result concerning the bias term $d_2^2(\sqrt{f},V)$ comes from the following lemma:

Lemma 15. *If $\sqrt{f}\in\mathcal{B}^{\alpha}_{2,\infty}(R)$ and $f\in\mathcal{T}_{1/2}(M)$, for $\alpha\in(0,\tau)$, $R>0$, $M\geq 1$,*

$$d_2^2\left(\sqrt{f},V\right)\leq c\left[R^2 2^{-2J\alpha}+\varepsilon_n\right],$$

where c only depends on α and the wavelets.

Proof of Lemma 15. We have,

$$\begin{aligned}d_2^2\left(\sqrt{f},V\right)\leq c_1&\left[\sum_{k\notin K_{-1}}\left(\alpha_{0,k}(\sqrt{f})\right)^2+\sum_{j=0}^{J}\sum_{k\notin K_j}\left(\beta_{j,k}(\sqrt{f})\right)^2\right.\\ &\left.+\sum_{j=J+1}^{\infty}\sum_{k\in\mathbb{Z}}\left(\beta_{j,k}(\sqrt{f})\right)^2\right].\end{aligned}$$

The smoothness condition ensures

$$\sum_{k\in\mathbb{Z}}\left(\beta_{j,k}(\sqrt{f})\right)^2\leq R^2 2^{-2j\alpha}.$$

Besides, we have $\left(\beta_{j,k}(\sqrt{f})\right)^2\leq c_2 f_{j,k}$ where $f_{j,k}$ is defined in (4.3). By doing as in the previous proof,

$$\sum_{k\notin K_{-1}}\left(\alpha_{0,k}(\sqrt{f})\right)^2\leq c_3 M n^{-1/2}\text{ and }\sum_{k\notin K_j}\left(\beta_{j,k}(\sqrt{f})\right)^2\leq c_4 M n^{-1/2}2^{j/2}.$$

Therefore,

$$\sum_{j=0}^{J}\sum_{k\notin K_j}\left(\beta_{j,k}(\sqrt{f})\right)^2\leq c_5\sum_{j=0}^{J}\min\left\{Mn^{-1/2}2^{j/2},R^2 2^{-2j\alpha}\right\},$$

hence the result. □

As to the $\mathbb{L}^2$ loss, the statements ensue from the following inequalities.

Lemma 16. *If $f \in \mathcal{B}^{\alpha}_{2,\infty}(R)$ with $\alpha \in (0,\tau)$, $R > 0$,*

$$\max\left\{d_2^2(f,\bar{f}), d_2^2(f,\bar{f}')\right\} \leq c_1\left[R^2 2^{-2J\alpha} + R^{2/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)} + n^{-1}\right] \tag{4.44}$$

$$\max\left\{\mathbb{E}\left[d_2^2(\tilde{f},\bar{f})\right], \mathbb{E}\left[d_2^2(\tilde{f}',\bar{f}')\right]\right\} \leq c_2 2^J/n, \tag{4.45}$$

where c_1, c_2 only depend on α and the wavelet basis.

Proof of Lemma 16. We have,

$$\max\left\{d_2^2(f,\bar{f}), d_2^2(f,\bar{f}')\right\} \leq c_1\left[\sum_{k\notin K_{-1}} (\alpha_{0,k}(f))^2 + \sum_{j=0}^{J}\sum_{k\notin K_j}(\beta_{j,k}(f))^2 + \sum_{j=J+1}^{\infty}\sum_{k\in\mathbb{Z}}(\beta_{j,k}(f))^2\right].$$

The last term can be bounded using $f \in \mathcal{B}^{\alpha}_{2,\infty}(R)$. For the second last, we note $(\beta_{j,k}(f))^2 \leq c_2 2^j f_{j,k}^2$ and do as in the proof of Lemma 14 by noticing that $|\overline{\mathbb{Z}}_{j,r}| \leq c_3 2^r$ as f is a density:

$$\sum_{k\notin K_j}(\beta_{j,k}(f))^2 \leq c_4 \sum_{2^r \geq n} 2^{-r} 2^j \leq c_5 2^j n^{-1}.$$

The reasoning is similar for the first term. Now,

$$\sum_{j=0}^{J}\sum_{k\notin K_j}(\beta_{j,k}(f))^2 \leq c_6 \sum_{j=0}^{J} \min\left\{2^j n^{-1}, R^2 2^{-2j\alpha}\right\} \leq c_7 R^{2/(2\alpha+1)} n^{-2\alpha/(2\alpha+1)}.$$

We put everything together to get (4.44). We now show (4.45):

$$\max\left\{\mathbb{E}\left[d_2^2(\tilde{f},\bar{f})\right], \mathbb{E}\left[d_2^2(\tilde{f}',\bar{f}')\right]\right\} \leq c_8\left[\sum_{k\in\mathbb{Z}} f_{-1,k} + \sum_{j=0}^{J} 2^j \sum_{k\in\mathbb{Z}} f_{j,k}\right] n^{-1} \leq c_9 2^J n^{-1}.$$

□

References

[BBS17] Yannick Baraud, Lucien Birgé, and Mathieu Sart. A new method for estimation and model selection: ρ-estimation. *Inventiones mathematicae*, 207(2):425–517, 2017. MR3595933

[BC21] Elodie Brunel and Fabienne Comte. Hazard regression with noncompactly supported bases. *Canadian Journal of Statistics*, 49(4):1273–1297, 2021. MR4349645

[Bir06a] Lucien Birgé. Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Annales de l'IHP Probabilités et statistiques*, 42(3):273–325, 2006. MR2219712

[Bir06b] Lucien Birgé. Statistical estimation with model selection. *Indagationes Mathematicae*, 17(4):497–537, 2006. MR2320111

[Bir14] Lucien Birgé. Model selection for density estimation with $\mathbb{L}_2$-loss. *Probability Theory and Related Fields*, 158:533–574, 2014. MR3176358

[BM98] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998. MR1653272

[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. MR1162365

[CGC21] Fabienne Comte and Valentine Genon-Catalot. Drift estimation on non compact support for diffusion models. *Stochastic Processes and their Applications*, 134:174–207, 2021. MR4206844

[CL20] Kaikai Cao and Youming Liu. Uncompactly supported density estimation with L^1 risk. *Communications on Pure & Applied Analysis*, 19(8), 2020. MR4112009

[CL23] Fabienne Comte and Claire Lacour. Non compact estimation of the conditional density from direct or noisy data. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 59(3): 1463–1507, 2023. MR4635716

[CVNRF15] René Erlin Castillo, Fabio Andrés Vallejo Narvaez, and Julio C. Ramos Fernández. Multiplication and composition operators on weak L_p spaces. *Bulletin of the Malaysian Mathematical Sciences Society*, 38(3):927–973, 2015. MR3352659

[DJ97] Bernard Delyon and Anatoli Juditsky. On the computation of wavelet coefficients. *Journal of Approximation Theory*, 88(1):47–79, 1997. MR1426459

[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. MR1394974

[DL93] Ronald A. DeVore and George G. Lorentz. *Constructive approximation*. Springer Berlin, Heidelberg, 1993. MR1261635

[DL14] Cécile Durot and Hendrik P. Lopuhaä. A Kiefer-Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. *Electronic Journal of Statistics*, 8(2):2479–2513, 2014. MR3285873

[DL18] Cécile Durot and Hendrik P. Lopuhaä. Limit theory in monotone function estimation. *Statistical Science*, 33(4):547–567, 2018. MR3881208

[DW16] Charles R. Doss and Jon A. Wellner. Global rates of convergence of the MLEs of log-concave and s-concave densities. *The Annals of statistics*, 44(3):954–981, 2016. MR3485950

[EL00] P.P.B. Eggermont and V.N. LaRiccia. Maximum likelihood estimation of smooth monotone and unimodal densities. *Annals of statistics*, 28:922–947, 2000. MR1792794

[GL11] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011. MR2850214

[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. MR3230001

[HK05] Peter Hall and Kee-Hoon Kang. Unimodal kernel density estimation by data sharpening. *Statistica Sinica*, 15:73–98, 2005. MR2125721

[JLL04] Anatoli Juditsky and Sophie Lambert-Lacroix. On minimax density estimation on $\mathbb{R}$. *Bernoulli*, 10(2):187–220, 2004. MR2046772

[KS16] Arlene K.H. Kim and Richard J. Samworth. Global rates of convergence in log-concave density estimation. *The Annals of statistics*, 44(6):2756–2779, 2016. MR3576560

[Lep13] Oleg Lepski. Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. *The Annals of Statistics*, 41(2):1005–1034, 2013. MR3099129

[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. MR3605364

[LM19] Hendrik P. Lopuhaä and Eni Musta. Central limit theorems for the L_p-error of smooth isotonic estimators. *Electronic Journal of Statistics*, 13, 2019. MR3935844

[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. MR3909933

[Mas07] Pascal Massart. *Concentration inequalities and model selection.* Springer, 2007. MR2319879

[RBR10] Patricia Reynaud-Bouret and Vincent Rivoirard. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electronic Journal of Statistics*, 4:172–238, 2010. MR2645482

[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. MR2719482

[Sar21] Mathieu Sart. Minimax bounds for Besov classes in density estimation. *Electronic Journal of Statistics*, 15(1):3184–3216, 2021. MR4280167

[Sar23] Mathieu Sart. Non linear wavelet density estimation on the real line. *Available on Hal*, 2023.

[Tsy08] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008. MR2724359

[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. MR1986176