Minimax bounds for Besov classes in density estimation

Mathieu Sart

Univ Lyon, Université Jean Monnet Saint-Étienne, CNRS UMR 5208, Institut Camille Jordan, F-42023 Saint-Etienne, France
e-mail: mathieu.sart@univ-st-etienne.fr

Abstract: We study the problem of density estimation on $[0,1]$ under $\mathbb{L}^p$ norm. We carry out a new piecewise polynomial estimator and prove that it is simultaneously (near)-minimax over a very wide range of Besov classes $\mathcal{B}_{\pi,\infty}^{\alpha}(R)$. In particular, we may deal with unbounded densities and shed light on the minimax rates of convergence when $\pi < p$ and $\alpha \in (1/\pi - 1/p, 1/\pi]$.

Keywords and phrases: Density estimation, minimax risk, Besov spaces.

Contents

1. Introduction

We consider n independent and identically distributed random variables $X_1, \ldots, X_n$ defined on an abstract probability space $(\Omega, \mathscr{E}, \mathbb{P})$. We suppose that X_i admits a density f with respect to the Lebesgue measure on $[0,1]$ and are interested in estimating f from the observations $X_1, \ldots, X_n$.

We consider $p \geq 1$, a subset $\mathcal{F}$ of the linear space $(\mathbb{L}^p([0,1]), \|\cdot\|_p)$, and define the celebrated minimax risk

$$\mathcal{R}_p(\mathcal{F}) = \inf_{\hat{f}} \mathcal{R}_p(\hat{f}) \quad \text{where} \quad \mathcal{R}_p(\hat{f}) = \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[d_p^p(f, \hat{f}) \right]^{1/p},$$

where d_p denotes the distance in the above space, and where the infimum is taken over all estimators $\hat{f}$ with values in $\mathcal{F}$. The minimax rate of convergence, that is the rate at which $\mathcal{R}_p(\mathcal{F})$ converges to 0 (if it converges), is the best possible for procedures based solely on assumptions modelled by $\mathcal{F}$. This rate can therefore be seen as a benchmark for statistical procedures.

In the present paper, our aim is to define a (near) minimax estimator under smoothness constraints. We will therefore pay special attention to bounded subsets $\mathcal{F}$ of Besov spaces $\mathcal{B}_{\pi,\infty}^\alpha$. The subscript π indicates here in which (quasi) norm the regularity α is measured. The smaller π is, the larger the class is, and the more difficult the estimation problem is. It is sometimes said that these Besov spaces allow taking into account spatially inhomogeneous smoothness. The minimax results rely heavily on π, p, α as described below.

The simplest case is $\pi \geq p$. The minimax rate of convergence is then $n^{-\alpha/(1+2\alpha)}$, and this rate can even be achieved by a linear estimator, see [DJKP96]. In the literature, the case $\pi < p$ is often associated with the restriction $\alpha > 1/\pi$. Despite this, linear estimators are no longer rate optimal and more sophisticated procedures need to be considered. We may cite for instance [DJKP96, DJ96] for wavelet thresholding estimators, [BM97] for penalized minimum contrast estimators and [Bir06] for T-estimators. The results are as follows: the previous rate $n^{-\alpha/(1+2\alpha)}$ still applies when $\alpha > (p - \pi)/(2\pi)$, but becomes $(\log n/n)^{(\alpha-1/\pi+1/p)/(1+2\alpha-2/\pi)}$ when $\alpha < (p - \pi)/(2\pi)$. At the boundary $\alpha = (p - \pi)/(2\pi)$, additional logarithmic factors may appear.

The situation appears to be more complicated to deal with when $\pi < p$ and $\alpha \in (1/\pi - 1/p, 1/\pi]$. The condition $\alpha > 1/\pi - 1/p$ ensures that $\mathcal{B}_{\pi,\infty}^\alpha \subset \mathbb{L}^p([0,1])$ and allows the use of a $\mathbb{L}^p$ loss. However, $\mathcal{B}_{\pi,\infty}^\alpha$ is not included in $\mathbb{L}^\infty([0,1])$ as $\alpha \leq 1/\pi$. A key paper for understanding the importance of the condition $\|f\|_\infty < +\infty$ in estimation procedures is that of [Bir14]. In this paper, [Bir14] showed a lower minimax bound. It reveals that the usual rate $n^{-\alpha/(1+2\alpha)}$ cannot be true over the whole range $(1/\pi - 1/p, 1/\pi]$ when $p = 2$. However, this rate is the right one when the density is bounded.

This problem of determining the optimal rates when the target function is smooth but not bounded has already been studied in the literature in other statistical settings such as, for instance, the Gaussian white noise model and the regression model with random design. We refer here to [Bar02, Bir04, Lep15] and the references therein. In the latter model, the minimax rate is known (up

to log factors) for the $\mathbb{L}^2$ loss and is $n^{-\min\{\alpha/(1+2\alpha),\alpha-1/\pi+1/2\}}$ when $\pi \in [1, 2)$ and $\alpha \in (1/\pi - 1/2, 1/\pi]$. We observe therefore a possible deterioration of the rates when the target is not bounded. This is similar to density estimation but contrasts with the white noise model where the boundedness assumption can be removed without changing the exponent in the optimal rates: it is $\alpha/(1 + 2\alpha)$ regardless of $\pi \in [1, 2)$, $\alpha \in (1/\pi - 1/2, 1/\pi]$ when $p = 2$.

Several procedures for unbounded densities have been proposed in the statistical literature. We may cite for instance the wavelet thresholding procedures of [BTWB10, RBRTM11]. They lead to nice oracle inequalities for the $\mathbb{L}^2$ loss without condition on the supremum norm of f. Also worth mentioning are the general procedures of [LW19, Bir06, Bir14]. The first is based on a pointwise selection scheme. It leads to local risk bounds, which can be integrated to become global. The other two are based on robust tests and allow for more general assumptions than smoothness. Despite this, the minimax rate does not seem to have attracted much attention in the literature when $\alpha \in (1/\pi - 1/p, 1/\pi]$ and $\pi < p$. In the present paper, we define a new estimator and show that it achieves the rate $(\log n)^\beta n^{-\min\{\alpha/(1+2\alpha),(\alpha-1/\pi+1/p)/(1+\alpha-1/\pi)\}}$ for some β independent of n. Moreover, we show that this rate is optimal, up to logarithmic factors. In particular, when $\pi < p - 2$, the minimax rate is never $n^{-\alpha/(1+2\alpha)}$. When $\pi > p - 2$, there is an elbow effect, the exponent being $(\alpha - 1/\pi + 1/p)/(1 + \alpha - 1/\pi)$ for "small values of α" and $\alpha/(1 + 2\alpha)$ for "large values".

We restrict our study to the estimation of the density on the unit interval $[0, 1]$ but other estimation domains have also been considered in the literature. For results concerning the estimation on the real line under smoothness constraints, we refer to [BH78, JLL04, Efr08, RBRTM11, GL11, GL14]. However, it should be noted that the minimax rates on Besov classes can be very different, depending on whether the estimation domain is compact or not. As far as we know, the rates on the real line are not fully known. In higher dimension, the whole point is to reduce the curse of dimensionality. A solution is to allow the smoothness of f to vary with the direction. For more informations on these issues of anisotropy, we refer to [Kle09, GL11, Aka12, GL14].

Our estimation strategy is based on projection estimators and on a new estimator selection rule. This procedure may be thought of as a mix between a Lespki-type procedure [Lep92] and the one of [Sar14]. It leads to a piecewise polynomial estimator of degree r that is (near) minimax and adaptive over the full scale of Besov classes $\pi \in (0, p)$ and $\alpha \in (1/\pi - 1/p, r + 1)$. In other terms, it achieves the rates given above, up to logarithmic factors, without the prior knowledge of α and π. Our estimator has also computational properties. We may build it in polynomial time when $p > 1$, and more precisely in around $\mathcal{O}(n^{p/(p-1)} \log n)$ operations. It is also possible to make our estimator computable in (nearly) linear time when the classical condition $\alpha > 1/\pi$ is met.

The paper is organized as follows. We make explicit the minimax rates of convergence in the next section. Section 3 is devoted to the construction of our estimator and contains intermediate results such as an oracle inequality and a result on the approximation of the elements of $\mathcal{B}_{\pi,\infty}^\alpha$ by piecewise polynomial functions. The proofs are postponed to Section 4.

2. Minimax results

2.1. Besov classes

We recall here the definition of Besov spaces to introduce our notations, and refer to [DL93] for more details.

We consider $\alpha, \pi > 0$, $\pi' \in (0, +\infty]$ and the smallest integer r larger than α. For any map f on $[0, 1]$, $h > 0$ and $t \in [0, 1 - rh]$, we set

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

Let then $\omega_p(f, x)$ be the modulus of smoothness

$$\omega_\pi(f, x) = \sup_{0 < h \leq x} \left[\int_{[0, 1-rh]} |\Delta_h^r f(t)|^\pi \, \mathrm{d}t \right]^{1/\pi},$$

and

$$|f|_{\mathcal{B}^\alpha_{\pi,\pi'}} = \begin{cases} \left(\int_0^\infty [x^{-\alpha} \omega_\pi(f, x)]^{\pi'} \frac{\mathrm{d}x}{x} \right)^{1/\pi'} & \text{if } \pi' < +\infty \\ \sup_{x > 0} x^{-\alpha} \omega_\pi(f, x) & \text{if } \pi' = +\infty. \end{cases}$$

When $\pi, \pi' \geq 1$, $|\cdot|_{\mathcal{B}^\alpha_{\pi,\pi'}}$ is a semi-norm, and a quasi semi-norm otherwise.

We define the Besov space $\mathcal{B}^\alpha_{\pi,\pi'}$ as the set of functions f on $[0, 1]$ with finite (quasi) semi-norm $|f|_{\mathcal{B}^\alpha_{\pi,\pi'}}$. We investigate in the next section the problem of minimax estimation over the class

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

2.2. Minimax results

We consider $\pi \in (0, p)$,

$$\bar{\alpha} = 1/(2\pi) - 1/p + \sqrt{1/\pi - 1/p + (1/(2\pi) - 1/p)^2}, \tag{1}$$

and set for $\alpha \in (1/\pi - 1/p, 1/\pi]$,

$$\psi = \min \left\{ (\alpha - 1/\pi + 1/p)/(1 + \alpha - 1/\pi), \alpha/(1 + 2\alpha) \right\},$$
$$= \begin{cases} (\alpha - 1/\pi + 1/p)/(1 + \alpha - 1/\pi) & \text{if } \alpha \in (1/\pi - 1/p, \bar{\alpha}], \\ \alpha/(1 + 2\alpha) & \text{if } \alpha \in (\bar{\alpha}, 1/\pi]. \end{cases}$$

Note that $\bar{\alpha}$ may be larger than $1/\pi$, in which case $(\bar{\alpha}, 1/\pi] = \emptyset$. In Section 3, we carry out an (adaptive) estimation procedure leading to the following result:

Theorem 1. *Let $p \in [1, +\infty)$, $\pi \in (0, p)$, $\alpha \in (1/\pi - 1/p, 1/\pi]$ and $R > 0$. Then, for n large enough,*

$$\mathcal{R}_p(\mathcal{B}_{\pi,\infty}^\alpha(R)) \leq c_1 (\log n)^\beta n^{-\psi}, \tag{2}$$

where β only depends on p, π, α and where c_1 only depends on R, p, π, α.

Up to log factors, the rate that is achieved when $\alpha > \bar{\alpha}$ cannot be improved as it corresponds to the minimax rate for Hölder semi-balls. In the literature, [Bir14] obtained a lower bound that matches with this upper bound (to within logarithmic factors) when $p = 2$ and $\alpha \leq \bar{\alpha}$. Slight modifications of his proof lead to:

Proposition 2. *Let $p \in [1, +\infty)$, $\pi \in (0, p)$ and $R > 0$. Then, if $\alpha \in (1/\pi - 1/p, 1/\pi]$, for n large enough,*

$$\mathcal{R}_p(\mathcal{B}_{\pi,\infty}^\alpha(R)) \geq c_2 n^{-\psi}.$$

Moreover, if $\alpha \leq 1/\pi - 1/p$ and $p > 1$, for n large enough,

$$\mathcal{R}_p\left(\left\{f \in \mathcal{B}_{\pi,\infty}^\alpha(R), \|f\|_p \leq 1 + R\right\}\right) \geq c_2.$$

In these inequalities, c_2 depends on R, p, π, α only.

Therefore, the optimal rate of convergence is $n^{-\psi}$, up to possible logarithmic factors. When α is smaller than $1/\pi - 1/p$, $\mathcal{B}_{\pi,\infty}^\alpha(R)$ is not included in $\mathbb{L}^p([0,1])$ and minimax results for $\mathcal{B}_{\pi,\infty}^\alpha(R)$ are meaningless. The second point of the proposition says that the minimax rate does not tend to zero even under an additional condition on the $\mathbb{L}^p$ norm when $p > 1$.

When $p = 1$, $\bar{\alpha} = 1/\pi - 1$, and we recover that $\psi = \alpha/(1 + 2\alpha)$ on the whole range $(1/\pi - 1, 1/\pi]$. This rate is actually free of logarithmic factors, see [Bir06]. The formula for ψ when $p > 1$, $\alpha < 1/\pi$ does not seem to appear in the literature. We suspect, however, that already existing estimators are (near) minimax. Of particular note are the wavelet estimators of [BTWB10, RBRTM11] and the selection rule of [LW19]. Unfortunately, the authors did not explicitly address this issue. But there is a likeness between the $\mathbb{L}^2$ oracle inequalities of [BTWB10, RBRTM11] and ours (see (8) or (16) below). What is missing is the computation of the risk from these inequalities. As for the very general paper of [LW19], the authors explain in their Theorem 2 how to control the maximal risk of their estimator over a collection $\mathbb{F}$. The functions of $\mathbb{F}$ need not to be bounded. The condition relates rather to the inclusion of $\mathbb{F}$ into a $\mathbb{L}^q$ space or more generally a Lorentz space. The proof of Theorem 1 also turns out to be based on a similar embedding. For each of these references, a logarithmic factor in the convergence rates is to be expected. In [LW19], this factor may be due to a too local approach to the estimation problem. In our paper, and in [BTWB10, RBRTM11], this factor appears in the penalties/thresholds. These could therefore be a little too large. Improving them would require controlling the underlying empirical process more accurately. The question of whether the optimal rates involve
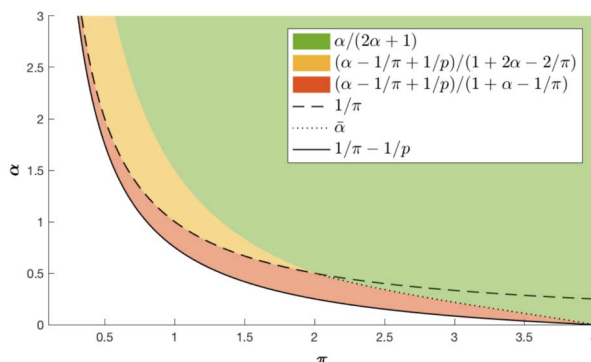
FIG 1. *Minimax rates for Besov classes and compact estimation domains*

logarithmic factors is currently open. The negative side of Theorem 1 is that it implies the suboptimality of the T-estimator of [Bir14] (see his Section 4.4.3).

When $\alpha > 1/\pi$, the minimax rate of convergence is, up to logarithmic factors, $n^{-\psi}$ where

$$\psi = \begin{cases} \alpha/(1+2\alpha) & \text{if } \alpha > (p-\pi)/(2\pi) \\ (\alpha - 1/\pi + 1/p)/(1+2\alpha - 2/\pi) & \text{if } \alpha \leq (p-\pi)/(2\pi). \end{cases}$$

This result has been known for a long time. We present in Figure 1 above a graph to summarize these rates when $p = 4$.

There are therefore three possible formula for the rate on $(1/\pi - 1/p, +\infty)$. The well-known elbow phenomenon refers in literature to the non-differentiability of ψ at $(p - \pi)/(2\pi)$ when $p > 2 + \pi$, i.e. when moving from the orange zone to the green zone. Actually, ψ is not differentiable at $\min\{\bar{\alpha}, 1/\pi\}$ either. Therefore, there is always at least one elbow effect on $(1/\pi - 1/p, +\infty)$ and sometimes two.

3. Estimation procedure

For all subset $\mathcal{F} \subset \mathbb{L}^p([0,1])$, and $f \in \mathbb{L}^p([0,1])$, we set $d_p(f, \mathcal{F}) = \inf_{g \in \mathcal{F}} d_p(f, g)$. The notation $|I|$ stands for both the cardinal of a finite set I, and the length of an interval I. We set $\mathbb{N}^\star = \mathbb{N} \setminus \{0\}$. The letters $c, c', C, C', \ldots$ denote quantities that may change from line to line.

3.1. Collection of partitions

We introduce here tree-structured partitions m of $[0,1]$ derived from the recursive algorithm of [DY90] that are frequently encountered when estimating smooth functions by histograms or more generally by piecewise polynomial estimators (see [BB09, Aka12, Sar14] among other references).

Consider a partition m of $[0, 1]$. We may refine m by dividing some intervals I of m into two equal parts. The collection of all partitions that can be constructed from m by this way is denoted by $\mathcal{M}(m)$. We then define collections $\mathcal{M}_\ell$ of partitions by induction by setting $\mathcal{M}_0 = \{\{[0, 1]\}\}$ and

$$\mathcal{M}_\ell = \mathcal{M}_{\ell-1} \cup \bigcup_{m \in \mathcal{M}_{\ell-1}} \mathcal{M}(m).$$

The collection $\mathcal{M}_\ell$ is therefore composed of partitions of $[0, 1]$ into intervals with endpoints of the form $k/2^\ell$. It does not only contain the regular partitions of $[0, 1]$ of size 2^k, $k \leq \ell$ but also partitions that are very thin locally, and wider elsewhere.

We moreover define the collection $\mathcal{M}_\infty = \bigcup_{\ell \geq 0} \mathcal{M}_\ell$ of partitions m that can be constructed by this algorithm in a finite number of steps.

3.2. *Projection estimators*

We estimate the density f by means of piecewise polynomial estimators. They are defined as projection estimators as described below.

Consider an integer r, a collection m of disjoint intervals, and the space

$$\mathcal{P}_r(m) = \left\{ \sum_{I \in m} P_I \mathbb{1}_I, \text{ where } P_I \text{ is a polynomial function of degree at most } r \right\}$$

of piecewise polynomial functions. Let $(\varphi_{I,j})_{I \in m, j \in \{0,\ldots,r\}}$ be the orthonormal basis of $\mathcal{P}_r(m)$ defined from the Legendre polynomials

$$Q_j(x) = \frac{1}{2^j} \sum_{k=0}^{j} \binom{j}{k}^2 (x-1)^{j-k}(x+1)^k,$$

by the formula

$$\varphi_{I,j}(x) = \sqrt{\frac{2j+1}{b-a}} Q_j\left(2\frac{x-a}{b-a} - 1\right) \mathbb{1}_I(x),$$

where $a < b$ denote the extremities of I. We define the projection estimator

$$\hat{f}_m = \sum_{\substack{I \in m \\ j \in \{0,\ldots,r\}}} \hat{\beta}_{I,j} \varphi_{I,j} \quad \text{with } \hat{\beta}_{I,j} = \frac{1}{n} \sum_{i=1}^{n} \varphi_{I,j}(X_i),$$

and omit the dependency in r to lighten the notations.

Since their introduction by [Cen62], projection estimators have received considerable interest in statistical estimation. They are at the heart of many model selection procedures, see [BM98, BBM99, Mas07] for key references. For pedagogical reasons, we present below a risk bound for this estimator when $p = 2$:

Proposition 3. *Let m be a (finite) partition of $[0, 1]$ into intervals of positive lengths. Then, if $f \in \mathbb{L}^2([0, 1])$,*

$$\mathbb{E}_f\left[d_2^2(f, \hat{f}_m)\right] \leq d_2^2(f, \mathcal{P}_r(m))$$

$$+ \inf_{q>1} \|f\|_q \frac{(r+1)^2 \left(\sum_{I \in m} |I|^{1-q/(q-1)}\right)^{(q-1)/q}}{n}. \quad (3)$$

The proof of this result is merely based on Hölder's inequality, see Section 4.2. It highlights what will allow us to estimate unbounded densities f of $\mathcal{B}_{\pi,\infty}^\alpha$.

When m is a regular partition, that is when the intervals $I \in m$ are of the same size, the term in the infimum is $\|f\|_q (r+1)^2 |m|/n$. In particular, by letting q tend to 1,

$$\mathbb{E}_f\left[d_2^2(f, \hat{f}_m)\right] \leq d_2^2(f, \mathcal{P}_r(m)) + \frac{(r+1)^2|m|}{n}. \quad (4)$$

For more general partitions m, and $q \to +\infty$,

$$\mathbb{E}_f\left[d_2^2(f, \hat{f}_m)\right] \leq d_2^2(f, \mathcal{P}_r(m)) + \|f\|_\infty \frac{(r+1)^2|m|}{n}. \quad (5)$$

These two inequalities apply either for regular partitions, without assumption on $\|f\|_\infty$, or for any partition when $\|f\|_\infty$ is finite. More generally, the risk bounds of projection estimators sometimes involve the supremum norm of f, and sometimes not, depending on whether or not the model checks a condition linking the $\mathbb{L}^2$ and $\mathbb{L}^\infty$ structure of the model (equation (7.16) in [Mas07]). For more details on this phenomenon, we refer to [Bir14].

It is worth mentioning that (4) and (5) are not suitable for estimating densities f in $\mathcal{B}_{\pi,\infty}^\alpha(R)$ when $\alpha < 1/\pi$ and $\pi < 2$. Indeed, such densities may not be bounded and may be poorly approached by piecewise polynomial functions based on regular grids. Making the bias term $d_2^2(f, \mathcal{P}_r(m))$ small requires working with partitions m adapted to the spatial inhomogeneity of f. In some sense, (3) fills the gap between (4) and (5). We can deal with irregular partitions and unbounded densities. The condition on the supremum norm is replaced by a condition on the $\mathbb{L}^q$ norm. The larger q is, the smaller $(\sum_{I \in m} |I|^{1-q/(q-1)})^{(q-1)/q}$ is and vice versa.

Two issues remain to be addressed. First, the risk of the estimator depends on the choice of m. We need to explain how to define a partition m realizing a good compromise between the two terms in (3). Second, we need to compute the risk of this estimator when the target f lies in a Besov class. In the section below, we start by answering the first point.

3.3. Selection rule

For all partitions m, m' and interval J, we set

$$m \vee J = \{I \cap J, I \in m\}$$

$$m \vee m' = \bigcup_{J \in m'} m \vee J = \{I \cap J, \ I \in m, \ J \in m'\}.$$

For all interval I, we define

$$\widehat{v}^p(I) = \left\| \sum_{j=0}^{r} |\varphi_{I,j}| \sqrt{\frac{1}{n} \sum_{i=1}^{n} \varphi_{I,j}^2(X_i)} \right\|_p^p,$$

$$\widehat{w}^p(I) = \frac{1}{|I|^{p/2}} \left\| \sum_{j=0}^{r} \sqrt{(2j+1)} |\varphi_{I,j}| \right\|_p^p.$$

We set for any collection m of intervals and $\xi > 0$,

$$\widehat{v}_\xi^p(m) = \sum_{I \in m} ((\xi + \log_2(1/|I|))/n)^{p/2} \, \widehat{v}^p(I)$$

$$\widehat{w}_\xi^p(m) = \sum_{I \in m} ((\xi + \log_2(1/|I|))/n)^p \, \widehat{w}^p(I),$$

where $\log_2$ denotes the logarithm base 2. We consider κ_1, κ_2, and

$$\widehat{\operatorname{pen}}_\xi^p(m) = \kappa_1 \widehat{v}_\xi^p(m) + \kappa_2 \widehat{w}_\xi^p(m).$$

We consider some $\ell \in \mathbb{N}^\star \cup \{\infty\}$ and define for $m \in \mathcal{M}_\ell$,

$$\gamma(m) = \sum_{J \in m} \sup_{m' \in \mathcal{M}_\ell} \left\{ d_p^p(\hat{f}_m \mathbb{1}_J, \hat{f}_{m \vee m'} \mathbb{1}_J) - \widehat{\operatorname{pen}}_\xi^p(m' \vee J) \right\}. \qquad (6)$$

We then define $\hat{m}$ as any partition of $\mathcal{M}_\ell$ satisfying

$$\gamma(\hat{m}) + (1 + 2^{1-p})\widehat{\operatorname{pen}}_\xi^p(\hat{m}) \leq \inf_{m \in \mathcal{M}_\ell} \left\{ \gamma(m) + (1 + 2^{1-p})\widehat{\operatorname{pen}}_\xi^p(m) \right\}$$
$$+ 1/n^{2p}, \qquad (7)$$

and shrink the resulting estimator $\hat{f} = \min\{1, n\|\hat{f}_{\hat{m}}\|_p^{-1}\}\hat{f}_{\hat{m}}$.

This criterion can be seen as a Lespki-type procedure [Lep92] (see also [BB09]) modified as in [Sar14] in order make possible the construction of the estimator by dynamic programming algorithms. We leave these computational aspects aside for the moment to focus on the theoretical properties of $\hat{f}$. They will be discussed in Section 3.7.

3.4. Risk bound

We prove in Section 4.3:

Theorem 4. *Let $r \in \mathbb{N}$, $\ell \in \mathbb{N}^\star \cup \{\infty\}$ and $p \in [1, +\infty)$. Then, there exists κ depending only on p such that if $\kappa_1, \kappa_2 \geq \kappa$, if $\xi = 2p \log n + \log(r+1)$, and if $f \in \mathbb{L}^p([0,1])$, the preceding estimator satisfies for $n \geq \|f\|_p$,*

$$\mathbb{E}_f \left[d_p^p(f, \hat{f}) \right]^{1/p} \leq C \inf_{m \in \mathcal{M}_\ell} \left\{ d_p(f, \mathcal{P}_r(m)) + \inf_{q \in [p, +\infty]} \sqrt{M_q(m) v_q(m) \frac{\ell_m + \log n}{n}} \right.$$

$$\left. + w(m) \frac{\ell_m + \log n}{n} \right\}, \tag{8}$$

where ℓ_m is the smallest value of $\mathbb{N}^\star$ such that $m \in \mathcal{M}_{\ell_m}$,

$$M_q(m) = \min\{\|f\|_q, \ell_m^{1/q} \|f\|_{q/(1+q/(p\ell_m))}\},$$

$$v_q(m) = \left[\sum_{I \in m} |I|^{1-pq/(2q-p)} \right]^{(2q-p)/(pq)},$$

$$w(m) = \left(\sum_{I \in m} |I|^{1-p} \right)^{1/p}.$$

Moreover, C only depends on r, p, κ_1 and κ_2.

Thereby, the risk of the estimator is controlled by the best possible compromise between a bias term $d_p(f, \mathcal{P}_r(m))$ and an estimation term

$$\inf_{q \in [p, +\infty]} \sqrt{M_q(m) v_q(m) \ (\ell_m + \log n)/n} + w(m)(\ell_m + \log n)/n,$$

up to a multiplicative factor. The partitions m can vary freely in $\mathcal{M}_\ell$, and can be very thin locally and thus well adapted to the target density f. Note that this result improves with ℓ (we can even set $\ell = +\infty$ in theory).

Our theorem applies for any density $f \in \mathbb{L}^p([0,1])$, and in particular for unbounded densities. We may always set $q = p$ in the infimum but playing with larger values of q will be worthwhile as $v_q(m)$ is non-increasing with q. The value of $q = +\infty$ is allowed using standard algebra in $\mathbb{R} \cup \{\infty\}$ to deal with bounded f. In that case $M_q(m) = \|f\|_{p\ell_m} \leq \|f\|_\infty$ and

$$v_q(m) = \left[\sum_{I \in m} |I|^{1-p/2} \right]^{2/p}.$$

There are two main differences between the present estimation term and that of Proposition 3. We have here an additional term $\ell_m + \log n$ which will be of the order of $\log n$ for best partitions m and Besov-type regularity constraints. This implies that our results will be slightly sub-optimal in some cases. But this term cannot be avoided in general as (8) leads to the optimal rates in the sparse zone (orange zone of Figure 1), see Section 3.6. The second difference lies in the presence of the additional term $w(m)(\ell_m + \log n)/n$. For adequate partitions m, and densities $f \in \mathcal{B}_{\pi, \infty}^\alpha(R)$, this last term will be, in the worst case, of the same order of magnitude as $\sqrt{M_q(m) v_q(m)(\ell_m + \log n)/n}$.

3.5. Approximation

We need to bound (8) from above when f lies in a Besov class to get our maximal risk bound. This problem falls under the theory of approximation and is treated below.

We know from the literature that functions $f \in \mathcal{B}^\alpha_{\pi,\infty}(R)$ can be well approximated by piecewise polynomial functions defined over a moderate number of intervals. We refer to [DY90, BM00, Aka12]. Let us observe that

$$v_q(m) = \left[\sum_{I \in m} |I|^{1-\delta} \right]^{1/\delta} \quad \text{with } \delta = pq/(2q - p). \tag{9}$$

When $\delta \leq 1$, a concavity argument entails $v_q(m) \leq |m|$. Moreover, if $p = 1$, $w(m) = |m|$ and Corollary 3.3 of [DY90] may be used to control the right-hand side of (8). In general, however, $v_q(m)$ and $w(m)$ may be much larger than $|m|$: we do not only need to control the size of m but also the thinness of the intervals. The following result, to be proved in Section 4.4, is tailored to solve this problem.

Proposition 5. *Let* $\delta \geq 1$, $p \in [1, +\infty)$, $\pi \in (0, p)$, $\alpha > 1/\pi - 1/p$, $R > 0$, $f \in \mathcal{B}^\alpha_{\pi,\infty}(R)$, $r > \alpha - 1$ *and* $k \geq 1$.

- *Suppose that* $\alpha > \delta(1/\pi - 1/p)$. *Then, there exist* ℓ *and* $m \in \mathcal{M}_\ell$ *such that*

$$d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k\alpha}, \quad \sum_{I \in m} |I|^{1-\delta} \leq C2^{k\delta},$$

 and

$$\sum_{I \in m} |I|^{1-p} \leq \begin{cases} C2^{kp} & \text{if } \alpha > p/\pi - \delta/p \\ C\ell 2^{kp} & \text{if } \alpha = p/\pi - \delta/p \\ C2^{k\pi(\alpha+\delta/p)+\ell\pi(p/\pi-\delta/p-\alpha)} & \text{if } \alpha < p/\pi - \delta/p. \end{cases}$$

 Moreover,

$$\ell \leq \frac{\alpha + \delta/p}{\delta/p + \alpha - 1/\pi} k.$$

- *Suppose that* $\alpha < \delta(1/\pi - 1/p)$. *Then, there exists a partition* $m \in \mathcal{M}_\ell$ *with* $\ell \leq k$ *such that*

$$d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k(\alpha+1/p-1/\pi)}, \quad \sum_{I \in m} |I|^{1-\delta} \leq C2^{k(\delta-1)},$$

 and if $p > \delta$,

$$\sum_{I \in m} |I|^{1-p} \leq C2^{k(p-1)}.$$

- *Suppose that $\alpha = \delta(1/\pi - 1/p)$. Then, there exists a partition $m \in \mathcal{M}_\ell$ with $\ell \leq k$ such that*

$$d_p(f, \mathcal{P}_r(m)) \leq CRk^{1/p}2^{-k(\alpha+1/p-1/\pi)}, \quad \sum_{I \in m} |I|^{1-\delta} \leq Ck2^{k(\delta-1)},$$

and if $p > \delta$,
$$\sum_{I \in m} |I|^{1-p} \leq C2^{k(p-1)}.$$

In the above inequalities, C depends on α, δ, p, r, π only.

3.6. Minimax bound

We may therefore define partitions m with moderate $v_q(m)$ and $w(m)$ that are well adapted to the variations of f. This makes it possible to bound the infimum in (8) from above. We state:

Theorem 6. *Let $r \geq 0$, $p \in [1, +\infty)$, and $\hat{f}$ be the estimator defined in Section 3.3 with $\kappa_1 = \kappa_2 = \kappa$, $\ell = +\infty$ and ξ given by Theorem 4. For all $\pi \in (0, p)$, $\alpha \in (1/\pi - 1/p, r+1)$ and $R > 0$, the estimator $\hat{f}$ satisfies for n large enough,*

$$\sup_{f \in \mathcal{B}^\alpha_{\pi,\infty}(R)} \mathbb{E}_f \left[d_p^p(f, \hat{f}) \right]^{1/p} \leq C(\log n)^\tau (\log n/n)^\psi, \tag{10}$$

where ψ is defined in Section 2.2 and where τ depends on p, π, α only. If $\alpha \in (\bar{\alpha}, 1/\pi]$, or if $\alpha > 1/\pi$ and $\alpha \neq (p-\pi)/(2\pi)$, then $\tau = 0$. If $\alpha > 1/\pi$ and $\alpha = (p-\pi)/(2\pi)$, $\tau = 1/p$. Moreover, C depends on R, p, π, α only.

Combined with Proposition 2, this theorem says that our estimator $\hat{f}$ is adaptive and near-minimax over a large scale of Besov classes. It implies Theorem 1. Note that the rate is optimal in the orange zone of Figure 1.

The precise value of $\bar{\alpha}$ can be explained by the following reasoning. A function $f \in \mathcal{B}^\alpha_{\pi,\infty}$ "almost belongs" to $\mathbb{L}^q$ where $1/q = 1/\pi - \alpha$. Assuming that this is true, we see from Proposition 5 that the change of rates occurs at $\alpha = \delta(1/\pi - 1/p)$ with $\delta = pq/(2q-p)$. This leads to the equation

$$\alpha = \frac{1/\pi - 1/p}{2/p - 1/\pi + \alpha} \tag{11}$$

for which $\bar{\alpha}$ is solution. By the way, slight modifications of the proof entail that $\tau = 0$ when $\alpha < 1/\pi$ and $\alpha < \bar{\alpha}$ when the supremum is taken over all the densities of $\mathcal{B}^\alpha_{\pi,\infty}(R)$ whose $\mathbb{L}^q$ norm is uniformly bounded (with $1/q = 1/\pi - \alpha$).

When $\pi \geq p$, the functions in $\mathcal{B}^\alpha_{\pi,\infty}(R)$ can be approached by piecewise polynomial functions based on regular partitions, see Lemma 12 of [BBM99]. More precisely, for every $\alpha > 0$ and $k \geq 1$, there is a regular partition m of size 2^k such that $d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k\alpha}$. For such a partition, $v_q(m) = w(m) = |m|$. By putting these results in (8), we deduce that our estimator achieves the rate $(\log n/n)^{\alpha/(1+2\alpha)}$ over the Besov classes $\mathcal{B}^\alpha_{\pi,\infty}(R)$ for all $\pi \geq p$ and $\alpha \in (0, r+1)$.

3.7. Computational statistics

To compute our estimator, we have to minimize $\gamma(m) + (1 + 2^{1-p})\widehat{\mathrm{pen}}^p_\xi(m)$ on the set of partitions m of $\mathcal{M}_\ell$. It is not advisable, in practice, to solve this optimization problem by an exhaustive search of $\hat{m}$ among all the partitions of $\mathcal{M}_\ell$. This is due to the very large cardinal of $\mathcal{M}_\ell$, even when ℓ is moderate. Actually, the calculation of $\gamma(m)$ itself is an optimization problem that can hardly be solved by a naive approach.

Fortunately, dynamic programming allows here to build the estimator more efficiently. We refer to [Don97, BSR04, AL11, Aka12, Sar14] for some examples of this technique in statistics. In particular, we may slightly adapt the algorithm of [Sar14] to perform the exact computation of $\hat{m}$ in at most $C\ell(2^\ell + n)$ operations (see his Proposition A.1). The term C does not depend on ℓ, n and by operations we mean elementary operations such as additions, multiplications etc., computation of elementary functions or integrals. Naturally, the numerical complexity of the procedure increases with ℓ. Since the theoretical results improve with ℓ, the choice of ℓ is, at first sight, a sort of compromise between the theoretical and computational properties of the estimator.

One may, however, in practice, be interested only in linear (or polynomial) time estimators. This is equivalent to setting ℓ and looking for the classes of functions on which our estimator is (nearly) rate optimal. We show:

Proposition 7. *Let $p \in [1, +\infty)$, $\pi \in (0, p)$, $r \geq 0$, $\alpha \in (1/\pi - 1/p, r + 1)$ and $R > 0$. Let $\mathfrak{c}$ and ℓ be the largest integer such that $2^\ell \leq n^{\mathfrak{c}}$, $\kappa_1 = \kappa_2 = \kappa$ and $\hat{f}$ be the estimator defined in Section 3.3.*

Then, if $p > 1$ and $\mathfrak{c} = p/(p-1)$, or if $\mathfrak{c} = 1$ and $\alpha > 1/\pi$, the estimator satisfies (10) for n large enough.

The numerical complexity of our procedure is of the order of $n^{\mathfrak{c}}(\log n)$. By setting $\mathfrak{c} = 1$, we get the preceding rates under the standard condition $\alpha > 1/\pi$ in nearly linear time. When $p > 1$, we may set $\mathfrak{c} = p/(p-1)$ to get all the rates in polynomial time.

Remark: the minimum length of an interval of a partition $m \in \mathcal{M}_\ell$ is $2^{-\ell}$. In particular, it is not necessary to go below $1/n$ to be (near) optimal when $\alpha > 1/\pi$. We need to cross this threshold to estimate unbounded densities corresponding to smaller values of α.

4. Proofs

4.1. Proof of Proposition 2

Due to the results on Hölder's classes, we only need to show

$$\mathcal{R}_p\left(\mathcal{B}^\alpha_{\pi,\infty}(R)\right) \geq c(1/n)^{(\alpha+1/p-1/\pi)/(1+\alpha-1/\pi)}.$$

This inequality can be shown by slightly adapting the arguments of [Bir14] to the case $p \neq 2$. It is actually attained on a two-point problem $\{f_1, f_2\}$.

Consider $\alpha \in [1/\pi - 1/p, 1/\pi]$, $p > 1$ and a density $\varphi \in \mathcal{B}_{\pi,\infty}^{\alpha}$ with support included in $(0, 1)$. We set $\varphi(x) = 0$ for $|x| > 1$, define

$$\rho = \min\{R, R/\|\varphi\|_p\}/\max\{|\varphi|_{\mathcal{B}_{\pi,\infty}^{\alpha}}, 1\},$$

a such that $a^{1+\alpha-1/\pi} = \rho(4n)^{1/\pi-\alpha}$ and $g(x) = a\varphi(4anx)$. We suppose from now on that n is large enough to ensure that g is compactly supported on $[0, 1/2]$. Moreover, $|g|_{\mathcal{B}_{\pi,\infty}^{\alpha}} = \rho|\varphi|_{\mathcal{B}_{\pi,\infty}^{\alpha}}$ and $\|g\|_p = \varepsilon$ where

$$\varepsilon = \rho^{(1-1/p)/(1+\alpha-1/\pi)}(1/(4n))^{(\alpha+1/p-1/\pi)/(1+\alpha-1/\pi)}\|\varphi\|_p.$$

Define now f_1 for all $x \in [0, 1]$ by

$$f_1(x) = g(x) + 1 - 1/(4n),$$

and set $f_2(x) = f_1(1 - x)$.

Then, f_1 and f_2 are two densities such that $|f_1|_{\mathcal{B}_{\pi,\infty}^{\alpha}} = |f_2|_{\mathcal{B}_{\pi,\infty}^{\alpha}} = |g|_{\mathcal{B}_{\pi,\infty}^{\alpha}} = \rho|\varphi|_{\mathcal{B}_{\pi,\infty}^{\alpha}}$ and

$$\max\{\|f_1\|_p, \|f_2\|_p\} \leq 1 + \|g\|_p \leq 1 + \varepsilon.$$

Since $g(x)$ and $g(1 - x)$ cannot be simultaneously non-zero,

$$\|f_2 - f_1\|_p = 2^{1/p}\|g\|_p = 2^{1/p}\varepsilon.$$

Let h be the Hellinger distance. Then,

$$\begin{aligned}
h^2(f_1, f_2) &= \frac{1}{2}\int_{\mathbb{R}}\left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 \mathrm{d}x \\
&= \int_0^{1/(4an)}\left(\sqrt{g(x) + 1 - 1/(4n)} - \sqrt{1 - 1/(4n)}\right)^2 \mathrm{d}x \\
&\leq \int_0^{1/(4an)} g(x)\,\mathrm{d}x \leq 1/(4n).
\end{aligned}$$

Our lower bound when $\alpha \geq 1/\pi - 1/p$ then follows from standard results in minimax estimation. We refer, for instance, to Proposition 5 of [Bir14]. As to the case, $\alpha < 1/\pi - 1/p$, we use $\mathcal{B}_{\pi,\infty}^{1/\pi-1/p}(R) \subset \mathcal{B}_{\pi,\infty}^{\alpha}(R)$ and apply the previous result with $\alpha = 1/\pi - 1/p$. $\qquad\square$

4.2. *Proof of Proposition 3*

Let f_m be the $\mathbb{L}^2$ projection of f on $\mathcal{P}_r(m)$ defined by

$$f_m = \sum \beta_{I,j}\varphi_{I,j} \quad \text{with } \beta_{I,j} = \int \varphi_{I,j}f = \mathbb{E}[\hat{\beta}_{I,j}],$$

where the sum runs over the intervals $I \in m$ and $j \in \{0, \ldots, r\}$. Pythagoras'

theorem implies

$$d_2^2(f, \hat{f}_m) = d_2^2(f, f_m) + \sum \left(\hat{\beta}_{I,j} - \beta_{I,j} \right)^2$$

$$= d_2^2(f, \mathcal{P}_r(m)) + \sum \left(\hat{\beta}_{I,j} - \mathbb{E}[\hat{\beta}_{I,j}] \right)^2.$$

Therefore,

$$\mathbb{E}\left[d_2^2(f, \hat{f}_m) \right] \leq d_2^2(f, \mathcal{P}_r(m)) + A/n \quad \text{with } A = \sum \int \varphi_{I,j}^2 f.$$

We use $\|\varphi_{I,j}\|_\infty = \sqrt{(2j+1)/|I|}$ to get

$$A \leq (r+1)^2 \sum_{I \in m} \frac{1}{|I|} \int_I f,$$

and apply Hölder's inequality twice:

$$A \leq (r+1)^2 \sum_{I \in m} \left(\int_I f^q \right)^{1/q} |I|^{-1/q}$$

$$\leq (r+1)^2 \|f\|_q \left(\sum_{I \in m} |I|^{-1/(q-1)} \right)^{(q-1)/q}. \qquad \square$$

4.3. Proof of Theorem 4

We define the collection $\mathcal{I}_\infty$ gathering all the dyadic intervals that appear in the partitions m of $\mathcal{M}_\infty$, that is, all the intervals with endpoints $k2^{-\ell}$, $(k+1)2^{-\ell}$ where $k \in \{0, \ldots, 2^\ell - 1\}$ and $\ell \geq 0$. Without loss of generality, we may assume that these intervals are of the form $[a, b)$ when $b < 1$ and $[a, b]$ when $b = 1$.

We set for $I \in \mathcal{I}_\infty$, collection m of intervals and $\xi > 0$,

$$v_{\mathrm{id}}^p(I) = \left\| \sum_{j=0}^r |\varphi_{I,j}| \sqrt{\mathbb{E}[\varphi_{I,j}^2(X)]} \right\|_p^p,$$

$$v_{\mathrm{id},\xi}^p(m) = \sum_{I \in m} \left((\xi + \log_2 (1/|I|))/n \right)^{p/2} v_{\mathrm{id}}^p(I),$$

$$\mathrm{pen}_{\mathrm{id},\xi}^p(m) = v_{\mathrm{id},\xi}^p(m) + \widehat{w}_\xi^p(m).$$

We begin by proving a uniform risk bound in deviation for the projection estimators:

Lemma 1. *For all $r \in \mathbb{N}$, $f \in \mathbb{L}^p([0,1])$, $\xi \geq \log 4 + \log(r+1)$ and probability larger than $1 - (1/2)e^{-\xi}$: for all $J \in \mathcal{I}_\infty$, and $m \in \mathcal{M}_\infty$,*

$$d_p^p(f\mathbb{1}_J, \hat{f}_{m \vee J}\mathbb{1}_J) \leq c_1 d_p^p(f\mathbb{1}_J, \mathcal{P}_r(m \vee J)) + c_2 \mathrm{pen}_{\mathrm{id},\xi}^p(m \vee J),$$

where c_1 only depends on p, r and where c_2 only depends on p.

Proof of Lemma 1. Let $f_{m \vee J}$ be the $\mathbb{L}^2$ projection of f on $\mathcal{P}_r(m \vee J)$ defined by

$$f_{m \vee J} = \sum_{\substack{I \in m \vee J \\ j \in \{0, \dots, r\}}} \beta_{I,j} \varphi_{I,j} \quad \text{with } \beta_{I,j} = \int f \varphi_{I,j} = \mathbb{E}[\hat{\beta}_{I,j}].$$

By using the triangle inequality, and the elementary inequality $(a + b)^p \leq 2^{p-1}(a^p + b^p)$,

$$d_p^p(f \mathbb{1}_J, \hat{f}_{m \vee J} \mathbb{1}_J) \leq 2^{p-1} \left[d_p^p(f \mathbb{1}_J, f_{m \vee J} \mathbb{1}_J) + d_p^p(f_{m \vee J} \mathbb{1}_J, \hat{f}_{m \vee J} \mathbb{1}_J) \right]. \quad (12)$$

It follows from standard results about the $\mathbb{L}^p$ norm of the $\mathbb{L}^2$ projection operator that

$$d_p(f \mathbb{1}_J, f_{m \vee J} \mathbb{1}_J) \leq c d_p(f \mathbb{1}_J, \mathcal{P}_r(m \vee J)), \quad (13)$$

where c only depends on r. We refer for instance to the argument below Lemma 3 of [BM00] for the proof of this inequality. We now tackle the second term in (12).

Consider $\ell \geq 0$, and note that the intervals of $\mathcal{I}_\ell = \{I \in \mathcal{I}_\infty, |I| = 2^{-\ell}\}$ are of the form $[k/2^\ell, (k+1)/2^\ell)$ (the interval is closed when $(k+1)/2^\ell = 1$). In particular, $|\mathcal{I}_\ell| = 2^\ell$. We deduce from Bernstein's inequality (Theorem 2.10 of [BLM13]) that there exists for each interval $I \in \mathcal{I}_\ell$ and j, an event $\Omega(I, j)$ of probability $1 - e^{-\xi_\ell}$ with $\xi_\ell = 2\xi + 2\ell \log 2$ on which

$$|\hat{\beta}_{I,j} - \beta_{I,j}| \leq \sqrt{2(\xi_\ell/n) \mathbb{E}_f[\varphi_{I,j}^2(X)]} + \xi_\ell/(3n) \|\varphi_{I,j}\|_\infty. \quad (14)$$

We then set

$$\Omega_\xi = \bigcap_{\substack{I \in \mathcal{I}_\ell \\ \ell \geq 0 \\ j \in \{0, \dots, r\}}} \Omega(I, j),$$

and deduce from a union bound

$$\begin{aligned}
\mathbb{P}\left[\Omega_\xi^c\right] &\leq (r+1) \sum_{\ell \geq 0} |\mathcal{I}_\ell| e^{-\xi_\ell} \\
&\leq (r+1) \sum_{\ell \geq 0} 2^\ell e^{-2\xi - 2\ell \log 2} \\
&\leq 2(r+1) e^{-2\xi},
\end{aligned}$$

which is not larger than $(1/2)e^{-\xi}$ as $\xi \geq \log 4 + \log(r+1)$.

On the event Ω_ξ, (14) holds true simultaneously for all $I \in \mathcal{I}_\infty$, $j \in \{0, \dots, r\}$ with ℓ such that $2^{-\ell} = |I|$. Now, for $m \in \mathcal{M}_\infty$, and $J \in \mathcal{I}_\infty$,

$$d_p^p(f_{m \vee J} \mathbb{1}_J, \hat{f}_{m \vee J} \mathbb{1}_J) = \int \left| \sum_{\substack{I \in m \vee J \\ j \in \{0, \dots, r\}}} \left(\hat{\beta}_{I,j} - \beta_{I,j} \right) \varphi_{I,j} \right|^p$$

$$\leq \sum_{I \in m \vee J} \int \left(\sum_{j \in \{0,\dots,r\}} \left| \hat{\beta}_{I,j} - \beta_{I,j} \right| |\varphi_{I,j}| \right)^p.$$

As $m \vee J \subset \mathcal{I}_\infty$, we may use (14), and then $(a+b)^p \leq 2^{p-1}(a^p + b^p)$, $\|\varphi_{I,j}\|_\infty = \sqrt{(2j+1)/|I|}$ to conclude the proof. $\qquad \square$

Lemma 2. *For all $r \in \mathbb{N}$, $\xi \geq \log 4 + \log(r+1)$ and $f \in \mathbb{L}^p([0,1])$, the following assertion holds true with probability larger than $1 - (1/2)e^{-\xi}$: for all $m \in \mathcal{M}_\infty$, $J \in \mathcal{I}_\infty$,*

$$\mathrm{pen}_{\mathrm{id},\xi}^p(m \vee J) \leq \widehat{\mathrm{pen}}_\xi^p(m \vee J) \leq c_3 \mathrm{pen}_{\mathrm{id},\xi}^p(m \vee J), \tag{15}$$

provided that $\kappa_1, \kappa_2 \geq \kappa$ for some κ depending only on p and where c_3 only depends on κ_1, κ_2, p.

Proof of Lemma 2. As in the previous proof, we may apply Bernstein's inequality to get with probability $1 - (1/2)e^{-\xi}$: for all $j \in \{0,\dots,r\}$, and $I \in \mathcal{I}_\infty$,

$$\left| \frac{1}{n} \sum_{i=1}^n \varphi_{I,j}^2(X_i) - \mathbb{E}_f[\varphi_{I,j}^2(X)] \right| \leq \sqrt{2(\xi_\ell/n)\mathbb{E}_f[\varphi_{I,j}^4(X)]} + (\xi_\ell/(3n))\|\varphi_{I,j}\|_\infty^2$$

$$\leq \sqrt{2(\xi_\ell/n)\|\varphi_{I,j}\|_\infty^2 \mathbb{E}_f[\varphi_{I,j}^2(X)]} \\ + (\xi_\ell/(3n))\|\varphi_{I,j}\|_\infty^2,$$

where ℓ is such that $2^{-\ell} = |I|$ and $\xi_\ell = 2\xi + 2\ell \log 2$. By using the elementary inequality $\sqrt{2ab} \leq a/2 + b$,

$$\left| \frac{1}{n} \sum_{i=1}^n \varphi_{I,j}^2(X_i) - \mathbb{E}_f[\varphi_{I,j}^2(X)] \right| \leq \frac{1}{2}\mathbb{E}_f[\varphi_{I,j}^2(X)] + (4/3)(\xi_\ell/n)\|\varphi_{I,j}\|_\infty^2.$$

Therefore,

$$\frac{1}{n} \sum_{i=1}^n \varphi_{I,j}^2(X_i) \leq (3/2)\mathbb{E}_f[\varphi_{I,j}^2(X)] + (4/3)(\xi_\ell/n)\|\varphi_{I,j}\|_\infty^2$$

$$\mathbb{E}_f[\varphi_{I,j}^2(X)] \leq \frac{2}{n} \sum_{i=1}^n \varphi_{I,j}^2(X_i) + (8/3)(\xi_\ell/n)\|\varphi_{I,j}\|_\infty^2,$$

which leads to the result after some computations. $\qquad \square$

Proposition 8. *Let $r \in \mathbb{N}$, $\xi > 0$, $\ell \in \mathbb{N}^\star \cup \{\infty\}$ and $\hat{m}$ be a partition satisfying (7). Then if $\kappa_1, \kappa_2 \geq \kappa$ for some κ depending only on p, and if $f \in \mathbb{L}^p([0,1])$, the following assertion holds true: for all $\xi \geq \log 4 + \log(r+1)$ with probability $1 - e^{-\xi}$:*

$$d_p^p(f, \hat{f}_{\hat{m}}) \leq C \inf_{m \in \mathcal{M}_\ell} \left\{ d_p^p(f, \mathcal{P}_r(m)) + \mathrm{pen}_{\mathrm{id},\xi}^p(m) + 1/n^{2p} \right\}. \tag{16}$$

In this inequality, C depends on p, r, κ_1, κ_2 only.

Proof of Theorem 8. Up to an increase of κ, the two preceding lemmas assert: with probability $1 - e^{-\xi}$, for all $m \in \mathcal{M}_\ell$, $J \in \mathcal{I}_\infty$, $\kappa_1, \kappa_2 \geq \kappa$,

$$2^{p-1} d_p^p(f \mathbb{1}_J, \hat{f}_{m \vee J} \mathbb{1}_J) \leq c d_p^p(f \mathbb{1}_J, \mathcal{P}_r(m \vee J)) + \widehat{\mathrm{pen}}_\xi^p(m \vee J), \qquad (17)$$

where c depends on p and r only. In the sequel, we repeatedly use the inequality $(a + b)^p \leq 2^{p-1}(a^p + b^p)$ without mentioning it again. The triangle inequality implies

$$\gamma(m) \leq 2^{p-1} d_p^p(f, \hat{f}_m) + \sum_{J \in m} \sup_{m' \in \mathcal{M}_\ell} \left\{ 2^{p-1} d_p^p(f \mathbb{1}_J, \hat{f}_{m' \vee J} \mathbb{1}_J) - \widehat{\mathrm{pen}}_\xi^p(m' \vee J) \right\},$$

and (17) ensures

$$\begin{aligned}
\gamma(m) &\leq 2^{p-1} d_p^p(f, \hat{f}_m) + \sum_{J \in m} \sup_{m' \in \mathcal{M}_\ell} \left\{ c d_p^p(f \mathbb{1}_J, \mathcal{P}_r(m' \vee J)) \right\} \\
&\leq 2^{p-1} d_p^p(f, \hat{f}_m) + c d_p^p(f, \mathcal{P}_r(m)) \\
&\leq 2c d_p^p(f, \mathcal{P}_r(m)) + \widehat{\mathrm{pen}}_\xi^p(m).
\end{aligned} \qquad (18)$$

The triangle inequality entails for all $m \in \mathcal{M}_\ell$,

$$\begin{aligned}
d_p^p(f, \hat{f}_{\hat{m}}) &\leq 2^{p-1} \left[d_p^p(f, \hat{f}_{\hat{m} \vee m}) + \widehat{\mathrm{pen}}_\xi^p(\hat{m} \vee m) \right. \\
&\qquad\qquad \left. + \left\{ d_p^p(\hat{f}_{\hat{m}}, \hat{f}_{\hat{m} \vee m}) - \widehat{\mathrm{pen}}_\xi^p(\hat{m} \vee m) \right\} \right] \\
&\leq 2^{p-1} \left[d_p^p(f, \hat{f}_{\hat{m} \vee m}) + \widehat{\mathrm{pen}}_\xi^p(\hat{m} \vee m) \right. \\
&\qquad\qquad \left. + \sum_{J \in \hat{m}} \left\{ d_p^p(\hat{f}_{\hat{m}} \mathbb{1}_J, \hat{f}_{\hat{m} \vee m} \mathbb{1}_J) - \widehat{\mathrm{pen}}_\xi^p(m \vee J) \right\} \right].
\end{aligned}$$

We apply (17) with the inequalities $\widehat{\mathrm{pen}}_\xi^p(\hat{m} \vee m) \leq \widehat{\mathrm{pen}}_\xi^p(\hat{m}) + \widehat{\mathrm{pen}}_\xi^p(m)$ (as $\hat{m} \vee m \subset \hat{m} \cup m$) and $d_p^p(f, \mathcal{P}_r(\hat{m} \vee m)) \leq d_p^p(f, \mathcal{P}_r(m))$ to deduce,

$$\begin{aligned}
d_p^p(f, \hat{f}_{\hat{m}}) &\leq 2^{p-1} \left[c' d_p^p(f, \mathcal{P}_r(m)) + (1 + 2^{1-p}) \widehat{\mathrm{pen}}_\xi^p(m) \right. \\
&\qquad \left. + \sum_{J \in \hat{m}} \left\{ d_p^p(\hat{f}_{\hat{m}} \mathbb{1}_J, \hat{f}_{\hat{m} \vee m} \mathbb{1}_J) - \widehat{\mathrm{pen}}_\xi^p(m \vee J) \right\} + (1 + 2^{1-p}) \widehat{\mathrm{pen}}_\xi^p(\hat{m}) \right].
\end{aligned}$$

By using (6), and taking the infimum over all the partitions $m \in \mathcal{M}_\ell$,

$$\begin{aligned}
d_p^p(f, \hat{f}_{\hat{m}}) &\leq 2^{p-1} \inf_{m \in \mathcal{M}_\ell} \left\{ c' d_p^p(f, \mathcal{P}_r(m)) + (1 + 2^{1-p}) \widehat{\mathrm{pen}}_\xi^p(m) \right\} \\
&\qquad + 2^{p-1} \left[\gamma(\hat{m}) + (1 + 2^{1-p}) \widehat{\mathrm{pen}}_\xi^p(\hat{m}) \right],
\end{aligned}$$

and (7) entails

$$d_p^p(f, \hat{f}_{\hat{m}}) \leq 2^{p-1} \inf_{m \in \mathcal{M}_\ell} \left\{ c' d_p^p(f, \mathcal{P}_r(m)) + 2(1 + 2^{1-p}) \widehat{\mathrm{pen}}_\xi^p(m) + \gamma(m) + 1/n^{2p} \right\}.$$

We then use (18) to get

$$d_p^p(f, \hat{f}_{\hat{m}}) \leq C \inf_{m \in \mathcal{M}_\ell} \left\{ d_p^p(f, \mathcal{P}_r(m)) + \widehat{\text{pen}}_\xi^p(m) + 1/n^{2p} \right\},$$

and apply (15). □

Lemma 3. *For all* $m \in \mathcal{M}_\infty$, $\xi > 0$ *and* $q \in [p, +\infty]$,

$$\text{pen}_{\text{id},\xi}^p(m) \leq C \left[((\xi + \ell_m)/n)^{p/2} M_q^{p/2}(m) v_q^{p/2}(m) + ((\xi + \ell_m)/n)^p w^p(m) \right],$$

where C *only depends on* p *and* r.

Proof. Since $|I| \geq 2^{-\ell_m}$ for all $I \in m$,

$$v_{\text{id},\xi}^p(m) \leq ((\xi + \ell_m)/n)^{p/2} \sum_{I \in m} \int_I \left| \sum_{j=0}^r |\varphi_{I,j}| \sqrt{\int_I f \varphi_{I,j}^2} \right|^p.$$

We use $\|\varphi_{I,j}\|_\infty = \sqrt{(2j+1)/|I|}$ to get

$$v_{\text{id},\xi}^p(m) \leq ((\xi + \ell_m)/n)^{p/2} \sum_{I \in m} \int_I \left| \sum_{j=0}^r \frac{2j+1}{|I|} \sqrt{\int_I f} \right|^p$$

$$\leq ((\xi + \ell_m)/n)^{p/2} (r+1)^{2p} A \quad \text{with } A = \sum_{I \in m} \left(\int_I f \right)^{p/2} |I|^{1-p}. \quad (19)$$

We now bound above A by repeatedly using Hölder's inequality as in the proof of Proposition 3.

First, we use it twice to get for every $q \geq p$,

$$A \leq \sum_{I \in m} \left(\int_I f^q \right)^{p/(2q)} |I|^{1-p/2(1+1/q)}$$

$$\leq \|f\|_q^{p/2} v_q^{p/2}(m). \quad (20)$$

Second, let $q' \geq p/2$ such that $1/q' = 1/q + 1/(p\ell_m)$ and for $j \geq 0$, $m_j = \{I \in m, |I| = 2^{-j}\}$, $I_j = \cup_{I \in m_j} I$. We deduce,

$$A \leq \sum_{j=0}^{\ell_m} \|f \mathbb{1}_{I_j}\|_{q'}^{p/2} |m_j|^{1-p/(2q')} 2^{-j(1-p/2(1+1/q'))}$$

$$\leq \|f\|_{q'}^{p/2} 2^{1/2} \sum_{j=0}^{\ell_m} |m_j|^{1-p/(2q)} 2^{-j(1-p/2(1+1/q))}$$

$$\leq \|f\|_{q'}^{p/2} 2^{1/2} (\ell_m + 1)^{p/(2q)} \left(\sum_{j=0}^{\ell_m} |m_j| 2^{-j(1-p/2(1+1/q))2q/(2q-p)} \right)^{1-p/(2q)}$$

$$\leq \|f\|_{q'}^{p/2} 2^{1/2} (\ell_m + 1)^{p/(2q)} v_q^{p/2}(m).$$

By combining this result together with (20) and (19), we obtain

$$v_{\mathrm{id},\xi}^p(m) \leq C \left((\xi + \ell_m)/n \right)^{p/2} M_q^{p/2}(m) v_q^{p/2}(m).$$

We now turn to the term $\widehat{w}_\xi^p(m)$. We have,

$$\widehat{w}_\xi^p(m) = \sum_{I \in m} \left((\xi + \log_2 (1/|I|))/n \right)^p \widehat{w}(I)^p$$

$$\leq \left((\xi + \ell_m)/n \right)^p \sum_{I \in m} \frac{1}{|I|^{p/2}} \left\| \sum_{j=0}^r \sqrt{(2j+1)} |\varphi_{I,j}| \right\|_p^p.$$

We simply use $|\varphi_{I,j}| \leq \sqrt{(2j+1)/|I|}$ to get

$$\widehat{w}_\xi^p(m) \leq C((\xi + \ell_m)/n)^p w^p(m),$$

which ends the proof. $\square$

Proof of Theorem 4. We denote by x_+ the positive part of a real number x. We have,

$$d_p(\hat{f}_{\hat{m}}, \hat{f}) = (\|\hat{f}_{\hat{m}}\|_p - n)_+$$
$$\leq (\|\hat{f}_{\hat{m}}\|_p - \|f\|_p)_+$$

as $\|f\|_p \leq n$. Therefore, $d_p(\hat{f}_{\hat{m}}, \hat{f}) \leq d_p(f, \hat{f}_{\hat{m}})$ and $d_p(f, \hat{f}) \leq 2d_p(f, \hat{f}_{\hat{m}})$. We apply Proposition 8 to define an event Ω_ξ of probability $1 - e^{-\xi}$ on which (16) holds true. By using moreover $d_p(f, \hat{f}) \leq \|f\|_p + \|\hat{f}\|_p \leq 2n$,

$$\mathbb{E}_f \left[d_p^p(f, \hat{f}) \right] \leq \mathbb{E}_f \left[d_p^p(f, \hat{f}) \mathbb{1}_{\Omega_\xi} \right] + (2n)^p e^{-\xi}$$
$$\leq 2^p \mathbb{E}_f \left[d_p^p(f, \hat{f}_{\hat{m}}) \mathbb{1}_{\Omega_\xi} \right] + (2n)^p e^{-\xi}.$$

Since $\xi \geq 2p \log n$, and $w(m) \geq 1$ for all m, $(2n)^p e^{-\xi} \leq 2^p (w(m)/n)^p$. We then bound $d_p^p(f, \hat{f}_{\hat{m}})$ thanks to (16) and Lemma 3. $\square$

4.4. Proof of Proposition 5

We introduce for each interval I the space

$$\mathcal{P}_r(I) = \{P \mathbb{1}_I, \text{ where } P \text{ is a polynomial function of degree at most } r\}$$

composed of polynomial functions on I. We need the following claim:

Claim 1. *Let $j \in \mathbb{N}$ and $\bar{\mathfrak{m}}_j$ be the regular partition of $[0,1]$ of size 2^j. Then, for all $p \in [1, +\infty]$, $\pi \in (0, p)$, $\alpha > 1/\pi - 1/p$, $r > \alpha - 1$, $R > 0$ and $f \in \mathcal{B}^\alpha_{\pi,\infty}(R)$,*

$$\left(\sum_{I \in \bar{\mathfrak{m}}_j} d_p^\pi(f \mathbb{1}_I, \mathcal{P}_r(I)) \right)^{1/\pi} \leq C\,(1 + 1/\tau)\,R2^{-j\tau},$$

where $\tau = \alpha + 1/p - 1/\pi$ and where C depends on π, r only.

This claim is a slight revisit of Lemma 1 of [Aka12] in a unidimensional context. Note that it holds for all $f \in \mathcal{B}^\alpha_{\pi,\infty}(R)$ whereas her result is restricted to functions $f \in \mathcal{B}^\alpha_{\pi,\pi}(R)$ when $\pi \in (1, p)$. The term in front of $R2^{-j\tau}$ is also made more explicit here (which will be of interest for the proof of Theorem 6).

Sketch of the proof of Claim 1. We make the following minor modification in the proof of [Aka12]. Instead of applying Hölder's inequality to her inequality (27), we apply Minkowski's integral inequality. By using her notations, this yields

$$2^{jd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \left(\sum_{K \in \mathcal{D}_j^\sigma} \mathcal{E}_{\boldsymbol{r}}^p(s, K)_q \right)^{1/p}$$

$$\leq C(d, \boldsymbol{r}, \boldsymbol{\sigma}, p, q) 2^{jd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \sum_{k \geq j} \left\{ \sum_{K \in \mathcal{D}_j^\sigma} \left(2^{-kd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} 2^{k\underline{\sigma}} e_k(s, K) \right)^p \right\}^{1/p}$$

where the term $C(d, \boldsymbol{r}, \boldsymbol{\sigma}, p, q)$ comes from her equation (19).

If we go back to her calculations, we can observe that $C(d, \boldsymbol{r}, \boldsymbol{\sigma}, p, q)$ does not depend on $\boldsymbol{\sigma}$ in a unidimensional setting (as $\underline{\boldsymbol{\sigma}} = H(\boldsymbol{\sigma})$). Moreover, the dependency on q comes from her equation (20), and can be removed thanks to Theorem 2.6 in Chapter 4 of [DL93]. We therefore write $C(r, p)$ in place of $C(d, \boldsymbol{r}, \boldsymbol{\sigma}, p, q)$ (as $d = 1$).

We thus have,

$$2^{jd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \left(\sum_{K \in \mathcal{D}_j^\sigma} \mathcal{E}_{\boldsymbol{r}}^p(s, K)_q \right)^{1/p}$$

$$\leq C(r, p) 2^{jd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \sum_{k \geq j} 2^{-kd\tau\underline{\sigma}/H(\boldsymbol{\sigma})} 2^{k\underline{\sigma}} e_k(s, [0,1]^d)$$

$$\leq C(r, p) \left(1 - 2^{-d\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \right)^{-1} \sup_{k \geq j} \left(2^{k\underline{\sigma}} e_k(s, [0,1]^d) \right),$$

which gives the result using

$$\left(1 - 2^{-d\tau\underline{\sigma}/H(\boldsymbol{\sigma})} \right)^{-1} = \left(1 - 2^{-\tau} \right)^{-1} \leq (1/\log(2))(1 + 1/\tau). \qquad \square$$

We now prove the following lemma.

Lemma 4. *Let $(\eta_j)_{j\geq 0}$ be a sequence of numbers that is bounded below by a positive constant. Let $p \geq 1$, $\pi \in (0,p)$, $\alpha > 1/\pi - 1/p$, $r > \alpha - 1$, $R > 0$ and $f \in \mathcal{B}^\alpha_{\pi,\infty}(R)$. Then, there exists $m \in \mathcal{M}_\infty$ such that $m_j = \{I \in m, |I| = 2^{-j}\}$ satisfies for all $j \geq 1$,*

$$|m_j| \leq \min\left\{ C\left(R\frac{2^{-(j-1)\tau}}{\eta_{j-1}}\right)^\pi, 2^j \right\}, \tag{21}$$

where $\tau = \alpha + 1/p - 1/\pi$ and where C depends on p, π, r, α only. Moreover,

$$d_p^p(f, \mathcal{P}_r(m)) \leq \sum_{j=0}^\infty |m_j| \eta_j^p. \tag{22}$$

Remark: the construction of the partition m is the same to that of [DY90] when the sequence $(\eta_j)_j$ is constant.

Proof of Lemma 4. We begin by defining two preliminary collections $\bar{m}_j$, $\bar{m}'_j$ of disjoint intervals of $[0,1]$ by induction.

When $d_p(f\mathbb{1}_{[0,1]}, \mathcal{P}_r([0,1])) \leq \eta_0$, we set $(\bar{m}_0, \bar{m}'_0) = (\bar{\mathfrak{m}}_0, \emptyset)$ and when $d_p(f\mathbb{1}_{[0,1]}, \mathcal{P}_r([0,1])) > \eta_0$, we set $(\bar{m}_0, \bar{m}'_0) = (\emptyset, \bar{\mathfrak{m}}_0)$. We then define $\bar{m}_j$ and $\bar{m}'_j$ from $\bar{m}_{j-1}$, $\bar{m}'_{j-1}$ by

$$\bar{m}_j = \{I \in \bar{\mathfrak{m}}_j, \exists I' \in \bar{m}'_{j-1} \text{ such that } I \subset I' \text{ and } d_p(f\mathbb{1}_I, \mathcal{P}_r(I)) \leq \eta_j\}$$
$$\bar{m}'_j = \{I \in \bar{\mathfrak{m}}_j, \exists I' \in \bar{m}'_{j-1} \text{ such that } I \subset I' \text{ and } d_p(f\mathbb{1}_I, \mathcal{P}_r(I)) > \eta_j\}.$$

Therefore, $\bar{m}_j$, $\bar{m}'_j$ are subsets of $\bar{\mathfrak{m}}_j$, $\bar{m}_j \cup \bar{m}'_j$ is a partition of $\cup_{I' \in \bar{m}'_{j-1}} I'$ and

$$d_p(f\mathbb{1}_I, \mathcal{P}_r(I)) \leq \eta_j \quad \text{for all } I \in \bar{m}_j$$
$$d_p(f\mathbb{1}_I, \mathcal{P}_r(I)) > \eta_j \quad \text{for all } I \in \bar{m}'_j.$$

For every k, $\bar{m}''_k = \left(\bigcup_{j=0}^k \bar{m}_j\right) \cup \bar{m}'_k$ is a partition of $[0,1]$. Moreover, we derive $\bar{m}''_k$ from $\bar{m}''_{k-1}$ by dividing some of its intervals in two equal parts (more precisely, we divide the intervals of $\bar{m}'_{k-1}$). Therefore, $\bar{m}''_k \in \mathcal{M}(\bar{m}''_{k-1})$ and $\bar{m}''_k \in \mathcal{M}_\ell$ for some ℓ, see Section 3.1.

Since $(\eta_j)_j$ is bounded below, we may define the smallest ℓ such that $\bar{m}'_\ell = \emptyset$. We then set $m = \bar{m}''_\ell = \bigcup_{j=0}^\ell \bar{m}_j$. Note that $\bar{m}_j = m_j = \{I \in m, |I| = 2^{-j}\}$ and $m \in \mathcal{M}_\ell$.

We now show the bound on $|m_j| = |\bar{m}_j|$ for $j \geq 1$. For all $I \in \bar{m}_j$, there exists $I' \in \bar{m}'_{j-1}$ such that $I \subset I'$. Therefore,

$$(\eta_{j-1})^\pi < d_p^\pi(f\mathbb{1}_{I'}, \mathcal{P}_r(I')),$$

and hence

$$(\eta_{j-1})^\pi |\bar{m}_j| < \sum_{I \in \bar{m}_j} d_p^\pi(f\mathbb{1}_{I'}, \mathcal{P}_r(I'))$$

$$< \sum_{\substack{J \in \bar{m}'_{j-1}}} \sum_{\substack{I \in \bar{m}_j \\ I' = J}} d_p^\pi(f \mathbb{1}_J, \mathcal{P}_r(J))$$

$$< 2 \sum_{J \in \bar{m}'_{j-1}} d_p^\pi(f \mathbb{1}_J, \mathcal{P}_r(J)).$$

By using $\bar{m}'_{j-1} \subset \bar{\mathfrak{m}}_{j-1}$ and Claim 1,

$$(\eta_{j-1})^\pi |\bar{m}_j| < 2 \left(C' R 2^{-(j-1)\tau} \right)^\pi,$$

which shows the first bound in (21). As to the second one, we merely use that $\bar{m}_j \subset \bar{\mathfrak{m}}_j$ and $|\bar{\mathfrak{m}}_j| = 2^j$.

Now,

$$d_p^p(f, \mathcal{P}_r(m)) = \sum_{I \in m} d_p^p(f \mathbb{1}_I, \mathcal{P}_r(I)) \le \sum_{j=0}^\infty \sum_{I \in m_j} d_p^p(f \mathbb{1}_I, \mathcal{P}_r(I)) \le \sum_{j=0}^\infty |m_j| \eta_j^p,$$

which shows (22). $\qquad\square$

Proof of Proposition 5. We apply Lemma 4 with

$$\eta_j = \begin{cases} C^{1/\pi} R 2^{(j+1)(\delta-1)/p + \tau - k(\alpha + \delta/p)} & \text{if } \alpha > \delta(1/\pi - 1/p) \\ C^{1/\pi} R 2^{(j+1)(\delta-1)/p + \tau - k(\tau + (\delta-1)/p)} & \text{if } \alpha \le \delta(1/\pi - 1/p). \end{cases} \tag{23}$$

We deduce a partition m with $|m_j|$ such that

$$|m_j|^{1/\pi} \le \begin{cases} 2^{k(\alpha + \delta/p)} 2^{-j(\alpha + \delta/p - 1/\pi)} & \text{if } \alpha > \delta(1/\pi - 1/p) \\ 2^{(k-j)(\tau + (\delta-1)/p)} & \text{if } \alpha \le \delta(1/\pi - 1/p). \end{cases} \tag{24}$$

In the first case, $|m_j| = 0$ when $j > (\alpha + \delta/p)/(\alpha + \delta/p - 1/\pi)k$, which shows that $m \in \mathcal{M}_\ell$. As to the second case, we have $|m_j| = 0$ when $j > k$ and thus $m \in \mathcal{M}_k$.

Note now that

$$\sum_{I \in m} |I|^{1-\delta} = \sum_{j=0}^\infty |m_j| 2^{j(\delta-1)}. \tag{25}$$

By using $|m_j| \le 2^j$ and (24), we get in the first case

$$\sum_{I \in m} |I|^{1-\delta} \le \sum_{j=0}^k 2^j 2^{j(\delta-1)} + 2^{k\pi(\alpha + \delta/p)} \sum_{j=k+1}^\infty 2^{-j(\alpha + \delta/p - 1/\pi)\pi} 2^{j(\delta-1)}$$

$$\le \frac{2^{(k+1)\delta}}{2^\delta - 1} + 2^{k\pi(\alpha + \delta/p)} \frac{2^{-(k+1)(\alpha - \delta(1/\pi - 1/p))\pi}}{1 - 2^{-(\alpha - \delta(1/\pi - 1/p))\pi}}$$

$$\le C' 2^{k\delta}, \tag{26}$$

where C' depends on α, δ, p, π.

As to the second case, we use that $|m_j| = 0$ for $j > k$ and (24) to get

$$\sum_{I \in m} |I|^{1-\delta} \le \sum_{j=0}^{k} 2^{(k-j)(\tau+(\delta-1)/p)\pi} 2^{j(\delta-1)}$$

$$\le 2^{k\pi(\tau+(\delta-1)/p)} \sum_{j=0}^{k} 2^{j((1-\pi/p)(\delta-1)-\pi\tau)}. \tag{27}$$

When the condition on α is strict, $(1 - \pi/p)(\delta - 1) - \pi\tau > 0$, and thus,

$$\sum_{I \in m} |I|^{1-\delta} \le 2^{k\pi(\tau+(\delta-1)/p)} \times C' 2^{k((1-\pi/p)(\delta-1)-\pi\tau)}$$

$$\le C' 2^{k(\delta-1)}, \tag{28}$$

where C' depends on p, π, α, δ only.

Suppose now that $\alpha = \delta(1/\pi - 1/p)$. Then, we put the inequalities $(1 - \pi/p)(\delta - 1) - \pi\tau = 0$ and $\pi(\tau + (\delta - 1)/p) = \delta - 1$ into (27) to derive

$$\sum_{I \in m} |I|^{1-\delta} \le C' k 2^{k(\delta-1)}. \tag{29}$$

We now prove the bounds on $d_p^p(f, \mathcal{P}_r(m))$. We use (22) and (23) to get in the first case,

$$d_p^p(f, \mathcal{P}_r(m)) \le C^{p/\pi} R^p 2^{-kp(\alpha+\delta/p)+p\tau+\delta-1} \sum_{j=0}^{\infty} |m_j| 2^{j(\delta-1)},$$

and in the second case

$$d_p^p(f, \mathcal{P}_r(m)) \le C^{p/\pi} R^p 2^{-kp(\tau+(\delta-1)/p)+p\tau+\delta-1} \sum_{j=0}^{\infty} |m_j| 2^{j(\delta-1)}.$$

We then bound the sum by using (25) (26), (28) and (29).

We finally prove the bounds on

$$\sum_{I \in m} |I|^{1-p} = \sum_{j=0}^{\infty} |m_j| 2^{j(p-1)}.$$

When $\alpha > \delta(1/\pi - 1/p)$, and $\alpha \ge p/\pi - \delta/p$,

$$\sum_{I \in m} |I|^{1-p} \le \sum_{j=0}^{k} 2^j 2^{j(p-1)} + 2^{k\pi(\alpha+\delta/p)} \sum_{j=k+1}^{\ell} 2^{-j\pi(\alpha+\delta/p-1/\pi)} 2^{j(p-1)}$$

$$\le \frac{2^{(k+1)p}}{2^p - 1} + 2^{k\pi(\alpha+\delta/p)} \sum_{j=k+1}^{\ell} 2^{-j\pi[\alpha-(p/\pi-\delta/p)]} \tag{30}$$

and

$$\sum_{j=k+1}^{\ell} 2^{-j\pi[\alpha-(p/\pi-\delta/p)]} \leq \begin{cases} C''2^{-k\pi[\alpha-(p/\pi-\delta/p)]} & \text{if } \alpha > p/\pi - \delta/p \\ \ell - k & \text{if } \alpha = p/\pi - \delta/p. \end{cases}$$

We put this inequality in (30) to get the bound. When $\alpha > \delta(1/\pi - 1/p)$, and $\alpha < p/\pi - \delta/p$, we have

$$\sum_{I \in m} |I|^{1-p} \leq 2^{k\pi(\alpha+\delta/p)} \sum_{j=0}^{\ell} 2^{-j\pi(\alpha+\delta/p-1/\pi)} 2^{j(p-1)}$$

$$\leq 2^{k\pi(\alpha+\delta/p)} \sum_{j=0}^{\ell} 2^{-j\pi[\alpha-(p/\pi-\delta/p)]}$$

$$\leq 2^{k\pi(\alpha+\delta/p)} \times C''2^{-\ell\pi[\alpha-(p/\pi-\delta/p)]},$$

which implies the bound. When $\alpha \leq \delta(1/\pi - 1/p)$,

$$\sum_{I \in m} |I|^{1-p} \leq \sum_{j=0}^{k} 2^{(k-j)(\tau+(\delta-1)/p)\pi} 2^{j(p-1)}$$

$$\leq 2^{k\pi(\alpha-1/\pi+\delta/p)} \sum_{j=0}^{k} 2^{j\pi(p/\pi-\delta/p-\alpha)}.$$

Since $\delta < p$, we have $p/\pi - \delta/p - \alpha > 0$ and

$$\sum_{I \in m} |I|^{1-p} \leq C''2^{k\pi(\alpha-1/\pi+\delta/p)} 2^{k\pi(p/\pi-\delta/p-\alpha)} \leq C''2^{(p-1)k}. \qquad \square$$

4.5. Proof of Theorem 6

It is well known that there is an embedding of Besov spaces into $\mathbb{L}^q$ spaces, see [DP88, DeV98]. We use here the following lemma. It will be proved after the present proof for the sake of completeness.

Lemma 5. *Let f be a density belonging to $\mathcal{B}_{\pi,\infty}^\alpha(R)$ for some $\pi, \alpha > 0$. Then, for all $q \in [1, +\infty]$ such that $1/q > 1/\pi - \alpha$,*

$$\|f\|_q \leq c_q(1 + R), \qquad (31)$$

where c_q depends on π, α, q only. Moreover $c_q \leq c(1 + 1/(\alpha - 1/\pi + 1/q))$ for some c depending only on π, α.

We recall that $\bar{\alpha}$ is the unique positive solution of (11) that is

$$\alpha = (1/\pi - 1/p)/(2/p - 1/\pi + \alpha).$$

Define, when $\alpha \leq 1/\pi$, $q_\alpha = \pi/(1 - \pi\alpha)$, $1/\delta_\alpha = 2/p - 1/q_\alpha = 2/p - 1/\pi + \alpha$. Every $\alpha \in (\bar{\alpha}, 1/\pi]$ satisfies $\alpha > \delta_\alpha(1/\pi - 1/p)$ and every $\alpha \in (1/\pi - 1/p, \bar{\alpha}]$ satisfies $\alpha \leq \delta_\alpha(1/\pi - 1/p)$.

Suppose that $\alpha \in (\bar{\alpha}, 1/\pi]$. Let $\varepsilon > 0$ be small enough such that

$$\alpha > (1/\pi - 1/p)/(2/p - (1/q_\alpha + \varepsilon)).$$

Let $q'_\alpha \geq p$ such that $1/q'_\alpha = 1/q_\alpha + \varepsilon$ and $1/\delta'_\alpha = 2/p - 1/q'_\alpha$. We apply Proposition 5 with k defined as the smallest integer larger than

$$\frac{1}{1 + 2\alpha} \log_2 \left(\frac{R^2}{1 + R} \frac{n}{\log n} \right), \tag{32}$$

and $\delta = \max\{1, \delta'_\alpha\}$. We therefore get a partition $m \in \mathcal{M}_\ell$ such that $d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k\alpha}$,

$$\left(\sum_{I \in m} |I|^{1-\delta} \right)^{1/\delta} \leq C2^k, \tag{33}$$

and

$$w(m) \leq \begin{cases} C2^k & \text{if } \alpha > p/\pi - \delta/p \\ C\ell2^k & \text{if } \alpha = p/\pi - \delta/p \\ C2^{(\pi/p)[k(\alpha+\delta/p)+\ell(p/\pi-\delta/p-\alpha)]} & \text{if } \alpha < p/\pi - \delta/p. \end{cases} \tag{34}$$

We use (33) with a concavity argument when $\delta'_\alpha < 1$ to get

$$v_{q'_\alpha}(m) = \left(\sum_{I \in m} |I|^{1-\delta'_\alpha} \right)^{1/\delta'_\alpha} \leq C2^k.$$

Moreover, we have

$$\ell \leq \frac{\alpha + \delta/p}{\delta/p + \alpha - 1/\pi} k. \tag{35}$$

Lemma 5 implies that $M_{q'_\alpha}(m) \leq \|f\|_{q'_\alpha} \leq c(1 + R)$ for some c. Due to our choice of k, we have for n large enough, using $\ell_m \leq c' \log n$ for some suitable c',

$$d_p(f, \mathcal{P}_r(m)) + \sqrt{M_{q'_\alpha}(m)v_q(m) \frac{\ell_m + \log n}{n}}$$
$$\leq C'R^{1/(1+2\alpha)} \left((1+R)\frac{\log n}{n} \right)^{1/(1+2\alpha)}.$$

We now claim that the term $w(m)(\ell_m + \log n)/n$ in (8) is negligible when n is large. This is indeed straightforward when $\alpha \geq p/\pi - \delta/p$ as our bounds for

$w(m)$ and $v_{q'_\alpha}(m)$ are of the same order of magnitude (up to a logarithmic term in case of equality). When $\alpha < p/\pi - \delta/p$, we have, using (34) and (35),

$$w(m) \le C' 2^{k(\alpha+\delta/p)(1-1/p)/(\delta/p+\alpha-1/\pi)}.$$

We apply the claim below (proved after the present proof) to get $w(m) \le C' 2^{k\beta(1+\alpha)}$ for some $\beta < 1$.

Claim 2. *The following inequality holds true: for all $\alpha \in (1/\pi - 1/p, 1/\pi]$, $q \in [p, \pi/(1-\pi\alpha)]$, $\pi \in (0,p)$, $1/\delta' = 2/p - 1/q$ such that $\alpha > \delta'(1/\pi - 1/p)$,*

$$(\alpha + \delta/p)(1 - 1/p) < (1+\alpha)(\delta/p + \alpha - 1/\pi),$$

where $\delta = \max\{1, \delta'\}$.

It follows from the definition of k that for n large enough,

$$w(m)\frac{\ell_m + \log n}{n} \le C''\big(R^2/(1+R)\big)^{\beta(1+\alpha)/(1+2\alpha)} \left(\frac{\log n}{n}\right)^{(1+2\alpha-\beta(1+\alpha))/(1+2\alpha)}.$$

As $\beta < 1$, the exponent is always larger than $\alpha/(1 + 2\alpha)$, and the term is negligible. Since $\|f\|_p \le \|f\|_{q'_\alpha} \le c(1 + R)$, we may apply Theorem 4 to get the result for n large enough.

Suppose now that $\alpha \in (1/\pi - 1/p, \bar{\alpha})$ and $\alpha \le 1/\pi$. Then, $\alpha < \delta_\alpha(1/\pi - 1/p)$. Since $\alpha > 1/\pi - 1/p$, we may suppose that $\delta_\alpha > 1$. Let k be the smallest integer larger than

$$\frac{1}{2(\alpha - 1/\pi + 1/p) + 1 - 1/\delta_\alpha} \log_2\left(\frac{R^2}{1+R}\frac{n}{\log n}\right). \tag{36}$$

We derive from the second point of Proposition 5 a partition $m \in \mathcal{M}_k$ such that

$$v_{q_\alpha}(m) \le C 2^{k(1-1/\delta_\alpha)}, \quad w(m) \le C 2^{k(1-1/p)}, \text{ and}$$
$$d_p(f, \mathcal{P}_r(m)) \le C R 2^{-k(\alpha+1/p-1/\pi)}.$$

Lemma 5 ensures that $\|f\|_{q_\alpha/(1+q_\alpha/(pk))} \le ck(1+R)$, for some c depending only on π, α and p. We may thus apply Theorem 4 to get when n is large enough,

$$\mathbb{E}_f\left[d_p^p(f, \hat{f})\right]^{1/p} \le C'\left\{ R 2^{-k(\alpha+1/p-1/\pi)} + \sqrt{(1+R)k^{1/q_\alpha+1}2^{k(1-1/\delta_\alpha)}\frac{k+\log n}{n}} \right.$$
$$\left. + 2^{k(1-1/p)}\frac{k+\log n}{n} \right\}.$$

By replacing k by (36), we observe that the first three terms are all of the same order of magnitude (up to log terms). We then use Lemma 5 to bound the $\mathbb{L}^p$ norm of f in the remaining term.

In borderline case $\alpha = \bar{\alpha}$, and $\alpha \le 1/\pi$, we do as above. We simply apply the third point of Proposition 5 instead of the second one, which leads to additional logarithmic factors in the final result.

The proof when $\alpha > 1/\pi$ is quite similar. We set from now on $q_\alpha = \infty$, $\delta_\alpha = p/2$. Suppose that $\alpha > \delta_\alpha(1/\pi - 1/p)$. We apply Proposition 5 with k defined as the smallest integer larger than (32) and $\delta = \max\{1, \delta_\alpha\}$. We therefore get a partition $m \in \mathcal{M}_\ell$ such that $d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k\alpha}$, $v_\infty(m) \leq C2^k$, and $w(m)$ bounded from above by (34). Lemma 5 ensures that $\|f\|_\infty \leq c(1 + R)$, and we apply Theorem 4. We finally use (35) and the claim below to show that $w(m)(\ell_m + \log n)/n$ is negligible when $\alpha < p/\pi - \delta/p$.

Claim 3. *The following inequality holds true: for all $\alpha > 1/\pi$, $\pi \in (0, p)$, such that $\alpha > (p/2)(1/\pi - 1/p)$,*

$$(\alpha + \delta/p)(1 - 1/p) < (1 + \alpha)(\delta/p + \alpha - 1/\pi),$$

where $\delta = \max\{1, p/2\}$.

When $\alpha < \delta_\alpha(1/\pi - 1/p)$, we may suppose that $\delta_\alpha = p/2 \geq 1$ and define k as the smallest integer such that (36) holds true (with the present value of δ_α). We derive from the second point of Proposition 5 a partition $m \in \mathcal{M}_k$ such that

$$v_\infty(m) \leq C2^{k(1-1/\delta_\alpha)}, w(m) \leq C2^{k(1-1/p)}, \text{ and}$$
$$d_p(f, \mathcal{P}_r(m)) \leq CR2^{-k(\alpha+1/p-1/\pi)}.$$

By using $\|f\|_p \leq \|f\|_\infty \leq c(1 + R)$, Lemma 5 and Theorem 4, we have for n large enough

$$\mathbb{E}_f \left[d_p^p(f, \hat{f}) \right]^{1/p} \leq C' \left\{ R2^{-k(\alpha+1/p-1/\pi)} + \sqrt{(1 + R)2^{k(1-1/\delta_\alpha)} \frac{k + \log n}{n}} \right.$$
$$\left. + 2^{k(1-1/p)} \frac{k + \log n}{n} \right\}.$$

By replacing k by (36), the first two terms are of the order of $((\log n)/n)^{(\alpha-1/\pi+1/p)/(2\alpha-2/\pi+1)}$ and the third one $((\log n)/n)^{(2\alpha-2/\pi+1/p)/(2\alpha-2/\pi+1)}$. It is therefore negligible compared to the first two terms as $\alpha > 1/\pi$.

Finally, when $\alpha = \delta_\alpha(1/\pi - 1/p)$, we do as above, but additional logarithmic factors appear as we apply the third point of Proposition 5. ☐

Proof of Lemma 5. We apply Claim 1 in Section 4.4 with $j = 0$, p replaced by q, and with r defined as the smallest integer larger than α. Then, there is a polynomial function g of degree at most r such that

$$d_q(f, g) \leq C \left(1 + \frac{1}{\alpha - 1/\pi + 1/q} \right) R, \tag{37}$$

where C depends on α and π only. Since g is a polynomial function, there exists C' depending only on r such that $\|g\|_q \leq C'\|g\|_1$ (see Theorem 2.6 in

Chapter 4 of [DL93]). Therefore, using the triangle inequality,

$$
\begin{aligned}
\|g\|_q &\leq C'\left[d_1(f,g) + \|f\|_1\right] \\
&\leq C'\left[d_q(f,g) + 1\right].
\end{aligned} \tag{38}
$$

We finally put $\|f\|_q \leq d_q(f,g) + \|g\|_q$, (37) and (38) together. □

Proof of Claim 2. We first suppose that $\delta' \geq 1$. Then,

$$
\begin{aligned}
(1+\alpha)(\delta/p+\alpha-1/\pi) &- (\alpha+\delta/p)(1-1/p) \\
&= (1+\alpha)(\delta'/p+\alpha-1/\pi) - (\alpha+\delta'/p)(1-1/p) \\
&= (\alpha+\delta'/p)(1/p+\alpha) - (1+\alpha)/\pi.
\end{aligned}
$$

We have $1/\delta' \leq 2/p - 1/\pi + \alpha$, and hence,

$$
\begin{aligned}
(1+\alpha)(\delta/p+\alpha-1/\pi) &- (\alpha+\delta/p)(1-1/p) \\
&\geq \left(\alpha + \frac{1}{p(2/p-1/\pi+\alpha)}\right)(1/p+\alpha) - (1+\alpha)/\pi.
\end{aligned}
$$

The right-hand side of this inequality is of the same sign as

$$
\begin{aligned}
\left[\alpha(2/p-1/\pi+\alpha)+1/p\right]&(1/p+\alpha) - (1/\pi)(1+\alpha)(2/p-1/\pi+\alpha) \\
&= (\alpha-1/\pi+1/p)\left(\alpha^2+(2/p-1/\pi)\alpha-(1/\pi-1/p)\right).
\end{aligned}
$$

Finding a root of the quadratic polynomial amounts to solving (11). The only positive root is therefore $\bar{\alpha}$. Since $\alpha > \delta'(1/\pi-1/p) \geq (1/\pi-1/p)/(2/p-1/\pi+\alpha)$, α is larger than $\bar{\alpha}$ and the above quantity is always positive.

Suppose now that $\delta' < 1$. Then, $\alpha > 1+1/\pi-2/p \geq 1/\pi-1/p$, and

$$
(1+\alpha)(\delta/p+\alpha-1/\pi) - (\alpha+\delta/p)(1-1/p) = (\alpha+1/p)^2 - (1+\alpha)/\pi.
$$

The right-hand side of this inequality is an increasing function of α on $[1/(2\pi)-1/p,+\infty) \supset [1/\pi-1/p,+\infty)$. It is therefore larger than when $\alpha = 1+1/\pi-2/p$. For such a value of α, it is equal to $(1-1/p)^2$ and is non-negative. □

Proof of Claim 3. We first suppose that $p \geq 2$. Then,

$$
\begin{aligned}
(1+\alpha)(\delta/p+\alpha-1/\pi) &- (\alpha+\delta/p)(1-1/p) \\
&= (1+\alpha)(1/2+\alpha-1/\pi) - (\alpha+1/2)(1-1/p) \\
&= (\alpha+1/2)(1/p+\alpha) - (1+\alpha)/\pi.
\end{aligned}
$$

The right-hand side is a quadratic polynomial of the form $\alpha^2+b\alpha+c$ with $b = 1/p-1/\pi+1/2 \geq 1/p-1/\pi$. It it therefore increasing on $[-b/2,+\infty) \supset [1/(2\pi)-1/(2p),+\infty)$. In particular, it is larger that when $\alpha = 1/\pi$ or when $\alpha = (p/2)(1/\pi-1/p)$. It is equal to $(-p+\pi+2)/(2\pi p)$ in the first case, and to $p(p-\pi-2)/(4\pi^2)$ in the second case. At least one of the two terms is non-negative.

We now suppose that $p < 2$. Then,

$$(1+\alpha)(\delta/p + \alpha - 1/\pi) - (\alpha + \delta/p)(1 - 1/p) = (1/p + \alpha)^2 - (1+\alpha)/\pi.$$

As in the proof of Claim 2, the right-hand side of this inequality is an increasing function of α on $[1/(2\pi) - 1/p, +\infty)$. It is therefore larger that when $\alpha = 1/\pi$, in which case it is equal to $(\pi + (2-p)p)/(p^2\pi) > 0$. $\qquad\square$

4.6. Proof of Proposition 7

We must show that the partitions m defined in the proof of Theorem 6 belong to $\mathcal{M}_\ell$ for some ℓ satisfying $2^\ell \leq n^{p/(p-1)}$ for n large enough. Moreover, we must have $2^\ell \leq n$ when $\alpha > 1/\pi$ and n large enough.

When $\alpha \leq \bar{\alpha}$ and $\alpha \leq 1/\pi$, or when $\alpha \leq (p-\pi)/(2\pi)$ and $\alpha > 1/\pi$, k is defined by (36) with $1/\delta_\alpha = 2/p - 1/\pi + \alpha$ or $1/\delta_\alpha = 2/p$. As $\ell \leq k$, we have,

$$2^\ell \leq C \left(\frac{n}{\log n}\right)^{1/(2(\alpha - 1/\pi + 1/p) + 1 - 1/\delta_\alpha)}.$$

When $1/\delta_\alpha$ is given by the first formula, the denominator in the exponent is

$$2(\alpha - 1/\pi + 1/p) + 1 - 1/\delta_\alpha = \alpha - 1/\pi + 1 \geq 1 - 1/p,$$

and $2^\ell \leq n^{p/(p-1)}$ for n large enough. When $1/\delta_\alpha = p/2$, we rather have

$$2(\alpha - 1/\pi + 1/p) + 1 - 1/\delta_\alpha = 2\alpha - 2/\pi + 1,$$

and using $\alpha > 1/\pi$, $2^\ell \leq n$ for n large enough.

Suppose now that $\alpha \in (\bar{\alpha}, 1/\pi]$, or $\alpha > 1/\pi$ and $\alpha > (p-\pi)/(2\pi)$. Then, k is defined by (32) and ℓ is bounded above by (35). In the first case, we use Claim 2 to get $\ell \leq (1+\alpha)/(1 - 1/p)$ and thus

$$2^\ell \leq C \left(\frac{n}{\log n}\right)^{(1+\alpha)/((1-1/p)(1+2\alpha))},$$

which is smaller than $n^{p/(p-1)}$ for n large. In the second case, we use the claim below to get $\ell \leq (1+2\alpha)k$ and hence $2^\ell \leq C\,(n/\log n)$. $\qquad\square$

Claim 4. *The following inequality holds true: for all $\alpha > 1/\pi$, $\pi \in (0, p)$, such that $\alpha > (p/2)(1/\pi - 1/p)$,*

$$\alpha + \delta/p < (1 + 2\alpha)(\delta/p + \alpha - 1/\pi),$$

where $\delta = \max\{1, p/2\}$.

Proof of Claim 4. We first suppose that $p \geq 2$. Then,

$$(1 + 2\alpha)(\delta/p + \alpha - 1/\pi) - (\alpha + \delta/p) = (1 + 2\alpha)(1/2 + \alpha - 1/\pi) - (\alpha + 1/2)$$
$$= 2(\alpha - 1/\pi)(\alpha + 1/2)$$
$$> 0.$$

We now suppose that $p < 2$. Then,

$$(1 + 2\alpha)(\delta/p + \alpha - 1/\pi) - (\alpha + \delta/p) = 2\alpha(\alpha + 1/p) - (1 + 2\alpha)/\pi.$$

This quadratic polynomial is increasing on $[(1/\pi - 1/p)/2, +\infty)$. The right-hand side of the above equality is therefore larger that when $\alpha = 1/\pi$, in which case it is equal to $(2 - p)/(p\pi) > 0$. □

References

[Aka12] Nathalie Akakpo. Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection. *Mathematical Methods of Statistics*, 21:1–28, 2012. MR2901269

[AL11] Nathalie Akakpo and Claire Lacour. Inhomogeneous and anisotropic conditional density estimation from dependent data. *Electronic journal of statistics*, 5:1618–1653, 2011. MR2870146

[Bar02] Yannick Baraud. Model selection for regression on a random design. *ESAIM: Probability and Statistics*, 6:127–146, 2002. MR1918295

[BB09] Yannick Baraud and Lucien Birgé. Estimating the intensity of a random measure by histogram type estimators. *Probability Theory and Related Fields*, 143:239–284, 2009. MR2449129

[BBM99] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probability theory and related fields*, 113(3):301–413, 1999. MR1679028

[BH78] Jean Bretagnolle and Catherine Huber. Estimation des densités: risque minimax. In *Séminaire de Probabilités XII*, pages 342–363. Springer, 1978. MR0520011

[Bir04] Lucien Birgé. Model selection for Gaussian regression with random design. *Bernoulli*, 10(6):1039–1051, 2004. MR2108042

[Bir06] 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

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

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013. MR3185193

[BM97] Lucien Birgé and Pascal Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, 1997. MR1462939

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

[BM00] Lucien Birgé and Pascal Massart. An adaptive compression algorithm in Besov spaces. *Constructive Approximation*, 16(1):1–36, 2000. MR1848840

[BSR04] Gilles Blanchard, Christin Schäfer, and Yves Rozenholc. Oracle bounds and exact algorithm for dyadic classification trees. In *International Conference on Computational Learning Theory*, pages 378–392. Springer, 2004. MR2177922

[BTWB10] Florentina Bunea, Alexandre B Tsybakov, Marten H Wegkamp, and Adrian Barbu. Spades and mixture models. *The Annals of Statistics*, 38(4):2525–2558, 2010. MR2676897

[Cen62] Nikolai N Cencov. Estimation of an unknown distribution density from observations. *Soviet Math.*, 3:1559–1566, 1962.

[DeV98] Ronald A DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998. MR1689432

[DJ96] Bernard Delyon and Anatoli Juditsky. On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, 3(3):215–228, 1996. MR1400080

[DJKP96] David L Donoho, Iain M Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, pages 508–539, 1996. MR1394974

[DL93] Ronald A DeVore and George G Lorentz. *Constructive approximation*, volume 303. Springer Science & Business Media, 1993. MR1261635

[Don97] David L Donoho. Cart and best-ortho-basis: a connection. *The Annals of Statistics*, 25(5):1870–1911, 1997. MR1474073

[DP88] Ronald A DeVore and Vasil A Popov. Interpolation of Besov spaces. *Transactions of the American Mathematical Society*, 305(1):397–414, 1988. MR0920166

[DY90] Ronald A DeVore and Xiang Ming Yu. Degree of adaptive approximation. *Mathematics of Computation*, 55:625–635, 1990. MR1035930

[Efr08] Sam Efromovich. Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *The Annals of Statistics*, 36(3):1127–1155, 2008. MR2418652

[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(3-4):479–543, 2014. MR3230001

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

[Kle09]   Jussi Klemelä. Multivariate histograms with data-dependent partitions. *Statistica sinica*, pages 159–176, 2009. MR2487883

[Lep92]   Oleg Lepski. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992. MR1147167

[Lep15]   Oleg Lepski. Adaptive estimation over anisotropic functional classes via oracle approach. *The Annals of Statistics*, 43(3):1178–1242, 2015. MR3346701

[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*, volume 6. Springer, 2007. MR2319879

[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

[Sar14]   Mathieu Sart. Estimation of the transition density of a Markov chain. *Annales de l'Institut Henri Poincaré. Probabilités et Statistique*, 50(3):1028–1068, 2014. MR3224298