



HAL
open science

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

Mathieu Sart

► **To cite this version:**

Mathieu Sart. About the optimal estimation of a density with infinite support under Hellinger loss. 2023. hal-04042886v1

HAL Id: hal-04042886

<https://hal.science/hal-04042886v1>

Preprint submitted on 23 Mar 2023 (v1), last revised 29 Mar 2024 (v2)

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

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

ABOUT THE OPTIMAL ESTIMATION OF A DENSITY WITH INFINITE SUPPORT UNDER HELLINGER LOSS

MATHIEU SART

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 $\mathfrak{B}_{p,\infty}^\alpha(R)$. We make them explicit without further conditions when $p < 2$, and under a tail dominance condition when p is larger.

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} , \mathcal{F} be a subset of $\mathcal{D}(\mathbb{R})$, and \mathcal{L} be a loss function. The minimax risk is

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

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

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

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

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

An optimal estimation procedure \hat{f} is therefore a procedure whose risk $\mathbb{E}[\mathcal{L}(f, \hat{f})]$ converges at the rate ε_n under the sole condition that f lies in \mathcal{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} .

Date: March, 2023.

2010 Mathematics Subject Classification. 62G05, 62G07.

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

To formalize things a little more, we state that f is smooth if f belongs to a ball $\mathfrak{B}_{p,\infty}^\alpha(R)$ of a Besov space. In a nutshell, the parameter R is an upper-bound of the (quasi) Besov norm of the elements f of $\mathfrak{B}_{p,\infty}^\alpha(R)$. 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 2.1. For the sake of rigour, we assume throughout this introduction that R is large enough ($\mathfrak{B}_{p,\infty}^\alpha(R)$ 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 $\mathcal{L} = d_q^q$. They are now fully known, up to log factors, when the density is also compactly supported, that is when it belongs to

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

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 $\mathcal{L} = h^2$. 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 $\sqrt{f} \in \mathfrak{B}_{p,\infty}^\alpha(R)$ 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 $f \in \mathfrak{B}_{p,\infty}^\alpha(R)$ 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 $[0, 1]$.

We present our results in the forthcoming section. The proofs are postponed to Section 3. 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 \mathcal{F} of functions,

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

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

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

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

2. MINIMAX RATES

2.1. Assumptions on the density. We present in this section the classes of functions we use to model the smoothness of f and the size of its tails.

2.1.1. 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 $\phi = 1_{[0,1]}$, 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 $\tau \in \mathbb{N}^*$.

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

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

where for any $j \geq -1, k \in \mathbb{Z}$,

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

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

$$\begin{aligned} \bar{\psi}_{-1,k}(x) &= \bar{\phi}(x - k), & \bar{\psi}_{j,k}(x) &= 2^{j/2} \bar{\psi}(2^j x - k) \\ \psi_{-1,k}(x) &= \phi(x - k), & \psi_{j,k}(x) &= 2^{j/2} \psi(2^j x - k). \end{aligned}$$

It is worthwhile to notice that the coefficients $\beta_{j,k}(f)$ are well defined if f is only supposed to be integrable. The decomposition (1) remains valid in this case (see Appendix H of [Sar23] for instance).

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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 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 .*

2.2. 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 $p \neq 2$,

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

When $p = 2$, we rather set

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

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

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

$$(3) \quad \mathcal{R}(\mathcal{F}_{\alpha,p,\theta}(R, M)) \leq c_1 \left[\varepsilon_n + \frac{\log n}{n} \right],$$

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 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,*

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

where ε_n is given in the preceding theorem, and where $\mathcal{F}'_{\alpha,p,\theta}(R, M)$ is a subset of $\mathcal{F}_{\alpha,p,\theta}(R, M)$. Moreover, any function $f \in \mathcal{F}'_{\alpha,p,\theta}(R, M)$ belongs to $\mathcal{T}_\theta(M)$ and satisfies $|x|f^\theta(x) \leq c_3 M$ for all $|x| \geq 1$. We also have $\sqrt{f} \in \mathcal{B}_{p,\infty}^\alpha(R)$ and even $\sqrt{f} \in \mathfrak{B}_{p,\infty}^\alpha(R)$ if $\theta < p/2$ or if $\theta = p/2$ with $M \leq R^p$. Above, c_2, c_3, 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 $\alpha \in ((1/p - 1/2)_+, \tau)$. We recall that τ can be taken arbitrarily large. 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 $q \geq 1$).

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 $\mathfrak{B}_{p,\infty}^\alpha(R)$ with R large enough (see the proof of Theorem 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,

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

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). We refer to the proof of Theorem 2 for more details.

In the previous results, 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 4. *For all $p \in (0, 2)$, $R > 0$, $\tau > 1/p - 1/2$ and $\alpha = 1/p - 1/2$,*

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

It is interesting to note that the exponent in the optimal rate does not tend to 0 when $\alpha \rightarrow 1/p - 1/2$. 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. PROOFS

3.1. Proof of Proposition 1. We only show that if \sqrt{f} belongs to $\mathfrak{B}_{p,\infty}^\alpha(R)$, 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

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

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

$$(4) \quad (F_k(f))^{p/2} \leq 2 \sum_{\substack{j_1 \geq -1 \\ j_2 \geq j_1}} (F_{k,j_1,j_2})^{p/2}.$$

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

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

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

$$(5) \quad |K_{j,j',k'}| \leq c_1 \left[1 + 2^{j-j'} \right]$$

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 $I_{j,k,j_1,k_1,j_2,k_2} \leq c_2 r_{j_1,j_2}$ 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,

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

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

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

A new application of Cauchy-Schwarz leads to

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

We now use (5) 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). \square

3.2. Proof of Theorem 2. Our proof relies on the result below that is due to [Bir06a] (see his Theorem 6 and Proposition 8).

Proposition 5. *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 an 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 $j \geq -1$ and $k \in \mathbb{Z}$,

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

and consider the (random) set

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

Let us order the sample $X'_{(1)} < X'_{(2)} < \dots < X'_{(n)}$ and define the smallest integer $\tilde{J} \geq 0$ satisfying

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

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

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

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

The dimension of this linear space is not larger than

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

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

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

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

$$\sum_{\mathbf{K} \in \tilde{\mathcal{K}}} e^{-\Delta(\mathbf{K})} \leq 1.$$

We apply Proposition 5 conditionally to the independent sample X'_1, \dots, X'_n and take the expectation of the result. By cleaning it a little, we get

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

where c_1 is universal, where $\log_+(x) = \log(e+x)$, and where $0 \times \log_+(|\tilde{\mathbb{Z}}_j|/0) = 0$. To simplify the notations, we set in the sequel $\beta_{j,k} = \beta_{j,k}(\sqrt{f})$. We deduce from the above inequality,

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

where

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

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 $\tilde{n}_{j,k} \leq 1$ if $k \notin \tilde{\mathbb{Z}}_j$ or if $j \geq \tilde{J} + 1$. Hence,

$$T \leq 2 \sum_{j=-1}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k}^2 \mathbb{P}[\tilde{n}_{j,k} \leq 1].$$

Set

$$\begin{aligned}
f_{j,k} &= \int f(x) 1_{\text{supp } \psi_{j,k}}(x) dx \\
\tilde{\mathbb{Z}}_j &= \{k \in \mathbb{Z}, f_{j,k} \geq 1/n\}.
\end{aligned}$$

We have,

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

where

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

Define $b_{j,k}$ such that $\beta_{j,k}^2 = 2^{-j/2} |b_{j,k}|$. For all $j \geq 0$, $\|b_{j,\cdot}\|_{p/2} \leq R^2 2^{-j(2\alpha+1/2-1/(p/2))}$ if $\sqrt{f} \in \mathcal{B}_{p,\infty}^\alpha(R)$. 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 . \square

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 $\tilde{J} \leq c_1 [1 + \log(1 + \xi)]$, with probability $1 - n\|f\|_q/\xi$. In this inequality, c_1 is a term only depending on q and ψ . We deduce,

$$\begin{aligned} \mathbb{E}[\tilde{J}] &\leq c_2 \left[1 + \int_0^\infty \mathbb{P} \left(\tilde{J} \geq c_1 [1 + \log(1 + \xi)] \right) (1 + \xi)^{-1} d\xi \right] \\ &\leq c_3 \left[1 + c_1 n \|f\|_q \int_{\max\{n\|f\|_q, 1\}}^\infty \xi^{-1} (1 + \xi)^{-1} d\xi + c_1 \int_0^{\max\{n\|f\|_q, 1\}} (1 + \xi)^{-1} d\xi \right] \\ &\leq c_4 [1 + \log(1 + n\|f\|_q)]. \end{aligned}$$

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 $\sum_{j=-1}^\infty 2^{(j/2)(1-1/q)} \|\beta_{j,\cdot}\|_{2q}$ is finite, $\sqrt{f} \in \mathbb{L}^{2q}(\mathbb{R})$, and

$$\begin{aligned} \|f\|_q^{1/2} &= \left\| \sqrt{f} \right\|_{2q} \\ (6) \quad &\leq c_5 \sum_{j=-1}^\infty 2^{(j/2)(1-1/q)} \|\beta_{j,\cdot}\|_{2q}. \end{aligned}$$

Note that $|\beta_{j,k}| \leq c_6 \sqrt{f_{j,k}} \leq c_6$ as f is a density and hence $\|\beta_{j,\cdot}\|_{2q}^{2q} \leq c_7 \sum_{k \in \mathbb{Z}} f_{j,k} \leq c_8$. We moreover have when p is finite and $j \geq 0$,

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

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

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

In both cases,

$$(7) \quad \sum_{j=0}^{\infty} 2^{(j/2)(1-1/q)} \|\beta_{j,\cdot}\|_{2q} < c_{12} R^r,$$

where $r = p/(2q)$ if p is finite, and $r = 1 - 1/q$ if $p = \infty$. We conclude by (6). \square

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

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

It follows from Lemma 23 of [Sar23] that

$$(8) \quad \mathbb{E} \left[|\tilde{\mathbb{Z}}_j| \right] \leq c_1 M n^{\theta} 2^{j(1-\theta)}.$$

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

$$\sum_{k \in \tilde{\mathbb{Z}}_j} \beta_{j,k}^2 \leq c_2 \|\beta_{j,\cdot}\|_{p,\infty}^2 |\tilde{\mathbb{Z}}_j|^{1-2/p}.$$

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

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

This last inequality is also true when $p = 2$ and $\sqrt{f} \in \mathcal{B}_{p,\infty}^{\alpha}(R)$. We deduce from (8) 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. \square

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

$$\tilde{K}_j = \left\{ k \in \tilde{\mathbb{Z}}_j, \beta_{j,k}^2 \geq 1/n \right\}$$

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

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

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

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

Therefore,

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

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

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

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

3.3. Proof of Theorem 3. 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 $g_0 \in \mathfrak{B}_{p,\infty}^\alpha(R_{g_0})$ 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 $(-1, 0)$, bounded by 1, and such that $\sqrt{\zeta} \in \mathfrak{B}_{p,\infty}^\alpha(R/\max\{2^{1/p}, 2\})$. We put

$$q = \int (g(x) + h_\delta(x))^2 dx,$$

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*

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

then, f_δ is a density belonging to $\mathcal{T}_\theta(M)$ such that $\sqrt{f_\delta} \in \mathcal{B}_{p,\infty}^\alpha(R)$. If $b2^{j/2}2^{j_0/p} \leq a_2R$, $\sqrt{f_\delta} \in \mathfrak{B}_{p,\infty}^\alpha(R)$. For all $|x| \geq 1$,

$$|x|f_\delta^\theta(x) \leq a_5M,$$

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

$$h^2(f_\delta, f_{\delta'}) = a_6b^2\Delta(\delta, \delta').$$

The terms $a_1, a_2, a_3, a_4, a_5, a_6$ 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

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

It follows from Assouad's lemma – see [Bir06b] – that if $b^2 = 1/(2a_6n)$,

$$\mathcal{R}(\mathcal{F}'_{\alpha,p,\theta}(R, M)) \geq c_0n^{-1}2^{j_0+j},$$

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

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

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

$$2^{j_0} \leq c_1Mn^\theta 2^{-j\theta}.$$

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 $2^{j_0} \leq c_2Mn^\theta$. All the conditions of the lemma are met, hence the result. \square

Proof of Lemma 4. First, observe that

$$q \leq 2 \left[\int g^2(x) dx + \int h_\delta^2(x) dx \right]$$

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

We have $\text{supp } h_\delta \subset [2^{\ell-j}(2\underline{k} + 1), 2^{\ell-j}(4\bar{k} + 2\underline{k} + 1)] \subset [2^{j_0-1}, 2^{j_0}]$, $\text{supp } g \subset [0, 2^{j_0+1}]$ 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

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

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

Note that ζ belongs to $\mathcal{T}_\theta(M/2)$ if M is large enough. Besides, $h_\delta^2 \in \mathcal{T}_\theta(\|h_\delta\|_\infty^2(2^{j_0+1} + 1))$, see Lemma 2.1 of [CL20]. A similar result holds true for g_δ^2 and hence $f_\delta \in \mathcal{T}_\theta(M)$ if a_4 is small enough.

As to the Hellinger distance, we have

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

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

Finally, for all $|x| \geq 1$,

$$\begin{aligned} f_\delta(x) &\leq 2 (\|g\|_\infty^2 + \|h_\delta\|_\infty^2) \\ &\leq c_2 b^2 2^j, \end{aligned}$$

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

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

□

3.4. Sketch of the proof of Proposition 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,

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

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

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

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 $\mathfrak{B}_{p,\infty}^\alpha(\varepsilon)$.*

We deduce that $\sqrt{f_\delta}$ lies in $\mathfrak{B}_{p,\infty}^\alpha(\mathcal{R})$ 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. □

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.
- [BC21] Elodie Brunel and Fabienne Comte. Hazard regression with noncompactly supported bases. *Canadian Journal of Statistics*, 49(4):1273–1297, 2021.
- [Bir06a] Lucien Birgé. Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Annales de l’IHP Probabilités et statistiques*, 42(3):273–325, 2006.
- [Bir06b] Lucien Birgé. Statistical estimation with model selection. *Indagationes Mathematicae*, 17(4):497–537, 2006.
- [BM98] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.
- [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.
- [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.
- [CL20] Kaikai Cao and Youming Liu. Uncompactly supported density estimation with l^1 risk. *Communications on Pure & Applied Analysis*, 19(8), 2020.
- [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 (to appear)*, 2023.
- [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.
- [DJ97] Bernard Delyon and Anatoli Juditsky. On the computation of wavelet coefficients. *Journal of Approximation Theory*, 88(1):47–79, 1997.
- [DJKP96] David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2):508–539, 1996.
- [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.
- [GL11] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.
- [GL14] Alexander Goldenshluger and Oleg Lepski. On adaptive minimax density estimation on \mathbb{R}^d . *Probability Theory and Related Fields*, 159(3):479–543, 2014.
- [JLL04] Anatoli Juditsky and Sophie Lambert-Lacroix. On minimax density estimation on \mathbb{R} . *Bernoulli*, 10(2):187–220, 2004.
- [KS16] Arlene KH Kim and Richard J Samworth. Global rates of convergence in log-concave density estimation. *The Annals of statistics*, 44(6):2756–2779, 2016.
- [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.

- [LW19] Oleg Lepski and Thomas Willer. Oracle inequalities and adaptive estimation in the convolution structure density model. *The Annals of Statistics*, 47(1):233–287, 2019.
- [Mas07] Pascal Massart. *Concentration inequalities and model selection: Ecole d’Eté de Probabilités de Saint-Flour XXXIII-2003*. Springer, 2007.
- [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.
- [RBRTM11] Patricia Reynaud-Bouret, Vincent Rivoirard, and Christine Tuleau-Malot. Adaptive density estimation: a curse of support? *Journal of Statistical Planning and Inference*, 141(1):115–139, 2011.
- [Sar21] Mathieu Sart. Minimax bounds for besov classes in density estimation. *Electronic Journal of Statistics*, 15(1):3184–3216, 2021.
- [Sar23] Mathieu Sart. Non linear wavelet density estimation on the real line. *Available on Hal*, 2023.

UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE, CNRS, INSTITUT CAMILLE JORDAN UMR 5208, F-42023, SAINT-ÉTIENNE, FRANCE.

Email address: `mathieu.sart@univ-st-etienne.fr`