Minimax optimality of the local multi-resolution projection estimator over Besov spaces

Jean-Baptiste Monnier

Minimax optimality of the local multi-resolution projection estimator over Besov spaces

Jean-Baptiste Monnier,
e-mail: monnier@math.jussieu.fr
Université Paris Diderot, Paris 7,
LPMA, office 5B01,
175 rue du Chevaleret,
75013, Paris, France.

February 27, 2012

Abstract

The local multi-resolution projection estimator (LMPE) has been first introduced in [7]. It was proved there that plug-in classifiers built upon the LMPE can reach super-fast rates under a margin assumption. As a by-product, the LMPE was also proved to be near minimax optimal in the regression setting over a wide generalized Lipschitz (or Hölder) scale. In this paper, we show that a direct treatment of the regression loss allows to generalize the minimax optimality of the LMPE to a much wider Besov scale. To be more precise, we prove that the LMPE is near minimax optimal over Besov spaces $B^s_{\tau,q}$, $s > 0$, $\tau \geq p$, $q > 0$, when the loss is measured in $\mathbb{L}_p$-norm, $p \in [2, \infty)$ (see Theorem 2.1), and over Besov spaces $B^s_{\tau,q}$, $s > d/\tau$, $\tau, q > 0$, when the loss is measured in $\mathbb{L}_\infty$-norm (see Theorem 2.2). Moreover, we show that an appropriate version of Lepski's method allows to make these results adaptive. Interestingly, all the proofs detailed here are largely different from the ones given in [7].

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62G05, 62G08.

KEY-WORDS: Nonparametric regression; Random design; Multi-resolution analysis

1 Introduction

1.1 The regression problem

Let us assume we dispose of a set $\mathcal{D}_n = \{(X_i, Y_i), i = 1, \ldots, n\}$ of n independent and identically distributed random vectors. They are constituted of a co-variable $X \in \mathcal{E} \subset \mathbb{R}^d$ and an associated observation $Y \in \mathbb{R}$ and are generated by the model,

$$Y = f(X) + \sigma\xi, \tag{1}$$

where f is an unknown map from $\mathcal{E}$ into $\mathbb{R}$, ξ is a standard normal random variable independent from X, which we write $\xi \sim \Phi(0, 1)$, and the law $\mathbb{P}_X$ of X admits an unknown density μ on $\mathcal{E}$.
In the regression setting, our aim is to best recover f from the data $\mathcal{D}_n$ under the assumption that

f belongs to a smoothness class $\mathcal{F}$. To be more precise, our aim is to build a map f_n from $\mathcal{E}$ into $\mathbb{R}$ upon the data $\mathcal{D}_n$ so as to minimize the loss $\mathbb{E}\|f_n - f\|^p_{\mathbb{L}_p(\mathcal{E},\mu)}$ uniformly over f. In other terms, we aim at constructing f_n such that,

$$\sup_{f\in\mathcal{F}} \mathbb{E}\|f - f_n\|^p_{\mathbb{L}_p(\mathcal{E},\mu)} \lesssim (\log n)^\delta \inf_{\theta_n} \sup_{f\in\mathcal{F}} \mathbb{E}\|f - \theta_n\|^p_{\mathbb{L}_p(\mathcal{E},\mu)}, \qquad \delta \geq 0, \qquad n \geq 0,$$

where $\lesssim$ means inferior or equal modulo a constant independent of n and the $\inf_{\theta_n}$ is taken over all the estimators θ_n of f, that is all measurable maps from $\mathcal{E}$ into $\mathbb{R}$. When $\delta = 0$ ($\delta > 0$), f_n will be said to be (near-) minimax optimal over the smoothness class $\mathcal{F}$.
Throughout the paper, we will work under the following assumption.

(**CS1**) There exist two universal constants $0 < \mu_{\min} \leq \mu_{\max} < \infty$ such that $\mu_{\min} \leq \mu(x) \leq \mu_{\max}$ for all $x \in \mathcal{E}$ and $\mathcal{E} = [0, 1]^d$.

This is a strong but classical assumption which can be found in previous works such as [2, 1, 7], for example.

1.2 Construction of the LMPE

We recall the construction of the LMPE, which was introduced in [7]. Assume we dispose of a multi-resolution analysis (MRA) of $\mathbb{L}_2(\mathbb{R}^d)$ constituted of nested approximation spaces $V_j \subset V_{j+1}$, $j \geq 0$, which reproduce polynomials of degree $r-1$ (meaning that $P_j p = p$, where P_j stands for the orthogonal projector from $\mathbb{L}_2(\mathbb{R}^d)$ into V_j and p is any polynomial of degree $r-1$). We assume that it is generated by the Daubechies' scaling function φ, meaning that $(\varphi_{j,k})_{k\in\mathbb{Z}^d}$ stands for an orthonormal basis of V_j, where we have written $\varphi_{j,k}(.) = 2^{j\frac{d}{2}}\varphi(2^j. - k)$. In the sequel, we will refer to such a MRA as a r-MRA. As described in [7], $[0, 1]^d$ can be partitioned into hypercubes of the form $\mathcal{O}_{j,k} = 2^{-j}(k + [0, 1]^d)$ for ks that belong to a well chosen subset $\mathcal{Z}^d_j$ of $\mathbb{Z}^d$. In addition, $\text{Supp}\varphi = [1 - r, r]^d$ so that there are only $m = (2r - 1)^d$ scaling functions which intercept a cell $\mathcal{O}_{j,k}$. It is crucial here to notice that m is a constant independent from both j and k. The LMPE of f at resolution level j is then obtained by performing 2^{jd} local regressions, that is one on each cell $\mathcal{O}_{j,k}$.
Under the assumption that f belongs to a Besov ball, the remainder $f - P_j f$ can be naturally controlled so that it remains to estimate $P_j f$ by localized projections onto the $\mathcal{O}_{j,k}$. As detailed above, we denote by $\varphi_{j,k_1}, \ldots, \varphi_{j,k_m}$ the scaling functions whose supports intercept the cell $\mathcal{O}_{j,k}$. For any $x \in \mathcal{O}_{j,k}$, we can write

$$P_j f(x) = \alpha_{j,k_1}\varphi_{j,k_1}(x) + \ldots + \alpha_{j,k_m}\varphi_{j,k_m}(x),$$

so that we must estimate the m coefficients α_{j,k_i} to obtain an estimator of $P_j f$ on $\mathcal{O}_{j,k}$. This is done by standard regression on $\mathcal{O}_{j,k}$. More precisely, we consider the least-squares problem,

$$LS(\mathcal{O}_{j,k}) = \arg \min_{(a_1,\ldots,a_t)\in\mathbb{R}^m} \sum_{i=1}^{n} \left(Y_i - \sum_{t=1}^{m} a_t \varphi_{j,k_t}(X_i) \right)^2 \mathbb{1}_{\mathcal{O}_{j,k}}(X_i).$$

Solving the above least-squares problem boils down to the inversion of the local regression matrix $G \in \mathbb{R}^m \times \mathbb{R}^m$ whose coefficients are the,

$$G_{\ell,\ell'} = \frac{1}{n} \sum_{i=1}^{n} \varphi_{j,k_\ell}(X_i)\varphi_{j,k_{\ell'}}(X_i)\mathbb{1}_{\mathcal{O}_{j,k}}(X_i). \tag{2}$$

Obviously, if G is invertible then $LS(\mathcal{O}_{j,k})$ holds one single element. Finally, we define the LMPE $f_j^{@}$ by spectral thresholding of the smallest eigenvalue of the local regression matrix $\lambda_{\min}(G)$ at level π_n^{-1}, where we can take $\pi_n = \log n$. We set indeed $f_j^{@} = 0$ if $\lambda_{\min}(G) < \pi_n^{-1}$ and $f_j^{@}(x) = \sum_{t=1}^m a_t^{@} \varphi_{j,k_t}(x)$ for all $x \in \mathcal{O}_{j,k}$ otherwise, where $a^{@}$ is the unique element of $LS(\mathcal{O}_{j,k})$.

2 Results

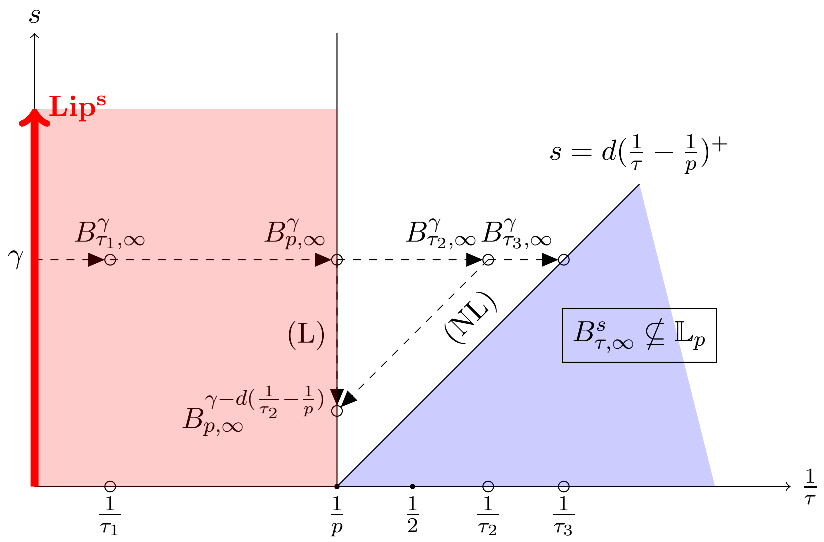

Figure 1: Each point of this figure at coordinates $(s, 1/\tau)$ stands for a Besov space $B^s_{\tau,\infty}(\mathcal{E})$. Assume that the regression loss is measured in $\mathbb{L}_p$-norm. Generalized Lipschitz spaces $Lip^s(\mathcal{E})$ are defined to be the Besov spaces $B^s_{\infty,\infty}(\mathcal{E})$ and thus correspond to the vertical coordinate axis $1/\tau = 0$ (see [8]). Besov spaces $B^s_{\tau,\infty}(\mathcal{E})$ get larger and larger along the arrows drawn on the figure. In [7, Corollary 7.1] the LMPE was proved to be near minimax optimal over generalized Lipschitz spaces (thick vertical red line). Here we show that these results can be extended to the whole domain $\{(1/\tau, s) : 1/\tau \leq 1/p\}$.

In what follows, we prove that the LMPE is minimax optimal in the regression setting over Besov spaces. This is a generalization of the results described in [7, Corollary 7.1], where the LMPE was proved to be minimax optimal over generalized Lipschitz spaces $Lip^s(\mathcal{E})$, which are slightly larger than Hölder spaces. To be more precise, we prove in what follows that the LMPE is near minimax optimal over Besov spaces $B^s_{\tau,q}(\mathcal{E})$, $s > 0$, $\tau \geq p$, $q > 0$, when the loss is measured in $\mathbb{L}_p$-norm, $p \in [2, \infty)$ (see Theorem 2.1), and over Besov spaces $B^s_{\tau,q}(\mathcal{E})$, $s > d/\tau$, $\tau, q > 0$, when the loss is measured in $\mathbb{L}_\infty$-norm (see Theorem 2.2). As described in Figure 1, these smoothness classes are much larger than the Lipschitz spaces. In addition, it is well-known that, outside this domain, linear estimators such as the LMPE are necessarily sub-optimal (see [4]). So that this is the largest domain on which we can obtain the near-minimax optimality of the LMPE (up to the constraint $p \geq 2$).
Interestingly, the proof of the initial result [7, Corollary 7.1] hinged on an integration of the exponential control of the probability of deviation of the LMPE $f_j^{@}(x)$ from $f(x)$ uniformly when $x \in \mathcal{E}$ (see [7, Theorem 7.1]). The latter exponential control is very stringent and was initially introduced to prove

results in the classification setting. It is however unnecessary in the regression setting and can thus be relaxed. The following proofs hinge in fact on a direct inspection of the integrated $\mathbb{L}_p$-loss and are thus obtained by very different means from the ones used in [7]. In particular, they rely on the recombination of the integrated approximation errors over the regression cells $\mathcal{O}_{j,k}$ of the partition of $\mathcal{E}$ (see proof of Theorem 2.1).

In what follows, and for any $\tau > 0$, we define $j_r, j_s, j_{s,\tau}, J$ such that

$$2^{j_r} \sim n^{\frac{1}{2r+d}}, \qquad 2^{j_s} \sim n^{\frac{1}{2s+d}}, \qquad 2^{j_{s,\tau}} \sim n^{\frac{1}{2(s+\frac{d}{2}-\frac{d}{\tau})}}, \qquad 2^{Jd} \sim \frac{n}{\kappa \pi_n^2 \log n},$$

where $\kappa > 0$ is a constant to be made more precise later. By $a_n \sim b_n$, we mean that there exist two absolute constants independent of n such that $ca_n \le b_n \le Ca_n$. Furthermore we define the grid $\mathcal{J}_n$ of resolution levels such that $\mathcal{J}_n = \{j_r, j_r+1, \dots, J\}$. Of course, $j_s, j_{s,\tau} \in \mathcal{J}_n$ under the assumption that $s \in (0, r)$.

In Theorem 2.1, we show that the LMPE is nearly minimax optimal in $\mathbb{L}_p$-norm, $p \in [2, \infty)$ over a wide Besov scale. This result contains [7, Corollary 7.1] as a particular case (take $\tau = q = \infty$) and thus generalizes it to larger smoothness classes.

Theorem 2.1. *Assume* (**CS1**) *holds true. Let $f_j^{@}$ stand for the local multi-resolution estimator of f defined in Section 1.2 and $B^s_{\tau,q}(\mathcal{E}, M)$ for the ball of radius $M > 0$ of $B^s_{\tau,q}(\mathcal{E})$. Then for all $j \in \mathcal{J}_n$, $p \in [2, \infty)$, $s > 0$, $\tau \ge p$ and $q > 0$, we have the following result,*

$$\sup_{f \in B^s_{\tau,q}(\mathcal{E},M)} \mathbb{E}\|f - f_j^{@}\|^p_{\mathbb{L}_p(\mathcal{E},\mu)} \le C(p)\pi_n^p \left[\left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}} \right)^p + 2^{-jsp} \right],$$

where the constant $C(p)$ is given explicitly in the proof. In particular, for $j = j_s$, we obtain

$$\sup_{f \in B^s_{\tau,q}(\mathcal{E},M)} \mathbb{E}\|f - f_{j_s}^{@}\|^p_{\mathbb{L}_p(\mathcal{E},\mu)} \le C(p)\pi_n^p n^{-\frac{sp}{2s+d}}.$$

The corresponding lower-bound result is well-known (see [10, 5, 7]), which finishes to prove the minimax optimality of the LMPE.

In Theorem 2.2, we show that the LMPE is nearly minimax optimal in sup-norm over a wide Besov scale. Notice interestingly that, until now, there were no results available in sup-norm for the LMPE.

Theorem 2.2. *Assume* (**CS1**) *holds true. Let $f_j^{@}$ stand for the local multi-resolution estimator of f defined in Section 1.2 and $B^s_{\tau,q}(\mathcal{E}, M)$ for the ball of radius $M > 0$ of $B^s_{\tau,q}(\mathcal{E})$. For all $f \in B^s_{\tau,q}(\mathcal{E}, M)$, $s > d/\tau$,*

$$\sup_{f \in B^s_{\tau,q}(\mathcal{E},M)} \mathbb{E}\|f - f_j^{@}\|^z_{\mathbb{L}_\infty(\mathcal{E})} \le C(z)\pi_n^z [\log n]^{\frac{z}{2}} \left(\left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}} \right)^z + 2^{-j(s-\frac{d}{\tau})z} \right),$$

where the constant $C(z)$ is given explicitly in the proof. In particular, for $j = j_{s,\tau}$, we obtain

$$\sup_{f \in B^s_{\tau,q}(\mathcal{E},M)} \mathbb{E}\|f - f_{j_{s,\tau}}^{@}\|^z_{\mathbb{L}_\infty(\mathcal{E})} \le C(z)\pi_n^z (\log n)^{\frac{z}{2}} n^{-\frac{z(s-\frac{d}{\tau})}{2(s+\frac{d}{2}-\frac{d}{\tau})}}.$$

The corresponding lower-bound result is well-known (see [10, 5, 7]), which finishes to prove the minimax optimality of the LMPE.

Both theorems prove the optimality of the estimators $f^{@}_{j_s}$ and $f^{@}_{j_{s,\tau}}$, whose resolution levels depend on the regularity s of the unknown regression function f. In the following theorem, we show that this dependence in s can be removed at the cost of a $\log n$ factor in the upper-bounds. The LMPE is thus said to be adaptive near minimax optimal. To be more explicit, the adaptation hinge on the data-driven choice $j^{@}$ of the global resolution level j of the LMPE (see Theorem 2.3), which itself hinges on a slightly modified version of Lepski's method (see [6]).

Theorem 2.3. *Assume* (**CS1**) *holds true. For all $f \in B^s_{\tau,q}(\mathcal{E}, M)$, $s, q > 0$, $\tau \geq p$, $p \geq 2$, we define the global resolution index $j^{@}$ from the data $\mathcal{D}_n$ as follows for $\gamma > 0$,*

$$j^{@} = \inf\left\{ j \in \mathcal{J}_n : \|f^{@}_j - f^{@}_k\|_{\mathbb{L}_p} \leq \gamma \log n \left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}} + \frac{2^{k\frac{d}{2}}}{\sqrt{n}} \right), \forall k > n, k \in \mathcal{J}_n \right\}.$$

Then, for γ large enough, we obtain,

$$\mathbb{E}\|f^{@}_{j^{@}} - f\|^p_{\mathbb{L}_p(\mathcal{E},\mu)} \lesssim \gamma^p (\log n)^p \pi_n^p n^{-\frac{sp}{2s+d}}, \qquad p \in [2, \infty),$$

$$\mathbb{E}\|f^{@}_{j^{@}} - f\|^z_{\mathbb{L}_\infty(\mathcal{E})} \lesssim \gamma^z (\log n)^z \pi_n^z n^{-\frac{z(s-\frac{d}{\tau})}{2(s+\frac{d}{2}-\frac{d}{\tau})}}.$$

So that $f^{@}_{j^{@}}$ is an adaptive near minimax optimal estimator of f.

In what follows, we detail the proofs of these three theorems. In order to alleviate notations, we will denote throughout the sequel by $\mathcal{H}$ a cell $\mathcal{O}_{j,k}$ of the partition of $[0,1]^d$ at level j and by $G_{\mathcal{H}}$ the associated local regression matrix (see eq. (2)). Furthermore, we will denote by $\mathcal{F}_j$ the partition of $[0,1]^d$ into hyper-cubes $\mathcal{O}_{j,k}$ at resolution level j.

3 Proofs

Remark 3.1. *Recall that for any $p \in [1, \infty]$, there exist two absolute constants c_p, C_p such that*

$$c_p 2^{jd(\frac{1}{2}-\frac{1}{p})} \leq \|\varphi_{j,k}\|_{\mathbb{L}_p} \leq C_p 2^{jd(\frac{1}{2}-\frac{1}{p})}.$$

Throughout the paper, constants of the form c_p, C_p will exclusively refer to these constants. Furthermore, we will write $\vartheta_p := \mathbb{E}|\xi|^p$, where the noise ξ has been defined in eq. (1).

3.1 Proof of Theorem 2.1

Proof of Theorem 2.1. Notice first that the regression loss can be broken up as follows,

$$\mathbb{E}\|f - f^{@}_j\|^p_{\mathbb{L}_P} \leq \sum_{\mathcal{H}} \mathbb{E}\|f - f^{@}_j\|^p_{\mathbb{L}_P(\mathcal{H})} \mathbb{1}_{\{\lambda_{\min}(G_{\mathcal{H}}) < \pi_n^{-1}\}},$$
$$+ 2^{p-1} \sum_{\mathcal{H}} \mathbb{E}\|P_j f - f^{@}_j\|^p_{\mathbb{L}_P(\mathcal{H})} \mathbb{1}_{\{\lambda_{\min}(G_{\mathcal{H}}) \geq \pi_n^{-1}\}} + 2^{p-1}\|R_j f\|^p_{\mathbb{L}_P}.$$

As described in the proof of [7, Theorem 7.1], we have, for $f \in B^s_{\tau,q}(\mathcal{E}, M)$,

$$\|f - f^{@}_j\|^p_{\mathbb{L}_p(\mathcal{H})} \mathbb{1}_{\{\lambda_{\min}(G_{\mathcal{H}}) < \pi_n^{-1}\}} \leq M^p \mu(\mathcal{H}) \mathbb{1}_{\{\lambda_{\min}(G_{\mathcal{H}}) < \pi_n^{-1}\}},$$
$$|P_j f(x) - f^{@}_j(x)|^p \mathbb{1}_{\{\lambda_{\min}(G_{\mathcal{H}}) \geq \pi_n^{-1}\}} \leq \pi_n^p \|W_{\mathcal{H}}\|^p_{\ell_2} \|\varphi_{\mathcal{H}}(x)\|^p_{\ell_2}, \qquad \forall x \in \mathcal{H}, \tag{3}$$

where $W_{\mathcal{H}} = (W_{j,k_1}, \dots, W_{j,k_m})$, $\varphi_{\mathcal{H}}(x) = (\varphi_{j,k_1}(x), \dots, \varphi_{j,k_m}(x))$, the φ_{j,k_i}, $1 \le i \le m$, are the scaling functions whose supports intercept $\mathcal{H}$ and we have written

$$W_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Z_i^{j,k}, \qquad Z_i^{j,k} = \varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)(R_j f(X_i) + \xi_i).$$

So that, putting everything together, we obtain,

$$\begin{aligned}
\mathbb{E}\|f - f_j^{@}\|_{\mathbb{L}_P}^p &\le M^p n \sup_{\mathcal{H}\in\mathcal{F}_j} \mathbb{P}(\lambda_{\min}(G_{\mathcal{H}}) < \pi_n^{-1}) \\
&\quad + 2^{p-1}\pi_n^p \sum_{\mathcal{H}} \left(\int_{\mathcal{H}} \|\varphi_{\mathcal{H}}(x)\|_{\ell_2}^p \mu(x)dx\right) \mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \\
&\quad + 2^{p-1}\|R_j f\|_{\mathbb{L}_P}^p \\
&= I + II + III.
\end{aligned}$$

I is of the right order since, according to Lemma 3.3, it can be made as small as desired for κ large enough. III is also of the right order according to Lemma 4.1. Let us now turn to II. Write $\mathcal{S}_j(\mathcal{H}) = \{k_i, i = 1, \dots, m\}$ and notice that, with $m = \#\mathcal{S}_j(\mathcal{H})$, we have,

$$\|\varphi_{\mathcal{H}}(x)\|_{\ell_2}^p \le \|\varphi_{\mathcal{H}}(x)\|_{\ell_1}^p \le m^{p-1}\|\varphi_{\mathcal{H}}(x)\|_{\ell_p}^p.$$

We thus obtain

$$\begin{aligned}
II &\le \mu_{\max} 2^{p-1}\pi_n^p \sum_{\mathcal{H}} \left(\sum_{k\in\mathcal{S}_j(\mathcal{H})} \|\varphi_{j,k}\|_{\mathbb{L}_p}^p\right) \mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \\
&\le \mu_{\max} 2^{p-1}\pi_n^p m C_p^p 2^{jd(\frac{p}{2}-1)} \mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \\
&\le C(p, \mu_{\max}, m, \vartheta_p, p)\pi_n^p \left(\left[\left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}}\right)^{2(p-1)} + \left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}}\right)^{p}\right](1 + \|R_j f\|_{\mathbb{L}_p}^p) + \|R_j f\|_{\mathbb{L}_p}^p\right),
\end{aligned}$$

where the last inequality follows from Proposition 3.1, eq. (5) and we can choose

$$\begin{aligned}
C(p, \mu_{\max}, m, \vartheta_p, p) &= C^{(1)} \vee C^{(2)} \vee C^{(3)}, \\
C^{(1)} &= \mu_{\max}^2 m^{p+1} 2^{2(p-1)} C_p^p C_{p'}^p, \\
C^{(2)} &= \mu_{\max}^2 m^{p+1} 2^{4(p-1)} C_p^p C_\infty^p (1 \vee \vartheta_p), \\
C^{(3)} &= \mu_{\max}^2 m^{p+1} 2^{\frac{7}{2}p-3} C_p^p (C_{2[p/2]'}^p \vee [C_2^p \vartheta_p^{\frac{p}{2}}]).
\end{aligned}$$

To conclude, notice that for $j \in \mathcal{J}_n$, $2^{jd/2}/\sqrt{n} \le 1$, $2(p-1) \ge p$ since $p \ge 2$ and $\|R_j f\|_{\mathbb{L}_p} \le M2^{-js}$ since $f \in B_{\tau,q}^s(\mathcal{E}, M)$ and $\tau \ge p$ (see Lemma 4.1). In particular, we can choose

$$C(p) = (C(p, \mu_{\max}, m, \vartheta_p, p) + 1)(1 \vee M^p).$$

□

3.2 Useful results for the proof of Theorem 2.1

Proposition 3.1. *Remark that we have*

$$\mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \lesssim n^{-p+1}\left(2^{jd\frac{p}{2}}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 2^{jd(\frac{p}{2}-1)}\right) + n^{-\frac{p}{2}}\left(2^{jd}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 1\right) + 2^{jd(1-\frac{p}{2})}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p, \tag{4}$$

where the upper-bound constants are detailed in the proof. We obtain the following result as a direct consequence of eq. (4),

$$\sum_{\mathcal{H}}\mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \lesssim n^{-p+1}\left(2^{jd\frac{p}{2}}\|R_jf\|_{\mathbb{L}_p}^p + 2^{jd(\frac{p}{2})}\right) + n^{-\frac{p}{2}}\left(2^{jd}\|R_jf\|_{\mathbb{L}_p}^p + 2^{jd}\right) + 2^{jd(1-\frac{p}{2})}\|R_jf\|_{\mathbb{L}_p}^p. \tag{5}$$

Proof. Recall first that we have written $m = \#\mathcal{S}_j(\mathcal{H}) = (2r-1)^d$. It is enough to notice that $\|.\|_{\ell^2} \leq \|.\|_{\ell^1}$ and use Jensen inequality to get,

$$\mathbb{E}\|W_{\mathcal{H}}\|_{\ell_2}^p \leq m^{p-1}\sum_{k\in\mathcal{S}_j(\mathcal{H})}\mathbb{E}|W_{j,k}|^p.$$

In addition, a triangular inequality leads to,

$$\mathbb{E}|W_{j,k}|^p \leq 2^{p-1}n^{-p}\mathbb{E}|\sum_{i=1}^n\left(Z_i^{j,k} - \mathbb{E}Z_i^{j,k}\right)|^p + 2^{p-1}|\mathbb{E}Z_i^{j,k}|^p.$$

Now, eq. (4) follows from eq. (10) and eq. (13) in Lemma 3.1 below, and the constants can be read off these latter equations. □

Lemma 3.1. *We have got the following four inequalities,*

$$|\mathbb{E}Z_i^{j,k}|^p \lesssim 2^{jd(1-\frac{p}{2})}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p, \tag{6}$$

$$\mathbb{E}|Z_i^{j,k}|^p \lesssim 2^{jd\frac{p}{2}}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 2^{jd(\frac{p}{2}-1)}, \tag{7}$$

$$\left(\mathbb{E}|Z_i^{j,k}|^2\right)^{\frac{p}{2}} \lesssim 2^{jd}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 1, \quad p \geq 2, \tag{8}$$

$$\mathbb{E}|\sum_{i=1}^n\left(Z_i^{j,k} - \mathbb{E}Z_i^{j,k}\right)|^p \lesssim n\left(2^{jd\frac{p}{2}}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 2^{jd(\frac{p}{2}-1)}\right) + n^{\frac{p}{2}}\left(2^{jd}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p + 1\right), \tag{9}$$

where the upper-bound constants can be read from eq. (10), eq. (11), eq. (12) and eq. (13) respectively. Notice interestingly that all upper bounds are independent of k.

Proof. The proof of eq. (6) follows from Hölder's inequality with $(p')^{-1} + p^{-1} = 1$. Notice indeed that,

$$\begin{aligned}|\mathbb{E}Z_i^{j,k}|^p &= |\mathbb{E}\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)R_jf(X_i)|^p \\ &\leq \mu_{\max}\|\varphi_{j,k}\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_1}^p \\ &\leq \mu_{\max}\|\varphi_{j,k}\|_{\mathbb{L}_{p'}}^p\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p \\ &\leq \mu_{\max}C_{p'}^p 2^{jd(1-\frac{p}{2})}\|\mathbb{1}_{\mathcal{H}}R_jf\|_{\mathbb{L}_p}^p. \end{aligned} \tag{10}$$

The proof of eq. (7) is a straightforward consequence of the triangular inequality. Notice indeed that,

$$\begin{aligned}\mathbb{E}|Z_i^{j,k}|^p &= \mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)(R_jf(X_i)+\xi_i)|^p\\ &\leq 2^{p-1}\mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)R_jf(X_i)|^p + 2^{p-1}\mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)\xi_i|^p\\ &\leq \mu_{\max}2^{p-1}\|\varphi_{j,k}\|^p_{\mathbb{L}_\infty}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p} + \vartheta_p\mu_{\max}\|\varphi_{j,k}\mathbb{1}_{\mathcal{H}}\|^p_{\mathbb{L}_p}\\ &\leq \mu_{\max}2^{p-1}C_\infty^p 2^{jd\frac{p}{2}}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p} + \vartheta_p\mu_{\max}2^{p-1}C_\infty^p2^{jd(\frac{p}{2}-1)}. \qquad (11)\end{aligned}$$

The proof of eq. (8) is a direct consequence of Hölder's inequality with $([p/2]')^{-1}+(p/2)^{-1}=1$, $p\geq 2$. Notice indeed that

$$\begin{aligned}\left(\mathbb{E}|Z_i^{j,k}|^2\right)^{\frac{p}{2}} &\leq 2^{p-1}\left(\mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)R_jf(X_i)|^2\right)^{\frac{p}{2}} + 2^{p-1}\vartheta_2^{\frac{p}{2}}\left(\mathbb{E}|\varphi_{j,k}(X_i)|^2\right)^{\frac{p}{2}}\\ &\leq \mu_{\max}2^{p-1}\|\varphi_{j,k}\|^p_{\mathbb{L}_{2[p/2]'}}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p} + \mu_{\max}C_2^p2^{p-1}\vartheta^{p/2} \qquad \text{(Hölder)}\\ &\leq \mu_{\max}2^{p-1}C^p_{2[p/2]'}2^{jd}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p} + \mu_{\max}C_2^p2^{p-1}\vartheta^{p/2}, \qquad (12)\end{aligned}$$

where the last inequality follows from the fact that

$$\|\varphi_{j,k}\|^p_{\mathbb{L}_{2[p/2]'}} \leq C^p_{2[p/2]'}2^{jd(\frac{1}{2}-\frac{1}{2[p/2]'})p} = C^p_{2[p/2]'}2^{jd}.$$

Let us now turn to the proof of eq. (9). Notice first that a direct application of Rosenthal's inequality leads to,

$$\begin{aligned}\mathbb{E}|\sum_{i=1}^n\left(Z_i^{j,k}-\mathbb{E}Z_i^{j,k}\right)|^p &\leq \sum_{i=1}^n\mathbb{E}|Z_i^{j,k}-\mathbb{E}Z_i^{j,k}|^p + \left(\sum_{i=1}^n\mathbb{E}|Z_i^{j,k}-\mathbb{E}Z_i^{j,k}|^2\right)^{p/2}\\ &\leq n2^{p-1}\mathbb{E}|Z_i^{j,k}|^p + n^{\frac{p}{2}}2^{\frac{p}{2}}\left(\mathbb{E}|Z_i^{j,k}|^2\right)^{p/2} \qquad \text{(Jensen)}\\ &\leq n\mu_{\max}2^{2(p-1)}C_\infty^p(1\vee\vartheta_p)(2^{jd\frac{p}{2}}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p^p} + 2^{jd(\frac{p}{2}-1)})\\ &+ n^{\frac{p}{2}}\mu_{\max}2^{\frac{3p}{2}-1}(C^p_{2[p/2]'}\vee[C_2^p\vartheta^{\frac{p}{2}}])(2^{jd}\|\mathbb{1}_{\mathcal{H}}R_jf\|^p_{\mathbb{L}_p^p}+1), \qquad (13)\end{aligned}$$

where the last inequality follows from both eq. (11) and eq. (12). ☐

3.3 Proof of Theorem 2.2

Proof of Theorem 2.2. Let us denote by $\mathcal{F}_j$ the partition of $[0,1]^d$ associated to the r-MRA at hand and by $\mathcal{H}$ a generic cell of this partition. Start with

$$\begin{aligned}\mathbb{E}\|f-f_j^@\|^z_{\mathbb{L}_\infty} &= \mathbb{E}\sup_{\mathcal{H}\in\mathcal{F}_j, x\in\mathcal{H}}|f(x)-f_j^@(x)|^z\\ &\leq \mathbb{E}\sup_{\mathcal{H}\in\mathcal{F}_j, x\in\mathcal{H}}|P_jf(x)-f_j^@(x)|^z + \|R_jf\|^z_{\mathbb{L}_\infty}\end{aligned}$$

The last term is of the right order according to Lemma 4.1. Focus now on the first one. Using the same reasoning (and notations) as at the beginning of the proof of Theorem 2.1, we obtain

$$\begin{aligned}\mathbb{E}\sup_{\mathcal{H}\in\mathcal{F}_j, x\in\mathcal{H}}|f_j(x)-f_j^@(x)|^z &\leq M\mathbb{P}(\inf_{\mathcal{H}\in\mathcal{F}_j}\lambda_{\min}(G_{\mathcal{H}})\leq\pi_n^{-1})\\ &+ \mathbb{E}\sup_{\mathcal{H}\in\mathcal{F}_j}\pi_n^z\|W_{\mathcal{H}}\|^z_{\ell_2}\sup_{x\in\mathcal{H}}\|\varphi_{\mathcal{H}}(x)\|^z_{\ell_2}\\ &= I+II.\end{aligned}$$

A direct application of Lemma 3.3 shows that I is of the right order for κ large enough. Let us now turn to II and notice that

$$\begin{aligned} II &\leq m^{\frac{z}{2}} C_\infty^z 2^{jd\frac{z}{2}} \pi_n^z \mathbb{E} \sup_{\mathcal{H}} \|W_{\mathcal{H}}\|_{\ell_2}^z \\ &\leq m^z C_\infty^z 2^{jd\frac{z}{2}} \pi_n^z \mathbb{E} \sup_{\mathcal{H}\in\mathcal{F}_j, k\in\mathcal{S}_j(\mathcal{H})} |W_{j,k}|^z. \end{aligned}$$

Notice besides that

$$|W_{j,k}|^z \leq 2^{z-1} |\frac{1}{n}\sum_{i=1}^n Z_i^{j,k} - \mathbb{E}Z_i^{j,k}|^z + 2^{z-1}|\mathbb{E}Z_i^{j,k}|^z.$$

The rhs is handled directly thanks to eq. (14). As regards the lhs, notice that $\#\{\mathcal{H}\in\mathcal{F}_j, k\in\mathcal{S}_j(\mathcal{H})\} = m2^{jd}$ and, given eq. (15), apply Proposition 4.1 (with $K = 2^{j\frac{d}{2}}$) conclude that

$$\mathbb{E} \sup_{\mathcal{H}\in\mathcal{F}_j, k\in\mathcal{S}_j(\mathcal{H})} |\frac{1}{n}\sum_{i=1}^n (Z_i^{j,k} - \mathbb{E}Z_i^{j,k})|^z \leq \left(\frac{2^{j\frac{d}{2}} \log(m2^{jd} + e^{z-1})}{n} + \sqrt{\frac{2\log(m2^{jd}+e^{z-1})}{n}} \right)^z.$$

Finally, putting everything together, we obtain

$$\mathbb{E}\|f - f_j^@\|_{\mathbb{L}_\infty}^z \lesssim \pi_n^z \|R_j f\|_{\mathbb{L}_\infty}^z + (\log n)^{\frac{z}{2}} \left(\frac{2^{j\frac{d}{2}}}{\sqrt{n}}\right)^z.$$

□

3.4 Useful results for the proof of Theorem 2.2

Lemma 3.2. *We have the following inequalities*

$$\mathbb{E}|Z_i^{j,k}|^z \leq C_1^z 2^{-jd\frac{z}{2}} \|R_j f\|_{\mathbb{L}_\infty}^z, \tag{14}$$

$$\mathbb{E}|Z_i^{j,k} - \mathbb{E}Z_i^{j,k}|^z \leq C(z)(2^{j\frac{d}{2}})^{z-2}(1 + \|R_j f\|_{\mathbb{L}_\infty}^z), \tag{15}$$

where we have written $C(z) = \mu_{\max} 2^{2z-1} C_z^z (\vartheta_z \vee 1)$.

Proof. For the first inequality,

$$\begin{aligned} |\mathbb{E}Z_i^{j,k}|^z &= |\mathbb{E}\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)R_j f(X_i)|^z \\ &\leq \mu_{\max} \|R_j f\|_{\mathbb{L}_\infty}^z \|\varphi_{j,k}\|_{\mathbb{L}_1}^z \\ &\leq \mu_{\max} C_1^z 2^{-jd\frac{z}{2}} \|R_j f\|_{\mathbb{L}_\infty}^z. \end{aligned}$$

For the second one,

$$\begin{aligned} \mathbb{E}|Z_i^{j,k} - \mathbb{E}Z_i^{j,k}|^z &\leq 2^z \mathbb{E}|Z_i^{j,k}|^z \\ &\leq 2^{2z-1}\mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)R_j f(X_i)|^z + 2^{2z-1}\mathbb{E}|\varphi_{j,k}(X_i)\mathbb{1}_{\mathcal{H}}(X_i)\xi_i|^z \\ &\leq \mu_{\max} 2^{2z-1} C_z^z 2^{jd(\frac{z}{2}-1)} \|R_j f\|_{\mathbb{L}_\infty}^z + \mu_{\max} 2^{2z-1} \vartheta_z C_z^z 2^{jd(\frac{z}{2}-1)} \\ &= \mu_{\max} 2^{2z-1} C_z^z (2^{j\frac{d}{2}})^{z-2} (1\vee\vartheta_z)(\|R_j f\|_{\mathbb{L}_\infty}^z + 1). \end{aligned}$$

□

Lemma 3.3. *Let us denote by $\mathcal{F}_j$ the partition of $[0,1]^d$ into hyper-cubes $\mathcal{O}_{j,k}$ at resolution level j. For $j \in \mathcal{J}_n = \{j_r, \dots, J\}$ where $2^{Jd} = n/(\kappa\pi_n^2 \log n)$, and any $\mathcal{H} \in \mathcal{F}_j$, and for*

$$\kappa \geq \widetilde{\kappa}(2\mu_{\max} m^4 + \frac{4}{3}m^2), \qquad\qquad \widetilde{\kappa} > 0,$$

we have got

$$\mathbb{P}(\lambda_{\min}(Q_{\mathcal{H}}) \leq \pi_n^{-1}) \leq 2m^2 n^{-\widetilde{\kappa}},$$
$$\mathbb{P}(\inf_{\mathcal{H}\in\mathcal{F}_j} \lambda_{\min}(Q_{\mathcal{H}}) \leq \pi_n^{-1}) \leq 2m^2 n^{-\widetilde{\kappa}+1}.$$

Proof. We know from [7, Proposition 12.3] that for all $0 < t \leq \frac{g_{\min}}{2}$,

$$\mathbb{P}(\lambda_{\min}(Q_{\mathcal{H}}) \leq t) \leq 2m^2 \exp(-n2^{-jd}\frac{t^2}{2\mu_{\max}m^4 + \frac{4}{3}m^2 t}).$$

For n large enough, we have $\pi_n \leq \frac{g_{\min}}{2}$ and the result follows directly from the latter inequality. □

3.5 Proof of Theorem 2.3

Proof of Theorem 2.3. Assume first that $p \in [2, \infty)$ and write,

$$\begin{aligned}\|f - f^{@}_{j^@}\|^p_{\mathbb{L}_p} &= \|f - f^{@}_{j^@}\|^p_{\mathbb{L}_p}\mathbb{1}_{\{j^@ \leq j_s\}} + \|f - f^{@}_{j^@}\|^p_{\mathbb{L}_p}\mathbb{1}_{\{j^@ > j_s\}}\\ &= I + II.\end{aligned}$$

As regards I, notice that

$$\begin{aligned}I &\leq 2^{p-1}\|f - f^{@}_{j_s}\|^p_{\mathbb{L}_p} + 2^{p-1}\|f^{@}_{j_s} - f^{@}_{j^@}\|^p_{\mathbb{L}_p}\mathbb{1}_{\{j^@ \leq j_s\}}\\ &\leq 2^{p-1}\|f - f^{@}_{j_s}\|^p_{\mathbb{L}_p} + 2^{p-1}2\gamma^p(\log n)^p\pi_n^p\left(\frac{2^{j_s\frac{d}{2}}}{\sqrt{n}}\right)^p.\end{aligned}$$

The left term is of the right order thanks to Theorem 2.1 and the last term as well by definition of j_s. As regards II, we proceed as in the proof of [7, Theorem 7.2] and notice that,

$$\begin{aligned}\mathbb{E}\|f - f^{@}_{j^@}\|^p_{\mathbb{L}_p}\mathbb{1}_{\{j^@ > j_s\}} &\leq \sum_{j > j_s}\mathbb{E}\|f - f^{@}_{j}\|^p_{\mathbb{L}_p}\mathbb{1}_{\{j^@ = j\}}\\ &\leq \sum_{j > j_s}\sqrt{\mathbb{E}\|f - f^{@}_{j}\|^{2p}_{\mathbb{L}_p}}\sqrt{\mathbb{P}(j^@ = j)}.\end{aligned}$$

The expectation is bounded by a constant according to Theorem 2.1. And we have besides

$$\begin{aligned}\mathbb{P}(j^@ = j) &\leq \sum_{k\geq j}\mathbb{P}\left(\|f^{@}_{j} - f^{@}_{k}\|_{\mathbb{L}_p} > \gamma(\log n)\pi_n\left[\frac{2^{j\frac{d}{2}}}{\sqrt{n}} + \frac{2^{k\frac{d}{2}}}{\sqrt{n}}\right]\right)\\ &\leq \sum_{k\geq j}\mathbb{P}\left(\|f^{@}_{j} - f\|_{\mathbb{L}_p} > \gamma(\log n)\pi_n\frac{2^{j\frac{d}{2}}}{\sqrt{n}}\right) + \mathbb{P}\left(\|f^{@}_{k} - f\|_{\mathbb{L}_p} > \gamma(\log n)\pi_n\frac{2^{k\frac{d}{2}}}{\sqrt{n}}\right),\end{aligned}$$

which can be made as small as desired for γ large enough according to Proposition 3.2. For $p = \infty$, the proof follows the same lines with $j_{s,\tau}$ in place of j_s. □

3.6 Useful results for the proof of Theorem 2.3

Proposition 3.2. *For all $f \in B^s_{\tau,q}(\mathcal{E}, M)$, $s, q > 0$, $\tau \geq p$, $p \geq 2$ and all $j \geq j_s$ (resp. $j \geq j_{s,\tau}$) for $p < \infty$ (resp. $p = \infty$), we have*

$$\mathbb{P}(\|f_j^{@} - f\|_{\mathbb{L}_p} \geq \gamma(\log n)\pi_n \frac{2^{j\frac{d}{2}}}{\sqrt{n}}) \leq mn^{-C(p,r,\sigma)\gamma} n^{\mathbb{1}_{\{p=\infty\}}},$$

where the constant $C(p, r, \sigma)$ can be found in the proof. In any case, the above term can thus be made as small as possible for γ large enough.

Proof. **Case where** $p \in [2, \infty)$**.** Recall that $\#\mathcal{F}_j = 2^{jd}$ and notice first that for $\delta/2 \geq \|Rf_j\|_{\mathbb{L}_p}$,

$$\begin{aligned}
\mathbb{P}(\|f_j^{@} - f\|_{\mathbb{L}_p} \geq \delta) &\leq \mathbb{P}(\|f_j^{@} - f_j\|^p_{\mathbb{L}_p} \geq \delta^p 2^{-p}) \\
&= \mathbb{P}(\sum_{\mathcal{H}} \|f_j^{@} - f_j\|^p_{\mathbb{L}_p(\mathcal{H})} \geq \delta^p 2^{-p}) \\
&= \sum_{\mathcal{H}} \mathbb{P}(\|f_j^{@} - f_j\|^p_{\mathbb{L}_p(\mathcal{H})} \geq \delta^p 2^{-jd} 2^{-p}) \\
&\leq \sum_{\mathcal{H}} \mathbb{P}(\|f_j^{@} - f_j\|^p_{\mathbb{L}_p(\mathcal{H})} \geq \delta^p 2^{-jd} 2^{-p}, \lambda_{\min}(G_{\mathcal{H}}) \geq \pi_n^{-1}) \\
&+ \sum_{\mathcal{H}} \mathbb{P}(\lambda_{\min_{\mathcal{H}}}(G_{\mathcal{H}}) < \pi_n^{-1}) \\
&= I + II.
\end{aligned}$$

II can be made as small as desired thanks to Lemma 3.3. For the first term, we deduce from eq. (3) that

$$\begin{aligned}
I &\leq \sum_{\mathcal{H}} \mathbb{P}(\pi_n^p \|W_{\mathcal{H}}\|^p_{\ell_2} \sum_{k \in \mathcal{S}_j(\mathcal{H})} \|\varphi_{j,k}\|^p_{\mathbb{L}_p} \geq 2^{-jd}\delta^p 2^{-p}) \\
&\leq \sum_{\mathcal{H}} \mathbb{P}(\pi_n^p \|W_{\mathcal{H}}\|^p_{\ell_2} C_p^p 2^{jd(\frac{p}{2}-1)} m \geq 2^{-jd}\delta^p 2^{-p}) \\
&\leq \sum_{\mathcal{H}} \mathbb{P}(\|W_{\mathcal{H}}\|_{\ell_2} \geq \frac{2^{-j\frac{d}{2}}\delta}{2\pi_n C_p m^{\frac{1}{p}}}) \\
&\leq m \sup_{k \in \mathcal{S}_j(\mathcal{H})} \mathbb{P}(|W_{j,k}| \geq \frac{2^{-j\frac{d}{2}}\delta}{2\pi_n C_p m^{\frac{1}{p}+1}}).
\end{aligned}$$

Now, notice that for $j \geq j_s$, we have $2^{j\frac{d}{2}}/\sqrt{n} \geq 2^{-js}$. Hence we can apply [7, Proposition 12.5] $\delta = \gamma(\log n)2^{j\frac{d}{2}}/\sqrt{n}$ to conclude that $I \leq mn^{-C(p,r,\sigma)\gamma}$, where $C(p, r, \sigma) = (1 \wedge \sigma^{-2})/C_p m^{\frac{1}{p}+1}$.

Case where $p = \infty$**.** The reasoning is similar as the one described above, with however a few slight modifications. For $\delta/2 > 2^{-j(s-\frac{d}{\tau})} \geq \|R_j f\|_{\mathbb{L}_\infty}$, we have,

$$\begin{aligned}
\mathbb{P}(\|f - f_j^{@}\|_{\mathbb{L}_\infty} \geq \delta) &\leq \mathbb{P}(\|f - f_j^{@}\|^z_{\mathbb{L}_\infty} \geq \delta^z 2^{-z}, \inf_{\mathcal{H}} \lambda_{\min}(G_{\mathcal{H}}) \geq \pi_n^{-1}) + \mathbb{P}(\inf_{\mathcal{H}} \lambda_{\min}(G_{\mathcal{H}}) < \pi_n^{-1}) \\
&= I + II.
\end{aligned}$$

II is of the right order thanks to Lemma 3.3. Moreover, similar arguments as the ones described in the proof of Theorem 2.2 allow to write,

$$\begin{aligned} I &\leq \mathbb{P}\left(\sup_{\mathcal{H}\in\mathcal{F}_j, k\in\mathcal{S}_j(\mathcal{H})} |W_{j,k}| \geq \frac{\delta 2^{-j\frac{d}{2}}}{2m\pi_n C_\infty} \right) \\ &\leq m2^{jd} \sup_{\mathcal{H}\in\mathcal{F}_j, k\in\mathcal{S}_j(\mathcal{H})} \mathbb{P}\left(|W_{j,k}| \geq \frac{\delta 2^{-j\frac{d}{2}}}{2m\pi_n C_\infty} \right). \end{aligned}$$

Now, notice that for $j \in \mathcal{J}_n$, $2^{jd} \leq n/\log n$. Notice besides that for $j \geq j_{s,\tau}$, we have $2^{j\frac{d}{2}}/\sqrt{n} \geq 2^{-j(s-\frac{d}{\tau})}$. Hence, following similar arguments as in [7, Proposition 12.5] and choosing $\delta = \gamma(\log n)2^{j\frac{d}{2}}/\sqrt{n}$, we can conclude that $I \leq mn^{-C(\infty,r,\sigma)\gamma+1}$, where $C(\infty,r,\sigma) = (1 \wedge \sigma^{-2})/mC_\infty$. □

4 Appendix

4.1 Properties of Besov spaces

Lemma 4.1. *We recall that for all $d \geq 1$, $\tau, q \in (0,\infty]$, $s > d/\tau$,*

$$f \in B^s_{\tau,q}(M) \Rightarrow \|R_j f\|_{\mathbb{L}_\infty} \leq M2^{-j(s-\frac{d}{\tau})}. \tag{16}$$

And for all $d \geq 1$, $q \in (0,\infty]$, $\tau \in [p,\infty)$, $s > 0$,

$$f \in B^s_{\tau,q}(M) \Rightarrow \|R_j f\|_{\mathbb{L}_p} \leq M2^{-js}. \tag{17}$$

Proof. These results can be found in [3] for example. Eq. (16) is a straight consequence of the embedding $B^s_{\tau,q}(M) \subset B^{s-\frac{d}{\tau}}_{\infty,\infty}(M)$. □

4.2 Uniform moment bound

Proposition 4.1 (see [9]). *Let $Z_1, \ldots, Z_n$ be a set of n independent random variables such that, for all i and $p \in \mathbb{N}$,*

$$\mathbb{E}\gamma_j(Z_i) = 0, \qquad \frac{1}{n}\sum_{i=1}^n \mathbb{E}|\gamma_j(Z_i)|^p \leq \frac{p!}{2}K^{p-2},$$

Then

$$\mathbb{E}\left(\max_{1\leq j\leq N} |\frac{1}{n}\sum_{i=1}^n \gamma_j(Z_i)|^p \right) \leq \left(\frac{K\log(N+e^{p-1})}{n} + \sqrt{\frac{2\log(N+e^{p-1})}{n}} \right)^p.$$

References

[1] Jean-Yves Audibert and Alexandre B. Tsybakov. Fast learning rates for plug-in classifiers. *Ann. Stat.*, 35:608–633, 2007.

[2] Yannick Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6:127–146, 2002.

[3] Albert Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in mathematics and its applications*. North-Holland, 2003.

[4] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Stat.*, 24(2):508–539, 1996.

[5] Alexander Korostelev and Alexandre B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer, New-York, 1993.

[6] Oleg V. Lepski, Enno Mammen, and Vladimir G. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Stat.*, 25(3):392–947, 1997.

[7] Jean-Baptiste Monnier. Classification via local multi-resolution projections. *Electronic Journal of Statistics*, 2011. URL http://arxiv.org/abs/1110.6427. To appear.

[8] Jean-Baptiste Monnier. *Quelques contributions en classification, régression et étude d'un problème inverse en finance*. PhD thesis, Université Paris Diderot, Paris 7, 12 2011. URL http://tel.archives-ouvertes.fr/tel-00650930/en/.

[9] Sara A. van de Geer. Uniform probability inequalities, with applications in high-dimensional statistical models. In *Cours de formation par la recherche, CREST-GENES*, 2010.

[10] Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates of convergence. *Ann. Stat.*, 27(5):1564–1599, 1999.