



**HAL**  
open science

# Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding

Eric Gautier, Erwan Le Pennec

► **To cite this version:**

Eric Gautier, Erwan Le Pennec. Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding. 2017. inria-00601274v4

**HAL Id: inria-00601274**

**<https://inria.hal.science/inria-00601274v4>**

Preprint submitted on 28 Nov 2017

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

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



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

# Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding

Eric Gautier<sup>\*†‡</sup> and Erwan Le Pennec<sup>§ ‡</sup>

November 2017

## Abstract

In the random coefficients binary choice model, a binary variable equals 1 iff an index  $X^\top \beta$  is positive. The vectors  $X$  and  $\beta$  are independent and belong to the sphere  $\mathbb{S}^{d-1}$  in  $\mathbb{R}^d$ . We prove lower bounds on the minimax risk for estimation of the density  $f_\beta$  over Besov bodies where the loss is a power of the  $L^p(\mathbb{S}^{d-1})$  norm for  $1 \leq p \leq \infty$ . We show that a hard thresholding estimator based on a needlet expansion with data-driven thresholds achieves these lower bounds up to logarithmic factors.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
2.1	Harmonic analysis . . . . .	5
2.2	The operator . . . . .	5
2.3	Needlets . . . . .	6
<b>3</b>	<b>Identification of <math>f_\beta</math></b>	<b>8</b>
<b>4</b>	<b>Lower bounds</b>	<b>9</b>
<b>5</b>	<b>Adaptive estimation by needlet thresholding</b>	<b>10</b>
5.1	Smoothed projections and linear needlet estimators . . . . .	10
5.2	Nonlinear estimator with data-driven thresholds . . . . .	11
<b>6</b>	<b>Simulation study</b>	<b>12</b>

---

<sup>\*</sup>Toulouse School of Economics, Université Toulouse Capitole, 21 allée de Brienne, 31015 Toulouse, France, [Eric.Gautier@tse-fr.eu](mailto:Eric.Gautier@tse-fr.eu)

<sup>†</sup>Eric Gautier acknowledges financial support from the grants ERC POEMH and ANR-13-BSH1-0004

<sup>‡</sup>The authors are grateful to Andrii Babii (TSE) for research assistance on the simulations.

<sup>§</sup>CMAP / Département de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau, France. [Erwan.Le-Pennec@polytechnique.edu](mailto:Erwan.Le-Pennec@polytechnique.edu)

<b>7</b>	<b>Appendix</b>	<b>13</b>
7.1	A preliminary lemma . . . . .	13
7.2	Proof of Proposition 5 . . . . .	15
7.3	Proof of Theorem 8 . . . . .	16
	7.3.1 Proof of the lower bound in the dense zone . . . . .	16
	7.3.2 Proof of the lower bound in the sparse zone . . . . .	18
7.4	Comparison between Besov ellipsoids of a function and its odd part . . . . .	19
7.5	A general inequality . . . . .	19
7.6	Proof of Lemma 14 . . . . .	21
	7.6.1 Preliminaries . . . . .	21
	7.6.2 Coefficientwise analysis . . . . .	22
	7.6.3 Scalewise analysis . . . . .	22
	7.6.4 Bernstein inequality and the term $\left  \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right ^z$ . . . . .	24
	7.6.5 Empirical Bernstein and the probabilities . . . . .	27
	7.6.6 The case $p = \infty$ . . . . .	27
	7.6.7 The case $p < \infty$ . . . . .	30
7.7	Proof of Theorem 9 . . . . .	33
	7.7.1 The case $1 \leq p < \infty$ . . . . .	33
	7.7.2 The case $p = \infty$ . . . . .	38

## 1 Introduction

Discrete choice models (see, *e.g.*, [21]) have applications in many areas ranging from planning of public transportation, economics of industrial organizations, evaluation of public policies, among others. This paper considers the binary choice model. There, agents (consumer, firm, country, etc.) choose between two exclusive alternatives 1 or -1 (*e.g.*, buying a good or not) the one that yields the highest utility. The utility that an agent  $i$  gets from choosing alternative -1 (resp. from choosing 1) is assumed to have the form

$$u_{-1,i} = z_{-1,i}^\top \gamma_i + \epsilon_{-1,i} \quad (\text{resp. } u_{1,i} = z_{1,i}^\top \gamma_i + \epsilon_{1,i}), \quad (1)$$

where  $z_{-1,i}$  (resp.  $z_{1,i}$ ) is a vector of  $d - 1$  characteristics of alternative -1 (resp. 1) for agent  $i$ ,  $d \geq 2$ ,  $\gamma_i$  are preferences of agent  $i$  for the characteristics, and  $\epsilon_{-1,i}$  and  $\epsilon_{1,i}$  absorb both the usual error terms and constants. In (1), the preferences are allowed to vary across individuals; namely, they are heterogeneous. This translates into a vector of coefficients  $\gamma$  indexed by  $i$  that we assume random. The characteristics of the alternatives are indexed by the agents, for example they can be characteristics of two goods that a consumer has to choose upon interacted with individual characteristics like age or distance. We assume that the random coefficients and errors are independent from the characteristics. The statistician observes a sample of characteristics and choices for agents  $i = 1, \dots, n$ , but  $\gamma_i$ ,  $u_{1,i}$ , and  $u_{-1,i}$  are not observed. Observing the choices corresponds to observing the sign  $y_i$  of the net utility  $u_{1,i} - u_{-1,i}$ . Indeed, agent  $i$  prefers 1 ( $y_i = 1$ ) if and only if the net utility for 1 is positive, *i.e.*,

$$u_{1,i} - u_{-1,i} = \epsilon_{1,i} - \epsilon_{-1,i} + (z_{1,i} - z_{-1,i})^\top \gamma_i > 0, \quad (2)$$

and prefers -1 ( $y_i = -1$ ) when

$$u_{1,i} - u_{-1,i} < 0.$$

We assume that the probability that  $|(\epsilon_{1,i} - \epsilon_{-1,i}, \gamma_i^\top)^\top|$  is the 0 and thus that agent  $i$  is indifferent (*i.e.*,  $u_{1,i} - u_{-1,i} = 0$ ) on a set of 0 probability. Hence, the linear random coefficients binary choice model is

$$y_i = \text{sign}(x_i^\top \beta_i), \quad (3)$$

where, for a real number  $a$ ,  $\text{sign}(a)$  is 1 if  $a > 0$ , -1 if  $a < 0$ , and is 0 if  $a = 0$ ,

$$\begin{aligned} x_i &= (1, (z_{1,i} - z_{-1,i})^\top)^\top / |(1, (z_{1,i} - z_{-1,i})^\top)^\top|, \\ \beta_i &= (\epsilon_{1,i} - \epsilon_{-1,i}, \gamma_i^\top)^\top / |(\epsilon_{1,i} - \epsilon_{-1,i}, \gamma_i^\top)^\top|, \end{aligned}$$

and  $|\cdot|$  is the Euclidean norm in  $\mathbb{R}^d$ . Like in [3, 4, 10, 13] among others, we consider a nonparametric specification of the joint distribution of  $\beta$  and this model is more general than the Logit, Probit, and Mixed-Logit models. Note that it is important to avoid restricting the dependence between the coordinates of  $(\epsilon_1 - \epsilon_{-1}, \gamma^\top)$  since they can be functions of a deep heterogeneity parameter (*e.g.*, the type of a consumer).

We denote by  $Y, Z_1, Z_{-1}, X, \epsilon_1, \epsilon_{-1}, \gamma$ , and  $\beta$  the population quantities corresponding to the lower cases letters indexed by  $i$ . The random vectors  $X$  and  $\beta$  are elements of the unit sphere  $\mathbb{S}^{d-1}$  of  $\mathbb{R}^d$ . For the main results of this paper we maintain the following restrictions on the distribution of  $(\beta^\top, X^\top)^\top$ .

**Assumption 1** (A1.1)  $X$  and  $\beta$  are independent,

(A1.2)  $X$  and  $\beta$  have densities  $f_X$  and  $f_\beta$  with respect to the spherical measure  $\sigma$ .

**Assumption 2** (A2.1)  $f_\beta(x)f_\beta(-x) = 0$  for a.e.  $x$  in  $\mathbb{S}^{d-1}$ ,

(A2.2) The support of  $X$ , denoted by  $\text{supp}(f_X)$ , is  $H^+ = \{x \in \mathbb{S}^{d-1} : x_1 \geq 0\}$ ,

(A2.3)  $f_X$  is known and we have  $A_X \stackrel{\text{def}}{=} \|f_X\|_{L^\infty(H^+)} < \infty$  and  $B_X \stackrel{\text{def}}{=} \|1/f_X\|_{L^\infty(H^+)} < \infty$ .

Under Assumption 1,  $f_\beta$  is solution of the ill-posed inverse problem: for a.e.  $x \in H^+$

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{S}^{d-1}} \text{sign}(x^\top y) f_\beta(y) d\sigma(y) \stackrel{\text{def}}{=} \mathcal{K}f_\beta(x). \quad (4)$$

The operator  $\mathcal{K}$  in (4) is a convolution on  $\mathbb{S}^{d-1}$ . Estimation of  $f_\beta$  in (4) is thus related to statistical deconvolution on  $\mathbb{S}^{d-1}$  (see, *e.g.*, [12, 16, 19]). However, the left-hand side of (4) is not a density but a regression function where the regressors are random. The identification issue in this model stems from the fact that: (1) the distribution of the observed data only characterizes  $\mathcal{K}f_\beta$  on  $\text{supp}(f_X)$  which is a proper subset of  $\mathbb{S}^{d-1}$  and (2) due to the sign function  $\mathcal{K}$  has an infinite dimensional null space. The support of  $X$  can only be as large as  $H^+$  because the first coordinate of  $X$  is positive. This is because we allow for the term  $\epsilon_{1,i} - \epsilon_{-1,i}$  in (2).

A simple estimator for the density of  $\beta$  in this model is given in [10]. There, rates of convergence for the  $L^p$ -losses for  $1 \leq p \leq \infty$  over Sobolev ellipsoids based on the same  $L^p$  space (as well as confidence intervals for the value of the density at a point, treatment of endogenous regressors, and of models where some coefficients are nonrandom) are obtained under similar assumptions for choices of the smoothing parameters which depend on unknown parameters of the Sobolev ellipsoids. It is assumed in [10] that the support of  $\beta$  lies in an (unknown) hemisphere, namely, that there exists  $n$  (unknown) in  $\mathbb{S}^{d-1}$  such that  $\mathbb{P}(n^\top \beta > 0) = 1$ . This assumption first appeared in [13]

and is stronger than (A2.1). It implies that for some difference of the characteristics, or taking a limit of these, everyone chooses the same alternative. In contrast, (A2.1) is much less restrictive and does not imply "unselected samples". However, everything in [10] also holds under (A2.1). Assumption (A2.2) requires that the support of  $Z_1 - Z_{-1}$  is  $\mathbb{R}^d$  and is also made in [10, 13]. [9] allows for continuous regressors which support is a proper subset at the expense of assuming some form of unselected samples and relying on integrability assumptions involving  $f_\beta$ . It is possible to obtain identification of  $f_\beta$  when we relax (A2.2) and the requirement that  $f_X$  exists (see (A1.2)). This is done in [8]. The estimation in this case is the subject of future work. (A2.3) strengthens (A2.2) and is used to obtain rates of convergence. It could be viewed as an assumption on the tails of  $X$ . It is relaxed in [10] and in this paper at the end of Section 5. Note as well that Assumption (A1.2) allows for one nonrandom coefficient in the original scale and that when there are more than two, one should proceed as in Section 5.2 in [10] with the estimator developed in this paper.

In this paper, we show that the estimator in [10] can be written as a plug-in of a linear needlet estimator. Needlets are a class of linear combinations of spherical harmonics which form a tight frame of localized functions on spheres (see [25]). Hard-thresholding of series estimators based on needlets have been successfully used in statistics for estimation of functions defined on spheres (see [2] for densities, [24] for regression functions, and [17, 18, 19] for some inverse problems) or compact manifolds (see [15]). This paper proves lower bounds on the minimax risk when the degree of integrability in the loss - specified by the statistician - can differ from the degree of integrability of the Besov body containing the unknown  $f_\beta$ , giving rise to sparse and dense regimes. The lower bounds correspond, up to logarithmic factors, to the upper bounds in [10] over Sobolev ellipsoids and matching degrees of integrability. This paper proposes to replace the linear needlet estimator in [10] by a nonlinear estimator based on hard-thresholding with data-driven thresholds and use the same plug-in strategy as in [10]. The upper bounds on the risk of the estimator also correspond to the lower bounds up to a logarithmic factor, but over all Besov bodies, including nonmatching degrees of integrability. Both the upper and lower bounds are also given for the sup-norm loss. The data-driven thresholds are similar in spirit to [5] for density estimation using the Dantzig selector (see also [6, 24] for other local thresholding procedures over the sphere), they are based on sharp concentration inequalities and make the implementation of the estimator feasible as it is independent of features of the unknown density. Proofs are given in the appendix.

## 2 Preliminaries

We use the notation  $x \wedge y$  and  $x \vee y$  for the minimum and the maximum between  $x$  and  $y$ . We write  $x \lesssim y$  when there exists  $c$  such that  $x \leq cy$ ,  $x \gtrsim y$  when there exists  $c$  such that  $x \geq cy$ , and  $x \simeq y$  when  $x \lesssim y$  and  $x \gtrsim y$ . We denote by  $|A|$  and  $\mathbf{1}_A$  the cardinal and indicator of the set  $A$ , by  $\mathbb{N}$  the nonnegative integers, by  $\mathbb{N}^*$  the positive integers, by a.e. almost every, and by a.s. almost surely. We denote for  $1 \leq p \leq \infty$  by  $\|\cdot\|_{\ell^p}$  the  $\ell^p$ -norm of a vector, by  $\|\cdot\|_p$  the usual norm on the space  $L^p(\mathbb{S}^{d-1})$  of  $p$  integrable real-valued functions with respect to the spherical measure  $\sigma$ . We write  $L^p_{\text{odd}}(\mathbb{S}^{d-1})$  (resp.  $L^p_{\text{even}}(\mathbb{S}^{d-1})$ ) the closure in  $L^p(\mathbb{S}^{d-1})$  of continuous functions on  $\mathbb{S}^{d-1}$  which are odd (*i.e.*, for every  $x \in \mathbb{S}^{d-1}$ ,  $f(-x) = -f(x)$ ) (resp. even). Every  $f \in L^p(\mathbb{S}^{d-1})$  can be uniquely decomposed as the sum of an odd and even function  $f^-$  and  $f^+$  in  $L^p(\mathbb{S}^{d-1})$ . The space  $L^2(\mathbb{S}^{d-1})$  is a Hilbert space with the scalar product  $\langle \cdot, \cdot \rangle$  derived from the norm, there  $f^-$  and  $f^+$  are orthogonal.  $\mathcal{D}$  is the set of densities and, as it will become clear after Proposition 5,  $\nu(d) = d/2$  is the degree of ill-posedness of the inverse problem.

## 2.1 Harmonic analysis

The basic element is the orthogonal decomposition  $L^2(\mathbb{S}^{d-1}) = \bigoplus_{k \in \mathbb{N}} H^{k,d}$ , where  $H^{k,d}$  are the eigenspaces of the Laplacian  $\Delta$  on  $\mathbb{S}^{d-1}$ , corresponding to the eigenvalues  $-\zeta_{k,d}$ , given by  $\zeta_{k,d} \stackrel{\text{def}}{=} k(k+d-2)$ , of dimension  $L(k,d) \stackrel{\text{def}}{=} (2k+d-2)(k+d-2)!/(k!(d-2)!(k+d-2))$ . The space  $H^{k,d}$  is spanned by an orthonormal basis  $(h_{k,l})_{l=1}^{L(k,d)}$  and  $H^{0,d}$  by 1. We also have  $L^2_{\text{odd}}(\mathbb{S}^{d-1}) = \bigoplus_{p \in \mathbb{N}} H^{2p+1,d}$  and  $L^2_{\text{even}}(\mathbb{S}^{d-1}) = \bigoplus_{p \in \mathbb{N}} H^{2p,d}$ . The projector  $L_{k,d}$  onto  $H^{k,d}$  is the operator with kernel

$$L_{k,d}(x,y) = \sum_{l=1}^{L(k,d)} h_{k,l}(x)h_{k,l}(y) = \frac{L(k,d)}{\sigma(\mathbb{S}^{d-1})P_k^{\mu(d)}(1)} P_k^{\mu(d)}(x^\top y), \quad (5)$$

where  $\mu(d) = (d-1)/2$ , the surface of  $\mathbb{S}^{d-1}$  is  $\sigma(\mathbb{S}^{d-1}) = 2\pi^{d/2}/\Gamma(d/2)$ , and  $C_k^\mu$  are the Gegenbauer polynomials. The Gegenbauer polynomials, defined for  $\mu > -1/2$ , are orthogonal in the space of square integrable functions on  $[-1, 1]$  with measure  $(1-t^2)^{\mu-1/2} dt$ . We have  $P_0^\mu(t) = 1$ ,  $P_1^\mu(t) = 2\mu t$  for  $\mu \neq 0$ ,  $P_1^0(t) = 2t$ , and for every  $k \in \mathbb{N}$

$$(k+2)P_{k+2}^\mu(t) = 2(\mu+k+1)tP_{k+1}^\mu(t) - (2\mu+k)P_k^\mu(t). \quad (6)$$

Clearly, for  $f \in L^2(\mathbb{S}^{d-1})$ , we have  $f = \sum_{k=0}^{\infty} L_{k,d}f$  and, due to (5),

$$\forall x \in \mathbb{S}^{d-1}, \|L_{k,d}(x, \cdot)\|_2^2 = \sum_{l=1}^{L(k,d)} |h_{k,l}(x)|^2 = \frac{L(k,d)}{\sigma(\mathbb{S}^{d-1})}. \quad (7)$$

Powers  $(-\Delta)^s f$  for  $s \in \mathbb{R}$  and  $f$  in a Banach space  $E_1$  are defined in a Banach space  $E_2$  when  $L_{k,d}f$  is defined in  $E_2$  and  $(-\Delta)^s f \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \zeta_{k,d}^s L_{k,d}f$  converges in  $E_2$ . The best approximation in  $L^r(\mathbb{S}^{d-1})$  of a function  $f$  by harmonics of degree less or equal to  $m$  is

$$E_m(f)_r = \inf_{P \in \bigoplus_{k=0}^m H^{k,d}} \|f - P\|_r.$$

**Definition 3** For  $s > 0$  and  $1 \leq r \leq \infty$ ,  $f$  belongs to the Sobolev space  $W_r^s(\mathbb{S}^{d-1})$  if

$$\|f\|_{r,s} = \|f\|_r + \left\| (-\Delta)^{s/2} f \right\|_r < \infty.$$

We denote by  $W_r^s(\mathbb{S}^{d-1})_{\text{odd}}$  the restriction of  $W_r^s(\mathbb{S}^{d-1})$  to odd functions.

**Definition 4** For  $s > 0$ ,  $1 \leq r \leq \infty$ , and  $0 < q \leq \infty$ ,  $f$  belongs to the Besov space  $B_{r,q}^s(\mathbb{S}^{d-1})$  if

$$\|f\|_{B_{r,q}^s}^A = \|f\|_r + \left\| (2^{js} E_{2^j}(f))_r \right\|_{\ell_q} < \infty.$$

## 2.2 The operator

**Proposition 5** The operator  $\mathcal{K}$  satisfies the following properties:

(P1.1) For every  $f \in L^1(\mathbb{S}^{d-1})$ ,  $\mathcal{K}f = \mathcal{K}(f^-)$ ,

(P1.2) If  $\mathcal{K}f = \mathcal{K}g$  with  $f, g \in L^1_{\text{odd}}(\mathbb{S}^{d-1})$  then  $g = f$ ,

(P1.3) For every  $1 \leq r \leq \infty$ ,

$$W_r^{\nu(d)+|1/r-1/2|(d-2)}(\mathbb{S}^{d-1}) \subseteq \mathcal{K}(L^r_{\text{odd}}(\mathbb{S}^{d-1})) \subseteq W_r^{\nu(d)-|1/r-1/2|(d-2)}(\mathbb{S}^{d-1}),$$

where the exponents  $\nu(d) \pm |1/r - 1/2|(d - 2)$  cannot be improved,

(P1.4) For every  $1 \leq r \leq \infty$ , there exists  $B(d, r)$  such that

$$\forall K \in \mathbb{N}, \forall P \in \bigoplus_{\substack{k=0 \\ k \text{ odd}}}^K H^{k,d}, \|\mathcal{K}^{-1}P\|_r \leq B(d, r)K^{\nu(d)}\|P\|_r. \quad (8)$$

Moreover,  $\mathcal{K}$  is a self-adjoint and compact operator on  $L^2(\mathbb{S}^{d-1})$  with null space  $L^2_{\text{even}}(\mathbb{S}^{d-1})$ , nonzero eigenvalues  $(\lambda_{2p+1,d})_{p \in \mathbb{N}}$  corresponding to the eigenspaces  $H^{2p+1,d}$  for  $p \in \mathbb{N}$

$$\lambda_{1,d} = \frac{2|\mathbb{S}^{d-2}|}{d-1}, \quad \forall p \in \mathbb{N}^* \quad \lambda_{2p+1,d} = \frac{2(-1)^p |\mathbb{S}^{d-2}| 1 \cdot 3 \cdots (2p-1)}{(d-1)(d+1) \cdots (d+2p-1)}.$$

For every  $d \in \mathbb{N} \setminus \{1\}$ , for every  $p \in \mathbb{N}$ , there exists  $c_\lambda(d), C_\lambda(d) > 0$  such that

$$c_\lambda^{-1}(d)p^{-\nu(d)} \leq |\lambda_{2p+1,d}| \leq C_\lambda(d)p^{-\nu(d)}. \quad (9)$$

$\mathcal{K}$  is a homeomorphism between  $L^2_{\text{odd}}(\mathbb{S}^{d-1})$  and  $W^{\nu(d)}_{2 \text{ odd}}(\mathbb{S}^{d-1})$ .

The fact that  $\nu(d)$  is the degree of ill-posedness of the inverse problem follows from (P1.4) and what follows, in particular (9).

Proposition 5 implies that every  $R \in W^{\nu(d)}_{2 \text{ odd}}(\mathbb{S}^{d-1})$  has a unique inverse given by

$$\mathcal{K}^{-1}(R) = \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} L_{k,d}(R) = \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} \sum_{l=1}^{L(k,d)} \langle R, h_{k,l} \rangle h_{k,l}. \quad (10)$$

### 2.3 Needlets

Smoothed projection operators (see [10]) have good approximation properties in all  $L^p(\mathbb{S}^{d-1})$  spaces and are uniformly bounded from  $L^p(\mathbb{S}^{d-1})$  to  $L^p(\mathbb{S}^{d-1})$ . One such operator, the delayed means, is the integral operator with kernel

$$K^{a,J}(x, y) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} a\left(\frac{k}{2^J}\right) L_{k,d}(x, y), \quad (11)$$

where  $J$  is an integer,  $a$  is a  $C^\infty$  and decreasing function on  $[0, \infty)$  supported on  $[0, 2]$  such that, for every  $0 \leq t \leq 2$ ,  $0 \leq a(t) \leq 1$  and, for every  $0 \leq t \leq 1$ ,  $a(t) = 1$ . The delayed means operator exhibits nearly exponential localization (see Theorem 2.2 in [25]) and is a building block for the construction of needlets.

Define  $b$  such that  $b^2(t) = a(t) - a(2t)$  for  $t \geq 0$ . It is nonzero only when  $1/2 \leq t \leq 2$ , satisfies  $b^2(t) + b^2(2t) = 1$  for  $1/2 \leq t \leq 1$  and thus for every  $t \geq 1$ ,  $\sum_{j=0}^{\infty} b^2\left(\frac{t}{2^j}\right) = 1$ , also  $b^2(t) = a(t)$  for  $1 \leq t \leq 2$ . Take  $a$  such that  $b$  is bounded away from 0 on  $3/5 \leq t \leq 5/3$ .

The second ingredient for the construction of needlets is a quadrature formula (Corollary 2.9 of [25]) with positive weights  $(\omega(j, \xi)^2)_{\xi \in \Xi_j}$  and nodes  $\xi \in \Xi_j$  which integrates functions in  $\bigoplus_{k=0}^{2^j} H^{k,d}$  and satisfy, for a constant  $C_{\Xi}$  which depends on  $d$ ,

$$\begin{aligned} \forall j \in \mathbb{N}, \forall \xi \in \Xi_j, \quad C_{\Xi}^{-1} 2^{j(d-1)} &\leq |\Xi_j| \leq C_{\Xi} 2^{j(d-1)} \\ C_{\Xi}^{-1} 2^{-j(d-1)/2} &\leq \omega(j, \xi) \leq C_{\Xi} 2^{-j(d-1)/2}. \end{aligned}$$

Needlets are defined as

$$\psi_{j,\xi}(x) \stackrel{\text{def}}{=} \omega(j, \xi) \sum_{k=0}^{\infty} b\left(\frac{k}{2^{j-1}}\right) L_{k,d}(\xi, x) \quad \text{if } j \in \mathbb{N}, \xi \in \Xi_j, \quad (12)$$

$$\psi_{0,\xi}(x) \stackrel{\text{def}}{=} L_{0,d}(\xi, x). \quad (13)$$

For  $j = 0$ ,  $\psi_{0,\xi}(x)$  is constant and  $\Xi_0$  is a singleton.

The  $L^p$ -norms of the needlets satisfy, for a constant  $C_p$  that can depend on  $d$ ,

$$\forall j \in \mathbb{N}, \forall \xi \in \Xi_j, \quad C_p^{-1} 2^{j(d-1)(1/2-1/p)} \leq \|\psi_{j,\xi}\|_p \leq C_p 2^{j(d-1)(1/2-1/p)}. \quad (14)$$

If  $f \in L^p(\mathbb{S}^{d-1})$  for  $1 \leq p \leq \infty$ , then  $f = \sum_{j=0}^{\infty} \sum_{\xi \in \Xi_j} \langle f, \psi_{j,\xi} \rangle \psi_{j,\xi}$ . The needlets form a tight frame, with unitary tightness constant, this means that for  $f \in L^2(\mathbb{S}^{d-1})$

$$\|f\|_2^2 = \sum_{j=0}^{\infty} \sum_{\xi \in \Xi_j} |\langle f, \psi_{j,\xi} \rangle|^2.$$

Needlets do not form a basis and there is redundancy. Lemma 6 (see [2]) relates  $L^p(\mathbb{S}^{d-1})$  norms at level  $j$  to  $\ell^p$  norms of needlet coefficients. Constants may depend on  $d$ .

**Lemma 6** (i) For every  $1 \leq p \leq \infty$ , there exists a constant  $C'_p$  such that for every  $j \in \mathbb{N}$  and  $(\beta_{\xi})_{\xi \in \Xi_j} \in \mathbb{R}^{\Xi_j}$

$$\left\| \sum_{\xi \in \Xi_j} \beta_{\xi} \psi_{j,\xi} \right\|_p \leq C'_p 2^{j(d-1)(1/2-1/p)} \left\| (\beta_{\xi})_{\xi \in \Xi_j} \right\|_{\ell^p}, \quad (15)$$

(ii) There exists constants  $c_A$  and  $c_{p,A}$  and sets  $A_j \subset \Xi_j$  with  $|A_j| \geq c_A 2^{j(d-1)}$  for  $j \in \mathbb{N}$  such that for every  $1 \leq p \leq \infty$ ,  $j \in \mathbb{N}$ , and  $(\beta_{\xi})_{\xi \in A_j} \in \mathbb{R}^{A_j}$ ,

$$\left\| \sum_{\xi \in A_j} \beta_{\xi} \psi_{j,\xi} \right\|_p \geq c_{p,A} 2^{j(d-1)(1/2-1/p)} \left\| (\beta_{\xi})_{\xi \in A_j} \right\|_{\ell^p}, \quad (16)$$

(iii) For every  $1 \leq p \leq \infty$ , there exists a constant  $C''_p$  such that for every  $j \in \mathbb{N}$

$$\left( \sum_{\xi \in \Xi_j} |\langle f, \psi_{j,\xi} \rangle|^p \right)^{1/p} 2^{j(d-1)(1/2-1/p)} \leq C''_p \|f\|_p. \quad (17)$$

Needlets are such that (see [25]), for all function  $a$  in the definition of the smoothed projection operators, the norm  $\|\cdot\|_{B_{r,q}^s}^A$  defining the Besov spaces is equivalent to

$$\|f\|_{B_{r,q}^s} = \left\| \left( 2^{j(s+(d-1)(1/2-1/r))} \left\| \langle f, \psi_{j,\xi} \rangle_{\xi \in \Xi_j} \right\|_{\ell^r} \right)_{j \in \mathbb{N}} \right\|_{\ell^q}.$$

The ball of radius  $M$  for this norm is denoted by  $B_{r,q}^s(M)$ .

Recall the following consequence of the proof of the continuous embeddings in [2].

**Lemma 7** (i) *If  $p \leq r \leq \infty$ , then we have  $B_{r,q}^s(M) \subseteq B_{p,q}^s(C_{\Xi}^{1/p-1/r} M)$ ,*

(ii) *If  $s > (d-1)(1/r - 1/p)$  and  $r \leq p \leq \infty$ , then we have  $B_{r,q}^s(M) \subseteq B_{p,q}^{s-(d-1)(1/r-1/p)}(M)$ ,*

(iii) *If  $f \in B_{r,q}^s(M)$  and  $(\beta_{j,\xi})_{\xi \in \Xi_j, j \in \mathbb{N}}$  are its needlet coefficients, then there exists  $(D_j)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  such that  $\|(D_j)_{j \in \mathbb{N}}\|_{\ell^q} \leq M$  and*

$$\forall z \geq 1, \forall j \in \mathbb{N}, \sum_{\xi \in \Xi_j} |\beta_{j,\xi}|^z \leq C_{\Xi}^{1-(z \wedge r)/r} D_j^z 2^{-jz(s+(d-1)(1/2-1/(z \wedge r)))}. \quad (18)$$

Finally recall that, when  $f \in B_{r,q}^s$  with  $s > (d-1)/r$ , then  $f$  is continuous.

### 3 Identification of $f_{\beta}$

Let us present the arguments for the identification of  $f_{\beta}$ . Proposition 5 (P1.1) implies that  $\mathcal{K}f_{\beta} = \mathcal{K}f_{\beta}^{-}$  is odd. Thus under (A2.2) we can define the odd function  $R$  as

$$R(x) = \begin{cases} \mathbb{E}[Y|X = x] & \text{for a.e. } x \in H^+ \\ -\mathbb{E}[Y|X = -x] & \text{for a.e. } x \in -H^+ \end{cases} \quad (19)$$

and we have, for a.e.  $x \in \mathbb{S}^{d-1}$ ,  $R(x) = \mathcal{K}f_{\beta}^{-}(x)$ . Uniqueness of  $f_{\beta}^{-}$  follows from (P1.2). Using, for a.e.  $x \in \mathbb{S}^{d-1}$   $f_{\beta}(x) \geq 0$  and  $f_{\beta}^{-}(x) = (f_{\beta}(x) - f_{\beta}(-x))/2$ , and condition (A2.1), yields that, for a.e.  $x \in \mathbb{S}^{d-1}$ , we have

$$f_{\beta}(x) = 2f_{\beta}^{-}(x)\mathbf{1}_{f_{\beta}^{-}(x) > 0}. \quad (20)$$

In this paper we normalize the vectors of random coefficients and covariates to have unit norm. Indeed, since only the sign of the net utility (2) matters for choosing between 1 and -1 and the index is linear, a scale normalization of  $(\epsilon_1 - \epsilon_{-1}, \gamma^{\top})$  is in order. Let us compare with the normalization in [9]. It is based on the following assumption, which is stronger than the condition in [13], that the support of  $\beta$  is a subset of some (unknown) hemisphere, which itself is stronger than (A2.1).

(H): a.s. there exists  $j \in \{1, \dots, d\}$ , the coordinate  $\gamma_j$  of  $\gamma$  has a sign (excluding 0).

Assumption (H) is likely to hold when  $Z_{1j}$  and  $Z_{-1j}$  are cost factors, since consumers dislike an increase in cost. If (H) holds we can identify for which index  $j$   $\gamma_j$  has a sign since it amounts to the finding for which coordinate  $z_j$  of  $z \rightarrow \mathbb{E}[Y|Z_1 - Z_{-1} = z]$  is (globally) monotone. We can identify the sign of the coefficient by assessing whether the function is increasing (positive) or decreasing (negative). If  $\gamma_j > 0$  then we normalize the vector of coefficients by dividing by  $\gamma_j$ . If  $\gamma_j < 0$  we

change the sign of  $Z_{1j} - Z_{-1j}$  to make it positive. A potential issue with this normalization is that if  $\beta_j$  can take small values then estimators could differ in finite samples depending on which coefficient is used for normalization. Also, monotonicity in one regressor of the conditional mean function implies a type of weak monotonicity (in the sense used to identify treatment effects, see, e.g., [9]) at the individual level as we now explain. Assuming that  $\gamma_j > 0$ ,  $z_{1i} - z_{-1i} = z$  for all  $i = 1, \dots, n$ , and that we change  $z_j$  to  $z'_j > z_j$  while leaving unchanged  $(\epsilon_{1i} - \epsilon_{-1i}, \gamma_i^\top)$  (the characteristics of the individuals) and the other components of  $z$ , then some people do not change their decision and some choose alternative 1 while originally they had chosen alternative -1, but no one changes from alternative 1 to alternative -1. Monotonicity of the conditional mean function implies monotonicity for every individual. This is sometimes not a realistic model of individuals making choices. Clearly (A2.1) allows both individuals to switch from 1 to -1 and individuals to switch from -1 to 1 after similar changes in  $z$  (or  $x$ ). On the other hand, if (H) holds then (A2.2) can be relaxed and we can consider an index which is nonlinear in  $X$  (cf. [9]).

## 4 Lower bounds

We take  $1 \leq p, r \leq \infty$ ,  $0 \leq q \leq \infty$ ,  $z \geq 1$ , and  $s > 0$ , and consider the minimax risk

$$\mathcal{R}_n^* \stackrel{\text{def}}{=} \inf_{\widehat{f}_\beta} \sup_{f_\beta \in B_{r,q}^s(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f}_\beta - f_\beta \right\|_p^z, \quad (21)$$

where the infimum is over all estimators based on the i.i.d. sample of size  $n$ . The degree of integrability  $r$  in the smoothness class  $B_{r,q}^s(M)$  is allowed to differ from the degree of integrability  $p$  in the loss function. We distinguish two zones for  $s, r, q, d$ , and  $p$ :

(1) the *dense zone* where  $s \geq p(\nu(d) + (d-1)/2)(1/r - 1/p)$  with the restriction  $q \leq r$  if  $s = p(\nu(d) + (d-1)/2)(1/r - 1/p)$ , where the rate involves

$$\mu_{\text{dense}}(d, p, r, s) \stackrel{\text{def}}{=} s / (s + \nu(d) + (d-1)/2),$$

(2) the *sparse zone* where  $(d-1)/r < s < p(\nu(d) + (d-1)/2)(1/r - 1/p)$ , where the rate involves

$$\mu_{\text{sparse}}(d, p, r, s) \stackrel{\text{def}}{=} (s - (d-1)(1/r - 1/p)) / (s + \nu(d) - (d-1)(1/r - 1/2)).$$

The terminology dense and sparse is justified by the following heuristic. The proofs of the lower bounds replace the infimum in (21) by a minimum over a set of functions which are difficult to estimate. The functions used to prove the lower bound in the dense zone are functions which could have many nonzero needlet coefficients for  $\xi \in A_j$  (see Lemma 6) and a well-chosen  $j$ . Those used to prove the lower bound in the sparse zone only have two nonzeros. In the dense zone, the rate is the same as for the matched case when  $r = p$  studied in [10].

**Theorem 8** (i) *In the dense zone we have*

$$\mathcal{R}_n^* \geq c_{\text{dense}}(d, M, p, r, s, z) \left( \frac{1}{\sqrt{nA_X}} \right)^{\mu_{\text{dense}}(d, p, r, s)z}, \quad (22)$$

(ii) *In the sparse zone we have*

$$\mathcal{R}_n^* \geq c_{\text{sparse}}(d, M, p, r, s, z) \left( \sqrt{\frac{\ln(nA_X)}{nA_X}} \right)^{\mu_{\text{sparse}}(d, p, r, s)z}, \quad (23)$$

where the constants  $c_{\text{dense}}$  and  $c_{\text{sparse}}$  depend on  $d, M, p, r, s$  and  $z$ .

The values of  $\mu_{\text{dense}}$  and  $\mu_{\text{sparse}}$  depend on  $d$  through the dimension of  $\mathbb{S}^{d-1}$ . This is the usual curse of dimensionality in nonparametric regression or density estimation. They also depend on  $d$  through the degree of ill-posedness  $\nu(d) = d/2$  of the inverse problem.

## 5 Adaptive estimation by needlet thresholding

Consider the estimator  $\widehat{f}_\beta = 2\widehat{f}_\beta^- \mathbf{1}_{\widehat{f}_\beta^- > 0}$ , where  $\widehat{f}_\beta^-$  is an estimator of  $f_\beta^-$ .

### 5.1 Smoothed projections and linear needlet estimators

A smoothed projection estimator of  $f_\beta^-$  with kernel (11), window  $a$ , and  $J \in \mathbb{N}$ , is given for  $x \in \mathbb{S}^{d-1}$  by

$$\widehat{f}_\beta^{-a,J}(x) = \sum_{k \text{ odd}} \frac{a\left(\frac{k}{2^J}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(x),$$

with the unbiased estimator of  $L_{k,d}R(x)$  (see Lemma 10):  $\widehat{L_{k,d}R}(x) = 0$  if  $k$  is even, else

$$\widehat{L_{k,d}R}(x) = \frac{2}{n} \sum_{i=1}^n \frac{y_i L_{k,d}(x_i, x)}{f_X(x_i)}.$$

Alternatively, we can estimate  $f_\beta^-$  using the needlet frame with smoothing window  $a$ . The coefficients  $\beta_{j,\xi}^a = \langle f_\beta^-, \psi_{j,\xi} \rangle$  are such that

$$\begin{aligned} \beta_{j,\xi}^a &= \omega(j, \xi) \sum_{k \text{ odd}} b\left(\frac{k}{2^{j-1}}\right) \langle f_\beta^-, L_{k,d}(\xi, \cdot) \rangle \\ &= \omega(j, \xi) \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \langle L_{k,d}R, L_{k,d}(\xi, \cdot) \rangle \\ &= \omega(j, \xi) \sum_{\substack{k \text{ odd} \\ 2^{j-2} < k < 2^j}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} L_{k,d}R(\xi). \end{aligned}$$

Using that  $a\left(\frac{k}{2^J}\right) = 1$  for  $k = 0, \dots, 2^j$  and denoting by  $f_\beta^{-a,J} = \mathbb{E}\left[\widehat{f}_\beta^{-a,J}\right]$ , we obtain that, for  $1 \leq j \leq J$ ,  $\beta_{j,\xi}^a = \langle f_\beta^{-a,J}, \psi_{j,\xi} \rangle$ , which can be estimated without bias by

$$\widehat{\beta}_{j,\xi}^a = \omega(j, \xi) \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(\xi) \stackrel{(\Delta_1)}{=} \left\langle \widehat{f}_\beta^{-a,J}, \psi_{j,\xi} \right\rangle.$$

Moreover, for  $x \in \mathbb{S}^{d-1}$ ,

$$\widehat{\beta}_{j,\xi}^a \psi_{j,\xi}(x) = \omega(j, \xi)^2 \left( \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(\xi) \right) \left( \sum_k b\left(\frac{k}{2^{j-1}}\right) L_{k,d}(\xi, x) \right)$$

belongs to  $\bigoplus_{k=0}^{2^j} H^{k,d}$ , thus by the quadrature formula

$$\sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^a \psi_{j,\xi}(x) = \sum_{k \text{ odd}} \frac{b^2\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}(x).$$

This yields  $\sum_{j=0}^J \sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^a \psi_{j,\xi} = \widehat{f}_\beta^{-a,J-1}$ , indeed

$$\begin{aligned} \sum_{j=0}^J \sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^a \psi_{j,\xi} &= \sum_{j=1}^J \sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^a \psi_{j,\xi} \quad (\text{due to } (\Delta_1) \text{ and because } \widehat{f}_\beta^{-a,J} \text{ is odd}) \\ &\stackrel{(\Delta_2)}{=} \sum_{\substack{1 \leq k < 2^{J-1} \\ k \text{ odd}}} \frac{1}{\lambda_{k,d}} \widehat{L_{k,d}R} + \sum_{\substack{2^{J-1} \leq k \leq 2^J \\ k \text{ odd}}} \frac{b^2\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R} \\ &\stackrel{(\Delta_3)}{=} \sum_{\substack{1 \leq k < 2^{J-1} \\ k \text{ odd}}} \frac{1}{\lambda_{k,d}} \widehat{L_{k,d}R} + \sum_{\substack{2^{J-1} \leq k \leq 2^J \\ k \text{ odd}}} \frac{a\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}, \end{aligned}$$

where  $(\Delta_2)$  uses that for  $1/2 \leq t \leq 1$ ,  $b^2(t) + b^2(2t) = 1$ , while  $(\Delta_3)$  that  $b^2(t) = a(t)$  for  $1 \leq t \leq 2$ . Thus, the smoothed projection and needlet estimators coincide.

## 5.2 Nonlinear estimator with data-driven thresholds

Consider, for  $\gamma \geq 1$  and  $\rho_{T_{j,\xi,\gamma}}(x) = x \mathbf{1}_{|x| > T_{j,\xi,\gamma}}$ , the nonlinear estimator of  $f_\beta^-$ :

$$\widehat{f}_\beta^{-a,\rho} = \sum_{j=0}^J \sum_{\xi \in \Xi_j} \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) \psi_{j,\xi}.$$

It is classical that the optimal choice of  $J$  for linear estimators depends on the parameters of the smoothness ellipsoid. In contrast, using a thresholded estimator allows to take  $J$  large and independent of the parameters. Thresholding induces additional bias compared to linear estimators which allows to reduce the variance incurred by taking  $J$  large.

The level of thresholding should depend on the size of the coefficients relative to their variance. This variance is proportional to  $1/\sqrt{n}$  so that the level of the threshold does not have to depend on the smoothness of the unknown function. Instead of using a conservative upper bound on their variance, as is usually the case in estimation using wavelets, we use data-driven levels of thresholding. These provide better estimators in small samples. Lemma 14 gives a theoretical guarantee that the performance is almost as good as that of an oracle which would know the variance of the estimators of the coefficients. The data-driven thresholding rule uses that  $\widehat{\beta}_{j,\xi}^a = \frac{1}{n} \sum_{i=1}^n G_{j,\xi}(x_i, y_i)$  with

$$G_{j,\xi}(x_i, y_i) \stackrel{\text{def}}{=} \frac{2}{n} \sum_{i=1}^n \omega(j, \xi) \frac{y_i}{f_X(x_i)} \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} L_{k,d}(x_i, \xi). \quad (24)$$

Define the estimator of the variance by

$$\widehat{\sigma}_{j,\xi} \stackrel{\text{def}}{=} \sqrt{\frac{1}{n(n-1)} \sum_{i=2}^n \sum_{k=1}^{i-1} (G_{j,\xi}(x_i, y_i) - G_{j,\xi}(x_k, y_k))^2}, \quad (25)$$

$t_n = \sqrt{\log n/n}$ , and the data-driven thresholds

$$T_{j,\xi,\gamma} \stackrel{\text{def}}{=} 2\sqrt{2\gamma}t_n\hat{\sigma}_{j,\xi} + \frac{28}{3}M_{j,\xi}\frac{\gamma\log n}{n-1},$$

where  $M_{j,\xi}$  is an upper bound on the sup-norm over  $H^+ \times \{\pm 1\}$  of  $G_{j,\xi}(x, y) - \mathbb{E}[G_{j,\xi}(X, Y)] = G_{j,\xi}(x, y) - \beta_{j,\xi}^a$  (e.g.,  $2\|G_{j,\xi}\|_\infty$ ). For example, using (14) and Proposition 5, we get

$$2\|G_{j,\xi}\|_\infty \leq 2\left\|\mathcal{K}^{-1}\left(\psi_{j,\xi}^-\right)\right\|_{B_X} \leq 2C_\infty B(d, \infty)2^{j(\nu(d)+(d-1)/2)}B_X \stackrel{\text{def}}{=} M_j. \quad (26)$$

The second term in  $T_{j,\xi,\gamma}$  controls the error in estimating the threshold.

**Theorem 9** For  $J$  such that  $2^{J(\nu(d)+(d-1)/2)}B_X^{1/2} \simeq t_n^{-1}$ ,  $M > 0$ , and  $s > (d-1)/r$ ,

(i) If  $z > 1$  and  $\gamma > z/2 + 1$ , we have

$$\sup_{f_\beta \in B_{r,q}^s(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f}_\beta^{a,\rho} - f_\beta \right\|_\infty^z \leq \tilde{c}(d, \infty, r, s, \gamma)(\log n)^{z-1} M^r (B_X t_n)^{\mu_{\text{sparse}}(d, \infty, r, s)z}. \quad (27)$$

(ii) If  $p < \infty$  and  $\gamma > p/2$ , we have

$$\sup_{f_\beta \in B_{r,q}^s(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f}_\beta^{a,\rho} - f_\beta \right\|_p^p \leq \tilde{c}(d, p, r, s, \gamma)(\log n)^{p-1} M^\varpi (B_X t_n)^{\mu(d,p,r,s)p}, \quad (28)$$

where  $\mu(d, p, r, s) = \mu_{\text{dense}}(d, p, r, s)$  and  $\varpi = r$  in the dense zone, while  $\mu(d, p, r, s) = \mu_{\text{sparse}}(d, p, r, s)$  and  $\varpi > p \frac{\nu(d)+(d-1)(1/2-1/p)}{s+\nu(d)-(d-1)(1/r-1/2)}$  is arbitrary in the sparse zone, and  $\tilde{c}(d, p, r, s, \gamma)$  is a constant which depends on  $d, p, r, s$ , and  $\gamma$ .

The upper bounds in Theorem 9 match the lower bound in Theorem 8 up to logarithmic factors. Hence, the proposed estimator is minimax adaptive (up to the log factors).

## 6 Simulation study

We study the performance of the estimator when  $d = 3$ ,  $n = 3000, 5000, 10000$ , and  $X$  is uniform on  $H^+$ . We use of the Von Mises-Fisher distribution  $\text{vMF}(\mu, \kappa)$  with density

$$f(\beta; \mu, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \mu^\top \beta)$$

with respect to  $\sigma$ . We take  $\beta = (\tilde{\beta}_1, \tilde{\beta}_2, |\tilde{\beta}_3|)$  in the cases:

- $\tilde{\beta}$  follows a  $\text{vMF}(\mu, \kappa)$  distribution where  $\mu = (0 \ 0 \ 1)^\top$  and  $\kappa = 10$ .
- $\tilde{\beta}$  follows a mixture  $\lambda \text{vMF}(\mu_1, \kappa) + (1 - \lambda) \text{vMF}(\mu_2, \kappa)$ , where  $\mu_1 = (2^{-1/2} \ 0 \ 2^{-1/2})^\top$ ,  $\mu_2 = (-2^{-1/2} \ 0 \ 2^{-1/2})^\top$ ,  $\kappa = 10$  and  $\lambda = 0.3$ .

We use the cubature defined in spherical coordinates as a product of the Gauss-Legendre quadrature with  $m$  nodes and trapezoid rule with  $2m$  subdivisions (see [1]). The resulting cubature has  $2m^2$  nodes and integrates exactly all polynomials on the sphere up to degree  $2m - 1$ . We take the same function  $a$  as in [2].

The threshold is driven by the parameter  $\gamma$ . The choice of  $\gamma$  slightly depends on the targeted norm. Here we focus on a simultaneous control of the  $L^1$ ,  $xL^2$ ,  $xL^4$  and  $L^\infty$  norm. According to our analysis,  $\gamma$  should be chosen strictly larger than 4. We have nevertheless chosen to use  $\gamma = 4$  which turns out to be sufficient in practice.

Figure 1 displays the distribution of estimates based on a Monte-Carlo experiments with 100 replications and  $n = 3000$ . We plot the Lambert equal-area projection on the disk which is defined (see [22])

$$(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)^\top \mapsto 2 \sin \left( \frac{\theta}{2} \right) (\cos \phi, \sin \phi)^\top.$$

Our main contribution is a control of the estimation error for all  $L^p$  norm. Table 1 displays the expected risk, approximated using Monte-Carlo and 100 replications, for some  $L^p$  norms. More precisely, we have approximated the following renormalized quantities:  $\left( \mathbb{E} \left[ \left\| \widehat{f}_\beta - f_\beta \right\|_p^p \right] / \|f_\beta\|_p^p \right)^{1/p}$

for  $p = \{1, 2, 4\}$  and  $\mathbb{E} \left[ \left\| \widehat{f}_\beta - f_\beta \right\|_\infty \right] / \|f_\beta\|_\infty$ . Figure 2 displays the decay of those error with respect to  $n$  in a logarithmic scales. As expected, we observe a simultaneous control over all norm and the error decays follows the power law given by the upper bounds. The results are similar to the one obtained in [10] except that our threshold does not depend on the unknown regularity of the function whereas the level used in [10] depends on it.

Risk	$n$	Unimodal					Mixture				
		1000	2000	3000	5000	10000	1000	2000	3000	5000	10000
$\mathbb{E} \left[ \left\  \widehat{f}_\beta - f_\beta \right\ _1 \right] / \ f_\beta\ _1$		0.89	0.64	0.53	0.43	0.32	0.92	0.68	0.57	0.46	0.34
$\left( \mathbb{E} \left[ \left\  \widehat{f}_\beta - f_\beta \right\ _2^2 \right] / \ f_\beta\ _2^2 \right)^{1/2}$		0.6	0.43	0.35	0.29	0.21	0.821	0.6	0.5	0.4	0.29
$\left( \mathbb{E} \left[ \left\  \widehat{f}_\beta - f_\beta \right\ _4^4 \right] / \ f_\beta\ _4^4 \right)^{1/4}$		0.49	0.36	0.29	0.24	0.17	0.8	0.58	0.48	0.38	0.27
$\mathbb{E} \left[ \left\  \widehat{f}_\beta - f_\beta \right\ _\infty \right] / \ f_\beta\ _\infty$		0.42	0.32	0.26	0.21	0.17	0.86	0.6	0.51	0.39	0.29

Table 1: Risk.

## 7 Appendix

### 7.1 A preliminary lemma

**Lemma 10** *The following equality holds for every  $g \in L^2(\mathbb{S}^{d-1})$ ,*

$$\langle R, g \rangle = 2\mathbb{E} \left[ \frac{Yg^-(X)}{f_X(X)} \right].$$

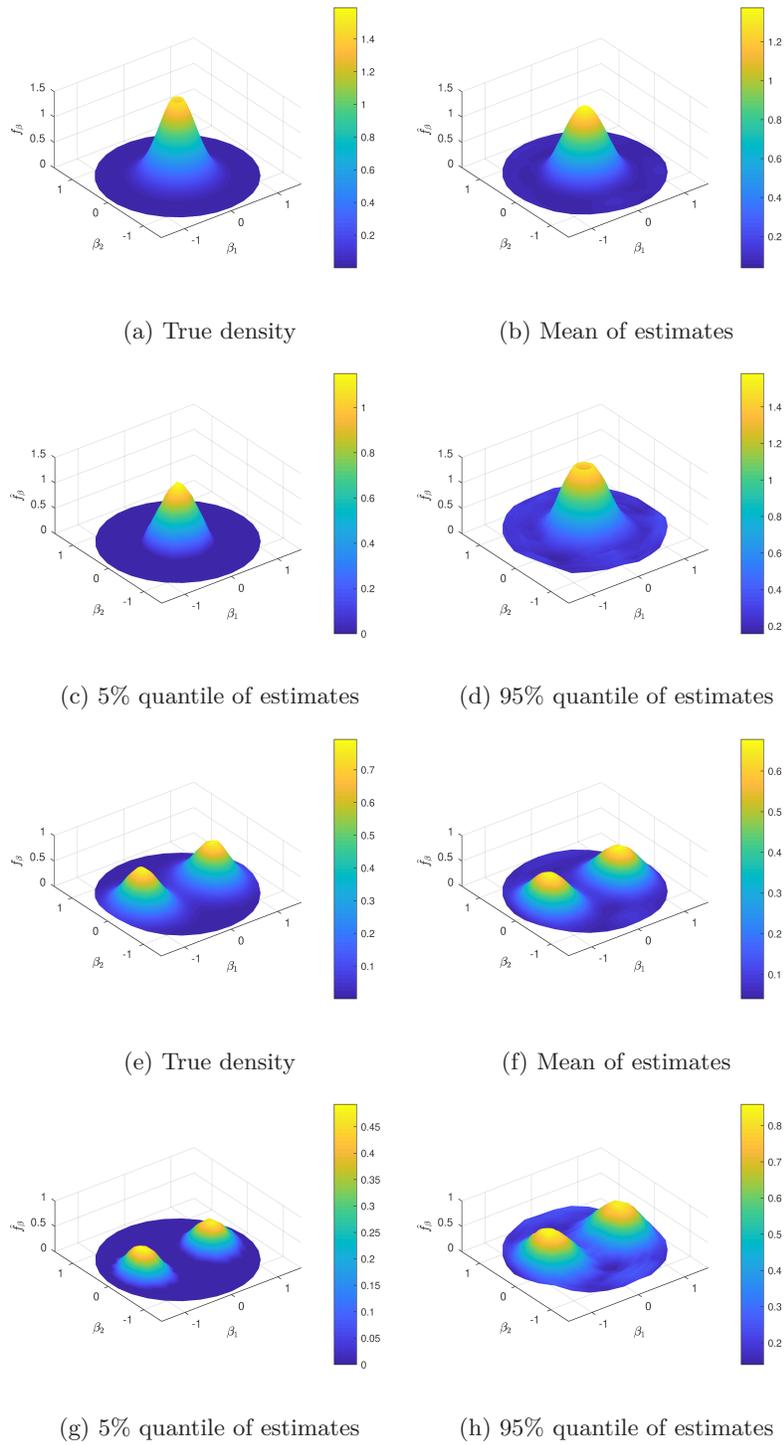


Figure 1: True density and distribution of the estimates.

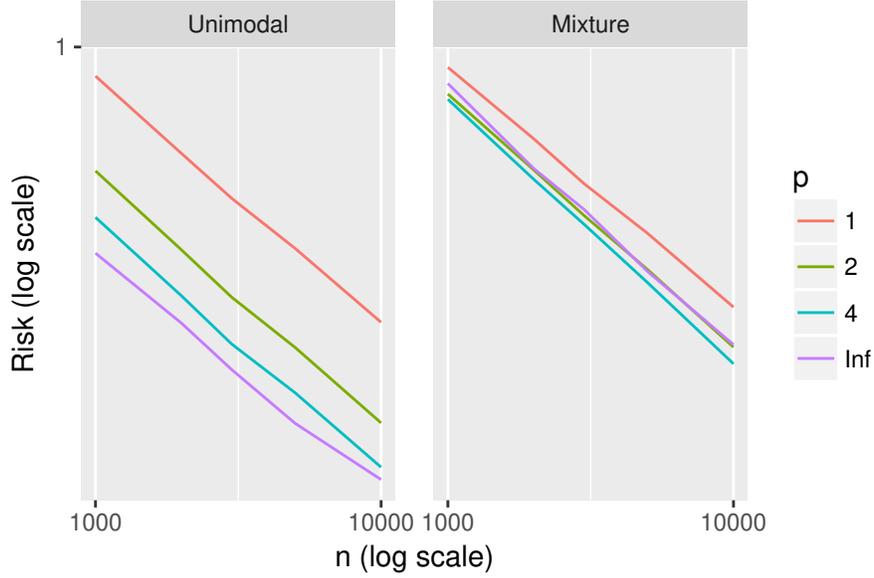


Figure 2: Decay of the risk with  $n$  in logarithmic scales.

**Proof.** The result is based on the following

$$\begin{aligned}
\langle R, g \rangle &= \langle R, g^- \rangle \quad (\text{because } R \text{ is odd}) \\
&= 2 \int_{H^+} \frac{R(x)g^-(x)}{f_X(x)} f_X(x) d\sigma(x) \\
&= 2\mathbb{E} \left[ \frac{R(X)g^-(X)}{f_X(X)} \right] \\
&= 2\mathbb{E} \left[ \frac{\mathbb{E}[Y|X]g^-(X)}{f_X(X)} \right]. \quad \square
\end{aligned}$$

## 7.2 Proof of Proposition 5

The operator  $\mathcal{K}$  is related to the Hemispherical transform (see [10, 26]) defined for  $f \in L^1(\mathbb{S}^{d-1})$  and a.e.  $x \in \mathbb{S}^{d-1}$  by

$$\mathcal{H}(f)(x) \stackrel{\text{def}}{=} \int_{\mathbb{S}^{d-1}} \mathbf{1}_{x^\top y > 0} f(y) d\sigma(y),$$

through

$$Kf = 2\mathcal{H}(f) - \int_{\mathbb{S}^{d-1}} f(y) d\sigma(y).$$

(P1.1) is a consequence of the fact that  $y \rightarrow x^\top y \in L^\infty_{\text{odd}}(\mathbb{S}^{d-1})$ . (P1.2) follows from Theorem 2 (ii), and (P1.3) follows from Theorem C in [26]. The second part of the proposition together with (P1.4) are consequences of the properties of  $\mathcal{H}$  detailed in [10]. The inequalities (9) correspond to

Lemma A.2. Note however that there is a typo in the proof and we should read  $1.3 \dots (2p - 1) \asymp p^{-1/2} 2.4 \dots (2p)$  but the result still holds.

### 7.3 Proof of Theorem 8

Start by noting that for every  $j \in \mathbb{N}$  and  $\xi \in \Xi_j$ ,

$$\int_{\mathbb{S}^{d-1}} \psi_{j,\xi}(x) dx = \omega(j, \xi) b(0) = \omega(j, \xi) (a(0) - a(0)) = 0.$$

This implies that the functions  $f_m$  that we introduce below integrate to 1.

#### 7.3.1 Proof of the lower bound in the dense zone

Consider the family  $(P_m)_{m=0}^{\mathcal{M}}$ , where  $\mathcal{M} \in \mathbb{N}^*$ , of distributions of an i.i.d. sample of  $(Y, X)$  of size  $n$  when  $f_\beta = f_m$  and the density of  $X$  is  $f_X$ . These probabilities are absolutely continuous with respect to the product of  $\delta_1 + \delta_{-1}$ , where  $\delta_y$  denotes the Dirac mass at  $y$  and  $\sigma$ . Take  $j \in \mathbb{N}$ ,  $f_0 = 1/\sigma(\mathbb{S}^{d-1})$ , and consider the set  $A_j$  from Lemma 6 (ii). By the Varshamov-Guilbert bound (Lemma 2.9 in [27]) there exists  $\Omega \subseteq \{0, 1\}^{A_j}$  containing  $(0, \dots, 0)$  such that  $|\Omega| = 2^{|A_j|/8}$  and  $\forall (\omega_1, \omega_2) \in \Omega^2$ ,  $\|\omega_1 - \omega_2\|_{\ell^1} \geq |A_j|/8$ . Enumerate the elements of  $\Omega$  from 0 (corresponding to the zero vector) to  $\mathcal{M} \stackrel{\text{def}}{=} |\Omega| - 1$  and define

$$f_m \stackrel{\text{def}}{=} f_0 + \gamma \sum_{\xi \in A_j} \omega_\xi \psi_{j,\xi}$$

when  $(\omega_\xi)_{\xi \in A_j}$  is the  $m$ th element of  $\Omega$  and  $\gamma = c C_{\Xi}^{-1/r} M 2^{-j(s+(d-1)/2)}$  for  $0 < c < 1$  such that all  $f_m$  are nonnegative. We now use the following result (see Theorem 2.5 in [27]).

**Lemma 11** *If for  $0 < \alpha < 1/8$  we have:*

- (i)  $f_m \in B_{r,q}^s(M) \cap \mathcal{D}$  for  $m = 0, \dots, \mathcal{M}$ ,
- (ii)  $\forall 0 \leq m < l \leq \mathcal{M}$ ,  $\|f_m - f_l\|_p \geq 2h > 0$ ,
- (iii)  $\frac{1}{\mathcal{M}} \sum_{m=1}^{\mathcal{M}} K(P_m, P_0) \leq \alpha \ln(\mathcal{M})$ ,

then for every  $z \geq 1$

$$\inf_{\widehat{f}_\beta} \sup_{f_\beta \in B_{r,q}^s(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f}_\beta - f_\beta \right\|_p^z \geq h^z \frac{\sqrt{\mathcal{M}}}{1 + \sqrt{\mathcal{M}}} \left( 1 - 2\alpha - \sqrt{\frac{2\alpha}{\ln(\mathcal{M})}} \right). \quad (29)$$

Start by checking (i) in Lemma 11. It is enough to show that  $f_m \in B_{r,q}^s(M)$ . Indeed, for  $r \geq 1$  and  $\omega \in \Omega$ , we have  $\left\| (\omega_\xi)_{\xi \in A_j} \right\|_{\ell^r} \leq \left\| (\omega_\xi)_{\xi \in A_j} \right\|_{\ell^1}^{1/r} \leq C_{\Xi}^{1/r} 2^{j(d-1)/r}$ , we obtain

$$\gamma 2^{j(s+(d-1)(1/2-1/r))} \left\| (\omega_\xi)_{\xi \in A_j} \right\|_{\ell^r} \leq \gamma C_{\Xi}^{1/r} 2^{j(s+(d-1)/2)} \leq M.$$

Lemma 6 (ii) now yields that for every  $1 \leq p \leq \infty$  and  $0 \leq m < l \leq \mathcal{M}$

$$\|f_m - f_l\|_p \geq \gamma c_{p,A} 2^{j(d-1)(1/2-1/p)} \left( \frac{c_A}{8} 2^{j(d-1)} \right)^{1/p} = 2h.$$

Thus (ii) in Lemma 11 follows with  $h = c_{p,A} \left(\frac{c_A}{8}\right)^{1/p} cC_{\Xi}^{-1/r} M2^{-js-1}$ .  
By independence, the Kullback-Leibler divergence between  $P_m$  and  $P_0$  is given by

$$K(P_m, P_0) = n\mathbb{E} \left[ \mathcal{H}(f_m)(X) \ln \left( \frac{\mathcal{H}(f_m)(X)}{\mathcal{H}(f_0)(X)} \right) + (1 - \mathcal{H}(f_m)(X)) \ln \left( \frac{1 - \mathcal{H}(f_m)(X)}{1 - \mathcal{H}(f_0)(X)} \right) \right].$$

Using that, for  $x > 0$ ,  $\ln(x) \leq x - 1$ , we obtain

$$K(P_m, P_0) \leq n\mathbb{E} \left[ \frac{\mathcal{H}(f_m - f_0)(X)^2}{\mathcal{H}(f_0)(X)(1 - \mathcal{H}(f_0)(X))} \right],$$

and thus

$$\begin{aligned} K(P_m, P_0) &\leq 4nA_X \|\mathcal{H}(f_m - f_0)\|_2^2 \\ &\leq 4nA_X \lambda_{2^j+1,d}^2 \|f_m - f_0\|_2^2, \end{aligned}$$

where the last display comes from the fact that  $f_m - f_0 \in \bigoplus_{2^j+1 \leq k \leq 2^{j+2}-1} H^{k,d}$ . From (9) we get

$$K(P_m, P_0) \leq 4C_{\lambda}(d)^2 nA_X 2^{-2j\nu(d)} \|f_m - f_0\|_2^2,$$

which yields using Lemma 6 (i)

$$\begin{aligned} K(P_m, P_0) &\leq (2C_{\lambda}(d)C'_2\gamma)^2 nA_X 2^{-2j\nu(d)} \left\| (\omega_{\xi})_{\xi \in A_j} \right\|_{\ell^2}^2 \\ &\leq (2C_{\lambda}(d)C'_2\gamma)^2 nA_X 2^{-2j\nu(d)} \left\| (\omega_{\xi})_{\xi \in A_j} \right\|_{\ell^1}^2 \\ &\leq (2C_{\lambda}(d)C'_2\gamma)^2 C_{\Xi} nA_X 2^{j(d-1-2\nu(d))} \\ &\leq (2C_{\lambda}(d)C'_2cM)^2 C_{\Xi}^{1-2/r} nA_X 2^{-2j(s+\nu(d))}. \end{aligned}$$

Condition (iii) of Lemma 11 is satisfied once

$$\frac{2^5 (C_{\lambda}(d)C'_2cM)^2}{\ln(2)} C_{\Xi}^{-2/r} nA_X 2^{-2j(s+\nu(d)+(d-1)/2)} \leq \alpha < \frac{1}{8}. \quad (30)$$

For  $\alpha < 1/8$ , the lower bound (29) yields that

$$\begin{aligned} \inf_{\widehat{f}_{\beta}} \sup_{f_{\beta} \in B_{r,q}^s(\mathcal{M})} \mathbb{E} \left\| \widehat{f}_{\beta} - f_{\beta} \right\|_p^z &\geq \left( c_{p,A} \left(\frac{c_A}{8}\right)^{1/p} cC_{\Xi}^{-1/r} M2^{-js-1} \right)^z \left( \frac{3}{4} - \frac{1}{2\sqrt{\ln(\mathcal{M})}} \right) \\ &\geq \frac{1}{2} \left( c_{p,A} \left(\frac{c_A}{8}\right)^{1/p} cC_{\Xi}^{-1/r} \frac{M}{2} \right)^z 2^{-jsz}, \end{aligned}$$

where the inequality leading to the second display holds when  $\ln(\mathcal{M}) \geq 4$ , for example for  $j(d-1) \geq \ln(5/c_A \ln(2))/\ln(2)$ . Now (30) is satisfied for

$$j \geq j_0 \stackrel{\text{def}}{=} 1 + \frac{\ln \left( 2^8 (C_{\lambda}(d)C'_2cM)^2 C_{\Xi}^{-2/r} nA_X / \ln(2) \right)}{2 \ln(2)(s + \nu(d) + (d-1)/2)},$$

which implies the lower bound

$$\begin{aligned} & \inf_{\widehat{f}_\beta} \sup_{f_\beta \in B_{r,q}^s(M)} \mathbb{E} \left\| \widehat{f}_\beta - f_\beta \right\|_p^z \\ & \geq \frac{1}{2} \left( c_{p,A} \left( \frac{c_A}{8} \right)^{1/p} c C_{\Xi}^{-1/r} M 2^{-s-1} \right)^z \left( \frac{2^8 (C_\lambda(d) C_2' c M)^2 C_{\Xi}^{-2/r} n A_X}{\ln(2)} \right)^{-\mu_{\text{dense}}(d,p,r,s)z/2}. \end{aligned}$$

### 7.3.2 Proof of the lower bound in the sparse zone

In this proof we consider asymptotic orders for simplicity. The various constants can be obtained like in Section 7.3.1. Consider the hypotheses

$$f_m = \frac{1}{\sigma(\mathbb{S}^{d-1})} + \gamma \psi_{j,\xi_m},$$

where  $\xi_m \in A_j$  and  $|\gamma| \lesssim 2^{-j(d-1)/2}$  to ensure the functions are positive. The constant is adjusted so that for one of the  $f_m$  that we denote  $f_0, \forall x \in H^+, |\mathcal{H}(f_0^-)(x)| \leq c_b$  with  $c_b \in (0, \frac{1}{2})$ . The function  $f_m$  also integrate to 1. We denote by  $\mathcal{M}$  the cardinality of  $A_j$  ( $\mathcal{M} \simeq 2^{j(d-1)}$ ),  $P_m$  the distributions of an i.i.d. sample of  $(Y, X)$  of size  $n$  when  $f_\beta = f_m$  and for a given  $f_X$ , and  $\Lambda(P_m, P_0)$  the likelihood ratio. Recall that  $K(P_m, P_0) = \mathbb{E}_{P_m} [\Lambda(P_m, P_0)]$ . We make use of the following Lemma from [20].

**Lemma 12** *If for  $\pi_0 > 0$  and  $\mathcal{M} \in \mathbb{N}^*$  the following three condition hold*

- (i)  $f_m \in B_{r,q}^s(M) \cap \mathcal{D}$  for  $m = 1, \dots, \mathcal{M}$ ,
- (ii)  $\forall m \neq l, \|f_m - f_l\|_p \geq 2h > 0$ ,
- (iii)  $\forall m = 1, \dots, \mathcal{M}, \Lambda(P_0, P_m) = \exp(z_n^m - v_n^m)$ , where  $z_n^m$  are random variables and  $v_n^m$  constants such that  $\mathbb{P}(z_n^m > 0) \geq \pi_0$  and  $\exp(\sup_{m=1, \dots, \mathcal{M}} v_n^m) \leq \mathcal{M}$ ,

then

$$\forall z \geq 1, \inf_{\widehat{f}_\beta} \sup_{f_\beta \in B_{r,q}^s(M) \cap \mathcal{D}} \mathbb{E} \left\| \widehat{f}_\beta - f_\beta \right\|_p^z \geq \frac{h^{-z} \pi_0}{2}.$$

Item (i) is satisfied when  $|\gamma| \leq M 2^{-j(s-(d-1)(1/r-1/2))}$ . This is more restrictive than the condition to ensure positivity because we assume that  $s \geq (d-1)/r$ . Thus, now we take  $\gamma = 2cM 2^{-j(s-(d-1)(1/r-1/2))}$  for a well-chosen constant  $c$ .

The constant  $h$  in (ii) is obtained as follows, if  $m \neq m'$ ,

$$\begin{aligned} \|f_m - f_{m'}\|_p &= \gamma \|\psi_{j,\xi_m} - \psi_{j,\xi_{m'}}\|_p \\ &\geq \gamma c_{p,A} 2^{j(d-1)(1/2-1/p)} \\ &\geq 2cM 2^{-j(s-(d-1)(1/r-1/p))}. \end{aligned}$$

Let us now consider item (iii), we obtain

$$\begin{aligned} P_m(\log(\Lambda(P_0, P_m))) &\geq -j(d-1) \log 2 \geq 1 - P_m(|\log(\Lambda(P_0, P_m))|) \geq j(d-1) \log 2 \\ &\geq 1 - \frac{\mathbb{E}_{P_m} [|\log(\Lambda(P_0, P_m))|]}{j(d-1) \log 2}. \end{aligned}$$

Thus, condition (iii) is satisfied when

$$\mathbb{E}_{P_m} [|\log(\Lambda(P_0, P_m))|] \leq \alpha j(d-1) \log 2,$$

for  $\alpha \in (0, 1)$ . The same computations as in the beginning of Section 5.1 yield that we need to impose  $n2^{-2j\nu(d)}\gamma^2 \lesssim j$ , thus

$$A_X n 2^{-2j(s+\nu(d)-(d-1)(1/r-1/2))} \lesssim j.$$

The desired rate is obtained by taking

$$2^j \simeq \left( \frac{nA_X}{\log(nA_X)} \right)^{\frac{1}{2(s+\nu(d)-(d-1)(1/r-1/2))}}.$$

## 7.4 Comparison between Besov ellipsoids of a function and its odd part

**Lemma 13** *For  $0 < s, q \leq \infty$  and  $1 \leq r \leq \infty$ , there exists a constant  $c_{\text{eq}}$  that can depend on  $d$  such that, for every  $f \in B_{r,q}^s$ ,  $\|f^-\|_{B_{r,q}^s} \leq c_{\text{eq}} \|f\|_{B_{r,q}^s}$ .*

**Proof.** In Definition 4 every  $f \in B_{r,q}^s(\mathbb{S}^{d-1})$  has same norm as  $x \rightarrow f(-x)$ , thus by the triangle inequality  $\|f^-\|_{B_{r,q}^s}^A \leq \|f\|_{B_{r,q}^s}^A$ . We conclude by equivalence of the norms.  $\square$

## 7.5 A general inequality

We make use of the constants  $c_{1,z}$  and  $c_{2,z}$  such that

$$\int_{\mathbb{R}^+} z\tau^{z-1} e^{-\beta\tau} d\tau \leq c_{1,z} \beta^{-z} \quad (31)$$

$$\int_{\mathbb{R}^+} z\tau^{z-1} e^{-\alpha\tau^2} d\tau \leq c_{2,z} \alpha^{-z/2}. \quad (32)$$

**Lemma 14** *For every  $\tau, \gamma, z > 1$  and*

$$T_{j,\xi,\gamma}^{s,++} \geq 3\sqrt{2}\gamma t_n \widehat{\sigma}_{j,\xi} + 26M_{j,\xi} \frac{\gamma \log n}{n-1} \stackrel{\text{def}}{=} T_{j,\xi,\gamma}^{s,+},$$

*the two following inequalities hold:*

*when  $p = \infty$ ,*

$$\begin{aligned} & \frac{1}{2^{z-1}} \mathbb{E} \left[ \left\| \widehat{f}_\beta^{a,\rho} - f_\beta \right\|_\infty^z \right] \\ & \leq \left\| f_\beta^{-a,J} - f_\beta^- \right\|_\infty^z + (J+1)^{z-1} C_\infty^z \left\{ \right. \\ & \quad a_{n,\infty,z,J} \sum_{j=0}^J 2^{j(d-1)z/2} \left( \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right] \right) \\ & \quad + \frac{4C_\Xi}{n^\gamma} \sum_{j=0}^J 2^{j(d-1)(z/2+1)} \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \\ & \quad \left. + \left( \frac{C_\Xi 4}{n^\gamma} \right)^{1-1/\tau} \left( \frac{1}{\sqrt{n}} B_X^{1/2} 2^{Jz(\nu(d)+(d-1)/2)} \right)^z 2^{J(d-1)(1-1/\tau)} b_{n,\infty,z,J,\tau} \right\}, \end{aligned}$$

where

$$\begin{aligned}
a_{n,\infty,z,J} &= 1 + \left( \frac{2}{\sqrt{\gamma \log n}} \right)^z \left( 2 + \left( \log \left( C_{\Xi} 2^{J(d-1)} c_{2,z} \right) \right)^{z/2} \right) \\
&\quad + \left( \frac{4}{\gamma \log n} \right)^z \left( 2 + \left( \log \left( C_{\Xi} 2^{J(d-1)} c_{1,z} \right) \right)^z \right) \\
b_{n,\infty,z,J,\tau} &= \frac{(2\sqrt{2}C_2B(d,2))^z \left( 2^{1/\tau} + \left( \log \left( C_{\Xi} 2^{J(d-1)} c_{2,z\tau} \right) \right)^{z/2} \right)}{1 - 2^{-(z\nu(d)+(d-1)(z/2+1-1/\tau))}} \\
&\quad + \frac{(8C_{\infty}B(d,\infty)/3)^z \left( 2^{1/\tau} + \left( \log \left( C_{\Xi} 2^{J(d-1)} c_{1,z} \right) \right)^z \right)}{1 - 2^{-(z\nu(d)+(d-1)(z+1-1/\tau))}} \left( \frac{2^{J(d-1)}}{n} B_X \right)^{z/2};
\end{aligned}$$

while, when  $1 \leq p \leq \infty$ ,

$$\begin{aligned}
&\frac{1}{2^{z-1}} \mathbb{E} \left[ \left\| \widehat{f}_{\beta}^{a,\rho} - f_{\beta} \right\|_p^z \right] \\
&\leq \left\| f_{\beta}^{-a,J} - f_{\beta}^{-} \right\|_p^z + (J+1)^{z-1} C_p'^z C_{\Xi}^{z/(p \wedge z) - 1} \left\{ \right. \\
&\quad a_{n,p,z,J} \sum_{j=0}^J 2^{j(d-1)(z/2-z/(p\nu z))} \sum_{\xi \in \Xi_j} \left( |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right) \\
&\quad + \frac{4}{n^{\gamma}} \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p\nu z))} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \\
&\quad \left. + \frac{2^{2-1/\tau}}{n^{\gamma(1-\frac{1}{p})}} C_{\Xi} \left( \frac{1}{\sqrt{n}} B_X^{1/2} 2^{J(\nu(d)+(d-1)/2)} \right)^z 2^{J(d-1)(1-z/(p\nu z))} b_{n,p,z,J,\tau} \right\},
\end{aligned}$$

where

$$\begin{aligned}
a_{n,p,z,J} &= 1 + 2 \left( \left( \frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma \log n}} \right)^z + \left( \frac{2c_{1,z}^{1/z}}{\gamma \log n} \right)^z \right) \\
b_{n,p,z,J,\tau} &= \frac{(2c_{2,z\tau}^{1/(z\tau)} C_2 B(d,2))^z}{1 - 2^{-(z\nu(d)+(d-1)(z/2+1-z/(p\nu z)))}} + \frac{\left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_{\infty} B(d,\infty) \right)^z}{1 - 2^{-(z\nu(d)+(d-1)(z+1-z/(p\nu z)))}} \left( \frac{2^{J(d-1)}}{n} B_X \right)^{z/2}.
\end{aligned}$$

The inequalities of Lemma 14 are similar to oracle inequalities, for a well-chosen  $J$  depending on  $n$  (see Theorem 9), where the oracle estimates  $\beta_{j,\xi}^a$  if and only if the error made by estimating this coefficient is smaller than the one made by discarding it. This *oracle* strategy would lead to a quantity of the form

$$|\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq (\mathbb{E} [ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|^z ] )^{1/z}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > (\mathbb{E} [ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|^z ] )^{1/z}}.$$

Proving such an oracle inequality would require to lower bound  $\left( \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \right)^{1/z}$ . In the inequalities of Lemma 14 the ideal quantity  $\left( \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \right)^{1/z}$  is replaced by  $T_{j,\xi,\gamma}^{s,++}$ , called

*quasi-oracle*. The remaining terms can be made as small as we want by taking  $\gamma$  large enough. The last term corresponds to the approximation error. Upper bounds of these types, uniform on Besov ellipsoids, yield an approximation error which can be expressed in terms of the regularity of the Besov class and is uniformly small for  $J$  large enough and allows to treat the bias/variance trade-off in the quasi-oracle term uniformly over the ellipsoid.

## 7.6 Proof of Lemma 14

### 7.6.1 Preliminaries

Recall from the proof of Theorem 4.1 in [10] that for every  $1 \leq p \leq \infty$

$$\left\| \widehat{f}_\beta^{a,\rho} - f_\beta \right\|_p \leq 2 \left\| \widehat{f}_\beta^{a,\rho} - f_\beta^- \right\|_p,$$

and that, for  $1 \leq z < \infty$ , we have

$$\left\| \widehat{f}_\beta^{a,\rho} - f_\beta^- \right\|_p^z \leq 2^{z-1} \left( \left\| \widehat{f}_\beta^{a,\rho} - f_\beta^{-a,J} \right\|_p^z + \left\| f_\beta^{-a,J} - f_\beta^- \right\|_p^z \right). \quad (33)$$

The first term corresponds to the error in the high dimensional space while the second term corresponds to the approximation error. Let us start by studying the first term.

Lemma 6 (i) yields

$$\begin{aligned} \left\| \widehat{f}_\beta^{a,\rho} - f_\beta^{-a,J} \right\|_p^z &\leq (J+1)^{z-1} \sum_{j=0}^J \left\| \sum_{\xi \in \Xi_j} \left( \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right) \psi_{j,\xi} \right\|_p^z \\ &\leq (J+1)^{z-1} \sum_{j=0}^J C_p'^z 2^{j(d-1)z(1/2-1/p)} \left\| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right\|_p^z. \end{aligned}$$

Thus, for  $p = \infty$ , we have

$$\left\| \widehat{f}_\beta^{a,\rho} - f_\beta^{-a,J} \right\|_p^z \leq (J+1)^{z-1} \sum_{j=0}^J C_\infty'^z 2^{j(d-1)z/2} \sup_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z,$$

while, for  $p < \infty$ , we have

$$\left\| \widehat{f}_\beta^{a,\rho} - f_\beta^{-a,J} \right\|_p^z \leq (J+1)^{z-1} C_p'^z C_\Xi^{z/(p \wedge z) - 1} \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p \vee z))} \sum_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z.$$

The last inequality is obtained by using that, when  $p \geq z$ , we have

$$\left( \sum_{\xi \in \Xi_j} |b_\xi|^p \right)^{z/p} \leq \sum_{\xi \in \Xi_j} |b_\xi|^z,$$

and by the Hölder inequality, when  $p \leq z$ , we have

$$\left( \sum_{\xi \in \Xi_j} |b_\xi|^p \right)^{z/p} \leq C_\Xi^{z/p-1} \sum_{\xi \in \Xi_j} |b_\xi|^z.$$

### 7.6.2 Coefficientwise analysis

For the simplicity of the notations we sometimes drop the dependence on  $\gamma$  in the sets of indices. We first consider the term

$$\delta_{j,\xi,z} \stackrel{\text{def}}{=} \left| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z.$$

By construction we have

$$\begin{aligned} \delta_{j,\xi,z} &= |\beta_{j,\xi}^a|^z \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| \leq T_{j,\xi,\gamma}} + \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| > T_{j,\xi,\gamma}} \\ &= \max \left( |\beta_{j,\xi}^a|^z \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| \leq T_{j,\xi,\gamma}}, \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| > T_{j,\xi,\gamma}} \right). \end{aligned}$$

We introduce two ‘‘phantom’’ random thresholds  $T_{j,\xi,\gamma}^b = T_{j,\xi,\gamma} - \Delta_{j,\xi,\gamma}$  and  $T_{j,\xi,\gamma}^s = T_{j,\xi,\gamma} + \Delta_{j,\xi,\gamma}$  for some  $\Delta_{j,\xi,\gamma}$  to be defined later. They are used to define *big* and *small* original needlet coefficients. We also use  $T_{j,\xi,\gamma}^{b,-}$  for a deterministic lower bound on  $T_{j,\xi,\gamma}^b$ ,  $T_{j,\xi,\gamma}^{s,+}$  and  $\Delta_{j,\xi,\gamma}^+$  for deterministic upper bounds on  $T_{j,\xi,\gamma}^s$  and  $\Delta_{j,\xi,\gamma}$ . These bounds will hold with high probability. We obtain almost surely

$$\begin{aligned} \delta_{j,\xi,z} &= \max \left( |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| \leq T_{j,\xi,\gamma}} \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| \leq T_{j,\xi,\gamma}} \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^s} \right), \right. \\ &\quad \left. \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| > T_{j,\xi,\gamma}} \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^b}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a| > T_{j,\xi,\gamma}} \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^b} \right) \right) \\ &\leq \max \left( |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right), \right. \\ &\quad \left. \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}}, \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^b} \right) \right) \\ &\leq \max \left( |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}}, \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right), \right. \\ &\quad \left. \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}}, \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}}, \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b} \right) \right). \end{aligned}$$

Sorting the terms according to the number of random terms we obtain

$$\begin{aligned} \delta_{j,\xi,z} &\leq \max \left( |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}}, |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right), \right. \\ &\quad \left. \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}}, \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}}, \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b} \right) \right). \end{aligned}$$

### 7.6.3 Scalewise analysis

Defining

$$\begin{aligned} M_{j,z} &\stackrel{\text{def}}{=} \sup_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z = \sup_{\xi \in \Xi_j} \delta_{j,\xi,z} \\ S_{j,z} &\stackrel{\text{def}}{=} \sum_{\xi \in \Xi_j} \left| \rho_{T_{j,\xi,\gamma}} \left( \widehat{\beta}_{j,\xi}^a \right) - \beta_{j,\xi}^a \right|^z = \sum_{\xi \in \Xi_j} \delta_{j,\xi,z}, \end{aligned}$$

we obtain

$$\begin{aligned}
M_{j,z} &\leq \max \left( \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}}, \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right), \right. \\
&\quad \left. \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}}, \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \right) \\
&\stackrel{\text{def}}{=} \max(M_{j,z}^{S0}, M_{j,z}^{S1}, M_{j,z}^{B1}, M_{j,z}^{B2}) \\
&\leq M_{j,z}^{S0} + M_{j,z}^{S1} + M_{j,z}^{B1} + M_{j,z}^{B2}, \\
S_{j,z} &\leq \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}} + \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \\
&\quad + \sum_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} + \sum_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \\
&\stackrel{\text{def}}{=} S_{j,z}^{S0} + S_{j,z}^{S1} + S_{j,z}^{B1} + S_{j,z}^{B2}.
\end{aligned}$$

We bound the expectations of the random terms as follows

$$\begin{aligned}
\mathbb{E} [M_{j,z}^{S1}] &\leq \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \right] \\
&\leq \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \left( \mathbb{P} \left( \bigcup_{\xi \in \Xi_j} \{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s\} \right) + \mathbb{P} \left( \bigcup_{\xi \in \Xi_j} \{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}\} \right) \right); \\
\mathbb{E} [M_{j,z}^{B1}] &\leq \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right]; \\
\mathbb{E} [M_{j,z}^{B2}] &\leq \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^{z\tau} \right]^{1/\tau} \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \right]^{1-1/\tau} \\
&\leq \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^{z\tau} \right]^{1/\tau} \left( \mathbb{P} \left( \bigcup_{\xi \in \Xi_j} \{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b\} \right) + \mathbb{P} \left( \bigcup_{\xi \in \Xi_j} \{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}\} \right) \right)^{1-1/\tau}; \\
\mathbb{E} [S_{j,z}^{S1}] &= \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbb{E} \left[ \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \right] \\
&\leq \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \left( \mathbb{P} \{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s\} + \mathbb{P} \{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}\} \right); \\
\mathbb{E} [S_{j,z}^{B1}] &= \sum_{\xi \in \Xi_j} \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right]; \\
\mathbb{E} [S_{j,z}^{B2}] &= \sum_{\xi \in \Xi_j} \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \max \left( \mathbf{1}_{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b}, \mathbf{1}_{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}} \right) \right] \\
&\leq \sum_{\xi \in \Xi_j} \left( \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^{z\tau} \right] \right)^{1/\tau} \left( \mathbb{P} \{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b\} + \mathbb{P} \{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| > \Delta_{j,\xi,\gamma}\} \right)^{1-1/\tau}.
\end{aligned}$$

The constant  $\tau > 1$  in the Hölder inequality will be specified later.

#### 7.6.4 Bernstein inequality and the term $|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|^z$

Let denote variance of  $G_{j,\xi}(X, Y)$

$$\sigma_{j,\xi}^2 \stackrel{\text{def}}{=} \mathbb{E} \left[ (G_{j,\xi}(X, Y) - \beta_{j,\xi}^a)^2 \right].$$

**Lemma 15** *We have*

$$\mathbb{E} \left[ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|^z \right] \leq 2 \left( c_{2,z} \left( \frac{2}{\sqrt{n}} \sigma_{j,\xi} \right)^z + c_{1,z} \left( \frac{4}{3n} M_{j,\xi} \right)^z \right).$$

**Proof.** The Bernstein inequality yields

$$\begin{aligned} \mathbb{P} \left\{ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| \geq u \right\} &\leq 2e^{-\frac{nu^2}{2((\sigma_{j,\xi})^2 + M_{j,\xi}u/3)}} \\ &\leq 2 \left( e^{-\frac{nu^2}{4(\sigma_{j,\xi})^2}} + e^{-\frac{3nu}{4M_{j,\xi}}} \right). \end{aligned}$$

Using now  $\mathbb{E} [|X|^z] = \int_{\mathbb{R}^+} zu^{z-1} \mathbb{P}\{|X| > u\} du$ , we obtain

$$\begin{aligned} \mathbb{E} \left[ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|^z \right] &\leq \int_{\mathbb{R}^+} zu^{z-1} \mathbb{P} \left\{ |\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a| \geq u \right\} du \\ &\leq \int_{\mathbb{R}^+} zu^{z-1} 2 \left( e^{-\frac{nu^2}{4(\sigma_{j,\xi})^2}} + e^{-\frac{3nu}{4M_{j,\xi}}} \right) du, \end{aligned}$$

hence the inequality from the lemma follows from (31) and (32).  $\square$

Lemma 15 is used to obtain a uniform upper bound of the power of the ratio between  $|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|$  and a threshold  $c_\sigma \sqrt{\log(n)/n} \sigma_{j,\xi} + c_M \log(n)/(n-1) M_{j,\xi}$ :

$$\begin{aligned} \mathbb{E} \left[ \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{c_\sigma \sqrt{\log(n)/n} \sigma_{j,\xi} + c_M \log(n)/(n-1) M_{j,\xi}} \right)^z \right] &\leq 2 \left( c_{2,z} \left( 2 \frac{1}{c_\sigma \sqrt{\log n} + c_M \frac{\sqrt{n} \log n}{n-1} \frac{M_{j,\xi}}{\sigma_{j,\xi}}} \right)^z \right. \\ &\quad \left. + c_{1,z} \left( \frac{4}{3} \frac{1}{c_\sigma \sqrt{n} \sqrt{\log n} \frac{\sigma_{j,\xi}}{M_{j,\xi}} + c_M \log(n) \frac{n}{n-1}} \right)^z \right) \\ &\leq 2 \left( c_{2,z} \left( 2 \frac{1}{c_\sigma \sqrt{\log n}} \right)^z + c_{1,z} \left( \frac{4}{3} \frac{1}{c_M \log n} \right)^z \right). \end{aligned} \tag{34}$$

The following similar lemma is useful to handle the case  $p = \infty$ .

**Lemma 16** For any  $\Xi'_j \subset \Xi_j$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{u_{j,\xi}} \right)^z \right] &\leq \left( \frac{2\sqrt{2}}{\sqrt{n}} \sup_{\xi \in \Xi'_j} \frac{\sigma_{j,\xi}}{c_{j,\xi}} \right)^z \left( 2 + (\log(c_{2,z} |\Xi'_j|))^{z/2} \right) \\ &\quad + \left( \frac{8}{3n} \sup_{\xi \in \Xi'_j} \frac{M_{j,\xi}}{c_{j,\xi}} \right)^z \left( 2 + (\log(c_{1,z} |\Xi'_j|))^{z/2} \right). \end{aligned} \quad (35)$$

**Proof.** A uniform union bound yields

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{\xi \in \Xi'_j} \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{u_{j,\xi}} \geq \tau \right\} \\ &\leq \min \left( 1, |\Xi'_j| 2 \left( e^{-\frac{1}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2} + e^{-\frac{3}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} \right) \right) \\ &\leq \min \left( 1, |\Xi'_j| 2e^{-\frac{1}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2} \right) + \min \left( 1, |\Xi'_j| 2e^{-\frac{3}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} \right) \end{aligned}$$

This yields

$$\begin{aligned} \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{u_{j,\xi}} \right)^z \right] &\leq \int_{\mathbb{R}^+} z\tau^{z-1} \min \left( 1, |\Xi'_j| 2e^{-\frac{1}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2} \right) d\tau \\ &\quad + \int_{\mathbb{R}^+} z\tau^{z-1} \min \left( 1, |\Xi'_j| 2e^{-\frac{3}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} \right) d\tau, \end{aligned}$$

and thus, for any  $\tau_1 \geq 0$  and  $\tau_2 \geq 0$ , we get

$$\begin{aligned} \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{u_{j,\xi}} \right)^z \right] &\leq \tau_2^z + \int_{\tau \geq \tau_2} z\tau^{z-1} |\Xi'_j| 2e^{-\frac{1}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2} d\tau \\ &\quad + \tau_1^z + \int_{\tau \geq \tau_1} z\tau^{z-1} |\Xi'_j| 2e^{-\frac{3}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} d\tau. \end{aligned}$$

Take

$$\tau_1 = \frac{8 \log(c_{1,z} |\Xi'_j|)}{3n \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}}} \quad \text{and} \quad \tau_2 = \frac{2\sqrt{2} \sqrt{\log(c_{2,z} |\Xi'_j|)}}{\sqrt{n} \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}}}.$$

Hence, by construction, we have:

$$\begin{aligned} \forall \tau \geq \tau_1, \quad |\Xi'_j| 2e^{-\frac{3}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} &\leq \frac{2}{c_{1,z}} e^{-\frac{3}{8}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}} \right) \tau} \\ \forall \tau \geq \tau_2, \quad |\Xi'_j| 2e^{-\frac{1}{4}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2} &\leq \frac{2}{c_{2,z}} e^{-\frac{1}{8}n \left( \inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}} \right)^2 \tau^2}. \end{aligned}$$

This implies

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{u_{j,\xi}} \right)^z \right] \\ & \leq \left( \frac{2\sqrt{2}}{\sqrt{n}} \frac{\sqrt{\log(c_{2,z} |\Xi'_j|)}}{\inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}}} \right)^z + 2 \left( \frac{2\sqrt{2}}{\sqrt{n}} \frac{1}{\inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{\sigma_{j,\xi}}} \right)^z \\ & \quad + \left( \frac{8}{3n} \frac{\log(c_{1,z} |\Xi'_j|)}{\inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}}} \right)^z + 2 \left( \frac{8}{3n} \frac{1}{\inf_{\xi \in \Xi'_j} \frac{u_{j,\xi}}{M_{j,\xi}}} \right)^z, \end{aligned}$$

which allows to establish the claimed result.  $\square$

Lemma 16 allows to obtain the upper bounds (37) and (38) below.

For  $u_{j,\xi} = \sigma_{j,\xi}$ , we obtain

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{\sigma_{j,\xi}} \right)^z \right] \\ & \leq \left( \frac{2\sqrt{2}}{\sqrt{n}} \right)^z \left( 2 + (\log(c_{2,z} |\Xi'_j|))^{z/2} \right) + \left( \frac{8}{3n} \sup_{\xi \in \Xi'_j} \frac{M_{j,\xi}}{\sigma_{j,\xi}} \right)^z \left( 2 + (\log(c_{1,z} |\Xi'_j|))^{z/2} \right). \end{aligned}$$

For future use, note that we can also use the uniform bounds  $M_j$  (see (26)) and

$$\sigma_{j,\xi} \leq C_2 B(d, 2) B_X^{1/2} 2^{j\nu(d)} \stackrel{\text{def}}{=} \sigma_j \quad (36)$$

instead of  $M_{j,\xi}$  and  $\sigma_{j,\xi}$ , and obtain

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \\ & \leq \left( \frac{2\sqrt{2}}{\sqrt{n}} \sigma_j \right)^z \left( 2 + (\log(c_{2,z} |\Xi'_j|))^{z/2} \right) + \left( \frac{8}{3n} M_j \right)^z \left( 2 + (\log(c_{1,z} |\Xi'_j|))^{z/2} \right). \quad (37) \end{aligned}$$

Along the same lines, with  $u_{j,\xi} = c_\sigma \sqrt{\log(n)/n} \sigma_{j,\xi} + c_M \log(n)/(n-1) M_{j,\xi}$ , we obtain

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\xi \in \Xi'_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{c'_\sigma \sqrt{\log n} \frac{\sigma_{j,\xi}}{\sqrt{n}} + c'_M \log n \frac{M_{j,\xi}}{n-1}} \right)^z \right] \\ & \leq \left( \frac{2\sqrt{2}}{c'_\sigma \sqrt{\log n}} \right)^z \left( 2 + (\log(c_{2,z} |\Xi'_j|))^{z/2} \right) + \left( \frac{8}{3c'_M \log n} \right)^z \left( 2 + (\log(c_{1,z} |\Xi'_j|))^{z/2} \right), \quad (38) \end{aligned}$$

recall that when  $\Xi'_j = \Xi_j$ ,  $|\Xi'_j| \leq C_{\Xi} 2^{j(d-1)}$ .

### 7.6.5 Empirical Bernstein and the probabilities

We take

$$\begin{aligned}\Delta_{j,\xi,\gamma} &= \sqrt{2\gamma}t_n\widehat{\sigma}_{j,\xi} + \frac{14}{3}M_{j,\xi}\frac{\gamma\log n}{n-1}; \\ T_{j,\xi,\gamma} &= 2\Delta_{j,\xi,\gamma}, \quad T_{j,\xi,\gamma}^b = \Delta_{j,\xi,\gamma}, \quad T_{j,\xi,\gamma}^s = 3\Delta_{j,\xi,\gamma}; \\ \Delta_{j,\xi,\gamma}^+ &= \sqrt{2\gamma}t_n\sigma_{j,\xi} + \frac{26}{3}M_{j,\xi}\frac{\gamma\log n}{n-1} \quad \text{and} \quad \Delta_{j,\xi,\gamma}^- = \sqrt{2\gamma}t_n\sigma_{j,\xi} + \frac{2}{3}M_{j,\xi}\frac{\gamma\log n}{n-1}; \\ T_{j,\xi,\gamma}^{b,-} &= \Delta_{j,\xi,\gamma}^- \quad \text{and} \quad T_{j,\xi,\gamma}^{s,+} = 3\Delta_{j,\xi,\gamma}^+.\end{aligned}$$

**Lemma 17** *The following upper bounds hold:*

$$\begin{aligned}\mathbb{P}\left\{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b\right\} &\leq \frac{1}{n\gamma}; \\ \mathbb{P}\left\{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s\right\} &\leq \frac{1}{n\gamma}; \\ \mathbb{P}\left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\} &\leq \frac{3}{n\gamma}; \\ \mathbb{P}\left(\bigcup_{\xi \in \Xi_j} \left\{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s\right\}\right) &\leq \sum_{\xi \in \Xi_j} \mathbb{P}\left\{T_{j,\xi,\gamma}^{s,+} < T_{j,\xi,\gamma}^s\right\} \leq C_{\Xi}2^{j(d-1)}\frac{1}{n\gamma}; \\ \mathbb{P}\left(\bigcup_{\xi \in \Xi_j} \left\{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b\right\}\right) &\leq \sum_{\xi \in \Xi_j} \mathbb{P}\left\{T_{j,\xi,\gamma}^{b,-} > T_{j,\xi,\gamma}^b\right\} \leq C_{\Xi}2^{j(d-1)}\frac{1}{n\gamma}; \\ \mathbb{P}\left(\bigcup_{\xi \in \Xi_j} \left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\}\right) &\leq \sum_{\xi \in \Xi_j} \mathbb{P}\left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \Delta_{j,\xi,\gamma}\right\} \leq C_{\Xi}2^{j(d-1)}\frac{3}{n\gamma}.\end{aligned}$$

**Proof.** Using the results of [23], we get:

$$\begin{aligned}\mathbb{P}\left\{\sigma_{j,\xi} > \widehat{\sigma}_{j,\xi} + 2\sqrt{2u}\frac{M_{j,\xi}}{\sqrt{n-1}}\right\} &\leq e^{-u}; \\ \mathbb{P}\left\{\sigma_{j,\xi} < \widehat{\sigma}_{j,\xi} - 2\sqrt{2u}\frac{M_{j,\xi}}{\sqrt{n-1}}\right\} &\leq e^{-u}; \\ \mathbb{P}\left\{\left|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a\right| > \sqrt{2u}\frac{\widehat{\sigma}_{j,\xi}}{\sqrt{n}} + \frac{14}{3}M_{j,\xi}\frac{u}{n-1}\right\} &\leq 3e^{-u},\end{aligned}$$

which yields the first inequalities. The others follow from the union bound.  $\square$

### 7.6.6 The case $p = \infty$

Let us consider the various terms one by one.

*Error in the high dimensional space.*

$$\mathbb{E}[M_{j,z}] \leq \mathbb{E}[M_{j,z}^{S0}] + \mathbb{E}[M_{j,z}^{S1}] + \mathbb{E}[M_{j,z}^{B1}] + \mathbb{E}[M_{j,z}^{B2}],$$

with

$$\begin{aligned}
\mathbb{E} [M_{j,z}^{S0}] &= \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}}; \\
\mathbb{E} [M_{j,z}^{S1}] &\leq C_{\Xi} 2^{j(d-1)} \frac{4}{n^\gamma} \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z; \\
\mathbb{E} [M_{j,z}^{B1}] &\leq \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right]; \\
\mathbb{E} [M_{j,z}^{B2}] &\leq \left( C_{\Xi} 2^{j(d-1)} \frac{4}{n^\gamma} \right)^{1-1/\tau} \left( \left( \frac{2\sqrt{2}}{\sqrt{n}} \sigma_j \right)^z \left( 2^{1/\tau} + \left( \sqrt{\log(|\Xi_j| c_{1,z\tau})} \right)^z \right) \right. \\
&\quad \left. + \left( \frac{8}{3n} M_j \right)^z \left( 2^{1/\tau} + \left( \log(|\Xi_j| c_{1,z\tau}) \right)^z \right) \right),
\end{aligned}$$

where we have used  $(a+b)^{1/\tau} \leq a^{1/\tau} + b^{1/\tau}$  for  $\tau \geq 1$ .

This yields

$$\begin{aligned}
&\frac{\mathbb{E} \left[ \left\| \widehat{f}_{\beta}^{-a,\rho} - f_{\beta}^{-a,J} \right\|_{\infty}^z \right]}{(J+1)^{z-1} C_{\infty}^z} \\
&\leq \sum_{j=0}^J 2^{j(d-1)z/2} \left( \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right] \right) \\
&\quad + \frac{4}{n^\gamma} C_{\Xi} \sum_{j=0}^J 2^{j(d-1)(z/2+1)} \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \\
&\quad + \left( C_{\Xi} \frac{4}{n^\gamma} \right)^{1-1/\tau} \sum_{j=0}^J 2^{j(d-1)(z/2+1-1/\tau)} \\
&\quad \quad \times \left( \left( \frac{2\sqrt{2}}{\sqrt{n}} \sigma_j \right)^z \left( 2^{1/\tau} + \left( \sqrt{\log(|\Xi_j| c_{1,z\tau})} \right)^z \right) + \left( \frac{8}{3n} M_j \right)^z \left( 2^{1/\tau} + \left( \log(|\Xi_j| c_{1,z\tau}) \right)^z \right) \right) \\
&\stackrel{\text{def}}{=} O'_{\infty,z} + R'_{1,\infty,z} + R'_{2,\infty,z}.
\end{aligned}$$

The terms  $R'_{1,\infty,z}$  and  $R'_{2,\infty,z}$ . The term  $R'_{1,\infty,z}$  is the term which appears in Theorem 14 and thus we only need to bound  $R'_{2,\infty,z}$ . As in the case  $p < \infty$ , we can use the uniform bounds on  $\sigma_{j,\xi}$  and

$M_{j,\xi}$ , namely, (26) and (36), and  $|\Xi_j| \leq |\Xi_J|$  to obtain

$$\begin{aligned}
& R'_{2,\infty,z} \\
& \leq \left( \frac{4C_{\Xi}}{n^{\gamma}} \right)^{1-1/\tau} \sum_{j=0}^J 2^{j(d-1)(z/2+1-1/\tau)} \\
& \quad \times \left( \left( \frac{2\sqrt{2}}{\sqrt{n}} C_2 B(d, 2) 2^{j\nu(d)} B_X^{1/2} \right)^z \left( 2^{1/\tau} + (\log(c_{2,z\tau} |\Xi_J|))^{z/2} \right) \right. \\
& \quad \left. + \left( \frac{8}{3n} C_{\infty} B(d, \infty) 2^{j(\nu(d)+(d-1)/2)} B_X \right)^z \left( 2^{1/\tau} + (\log(c_{1,z\tau} |\Xi_J|))^z \right) \right) \\
& \leq \left( \frac{4C_{\Xi}}{n^{\gamma}} \right)^{1-1/\tau} \left[ \left( \frac{2\sqrt{2}}{\sqrt{n}} C_2 B(d, 2) B_X^{1/2} \right)^z \left( 2^{1/\tau} + (\log(|\Xi_J| c_{2,z\tau}))^{z/2} \right) \sum_{j=0}^J 2^{j(\nu(d)z+(d-1)(z/2+1-1/\tau))} \right. \\
& \quad \left. + \left( \frac{8}{3n} C_{\infty} B(d, \infty) B_X \right)^z \left( 2^{1/\tau} + (\log(|\Xi_J| c_{1,z\tau}))^z \right) \sum_{j=0}^J 2^{j(\nu(d)z+(d-1)(z+1-1/\tau))} \right] \\
& \leq \left( \frac{4C_{\Xi}}{n^{\gamma}} \right)^{1-1/\tau} \left[ \left( \frac{2\sqrt{2}}{\sqrt{n}} C_2 B(d, 2) B_X^{1/2} \right)^z \left( 2^{1/\tau} + (\log(c_{2,z\tau} |\Xi_J|))^{z/2} \right) \frac{2^{J(\nu(d)z+(d-1)(z/2+1-1/\tau))}}{1 - 2^{-(\nu(d)z+(d-1)(z/2+1-1/\tau))}} \right. \\
& \quad \left. + \left( \frac{8}{3n} C_{\infty} B(d, \infty) B_X \right)^z \left( 2^{1/\tau} + (\log(c_{1,z\tau} |\Xi_J|))^z \right) \frac{2^{J(\nu(d)z+(d-1)(z+1-1/\tau))}}{1 - 2^{-(\nu(d)z+(d-1)(z+1-1/\tau))}} \right].
\end{aligned}$$

The term  $O'_{\infty,z}$ . Denote by

$$O'_{z,j} = \sup_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right].$$

Because  $T_{j,\xi,\gamma}^{s,++} \geq T_{j,\xi,\gamma}^{s,+}$ , we get

$$\begin{aligned}
& \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right] \\
& = \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right] + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq |\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right] \\
& \leq \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right] \\
& \quad + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left( \frac{\left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|}{T_{j,\xi,\gamma}^{b,-}} \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq |\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right)^z \right] \sup_{\xi \in \Xi_j} \left\{ \left| \beta_{j,\xi}^a \right|^z \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq |\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right\},
\end{aligned}$$

thus

$$O'_{z,j} \leq \left( 1 + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left( \frac{|\widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a|}{T_{j,\xi,\gamma}^{b,-}} \right)^z \right] \right) \sup_{\xi \in \Xi_j} \left\{ |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,++}} \right\} \\ + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right].$$

Using now (38), with  $c'_\sigma = \sqrt{2\gamma}$  and  $c'_M = \frac{2}{3}\gamma$ , and  $|\Xi_j| \leq C_\Xi 2^{j(d-1)}$ , we get the upper bound in Theorem 14.

### 7.6.7 The case $p < \infty$

Let us consider the various terms one by one.

*Error in the high dimensional space.* We obtain

$$\mathbb{E} [S_{j,z}] = \mathbb{E} [S_{j,z}^{S0}] + \mathbb{E} [S_{j,z}^{S1}] + \mathbb{E} [S_{j,z}^{B1}] + \mathbb{E} [S_{j,z}^{B2}].$$

with

$$\mathbb{E} [S_{j,z}^{S0}] = \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}}; \\ \mathbb{E} [S_{j,z}^{S1}] \leq \frac{4}{n\gamma} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z; \\ \mathbb{E} [S_{j,z}^{B1}] \leq \sum_{\xi \in \Xi_j} \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right]; \\ \mathbb{E} [S_{j,z}^{B2}] \leq \frac{4^{1-1/\tau}}{n\gamma(1-1/\tau)} \sum_{\xi \in \Xi_j} 2^{1/\tau} \left( \left( 2c_{2,z\tau}^{1/(z\tau)} \frac{\sigma_{j,\xi}}{\sqrt{n}} \right)^z + \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} \frac{M_{j,\xi}}{n} \right)^z \right),$$

where we have used  $(a + b)^{1/\tau} \leq (a^{1/\tau} + b^{1/\tau})$ . This yields

$$\begin{aligned}
& \frac{\mathbb{E} \left[ \left\| \widehat{f}_\beta^{a,\rho} - f_\beta^{-a,J} \right\|_p^z \right]}{(J+1)^{z-1} C_p'^z C_\Xi^{z/(p \wedge z) - 1}} \\
& \leq \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p \vee z))} \mathbb{E} [S_{j,z}] \\
& \leq \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p \vee z))} \sum_{\xi \in \Xi_j} \left( |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right) \\
& \quad + \frac{4}{n^\gamma} \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p \vee z))} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \\
& \quad + \frac{2^{2-1/\tau}}{n^{\gamma(1-1/\tau)}} \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p \vee z))} \sum_{\xi \in \Xi_j} \left( \left( 2c_{2,z\tau}^{1/(z\tau)} \frac{\sigma_{j,\xi}}{\sqrt{n}} \right)^z + \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} \frac{M_{j,\xi}}{n} \right)^z \right) \\
& \stackrel{\text{def}}{=} O_{p,z} + R_{1,p,z} + R_{2,p,z}.
\end{aligned}$$

The terms  $R_{1,p,z}$  and  $R_{2,p,z}$ . The term  $R_{1,p,z}$  appears as is in Lemma 14. To bound the term  $R_{2,p,z}$ , we rely on (26). We obtain

$$\begin{aligned}
& \sum_{\xi \in \Xi_j} 2^{1/\tau} \left( \left( 2c_{2,z\tau}^{1/(z\tau)} \frac{\sigma_j}{\sqrt{n}} \right)^z + \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} \frac{M_j}{n} \right)^z \right) \\
& \leq \sum_{\xi \in \Xi_j} 2^{1/\tau} \left( 2c_{2,z\tau}^{1/(z\tau)} C_2 B(d, 2) 2^{j\nu(d)} B_X^{1/2} \frac{1}{\sqrt{n}} \right)^z \\
& \quad + \sum_{\xi \in \Xi_j} 2^{1/\tau} \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d, \infty) 2^{j(\nu(d)+(d-1)/2)} B_X \frac{1}{n} \right)^z \\
& \leq C_\Xi 2^{1/\tau} \left( 2c_{2,z\tau}^{1/(z\tau)} C_2 B(d, 2) \right)^z B_X^{z/2} \frac{1}{n^{z/2}} 2^{j((d-1)+z\nu(d))} \\
& \quad + C_\Xi 2^{1/\tau} \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_\infty B(d, \infty) \right)^z B_X^z \frac{1}{n^z} 2^{j((d-1)+z(\nu(d)+(d-1)/2))};
\end{aligned}$$

this yields

$$\begin{aligned}
& \sum_{\bar{j}=0}^J 2^{j(d-1)z(1/2-1/(p\nu z))} \sum_{\xi \in \Xi_j} 2^{1/\tau} \left( \left( 2c_{2,z\tau}^{1/(z\tau)} \frac{\sigma_j}{\sqrt{n}} \right)^z + \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} \frac{M_j}{n} \right)^z \right) \\
& \leq \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p\nu z))} C_{\Xi} 2^{1/\tau} \left( 2c_{2,z\tau}^{1/(z\tau)} C_2 B(d, 2) \right)^z B_X^{z/2} \frac{1}{n^{z/2}} 2^{j((d-1)+z\nu(d))} \\
& \quad + \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p\nu z))} C_{\Xi} 2^{1/\tau} \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_{\infty} B(d, \infty) \right)^z B_X^z \frac{1}{n^z} 2^{j((d-1)+z(\nu(d)+(d-1)/2))} \\
& \leq C_{\Xi} 2^{1/\tau} \left( 2c_{2,z\tau}^{1/(z\tau)} C_2 B(d, 2) \right)^z B_X^{z/2} \frac{1}{n^{z/2}} \sum_{j=0}^J 2^{jz(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\nu z)))} \\
& \quad + C_{\Xi} 2^{1/\tau} \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_{\infty} B(d, \infty) \right)^z B_X^z \frac{1}{n^z} \sum_{j=0}^J 2^{jz(\nu(d)+(d-1)/z+(d-1)(1-1/(p\nu z)))} \\
& \leq \frac{C_{\Xi} 2^{1/\tau} \left( 2c_{2,z\tau}^{1/(z\tau)} C_2 B(d, 2) \right)^z}{1 - 2^{-z(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\nu z)))}} B_X^{z/2} \frac{1}{n^{z/2}} 2^{Jz(\nu(d)+(d-1)/z+(d-1)(1/2-1/(p\nu z)))} \\
& \quad + \frac{C_{\Xi} 2^{1/\tau} \left( \frac{4}{3} c_{1,z\tau}^{1/(z\tau)} C_{\infty} B(d, \infty) \right)^z}{1 - 2^{-z(\nu(d)+(d-1)/z+(d-1)(1-1/(p\nu z)))}} B_X^z \frac{1}{n^z} 2^{Jz(\nu(d)+(d-1)/z+(d-1)(1-1/(p\nu z)))}.
\end{aligned}$$

The term  $O_{p,z}$ . Denote by

$$O_{z,j,\xi} = |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,+}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}}.$$

Because  $T_{j,\xi,\gamma}^{s,++} \geq T_{j,\xi,\gamma}^{s,+}$ , we get

$$\begin{aligned}
& \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right] \\
& = \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right] + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq |\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}} \right] \\
& \leq \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right] + \frac{\mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right]}{\left( T_{j,\xi,\gamma}^{b,-} \right)^z} |\beta_{j,\xi}^a|^z \mathbf{1}_{T_{j,\xi,\gamma}^{s,++} \geq |\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{b,-}}, \\
O_{z,j,\xi} & \leq \left( 1 + \frac{\mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right]}{\left( T_{j,\xi,\gamma}^{b,-} \right)^z} \right) |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}}.
\end{aligned}$$

Now using the results of Section 7.6.4, with  $T_{j,\xi,\gamma}^{b,-} = \sqrt{2\gamma}t_n\sigma_{j,\xi} + \frac{2}{3}\gamma\frac{\log n}{n-1}M_{j,\xi}$ , we obtain

$$\begin{aligned} \sup_{j,\xi} \frac{\mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right]}{\left( T_{j,\xi,\gamma}^{b,-} \right)^z} &\leq 2 \left( c_{2,z} \left( 2 \frac{1}{\sqrt{2\gamma}\sqrt{\log n}} \right)^z + c_{1,z} \left( \frac{4}{3} \frac{1}{(2/3)\gamma \log n} \right)^z \right) \\ &\leq 2 \left( \left( \frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma \log n}} \right)^z + \left( \frac{2c_{1,z}^{1/z}}{\gamma \log n} \right)^z \right). \end{aligned}$$

This yields

$$\begin{aligned} O_{p,z} &\leq \left( 1 + 2 \left( \left( \frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma \log n}} \right)^z + \left( \frac{2c_{1,z}^{1/z}}{\gamma \log n} \right)^z \right) \right) \sum_{j=0}^J 2^{j(d-1)z(1/2-1/(p\vee z))} \\ &\quad \sum_{\xi \in \Xi_j} \left( |\beta_{j,\xi}^a|^z \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\xi,\gamma}^{s,++}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^z \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\xi,\gamma}^{s,++}} \right). \end{aligned}$$

## 7.7 Proof of Theorem 9

This proof requires an upper bound on: the approximation error,  $R_{1,p,z}$ ,  $R_{1,p,z}$ , and  $O_{p,z}$ . We use that because  $f_\beta \in B_{r,q}^s(M)$ , we have, by Lemma 13,  $f_\beta^- \in B_{r,q}^s(c_{\text{eq}}M)$ .

### 7.7.1 The case $1 \leq p < \infty$

Let us consider the terms one by one.

*The approximation error.* Start with

$$\left\| f_\beta^{-a,J} - f_\beta^- \right\|_p = \left\| \sum_{j>J} \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_p.$$

From Lemma 6 (i) and the definition of the Besov spaces as a sequence space, with  $1/q + 1/\bar{q} = 1$ , we obtain

$$\left\| \sum_{j>J} \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_p \leq C'_p \sum_{j>J} 2^{-js} 2^{j(s+(d-1)(1/2-1/p))} \left\| (\beta_{j,\xi}^a)_{\xi \in \Xi_j} \right\|_{\ell^p},$$

which yields

$$\begin{aligned} \left\| \sum_{j>J} \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_p &\leq C'_p 2^{-Js} (2^{s\bar{q}} - 1)^{-1/\bar{q}} \left\| f_\beta^- \right\|_{B_{p,q}^s} \\ &\leq \begin{cases} C'_p c_{\text{eq}} M C_{\Xi}^{1/p-1/r} (2^{s\bar{q}} - 1)^{-1/\bar{q}} \left\| f_\beta^- \right\|_{B_{p,q}^s} 2^{-Js} & \text{if } r \geq p \\ C'_p c_{\text{eq}} M (2^{s\bar{q}} - 1)^{-1/\bar{q}} \left\| f_\beta^- \right\|_{B_{p,q}^s} 2^{-J(s-(d-1)(1/r-1/p))} & \text{if } r \leq p. \end{cases} \end{aligned}$$

It is enough to consider the worst case where  $r \leq p$  and to check that  $\frac{s-(d-1)(1/r-1/p)}{\nu(d)+(d-1)/2} \geq \mu$  in the two zones.

In the dense zone, we have

$$s + \nu(d) + \frac{d-1}{2} \geq \left( \nu(d) + \frac{d-1}{2} \right) \frac{p}{r},$$

which yields

$$\frac{s}{s + \nu(d) + \frac{d-1}{2}} \leq \frac{s}{\left( \nu(d) + \frac{d-1}{2} \right) \frac{p}{r}}.$$

Because  $s > (d-1)/r$  and  $p \geq r$ , we have

$$s - \frac{d-1}{r} + \frac{d-1}{p} - \frac{sr}{p} = (d-1) \left( \frac{sr}{d-1} - 1 \right) \left( \frac{1}{r} - \frac{1}{p} \right) \geq 0,$$

which yields  $s - (d-1)(1/r - 1/p) \geq \frac{sr}{p}$  and gives the result.

In the sparse zone, because  $s > (d-1)/r$ , we have

$$\frac{s - (d-1)(1/r - 1/p)}{\nu(d) + (d-1)/2} \geq \frac{s - (d-1)(1/r - 1/p)}{s + \nu(d) - (d-1)(1/r - 1/2)}.$$

The terms  $R_{1,p,p}$  and  $R_{2,p,p}$ . Using Lemma 7 (iii) we obtain

$$R_{1,p,p} \leq \frac{4}{n^\gamma} (c_{\text{eq}} M)^p C_{\Xi}^{1-(p \wedge r)/r} \sum_{j=0}^J 2^{-jp(s+(d-1)(1/p-1/(p \wedge r)))},$$

where the exponent is nonpositive because  $s > (d-1)/r$ , thus

$$R_{1,p,p} \leq \frac{4(c_{\text{eq}} M)^p C_{\Xi}^{1-(p \wedge r)/r}}{n^\gamma (1 - 2^{-p(s+(d-1)(1/p-1/(p \wedge r)))})}.$$

With  $\gamma > p/2$ ,  $R_{1,p,p}$  is of lower order than  $t_n^p$ .

We also have

$$R_{2,p,p} \leq \frac{2^{2-1/\tau}}{n^{\gamma(1-1/\tau)}} C_{\Xi} b_{n,p,p,J,\tau}.$$

With the aforementioned choice of  $J$ ,

$$\begin{aligned} \frac{1}{\sqrt{n}} 2^{J(\nu(d)+(d-1)/2)} B_X^{1/2} &\lesssim 1; \\ \frac{2^{J(d-1)}}{n} B_X &\lesssim 1. \end{aligned}$$

Together, these yield that  $b_{n,p,p,J,\tau}$  is of the order of a constant.

This term is also of lower order than  $t_n^p$  for  $\tau$  large enough such that  $\gamma(1-1/\tau) > p/2$ .

The term  $O_{p,p}$ . First note that  $a_{n,p,p,J} = 1 + o(1)$ .  
We take  $T_{j,\xi,\gamma}^{s,++}$  uniform in  $\xi$ :

$$\begin{aligned} T_{j,\xi,\gamma}^{s,++} &= 3\sqrt{2\gamma}t_n C_2 B(d, 2) 2^{j\nu(d)} B_X^{1/2} \\ &\quad + 52C_\infty B(d, \infty) 2^{j(\nu(d)+(d-1)/2)} B_X \frac{\gamma \log n}{n-1} \\ &\leq 2^{j\nu(d)} \sqrt{\gamma} t_n B_X^{1/2} \left( 3\sqrt{2}C_2 B(d, 2) + 52C_\infty B(d, \infty) \frac{n\sqrt{\gamma}}{n-1} \right), \end{aligned}$$

where the last display uses the upper bound on  $J$ , this yields, for  $n \geq 2$ ,

$$T_{j,\xi,\gamma}^{s,++} \leq 2^{j\nu(d)} \sqrt{\gamma} t_n B_X^{1/2} \left( 3\sqrt{2}C_2 B(d, 2) + 104C_\infty B(d, \infty) \right) \stackrel{\text{def}}{=} T_{j,\gamma}^{s,++}.$$

As a consequence of Lemma 15, we get

$$\begin{aligned} \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^p \right] &\leq 2 \left( \left( 2c_{2,p}^{1/p} \frac{\sigma_j}{\sqrt{n}} \right)^p + \left( \frac{4}{3} c_{1,p}^{1/p} \frac{M_j}{n} \right)^p \right) \\ &\leq 2 \left( 2c_{2,p}^{1/p} C_2 B(d, 2) 2^{j\nu(d)} B_X^{1/2} \frac{1}{\sqrt{n}} \right)^p \\ &\quad + 2 \left( \frac{8}{3} c_{1,p}^{1/p} C_\infty B(d, \infty) 2^{j(\nu(d)+(d-1)/2)} B_X \frac{1}{n} \right)^p \\ &\leq 2^{j\nu(d)} \frac{1}{n^{p/2}} B_X^{p/2} 2^{p+1} \left( c_{2,p}^{1/p} C_2 B(d, 2) + \frac{4}{3} c_{1,p}^{1/p} C_\infty B(d, \infty) \right)^p \\ &\leq \frac{(T_{j,\gamma}^{s,++})^p}{(\gamma \log n)^{p/2}} 2 \left( \frac{c_{2,p}^{1/p} C_2 B(d, 2) + \frac{4}{3} c_{1,p}^{1/p} C_\infty B(d, \infty)}{3\sqrt{2}C_2 B(d, 2) + 104C_\infty B(d, \infty) (\sqrt{\gamma})} \right)^p \\ &\leq \frac{(T_{j,\gamma}^{s,++})^p}{(\gamma \log n)^{p/2}} 2 \left( \frac{\sqrt{2}}{3} c_{2,p}^{1/p} + \frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}} \right)^p. \end{aligned}$$

$$\text{Let } C_\gamma = 3\sqrt{2}C_2 B(d, 2) + 104C_\infty B(d, \infty)\sqrt{\gamma} \quad \text{and} \quad C_{\sigma,p} = 2^{1/p} \left( \frac{\sqrt{2}}{3} c_{2,p}^{1/p} + \frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}} \right).$$

For any  $0 < z < p$ , we have

$$\begin{aligned} &\sum_{\xi \in \Xi_j} \left( |\beta_{j,\xi}^a|^p \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\gamma}^{s,++}} + \mathbb{E} \left[ \left| \widehat{\beta}_{j,\xi}^a - \beta_{j,\xi}^a \right|^p \right] \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\gamma}^{s,++}} \right) \\ &\leq \sum_{\xi \in \Xi_j} \left( |\beta_{j,\xi}^a|^p \mathbf{1}_{|\beta_{j,\xi}^a| \leq T_{j,\gamma}^{s,++}} + \frac{(T_{j,\gamma}^{s,++})^p}{(\gamma \log n)^{p/2}} C_{\sigma,p}^p \mathbf{1}_{|\beta_{j,\xi}^a| > T_{j,\gamma}^{s,++}} \right) \\ &\leq \left( 1 + \frac{C_{\sigma,p}^p}{(\gamma \log n)^{p/2}} \right) (T_{j,\gamma}^{s,++})^{p-z} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z \\ &\leq \left( 1 + \frac{C_{\sigma,p}^p}{(\gamma \log n)^{p/2}} \right) \left( \sqrt{\gamma} t_n B_X^{1/2} C_\gamma \right)^{p-z} 2^{j\nu(d)(p-z)} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^z. \end{aligned}$$

We need to sum over  $j$  and take two different values for  $z$ , one that we denote  $z_1$  for  $j \leq j_0$  and one that we denote  $z_2$  for  $j_0 < j \leq J$ . The values of  $z_1, z_2, j_0$  will be specified later, depending on the value of the parameters  $r, q, s$  and  $p$  such that we are in the dense or sparse zone. Up to a multiplying constant, we thus need to control

$$\begin{aligned} A + B &= \left(B_X^{1/2} t_n\right)^{p-z_1} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-z_1)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^{z_1} \\ &\quad + \left(B_X^{1/2} t_n\right)^{p-z_2} \sum_{j=j_0+1}^J 2^{j[\nu(d)(p-z_2)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^{z_2}, \end{aligned}$$

where we choose adequately  $z_1, z_2$  and  $j_0$  in the two zones. Because of Lemma 7 (i), we only consider  $p \geq r$ .

Let us first consider the dense zone. We define

$$\tilde{r} = \frac{p(\nu(d) + (d-1)/2)}{s + \nu(d) + (d-1)/2}.$$

In the dense zone,  $\tilde{r} \leq r, p > \tilde{r}$  and

$$s = \left(\nu(d) + \frac{d-1}{2}\right) \left(\frac{p}{\tilde{r}} - 1\right). \quad (39)$$

With  $z_2 = r$ , we get

$$B \leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=j_0+1}^J 2^{j[\nu(d)(p-r)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^r.$$

Lemma 7 (iii) gives that

$$\sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^r \leq D_j^r 2^{-jr(s+(d-1)(1/2-1/r))},$$

where  $\forall j \in \mathbb{N}, D_j \geq 0, (D_j)_{j \in \mathbb{N}} \in \ell_q$ . Note that

$$s + (d-1) \left(\frac{1}{2} - \frac{1}{r}\right) = \frac{(d-1)p}{2\tilde{r}} - \frac{d-1}{r} + \nu(d) \left(\frac{p}{\tilde{r}} - 1\right), \quad (40)$$

thus

$$\begin{aligned} B &\leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=j_0+1}^J 2^{jp(1-\frac{p}{\tilde{r}})(\nu(d)+\frac{d-1}{2})} D_j^r \\ &\lesssim (c_{\text{eq}} M)^r \left(B_X^{1/2} t_n\right)^{p-r} 2^{j_0 p(1-\frac{p}{\tilde{r}})(\nu(d)+\frac{d-1}{2})}, \end{aligned}$$

for  $q \geq 1$  if  $r > \tilde{r}$  and for  $q \leq r$  if  $r = \tilde{r}$  (i.e.,  $s = p(\nu(d) + \frac{d-1}{2}) \left(\frac{1}{r} - \frac{1}{p}\right)$ ).

Taking  $2^{j_0 \frac{p}{\tilde{r}}(\nu(d)+\frac{d-1}{2})} \simeq \left(B_X^{1/2} t_n\right)^{-1}$ , we get

$$B \lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\tilde{r}},$$

which is the rate that we expect in that zone.

As for  $A$ , we take  $z_1 = \bar{r} < \tilde{r} \leq r$ , this yields, using Lemma 7 (iii),

$$\begin{aligned}
A &\leq \left(B_X^{1/2} t_n\right)^{p-\bar{r}} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-\bar{r})+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^{\bar{r}} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\bar{r}} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-\bar{r})+(d-1)(p/2-1)-\bar{r}(s+(d-1)(1/2-1/\bar{r}))]} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\bar{r}} \sum_{j=0}^{j_0} 2^{jp(\nu(d)+(d-1)/2)(1-\bar{r}/\tilde{r})} \quad (\text{using (39)}) \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\bar{r}} 2^{j_0 p(\nu(d)+(d-1)/2)(1-\bar{r}/\tilde{r})} \\
&\lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\tilde{r}} \quad (\text{from the definition of } j_0).
\end{aligned}$$

Let us now consider the sparse zone. We define by

$$\tilde{r} = p \frac{\nu(d) + (d-1)(1/2 - 1/p)}{s + \nu(d) - (d-1)(1/r - 1/2)},$$

in a such a way that

$$\begin{aligned}
p - \tilde{r} &= p \frac{s - (d-1)(1/r - 1/p)}{s + \nu(d) - (d-1)(1/r - 1/2)}; \\
\tilde{r} - r &= \frac{(p-r)((d-1)/2 + \nu(d)) - rs}{s + \nu(d) - (d-1)(1/r - 1/2)} > 0; \\
s + (d-1) \left(\frac{1}{2} - \frac{1}{r}\right) &= \frac{(d-1)p}{2\tilde{r}} - \frac{d-1}{\tilde{r}} + \nu(d) \left(\frac{p}{\tilde{r}} - 1\right). \tag{41}
\end{aligned}$$

For the term  $A$ , we take  $z_1 = r$  and obtain

$$\begin{aligned}
A &\leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=0}^{j_0} 2^{j[\nu(d)(p-r)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^r \\
&\leq \left(B_X^{1/2} t_n\right)^{p-r} \sum_{j=0}^{j_0} 2^{j[\nu(d)+(d-1)(1/2-1/p)\frac{p}{\tilde{r}}(\tilde{r}-r)]} D_j^r \quad (\text{using (41)}) \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-r} 2^{j_0[\nu(d)+(d-1)(1/2-1/p)\frac{p}{\tilde{r}}(\tilde{r}-r)]} M^r,
\end{aligned}$$

the last inequality holds because  $\nu(d) + (d-1)/2 - (d-1)/p > 0$ , indeed, because we are in the sparse zone  $\nu(d) + (d-1)/2 \geq s/(p/r - 1) = sr/(p-r) \geq 2/(p-r) \geq (d-1)/p$ . Taking  $2^{j_0[\nu(d)+(d-1)(1/2-1/p)\frac{p}{\tilde{r}}(\tilde{r}-r)]} \simeq \left(B_X^{1/2} t_n\right)^{-1}$ , yields

$$A \lesssim M^r \left(B_X^{1/2} t_n\right)^{p-\tilde{r}}.$$

For the term  $B$ , we take  $z_2 = \bar{r} > \tilde{r} > r$  and obtain

$$\begin{aligned}
B &\leq \left(B_X^{1/2} t_n\right)^{p-\bar{r}} \sum_{j=j_0+1}^J 2^{j[\nu(d)(p-\bar{r})+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} |\beta_{j,\xi}^a|^{\bar{r}} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\bar{r}} \sum_{j=j_0+1}^J 2^{j(\nu(d)+(d-1)(1/2-1/p))p(r-\bar{r})/\bar{r}} D_j^{\bar{r}} \quad (\text{using (41)}) \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\bar{r}} 2^{j_0(\nu(d)+(d-1)(1/2-1/p))p(r-\bar{r})/\bar{r}} M^{\bar{r}} \\
&\lesssim \left(B_X^{1/2} t_n\right)^{p-\tilde{r}} M^{\bar{r}}.
\end{aligned}$$

### 7.7.2 The case $p = \infty$

Consider  $r = \infty$ . The general case follows by Lemma 7 (ii).

*The approximation error.* Because  $f_\beta \in B_{\infty,q}^s(M)$ , we have by Lemma 6 (i)

$$\begin{aligned}
\left\| \sum_{j>J} \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_\infty &\leq \sum_{j>J} \left\| \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_\infty \\
&\leq C'_\infty c_{\text{eq}} M \sum_{j>J} 2^{j(d-1)/2} 2^{-j(s+(d-1)/2)} D_j \quad (\text{where } \|(D_j)_{j \in \mathbb{N}}\|_q \leq c_{\text{eq}} M) \\
&\leq C'_\infty c_{\text{eq}} M 2^{-Js} (2^{s\bar{q}} - 1)^{-1/\bar{q}}.
\end{aligned}$$

From the choice of  $J$ , we get

$$\left\| \sum_{j>J} \sum_{\xi \in \Xi_j} \beta_{j,\xi}^a \psi_{j,\xi} \right\|_\infty \lesssim C'_\infty c_{\text{eq}} M (2^{s\bar{q}} - 1)^{-1/\bar{q}} \left(t_n B_X^{1/2}\right)^{s/(\nu(d)+(d-1)/2)}.$$

This term is negligible because  $s/(\nu(d) + (d-1)/2) \geq s/(s\nu(d) + (d-1)/2)$ .

*The terms  $R'_{1,\infty,z}$  and  $R'_{2,\infty,z}$ .* Using the definition of the Besov norm, we obtain

$$\begin{aligned}
R'_{1,\infty,z} &\leq \frac{4}{n^\gamma} (c_{\text{eq}} M)^z C_\Xi \sum_{j=0}^J 2^{-jzs} 2^{j(d-1)} \\
&\lesssim \frac{4}{n^\gamma} 2^{J(d-1)} M^z.
\end{aligned}$$

With  $\gamma > z/2 + 1$ , which holds if  $2(\gamma-1)(1-1/\tau) > z$ ,  $R_{1,\infty,z}$  is of lower order than  $t_n^z$ .

Due to the choice of  $J$ , the term in bracket in the expression of  $R'_{2,\infty,z}$  in Theorem 14 is less than 1. The second term in the expression of  $b_{n,\infty,z,J,\tau}$  is of smaller order than the first term. The order of  $b_{n,\infty,z,J,\tau}$  is finally  $(\log n)^{z/2}$ . Thus, we have

$$R'_{2,\infty,z} \lesssim \left(n^{-\gamma} 2^{J(d-1)}\right)^{1-1/\tau} (\log n)^{z/2}.$$

This term is also of lower order than  $t_n^z$  when  $\tau$  is such that  $2(\gamma - 1)(1 - 1/\tau) > z$ .

The term  $O'_{\infty, z}$ . Note that here  $a_{n, \infty, z, J}$  is of the order of a constant. We now proceed like for the term  $O_{p, p}$ . Using (37), we obtain for arbitrary  $\bar{z} \in [0, z]$

$$\begin{aligned} & \sup_{\xi \in \Xi_j} |\beta_{j, \xi}^a|^z \mathbf{1}_{|\beta_{j, \xi}^a| \leq T_{j, \xi, \gamma}^{s, ++}} + \mathbb{E} \left[ \sup_{\xi \in \Xi_j} \left| \widehat{\beta}_{j, \xi}^a - \beta_{j, \xi}^a \right|^z \mathbf{1}_{|\beta_{j, \xi}^a| > T_{j, \xi, \gamma}^{s, ++}} \right] \\ & \lesssim \left( \sqrt{\gamma} t_n B_X^{1/2} \right)^{z - \bar{z}} 2^{j\nu(d)(z - \bar{z})} \sup_{\xi \in \Xi_j} |\beta_{j, \xi}^a|^{\bar{z}}. \end{aligned}$$

We use an upper bound on  $A + B$ , where:

$$\begin{aligned} A &= \left( B_X^{1/2} t_n \right)^{z - z_1} \sum_{j=0}^{j_0} 2^{j[\nu(d)(z - z_1) + (d-1)z/2]} \sup_{\xi \in \Xi_j} |\beta_{j, \xi}^a|^{z_1}; \\ B &= \left( B_X^{1/2} t_n \right)^{z - z_2} \sum_{j=j_0+1}^J 2^{j[\nu(d)(z - z_2) + (d-1)z/2]} \sup_{\xi \in \Xi_j} |\beta_{j, \xi}^a|^{z_2}, \end{aligned}$$

for well-chosen  $0 \leq j_0 \leq J$ ,  $z_1$  and  $z_2$ . Because  $f \in B_{\infty, q}^s(M)$ , we have

$$\forall \bar{z} \geq 1, \sup_{\xi \in \Xi_j} |\beta_{j, \xi}^a|^{\bar{z}} \leq (c_{\text{eq}} M)^{\bar{z}} 2^{-j(s + (d-1)/2)\bar{z}}.$$

The result follows taking  $z_1 = 0$ ,  $j_0$  such that  $2^{j_0} \simeq t_n^{-1/(s + \nu(d) + (d-1)/2)}$ , and  $z_2 = z$ .

## References

- [1] Atkinson, K. and Han, W. (2012). *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction*. Springer.
- [2] Baldi, P., Kerkycharian, G., Marinucci, D., and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395.
- [3] Beran, R., Feuerverger, A., and Hall, P. (1996). Estimating coefficient distributions in random coefficient regression. *Ann. Statist.* **24** 2569–2592.
- [4] Beran, R. and Millar, P. W. (1994). Minimum distance estimation in random coefficient regression models. *Ann. Statist.* **22** 1976–1992.
- [5] Bertin, K., Le Pennec, E., and Rivoirard, V. (2011). Adaptive Dantzig density estimation. *Ann. Inst. H. Poincaré Probab. Statist.* **47** 43–74.
- [6] Durastanti, C., Geller, D., and Marinucci, D. (2012). Adaptive nonparametric regression on spin fiber bundles. *J. Multivariate Anal.* **104** 16–38.
- [7] Funk, P. (1916). Über eine geometrische anwendung der abelschen integralgleichung. *Math. Ann.* **77** 129–135.

- [8] Gaillac, C., and Gautier, E. (2017). Nonparametric identification in some random coefficients models when regressors have limited variation. Working paper.
- [9] Gautier, E., and Hoderlein, S. (2011, 2015). A triangular treatment effect model with random coefficients in the selection equation. Preprint arXiv:1109.0362v4.
- [10] Gautier, E., and Kitamura, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica* **81** 581–607.
- [11] Groemer, H. (1996). *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge University Press, Cambridge, *Encyclopedia of Mathematics and its Applications*.
- [12] Healy, D. M., and Kim, P. T. (1996). An empirical bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.* **24** 232–254.
- [13] Ichimura, H., and Thompson, T. S. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *J. Econometrics* **86** 269–295.
- [14] Kamzolov, A. I. (1983). The best approximation of the class of functions  $\mathbf{W}_p^\alpha(\mathbb{S}^n)$  by polynomials in spherical harmonics. *Mat. Zametki* **32** 285–293.
- [15] Kerkyacharian, G., Nickl, R., Picard, D. (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Relat. Fields* **153** 363–404.
- [16] Kim, P. and Koo, J. (2000). Directional mixture models and optimal estimation of the mixing density. *Canad. J. Statist.* **28** 383–398.
- [17] Kerkyacharian, G., Kyriazis, G., Le Pennec, E., Petrushev, P., and Picard, D. (2010). Inversion of noisy radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.* **28** 24–45.
- [18] Kerkyacharian, G., Petrushev, P., Picard, D., and Willer, T. (2007). Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.* **1** 30–76.
- [19] Kerkyacharian, G., Phan Ngoc, T. M., and Picard, D. (2009). Localized deconvolution on the sphere. *Ann. Statist.* **39** 1042–1068.
- [20] Korostelev, V. and Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction*. Springer.
- [21] Mc Fadden, D. (2001). Economic choices - Nobel Lecture, December 2000. *American Economic Review* **91** 351–378.
- [22] Mardia, K. V. and Jupp, E. (2000). *Directional Statistics*. Wiley.
- [23] Maurer, A. and Pontil, M. (2009). Empirical Bernstein bounds and sample variance penalization. *COLT 2009*.
- [24] Monnier, J.-B. (2011). Nonparametric regression on the hyper-sphere with uniform design. *TEST* **20**, 412–446
- [25] Narcowich, F., Petrushev, P., and Ward, J. (2006). Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564.

- [26] Rubin, B. (1999). Inversion and characterization of the hemispherical transform. *J. Anal. Math.* **77** 105–128.
- [27] Tsybakov, A. (2009). *Introduction to Nonparametric Estimation*. Springer.