Centre de Recherche en Économie et Statistique

CREST

Série des Documents de Travail

n° 2011-20

Adaptive Estimation in the Nonparametric Random Coefficients Binary Choice Model by Needlet Thresholding

E. GAUTIER[1]

E. LE PENNEC[2]

1 CREST (ENSAE), 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France. eric.gautier@ensae.fr

2 SELECT / INRIA Saclay, Bâtiment 440, Faculté des Sciences d'Orsay, Université Paris-Sud, 91405 Orsay Cedex, France. erwan.le_pennec@inria.fr

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

E. Gautier[1] and E. Le Pennec[2]

June 16, 2011

[1] CREST (ENSAE), 3 avenue Pierre Larousse, 92 245 Malakoff Cedex, France; eric.gautier@ensae-paristech.fr.

[2] SELECT / INRIA Saclay, Bâtiment 440, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex, France; erwan.le_pennec@inria.fr.

Abstract

In this article we consider the estimation of the joint distribution of the random coefficients and error term in the nonparametric random coefficients binary choice model. In this model from economics, each agent has to choose between two mutually exclusive alternatives based on the observation of attributes of the two alternatives and of the agents, the random coefficients account for unobserved heterogeneity of preferences. Because of the scale invariance of the model, we want to estimate the density of a random vector of Euclidean norm 1. If the regressors and coefficients are independent, the choice probability conditional on a vector of $d-1$ regressors is an integral of the joint density on half a hyper-sphere determined by the regressors. Estimation of the joint density is an ill-posed inverse problem where the operator that has to be inverted in the so-called hemispherical transform. We derive lower bounds on the minimax risk under L^p losses and smoothness expressed in terms of Besov spaces on the sphere $\mathbb{S}^{d-1}$. We then consider a needlet thresholded estimator with data-driven thresholds and obtain adaptivity for L^p losses and Besov ellipsoids under assumptions on the random design.

Key Words: Discrete choice models, random coefficients, inverse problems, minimax rate optimality, adaptation, needlets, data-driven thresholding.

AMS 2010 Subject Classifications. Primary 91G70; secondary 42C15, 62C20, 62G05, 62G08, 62G20.

1 Introduction

Discrete choice models are important models in economics for the choice of agents between a number of exhaustive and mutually exclusive alternatives. They have applications in many areas ranging from empirical industrial organizations, labor economics, health economics, planning of public transportation, evaluation of public policies, etc. For a review, the interested reader can refer to the Nobel lectures of D. Mc Fadden [24]. We consider here a binary choice model where individuals only have two options. In a random utility framework, an agent chooses the alternative that yields the higher utility. Assume that the utility for each alternative is linear in regressors which are observed by the statistician. The regressors are typically attributes of the alternative faced by the individuals, *e.g.* the cost or time to commute from home to one's office for each of the two transport alternatives. Because this linear structure is an ideal situation and because the statistician is missing some factors, the utilities are written as the linear combination of the regressors plus some random error term. When the utility difference is positive the agent chooses the first alternative, otherwise he chooses the second. The Logit, Probit or Mixed-Logit models are particular models of this type. We consider the case where the coefficients of the regressors are random. This accounts for heterogeneity or taste variation: each individual is allowed to have his own set of coefficients (the preferences or tastes). Like in [8, 13], we consider a nonparametric treatment of the joint distribution of the error term and vector of random coefficients.

Nonparametric treatment of unobserved heterogeneity is very important in economics, references include [2, 4, 7, 8, 11, 12, 13]. It allows to be extremely flexible about the joint distribution of the preferences (as well as the error term). [7] considers treatment effects models with random coefficients in the case where the allocation to treatment corresponds to a decision mechanism formulated in the form of model (1) below. Random coefficients models can be viewed as mixture models. They also have a Bayesian interpretation, see for example [10] for a model similar to (1) on the sphere. Nonparametric estimation of the density of the vector of random coefficients corresponds to nonparametric estimation of a prior in the empirical Bayes setting.

In the nonparametric random coefficients binary choice model we assume that we have n i.i.d. observations (x_i, y_i) of (X, Y) where X is a random vector of Euclidean norm 1 in $\mathbb{R}^d$ and Y is a discrete random variable and Y and X are related through a non observed random vector β of norm 1 by

$$Y = 2\mathbf{1}_{\langle X,\beta\rangle>0} - 1 = \begin{cases} 1 & \text{if } X \text{ and } \beta \text{ are in the same hemisphere} \\ -1 & \text{otherwise.} \end{cases} \tag{1}$$

In (1), $\langle \cdot, \star \rangle$ is the scalar product in $\mathbb{R}^d$. We make the assumption that X and β are independent. This assumption corresponds to the exogeneity of the regressors. It could be relaxed using instrumental variables (see [8]). -1 and 1 are labels for the two choices. They correspond to the sign of $\langle X, \beta\rangle$. X and β are assumed to be of norm 1 because only the sign of $\langle X, \beta\rangle$ matters in the choice mechanism. The regressors in the latent variable model are thus assumed to be properly rescaled. Model (1) allows for arbitrary dependence between the random unobservables. In this model, X corresponds to a vector of regressors where, in an original scale, the first component is 1 and the remaining components are the regressors in the binary choice model. The 1 stands because in applications we always include a constant in the latent variable model for the binary choice model. The first element of β in this formulation absorbs the usual error term as well as the constant in standard binary choice models with non-random coefficients. We assume that X and β have densities f_X and f_β with respect to the spherical measure σ on the unit sphere $\mathbb{S}^{d-1}$ of the Euclidean space $\mathbb{R}^d$. Because in the original scale the first component of X is 1, the support of X is included in $H^+ = \{x \in \mathbb{S}^{d-1} : < x, (1, 0, \ldots, 0) >\geq 0\}$. We assume, for simplicity, through out this paper, that the support of X satisfies supp $f_X = H^+$. In [8], the case of regressors with limited support, including dummy variables is also studied but identification requires that these variables, as well as one continuously distributed regressor, are not multiplied by random coefficients.

The estimation of the density of the random coefficient can be viewed as a linear ill-posed inverse problem. We can write for $x \in H^+$,

$$\mathbb{E}[Y|X = x] = \int_{b\in\mathbb{S}^{d-1}} \operatorname{sign}\left(\langle x, b\rangle\right) f_\beta(b) d\sigma(b) \tag{2}$$

where sign denotes the sign. As recalled in [27], if φ is homogeneous of degree $-d$, *i.e.* there exists a function f on $\mathbb{S}^{d-1}$ such that $\varphi(x) = |x|^{-d} f(x/|x|)$, where $|\cdot|$ is the euclidean norm, then

$$\frac{2}{\pi}\text{Im}\int_{\mathbb{R}^d}\varphi(x)e^{i\langle x,y\rangle}dy = \int_{b\in\mathbb{S}^{d-1}}\text{sign}\left(\left\langle\frac{x}{|x|},b\right\rangle\right)f(b)d\sigma(b). \tag{3}$$

We can rewrite this in terms of another operator from integral geometry:

$$\mathbb{P}(Y=1|X=x) = \frac{\mathbb{E}[Y|X=x]+1}{2} = \int_{b\in\mathbb{S}^{d-1}}\mathbf{1}_{\langle x,b\rangle>0}f_\beta(b)d\sigma(b) \triangleq \mathcal{H}(f_\beta)(x). \tag{4}$$

The operator $\mathcal{H}$ is called the hemispherical transform. $\mathcal{H}$ is a special case of the Pompeiu operator (see, *e.g.*, [30]). The operator $\mathcal{H}$ arises when one wants to reconstruct a star-shaped body from its *half-volumes* (see [5]). Inversion of this operator was studied in [5, 27], it can be achieved in the spherical harmonic basis (also called the Fourier Laplace basis as the extension of the Fourier basis on $\mathbb{S}^1$ and the Laplace basis in $\mathbb{S}^2$), using polynomials in the Laplace-Beltrami operator for certain dimensions and using a continuous wavelet transform. [27], and in a certain extent [9], also discuss some of its properties. It is an operator which is diagonal in the spherical harmonic basis and which eigenvalues are known explicitly. The estimation problem is a deconvolution problem on the sphere where the left hand side is not a density but a regression function with random design. Deconvolution on the sphere has been studied by various authors among which [10, 16, 21]. Because of the indicator function, this is a type of boxcar deconvolution. Boxcar deconvolution has been studied in specific cases in [14, 20]. There are two important difficulties regarding identification: (1) because of the intercept in the latent variable model, the left hand side of (4) is not a function defined on the whole sphere, (2) $\mathcal{H}$ is not injective (this can easily be seen from (3) where φ cannot be identified from only the imaginary part of its Fourier transform, even less when X has limited support). Proper restrictions are imposed to identify f_β. Treatment of the random design (possibly inhomogeneous) with unknown distribution appearing in the regression function that has to be inverted is an important difficulty. Regression with random design is a difficult problem, see for example [18, 23] for the case of wavelet thresholding estimation using warped wavelet for a regression model on an interval, or [6] in the case of inhomogeneous designs. [8] propose an estimator using smoothed projections on the finite dimensional spaces spanned by the first vectors of the spherical harmonics basis. It is straightforward to compute in every dimension d (the specific tools are recalled in Section 2.1). Convergence rates for the L^p-losses for $p\in[1,\infty]$ and CLT are obtained in [8]. They depend on the degree of smoothing of the operator which is $\nu = d/2$ in the Sobolev spaces based on L^2, the smoothness of the unknown function, the smoothness of f_X as well as its degeneracy (when it takes small values or is 0, in particular when x is approaching the boundary of H^+). The treatment of the random design is a major difficulty that we deal with in this paper.

The goal of this paper is to provide an estimator of f_β which is adaptive in the unknown smoothness of the function. Needlets are localized frames built on the spherical harmonic basis, they were introduced in [26]. They were successfully used in statistics to provide adaptive estimation procedures in [1, 17, 19]. As they are built on the spherical harmonic basis, they are very well suited for deconvolution on the sphere, this was used in [21]. Unlike these articles, and in the spirit of [3], we propose a method with a more accurate data-driven thresholding method.

2 Preliminaries

We use the notation $x \wedge y$ and $x \vee y$ for respectively the minimum and the maximum between x and y. We write $x \lesssim y$ when there exists c such that $x \leq cy$ and $x \gtrsim y$ when there exists c such that $x \geq cy$. We also write $x \simeq y$ when $x \lesssim y$ and $x \gtrsim y$.

2.1 Harmonic analysis on the sphere

We denote by $\mathrm{L}^p(\mathbb{S}^{d-1})$ the space of real valued p integrable functions with respect to the spherical measure σ, we denote the L^p-norm by $\|\cdot\|_p$. $\mathrm{L}^2(\mathbb{S}^{d-1})$ is a Hilbert space with the classical L^2 scalar product.
Every function in $\mathrm{L}^2(\mathbb{S}^d)$ can be decomposed in the following way:

$$f = f^+ + f^-$$

where

$$f^+(b) = (f(b) + f(-b))/2$$

and

$$f^-(b) = (f(b) - f(-b))/2$$

f^+ (resp. f^-) is the even (resp. odd) part of the function f (taking L^2 limits of functions which are well defined pointwise). We can write the orthogonal sum

$$\mathrm{L}^2(\mathbb{S}^d) = \mathrm{L}^2_{\text{odd}}(\mathbb{S}^d) \bigoplus \mathrm{L}^2_{\text{even}}(\mathbb{S}^d).$$

It can be further decomposed as the orthogonal sum

$$\mathrm{L}^2(\mathbb{S}^{d-1}) = \bigoplus_{k\in\mathbb{N}} H^{k,d}$$

where $H^{k,d}$ are the eigenspaces of the Laplace-Beltrami operator on the sphere, corresponding to the eigenvalues $\zeta_{k,d} \triangleq k(k+d-2)$. The spaces $H^{k,d}$ are of dimension

$$L(k,d) \triangleq \frac{(2k+d-2)(k+d-2)!}{k!(d-2)!(k+d-2)}.$$

Each such finite dimensional space is generated by an orthonormal basis of spherical harmonics of degree k that we denote by $(h_{k,l})_{l=1}^{L(k,d)}$. $\mathrm{L}^2_{\text{odd}}(\mathbb{S}^d)$ (resp. $\mathrm{L}^2_{\text{even}}(\mathbb{S}^d)$), is the orthogonal sum of the $H^{k,d}$ for k odd (resp. even). The space $H^{0,d}$ of spherical harmonics of degree 0 is the one dimensional space spanned by 1. The projector $L_{k,d}$ onto $H^{k,d}$ is a kernel operator with kernel

$$L_{k,d}(x,y) = \sum_{l=1}^{L(k,d)} h_{k,l}(x)h_{k,l}(y) \tag{5}$$

having the simple expression

$$L_{k,d}(x,y) = {}^\flat L_{k,d}(< x,y >), \quad {}^\flat L_{k,d}(t) \triangleq \frac{L(k,d) C_k^{\mu(d)}(t)}{|\mathbb{S}^{d-1}| C_k^{\mu(d)}(1)} \quad \text{(Addition Formula)} \tag{6}$$

where C_k^μ are the Gegenbauer polynomials and $\mu(d) = (d-2)/2$. The Gegenbauer polynomials are defined for $\mu > -1/2$ and are orthogonal with respect to the weight function $(1-t^2)^{\mu-1/2}dt$ on $[-1,1]$. $C_0^\mu(t) = 1$ and $C_1^\mu(t) = 2\mu t$ for $\mu \neq 0$ while $C_1^0(t) = 2t$. They satisfy the recursion relation

$$(k+2)C_{k+2}^\mu(t) = 2(\mu+k+1)tC_{k+1}^\mu(t) - (2\mu+k)C_k^\mu(t). \tag{7}$$

It is classical (and follows easily from (5)) that the squared L^2-norm with respect to either one of the argument of the kernel is a constant:

$$\forall x \in \mathbb{S}^{d-1}, \ \|L_{k,d}(x,\cdot)\|_2^2 = \sum_{l=1}^{L(k,d)} |h_{k,l}(x)|^2 = \frac{L(k,d)}{|\mathbb{S}^{d-1}|}. \tag{8}$$

Recall that $|\mathbb{S}^{d-1}| = 2\pi^{d/2}/\Gamma(d/2)$. The condensed harmonic expansion of a function f in $\mathrm{L}^2(\mathbb{S}^d)$ is the expansion $f = \sum_{k=0}^\infty L_{k,d} f$.

In [8], smoothed projection operators are used, they have good approximation properties in all $\mathrm{L}^p(\mathbb{S}^{d-1})$ spaces and are uniformly bounded from L^p to L^p (the L^1- norm of the kernel is uniformly bounded). They are obtained using a proper damping of the high frequencies. One such operator is the delayed means ([8] also considers the Riesz means). It is obtained via a C^∞ and decreasing function a on $\mathbb{R}^+$ supported on $[0,2]$, such that $\forall t \in [0,2], \ 0 \le a(t) \le 1$ and $\forall t \in [0,1], \ a(t) = 1$. The delayed means are defined through the kernels

$$K^{a,J}(x,y) \triangleq \sum_{k=0}^\infty a\left(\frac{k}{2^J}\right) L_{k,d}(x,y). \tag{9}$$

These kernels have nearly exponential localization properties (see Theorem 2.2 in [26]). They are building blocks for the construction of needlets in [26].

2.2 Needlets and Besov spaces

Define b such that

$$\forall t \in \mathbb{R}^+, \ b^2(t) = a(t) - a(2t).$$

It is nonzero only when $1/2 \le t \le 2$ and satisfies $\forall t \in [1/2,1], \ b^2(t) + b^2(2t) = 1$ and thus

$$\forall t \ge 1, \ \sum_{j=0}^\infty b^2\left(\frac{t}{2^j}\right) = 1.$$

We assume as well that for some positive c, $b(t) > c$ if $t \in [3/5, 5/3]$. The needlets are the functions

$$\psi_{j,\xi}(x) \triangleq \omega(j,\xi) \sum_{k=0}^\infty b\left(\frac{k}{2^{j-1}}\right) L_{k,d}(\xi,x) \quad \text{if } j \in \mathbb{N}, \ \xi \in \Xi_j \tag{10}$$

$$\psi_{0,\xi}(x) \triangleq L_{0,d}(\xi, x), \tag{11}$$

where for all $j \in \mathbb{N}$, $\xi \in \Xi_j$ and $\left(\omega(j,\xi)^2\right)_{\xi\in\Xi_j}$ are respectively the nodes and positive weights of a quadrature formula on the sphere that integrates exactly all functions in $\bigoplus_{k=0}^{2^{j+1}} H^{k,d}$, and satisfy, for some positive C_Ξ, $\forall j \in \mathbb{N},\ \frac{1}{C_\Xi}2^{j(d-1)} \le |\Xi_j| \le C_\Xi 2^{j(d-1)}$, $\quad \forall j \in \mathbb{N}, \forall \xi \in \Xi_j,\ \frac{1}{C_\Xi}2^{-j(d-1)/2} \le \omega(j,\xi) \le C_\Xi 2^{-j(d-1)/2}$ where $|\Xi_j|$ denotes the cardinal of the set Ξ_j. The quadrature formula is given in Corollary 2.9 of [26]. Note that for $j=0$, $\psi_{0,\xi}(x)$ is constant and one takes Ξ_0 as a singleton. Note that the Addition Formula is a very useful tool because the needlets, unlike the spherical harmonics, have a simple expression in every dimension. The L^p-norms of the needlets satisfy, for constants c_p and C_p uniform in j and ξ,

$$c_p 2^{j(d-1)(1/2-1/p)} \le \|\psi_{j,\xi}\|_p \le C_p 2^{j(d-1)(1/2-1/p)}, \tag{12}$$

this is a consequence of the following localization property around the nodes of the quadrature formula

$$\forall \eta \in \mathbb{S}^{d-1}, \forall \xi \in \Xi_j,\ |\psi_{j,\xi}(\eta)| \le C'_k \frac{2^{j(d-1)/2}}{\left(1+2^j \arccos(\langle \xi, \eta\rangle)\right)^k}. \tag{13}$$

If $f \in \mathrm{L}^p(\mathbb{S}^{d-1})$ for $p \in [1,\infty]$, then

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

in $\mathrm{L}^p(\mathbb{S}^{d-1})$. The needlets form a tight frame:

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

In the sequel, we denote by $\|\cdot\|_{\ell^p}$ the ℓ^p-norm of a vector. The following lemma from [1] is useful in the analysis.

Lemma 1 (i) *For every $p \in (0,\infty]$, there exists a positive constant C''_p such that*

$$\left\| \sum_{\xi\in\Xi_j} \beta_\xi \psi_{j,\xi} \right\|_p \le C''_p 2^{j(d-1)(1/2-1/p)} \left\| (\beta_\xi)_{\xi\in\Xi_j} \right\|_{\ell^p}. \tag{14}$$

(ii) *There exist a constant c_A and subsets $A_j \subset \Xi_j$ with $|A_j| \ge c_A 2^{j(d-1)}$ such that for every $p \in (0,\infty]$, there exists a positive constant $c''_{I,A}$ such that*

$$\left\| \sum_{\xi\in A_j} \beta_\xi \psi_{j,\xi} \right\|_p \ge c''_{I,A} 2^{j(d-1)(1/2-1/p)} \left\| (\beta_\xi)_{\xi\in\Xi_j} \right\|_{\ell^p}. \tag{15}$$

(iii) *For every $p \in [1,\infty]$, there exists a positive constant C'''_p such that*

$$\left(\sum_{\xi\in\Xi_j} |\langle f, \psi_{j,\xi}\rangle|^p \right)^{1/p} 2^{j(d-1)(1/2-1/p)} \le C'''_p \|f\|_p. \tag{16}$$

[26] discuss three formulations of the Besov spaces $B^s_{p,q}$ on the sphere. The Besov spaces will be our scales of smoothness for the adaptive estimation. One characterization is in terms of the approximation error. If $s>0$, $p\in[1,\infty]$ and $q\in(0,\infty]$, f belongs to $B^s_{p,q}$ if and only if f is in $\mathrm{L}^p(\mathbb{S}^{d-1})$ and

$$\|f\|^A_{B^s_{p,q}}=\|f\|_p+\left\|\left(2^{js}E_{2^j}(f)_p\right)_{j\in\mathbb{N}}\right\|_{\ell^q}<\infty$$

where

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

Whatever the function a in the definition of the smoothed projection operators, the above norm is equivalent to the following sequence space norm

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

We denote by $B^s_{p,q}(M)$ the ball of radius M for the above norm in $B^s_{p,q}$. From the proof of the continuous embeddings in [1] we can get easily:

Lemma 2 *(i) If $p\le r\le\infty$, $B^s_{r,q}(M)\subset B^s_{I,q}(C_\Xi^{1/p-1/r}M)$*

(ii) If $s>(d-1)(1/r-1/p)$ and $r\le p\le\infty$, $B^s_{r,q}(M)\subset B^{s-(d-1)(1/r-1/p)}_{p,q}(M)$.

(iii) If $f\in B^s_{r,q}(M)$ and $\left((\beta_{j,\xi})_{\xi\in\Xi_j}\right)_{j\in\mathbb{N}}$ are its needlet coefficients, then

$$\forall z\ge1,\ \sum_{\xi\in\Xi_j}|\beta_{j,\xi}|^z\le C_\Xi^{1-(z\wedge r)/r}D_j^z2^{-jz(s+(d-1)(1/2-1/(z\wedge r)))}\tag{17}$$

where $\forall j\in\mathbb{N},\ D_j\ge0$, $(D_j)_{j\in\mathbb{N}}\in\ell_q$ and $\|(D_j)_{j\in\mathbb{N}}\|_{\ell^q}\le M$.

Note that $\|D_j\|_q\le M$ implies that $\forall j\in\mathbb{N},\ D_j\le M$. Recall as well that, when f belongs to $B^s_{I,q}$ with $s>(d-1)/p$, then f is continuous and bounded.

2.3 The hemispherical transform

The hemispherical transform is a mapping from $\mathrm{L}^2(\mathbb{S}^{d-1})$ to $\mathrm{L}^2(\mathbb{S}^{d-1})$ which maps a function f to a function which, evaluated at $x\in\mathbb{S}^{d-1}$, is the integral of the original function on the hemisphere $\{y\in\mathbb{S}^d:\ \langle x,y\rangle>0\}$. It is a special case of the Pompeiu operator and is strongly related to the spherical Radon transform. Several inversion formulas as well as properties of this mapping are given in [27]. These inversion formulas include polynomials in the spherical Laplacian (for certain dimensions) and a continuous wavelet transform, the known inversion formula in the spherical harmonic basis is recalled. We make use of this latter because the needlet frame is very well suited to this decomposition.

A consequence of the Funck-Hecke theorem (see, *e.g.*, [9]), is that $\mathcal{H}$ is a diagonal operator in the spherical harmonic basis $(h_{k,l})_{l=1,\dots,L(k,d),\ k\in\mathbb{N}}$ with the same eigenvalue on the spaces $H^{k,d}$. We thus only index them by the degree of the harmonics.

Proposition 3 *$\mathcal{H}$ is a self-adjoint operator on $\mathrm{L}^2(\mathbb{S}^{d-1})$ with null space*

$$ker\ \mathcal{H} = \bigoplus_{p=1}^{\infty} H^{2p,d} = \left\{ f \in \mathrm{L}^2_{even}(\mathbb{S}^{d-1}) :\ \int_{\mathbb{S}^{d-1}} f(x)d\sigma(x) = 0 \right\}.$$

Its nonzero eigenvalues $(\lambda_{k,d})_{k\in\mathbb{N}}$ (k indices the degree of the harmonics) are

$$\begin{aligned}
&\lambda_{0,d} = \frac{2}{|\mathbb{S}^{d-1}|},\\
&\lambda_{1,d} = \frac{|\mathbb{S}^{d-2}|}{d-1},\\
&\forall p \in \mathbb{N},\ \lambda_{2p+1,d} = \frac{(-1)^p|\mathbb{S}^{d-2}|1\cdot 3\cdots(2p-1)}{(d-1)(d+1)\cdots(d+2p-1)}.
\end{aligned}$$

Note that $\forall p \in \mathbb{N}\setminus\{0\},\ \lambda_{2p,d} = 0$. It is easy to check (see, *e.g.*, [27]) that $\mathcal{H}$ is continuous from $\mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1})$ to $\mathrm{H}^{d/2}_{\mathrm{odd}}$ and that its inverse is continuous from $\mathrm{H}^{d/2}_{\mathrm{odd}}$ to $\mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1})$, where $\mathrm{H}^{d/2}_{\mathrm{odd}}$ is the restriction to odd functions of the Sobolev space $\mathrm{H}^{d/2}$. H^s is defined, for arbitrary s, by

$$\mathrm{H}^s = \left\{ f \in \mathrm{L}^2(\mathbb{S}^{d-1}) :\ (-\Delta)^{s/2} f \triangleq \sum_{k\in\mathbb{N}} \zeta_{k,d}^{s/2} L_{k,d} f \in \mathrm{L}^2(\mathbb{S}^{d-1}) \right\}$$

equipped with the norm

$$\|f\|_{2,s} = \|f\|_2 + \left\| (-\Delta)^{s/2} f \right\|_2.$$

The inverse of a function R in $\mathrm{H}^{d/2}_{\mathrm{odd}}$ is

$$\mathcal{H}^{-1}(R) = \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} L_{k,d}(R) \left(= \sum_{k \text{ odd}} \frac{1}{\lambda_{k,d}} \sum_{l=1}^{L(k,d)} \langle R, h_{k,l}\rangle h_{k,l} \right), \tag{18}$$

we use a parenthesis to stress that the last equality is not practical if we work in arbitrary dimensions but can nevertheless be used in proofs. Let us also recall the following Bernstein type inequality from [8].

Proposition 4

$$\forall d \geq 2,\ \forall p \in [1,\infty],\ \exists B(d,p) > 0 :\ \forall P \in \bigoplus_{\substack{k=0\\ k\ odd}}^{K} H^{k,d},\ \|\mathcal{H}^{-1}P\|_p \leq B(d,p)K^{d/2}\|P\|_p. \tag{19}$$

Throughout the paper, we denote by $\nu = d/2$ the degree of ill-posedness of the inverse problem. It is the same degree of ill-posedness as that of the Radon transform in $\mathbb{R}^d$ which appears in tomography and in [12] for the estimation of the vector of random coefficients in the linear regression problem.

2.4 Identification of f_β

Let us review the main arguments for the identification of f_β that are taken from [8]. Imposing, as we do, that β belongs to $\mathbb{S}^{d-1}$ is not sufficient. First, the left hand side of (4) is only defined on the support of f_X. Through out the article we make the following assumption. [8] present cases where it could be relaxed when we do not assume that all coefficients are random.

Assumption 5 $\mathrm{supp} f_X = H^+$ *and* $\mathbb{E}[Y|X=x]$ *is well defined pointwise on* $\mathrm{supp} f_X$*.*

First note that, because f_β is a density,

$$\mathcal{H}(f_\beta) = \mathcal{H}(f_\beta^-) + \frac{1}{2}. \tag{20}$$

We can now introduce the function R such that

$$R(x) = \begin{cases} \mathbb{E}[Y|X=x] & \text{when } x \in H^+ \\ -\mathbb{E}[Y|X=-x] & \text{when } -x \in H^+ \end{cases}. \tag{21}$$

It is the unique extension of the regression function which is compatible with (1) and (20). We can now write

$$\frac{R}{2} = \mathcal{H}(f_\beta^-).$$

Thus, (1) implies implicitly, if f_β belongs to $\mathrm{L}^2(\mathbb{S}^{d-1})$, that $R \in \mathrm{H}^{d/2}_{\mathrm{odd}}(\mathbb{S}^{d-1})$ and is thus continuous on the whole sphere. Also, from properties of Section 2.3, there exists a unique f_β^- in $\mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1})$ such that $R = 2\mathcal{H}(f_\beta^-)$. The function f_β^- can be retrieved via the inversion formula (18). We need yet another assumption to identify f_β, this is due to the non invertibility of $\mathcal{H}$ in the whole $\mathrm{H}^{d/2}(\mathbb{S}^{d-1})$ space.

Assumption 6 f_β *is defined pointwise and has a support included in some hemisphere.*

Assumption 6 appears in both [8, 13]. In many applications this is a plausible assumption. It is the case for example if one coefficient has a sign or if some coefficients are non random. For example, if one regressor is the price difference, then the price coefficient is negative in the binary choice model. Indeed, when the price difference increases there is substitution from the good labeled 1 to good labeled -1 and the choice probability for good 1 decreases.

Using Assumption 6, we can recover uniquely f_β via

$$f_\beta = 2 f_\beta^- \mathbf{1}_{f_\beta^- > 0}.$$

Note that we do not need to know which hemisphere contains $\mathrm{supp} f_\beta$. Given an estimator $\widehat{f_\beta^-}$ of f_β^-, we shall always use $2\widehat{f_\beta^-}\mathbf{1}_{\widehat{f_\beta^-}>0}$ as an estimator of f_β. The first stage of the proof of Proposition 4.2 in [8] tells us how to relate the loss in the estimation of f_β with that of the estimation of f_β^-.

2.5 Random design

For the purpose of estimation, we also exploit the following relation which is valid for any g in $\mathrm{L}^2(\mathbb{S}^{d-1})$.

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

The expectation could be approximated by $\frac{2}{n}\sum_{i=1}^n \frac{Y_i g^-(X_i)}{\widehat{f_X}(X_i)}$ where $\widehat{f_X}$ is an estimator of the unknown f_X, possibly trimmed to avoid the division by quantities close to zero.

Like in [8], we rely on a plug-in estimator of f_X. Many such estimators exists and we would like to mention one particular estimator which is the needlet thresholding estimator of the density of [1].

3 Lower Bound

The following theorem gives lower bounds on the minimax risk.

Theorem 7 *Assume that $f_X \in \mathrm{L}^\infty(H^+)$.*

(i) When $p \geq 1$, $z \geq 1$, $q \geq 1$ (with the restriction $q \leq r$ is $s = p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$) and $s \geq p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$ (the parameters are in the dense zone),

$$\inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)} \mathbb{E}\left\|\widehat{f_\beta} - f_\beta\right\|_p^z \gtrsim \left(\frac{1}{\sqrt{n\|f_X\|_{\mathrm{L}^\infty(H^+)}}}\right)^{\frac{sz}{s+\nu+(d-1)/2}}. \tag{22}$$

(ii) When $p \geq 1$, $z \geq 1$, $q \geq 1$ and $\frac{d-1}{r} < s < p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$ (the parameters are in the sparse zone),

$$\inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)} \mathbb{E}\left\|\widehat{f_\beta} - f_\beta\right\|_p^z \gtrsim \left(\sqrt{\frac{\log\left(n\|f_X\|_{\mathrm{L}^\infty(H^+)}\right)}{n\|f_X\|_{\mathrm{L}^\infty(H^+)}}}\right)^{\left(\frac{s-(d-1)(1/r-1/p)}{s+\nu-(d-1)(1/r-1/2)}\right)z} \tag{23}$$

The proof of this result is given in Section 4. As discussed in Section 4.3 of [8], the classical assumption that f_X is bounded from below is very restrictive for the model at hand. In the $d = 2$ case, it would imply that, in the original scale, X has tails larger than the Cauchy tails. It is therefore important for applications to allow for densities which are unbounded from below. We make the dependence on $\|f_X\|_{\mathrm{L}^\infty(H^+)}$ explicit. However this does not give that the estimation problem is more difficult when f_X can take values arbitrary close to 0, it does not even take into account that f_X is a density as the larger $\|f_X\|_{\mathrm{L}^\infty(H^+)}$ the greater the lower bound. We therefore expect these lower bounds to properly characterize the difficulty of the estimation problem when f_X is bounded from below but to be too optimistic otherwise.

[6] introduces, in the case of the estimation of the regression function and inhomogeneous designs, risks where the rate is a function and can vary with the points in the support of the density of the design. There are no extensions to inverse problems up to our knowledge.

It will also appear in Section 4 that even if f_X were known but unbounded from below, good rates require trimming of the density f_X for design points where the density is low. Not knowing f_X might degrade the optimal rates in one step procedures. This is discussed in Section 5.

4 A needlet thresholded estimator when f_X is known and bounded from below

4.1 Smoothed projections and needlet estimators

In this section we present an ideal benchmark estimator. We assume that the density of the design is known and bounded from below. In practice it is unknown and in most cases unbounded from below (see the discussion in Section 3).

Using the identity of Section 2.5 with $g(\cdot) = L_{k,d}(\cdot, x)$ for fixed x, we estimate $L_{k,d}R(x)$ by

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

where $L_{k,d}^-(x_i, x) = 0$ if k is even and $L_{k,d}^-(x_i, x) = L_{k,d}(x_i, x)$ if k is odd. The subscript I stands for the ideal estimator where the density of the random design is known. Because $H^{k,d}$ is a vector space, $\widehat{L_{k,d}R}^I \in H^{k,d}$. A smoothed projection estimator with kernel (9) and smoothing window a (in the ideal case where f_X is known) can be written as

$$\widehat{f_\beta^-}^{I,a,J} = \frac{1}{2}\sum_{k \text{ odd}} \frac{a\left(\frac{k}{2^J}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}^I(x).$$

We can also estimate f_β^- using the needlet frame with the same smoothing window a. The needlet coefficients are equal to

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

$$= \omega(j,\xi) \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{2\lambda_{k,d}} L_{k,d} R(\xi)$$

$$= \omega(j,\xi) \sum_{\substack{2^{j-2}<k<2^j \\ k \text{ odd}}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{2\lambda_{k,d}} L_{k,d} R(\xi)$$

$$= \langle f_\beta^{-I,a,J}, \psi_{j,\xi} \rangle \quad \forall j \le J, \text{ (collecting back the terms using that } a\left(\tfrac{k}{2^j}\right) = 1 \text{ for } k = 0, \ldots, 2^j)$$

where $f_\beta^{-I,a,J}$ is the expected value of $\widehat{f_\beta^-}^{I,a,J}$ (the spherical convolution $K_a \star f_\beta^-$). The needlet coefficients can be estimated by

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

Moreover

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

which belongs to $\bigoplus_{k=0}^{2^{j+1}} H^{k,d}$, thus, from the quadrature formula,

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

and

$$\sum_{j=0}^{J} \sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^{I,a} \psi_{j,\xi} = \sum_{j=1}^{J} \sum_{\xi \in \Xi_j} \widehat{\beta}_{j,\xi}^{I,a} \psi_{j,\xi} \quad (\widehat{f_\beta^-}^{I,a,J} \text{ is odd and thus of integral 0 on the sphere})$$

$$= \frac{1}{2} \sum_{\substack{k=1 \\ k \text{ odd}}}^{2^{J-1}-1} \frac{1}{\lambda_{k,d}} \widehat{L_{k,d}R}^I + \frac{1}{2} \sum_{\substack{k=2^{J-1}+1 \\ k \text{ odd}}}^{2^J-1} \frac{b^2\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}^I \quad (\text{for } t \in [1/2, 1],\ b^2(t) + b^2(2t) = 1)$$

$$= \frac{1}{2} \sum_{\substack{k=1 \\ k \text{ odd}}}^{2^{J-1}-1} \frac{1}{\lambda_{k,d}} \widehat{L_{k,d}R}^I + \frac{1}{2} \sum_{\substack{k=2^{J-1}+1 \\ k \text{ odd}}}^{2^J-1} \frac{a\left(\frac{k}{2^{J-1}}\right)}{\lambda_{k,d}} \widehat{L_{k,d}R}^I \quad (b^2(t) = a(t) - a(2t))$$

$$= \widehat{f_\beta^-}^{I,a,J-1}$$

The smoothed projection and needlet estimators coincide. They are biased and the bias corresponds to the approximation error.

4.2 A data driven thresholding scheme

The unbiased estimators of the needlet coefficients

$$\widehat{\beta}_{j,\xi}^{I,a} = \frac{1}{n} \sum_{i=1}^{n} \omega(j,\xi) \frac{Y_i}{f_X(X_i)} \sum_{k \text{ odd}} \frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}} L_{k,d}(X_i, \xi)$$

$$\triangleq \frac{1}{n}\sum_{i=1}^{n} G^I_{j,\xi}(X_i, Y_i).$$

are used to define the needlet thresholded estimator

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

where $\rho_{T_{j,\xi,\gamma}}$ is a suitable level and local dependent thresholding function depending on some $\gamma \geq 1$. In the subsequent analysis, we consider the hard thresholding function

$$\rho_{T_{j,\xi,\gamma}}(x) = \mathbf{1}_{|x|>T_{j,\xi,\gamma}}.$$

The highest resolution level J that should be used to obtain a needlet estimator of section 4.1 that achieves the minimax rate of convergence depends on a prior knowledge of the smoothness of the unknown density of the random coefficient. Hard-thresholding is a nonlinear estimation method where we allow for a larger highest resolution level J, independent of the smoothness of the unknown function, but where thresholding allows to perform a bias/variance trade-off at the level of the coefficients in the high-dimensional space. As we will see this yields an adaptive procedure. We define the empirical variance estimator

$$\hat{\sigma}^I_{j,\xi} = \sqrt{\frac{1}{n(n-1)}\sum_{i=2}^{n}\sum_{k=1}^{i-1}\left(G^I_{j,\xi}(Y_i, X_i) - G^I_{j,\xi}(Y_k, X_k)\right)^2}$$

and the data driven thresholds

$$T_{j,\xi,\gamma} \triangleq T^I_{j,\xi,\gamma} = 2\sqrt{2\gamma}t_n\widehat{\sigma}^I_{j,\xi} + \frac{28}{3}M^I_{j,\xi}\frac{\gamma\log n}{n-1}$$

where $M^I_{j,\xi}$ is some upper bound on the sup-norm over of $\{\pm 1\}\times H^+$ of $G^I_{j,\xi}(x,y) - \mathbb{E}\left[G^I_{j,\xi}(X_i, Y_i)\right] = G^I_{j,\xi}(x,y) - \beta^a_{j,\xi}$ (remark that $M^I_{j,\xi}$ can be chosen equal to $2\|G^I_{j,\xi}\|_\infty$) and we use the short hand notation

$$t_n = \sqrt{\frac{\log n}{n}}.$$

Using (12) and Proposition 4, we get the following upper bound which is uniform in ξ

$$2\|G^I_{j,\xi}\|_\infty \leq 2\left\|\mathcal{H}^{-1}\left(\psi^-_{j,\xi}\right)\right\|_\infty \left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)} \leq 2C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)} \triangleq M^I_j. \quad (24)$$

It depends on some prior knowledge of $\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}$. The higher order term in the definition of $T_{j,\xi,\gamma}$ which involves $M^I_{j,\xi}$ allows to control the fluctuations of this estimated threshold.

The estimator of f_β that we consider is

$$\widehat{f_\beta}^{I,a,\rho} = 2\widehat{f_\beta^-}^{I,a,\rho}\mathbf{1}_{\widehat{f_\beta^-}^{I,a,\rho}>0}$$

where ρ is the above hard thresholding function with the data driven threshold.

4.3 Two general inequalities

We shall use below the constants $c_{1,z}$ and $c_{2,z}$ defined by

$$\int_{\mathbb{R}^+} z\tau^{z-1}e^{-\beta\tau}d\tau \leq c_{1,z}\beta^{-z} \tag{25}$$

$$\int_{\mathbb{R}^+} z\tau^{z-1}e^{-\alpha\tau^2}d\tau \leq c_{2,z}\alpha^{-z/2}. \tag{26}$$

Theorem 8 *For all $\tau > 1$, $\gamma > 1$, $z > 1$,*

$$T^{s,++}_{j,\xi,\gamma} \geq 3\sqrt{2\gamma}t_n\sigma^I_{j,\xi} + 26M^I_{j,\xi}\frac{\gamma\log n}{n-1} \triangleq T^{s,+}_{j,\xi,\gamma},$$

the two following inequalities hold:
when $p = \infty$,

$$\frac{1}{2^{z-1}}\mathbb{E}\left[\left\|\widehat{f_\beta}^{I,a,\rho} - f_\beta\right\|^z_\infty\right] \leq \left\|f_\beta^{-I,a,J} - f_\beta^-\right\|^z_\infty$$
$$+ (J+1)^{z-1}C''^z_\infty\left\{a_{n,\infty,z,J}\sum_{j=0}^J 2^{j(d-1)z/2}\left(\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,++}_{j,\xi,\gamma}} + \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^{s,++}_{j,\xi,\gamma}}\right]\right)\right.$$
$$+ \frac{4}{n^\gamma}C_\Xi\sum_{j=0}^J 2^{j(d-1)(z/2+1)}\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z$$
$$\left. + \left(\frac{C_\Xi 4}{n^\gamma}\right)^{1-1/\tau}\left(\frac{1}{\sqrt{n}}\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)}2^{Jz(\nu+(d-1)/2)}\right)^z 2^{J(d-1)(1-1/\tau)}b_{n,\infty,z,J,\tau}\right\}$$

where

$$a_{n,\infty,z,J} = 1 + \left(\frac{2}{\sqrt{\gamma\log n}}\right)^z\left(2 + \left(\log\left(C_\Xi 2^{J(d-1)}c_{2,z}\right)\right)^{z/2}\right) + \left(\frac{4}{\gamma\log n}\right)^z\left(2 + \left(\log\left(C_\Xi 2^{J(d-1)}c_{1,z}\right)\right)^z\right)$$

$$b_{n,\infty,z,J,\tau} = \frac{\left(2\sqrt{2}C_2B(d,2)\right)^z\left(2^{1/\tau} + \left(\log\left(C_\Xi 2^{J(d-1)}c_{2,z\tau}\right)\right)^{z/2}\right)}{1-2^{-(z\nu+(d-1)(z/2+1-1/\tau))}}$$
$$+ \frac{(8C_\infty B(d,\infty)/3)^z\left(2^{1/\tau} + \left(\log\left(C_\Xi 2^{J(d-1)}c_{1,z}\right)\right)^z\right)}{1-2^{-(z\nu+(d-1)(z+1-1/\tau))}}\left(\frac{2^{J(d-1)}}{n}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\right)^{z/2};$$

while when $p \in [1,\infty)$,

$$\frac{1}{2^{z-1}}\mathbb{E}\left[\left\|\widehat{f_\beta}^{I,a,\rho} - f_\beta\right\|^z_p\right] \leq \left\|f_\beta^{-I,a,J} - f_\beta^-\right\|^z_p$$
$$+ (J+1)^{z-1}C''^z_p C_\Xi^{z/(p\wedge z)-1}\left\{a_{n,p,z,J}\sum_{j=0}^J 2^{j(d-1)(z/2-z/(p\vee z))}\sum_{\xi\in\Xi_j}\left(\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,++}_{j,\xi,\gamma}} + \mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z\right]\mathbf{1}_{\left|\beta^a_{j,\xi}\right|> T^{s,++}_{j,\xi,\gamma}}\right)\right.$$

$$+\frac{4}{n^{\gamma}}\sum_{j=0}^{J}2^{j(d-1)z(1/2-1/(p\vee z))}\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z$$
$$\left.+\frac{2^{2-1/\tau}}{n^{\gamma(1-\frac{1}{\tau})}}C_{\Xi}\left(\frac{1}{\sqrt{n}}\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)}2^{J(\nu+(d-1)/2)}\right)^z 2^{J(d-1)(1-z/(p\vee z))}b_{n,p,z,J,\tau}\right\}.$$

where

$$a_{n,p,z,J}=1+2\left(\left(\frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma\log n}}\right)^z+\left(\frac{2c_{1,z}^{1/z}}{\gamma\log n}\right)^z\right)$$
$$b_{n,p,z,J,\tau}=\frac{\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z}{1-2^{-(z\nu+(d-1)(z/2+1-z/(p\vee z)))}}+\frac{\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z}{1-2^{-(z\nu+(d-1)(z+1-z/(p\vee z)))}}\left(\frac{2^{J(d-1)}}{n}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\right)^{z/2}.$$

The inequalities of Theorem 8 provide some theoretical guaranty valid without any assumptions on the function f_β. When J is well chosen depending on n and under some minimal regularity assumption on f_β (see for instance Theorem 9), the only two meaningful terms are the approximation term $\left\|f_\beta^{-I,a,J}-f_\beta^-\right\|_p^z$ and the term involving $\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,++}_{j,\xi,\gamma}}$ and $\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}$. This second term can be interpreted in term of oracle inequality, where the oracle estimates $\beta^a_{j,\xi}$ if and only if the error made by estimating this coefficient is smaller than the one made by discarding it. Indeed, such an *oracle* strategy would lead (when $p<\infty$) to a quantity of the form

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

Proving that such an oracle inequality holds would require to lower bound $\left(\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\right)^{1/z}$. In the inequalities of Theorem 8 the ideal quantity $\left(\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\right)^{1/z}$ is replaced by $T^{s,++}_{j,\xi,\gamma}$, we call this term a *quasi-oracle* term. The remaining terms can be made as small as we wish by taking γ large enough. 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.

Data driven thresholds are known to perform much better than thresholds involving deterministic upper bounds on the variance of the coefficients in finite samples. The inequalities of the Theorem 8 show that they work at least as well as a deterministic one using the unknown variance of each coefficient.

4.4 Adaptive estimation over Besov ellipsoids

The general inequalities of the previous section can be used to derive minimax results. We condider here some Besov ellipsoids and obtain

Theorem 9 *Take J such that $2^{J(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)}\le t_n^{-1}$. If $M>0$, $r\ge1$, $s>(d-1)/r$ and $q\ge1$ we have* (*i*) *For any* $z > 1$, *there exists a constant* $\tilde{c}_\infty = \tilde{c}_\infty(s, r, \gamma)$ *such that if* $\gamma > z/2 + 1$,

$$\sup_{f_\beta^- \in B^s_{r,q}(M)} \mathbb{E} \left\| \widehat{f_\beta}^{I,a,\rho} - f_\beta \right\|_\infty^z \leq \tilde{c}_\infty (\log n)^{z-1} M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)} t_n \right)^{\mu_{s,\infty} z} \tag{27}$$

where

$$\mu_{s,\infty} = \frac{s - (d-1)/r}{s + \nu - (d-1)(1/r - 1/2)}.$$

(*ii*) *For* $p \in [1, \infty)$, $q \geq 1$ *(with the restriction* $q \leq r$ *is* $s = p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$*), there exists some constant* $\tilde{c}_p = \tilde{c}_p(s, r, p, \gamma)$ *such that if* $\gamma > p/2$,

$$\sup_{f_\beta^- \in B^s_{r,q}(M)} \mathbb{E} \left\| \widehat{f_\beta}^{I,a,\rho} - f_\beta \right\|_p^p \leq \tilde{c}_p (\log n)^{p-1} M^{\varpi} \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)} t_n \right)^{\mu p} \tag{28}$$

where $\mu = \mu_d$ *with*

$$\mu_d = \frac{s}{s + \nu + (d-1)/2}$$

and $\varpi = \varpi_d = r$ *in the dense zone*

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

and $\mu = \mu_s$ *with*

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

and $\varpi = \varpi_s$ *is arbitrary such that* $\varpi > p\frac{\nu + (d-1)(1/2 - 1/(p \vee z))}{s + \nu - (d-1)(1/r - 1/2)}$ *in the sparse zone*

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

The constant $\mu_{s,\infty}$ corresponds to the limit of μ_s as p goes to infinity. It should be noted that these upper bounds blow-up when f_X is unbounded from below. We will see in the next section that trimming allows to avoid this problem, at the expense of a more complicated control of the expected loss.

5 The case where the density of the design is unknown and possibly unbounded from below

In this section, we consider a modified estimator to handle the case where the density of the design is unknown and possibly unbounded from below. We show a modified version of Theorem 9 in that case.

5.1 Plug-in strategy

We assume now that one has a preliminary estimator $\widehat{f_X}$ of f_X, based on a different sample. Expectations are taken conditional on that first sample. The estimator can be trimmed by a proper constant t to allow for designs with density approaching zero. This is particularly useful in the neighborhood of the boundary of H^+, in order to avoid too stringent assumptions on the distribution of the design.

Using a simple plug-in rule, f_X can be replaced in the previous estimators by $\widehat{f_X}$ yielding the estimated harmonic projection of the extended regression function

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

of expectation

$$L_{k,d}R^P(x) = 2\mathbb{E}_{(X,Y)}\left[\frac{Y L_{k,d}^-(X,x)}{\widehat{f_X}(X)}\right] = \left\langle R\left(\frac{f_X}{\widehat{f_X}}\right)^+, L_{k,d}^-(\cdot,x)\right\rangle$$

where $\left(f_X/\widehat{f_X}\right)^+$ is the even extension to the whole sphere of $f_X/\widehat{f_X}$ (initially defined on H^+). This gives rise to the following linear estimator

$$\widehat{f_\beta^-}^{P,a,J} = \frac{1}{2}\sum_{k \text{ odd}}\frac{a\left(\frac{k}{2^J}\right)}{\lambda_{k,d}}\widehat{L_{k,d}R}^P$$

whose mean is

$$f_\beta^{-P,a,J} = \frac{1}{2}\sum_{k \text{ odd}}\frac{a\left(\frac{k}{2^J}\right)}{\lambda_{k,d}}L_{k,d}R^P.$$

The plug-in estimators of the needlet coefficients are

$$\begin{aligned}\widehat{\beta}_{j,\xi}^{P,a} &= \frac{1}{2}\omega(j,\xi)\sum_{k \text{ odd}}\frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}}\widehat{L_{k,d}R}^P(\xi)\\ &= \langle \widehat{f_\beta^-}^{P,a,J}, \psi_{j,\xi}\rangle \quad \forall j \le J\end{aligned}$$

which yields the thresholded estimator

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

In this section we consider the data driven thresholds

$$T_{j,\xi,\gamma} \triangleq T_{j,\xi,\gamma}^P = 2\sqrt{2\gamma}t_n\widehat{\sigma}_{j,\xi}^P + \frac{28}{3}M_{j,\xi}^P\frac{\gamma\log n}{n-1}$$

where

$$\widehat{\sigma}^P_{j,\xi} = \sqrt{\frac{1}{n(n-1)}\sum_{i=2}^{n}\sum_{k=1}^{i-1}\left(G^P_{j,\xi}(Y_i,X_i) - G^P_{j,\xi}(Y_k,X_k)\right)^2} \tag{29}$$

$$G^P_{j,\xi}(Y_i,X_i) = \omega(j,\xi)\frac{Y_i}{\widehat{f_X}(X_i)}\sum_{k \text{ odd}}\frac{b\left(\frac{k}{2^{j-1}}\right)}{\lambda_{k,d}}L_{k,d}(X_i,\xi) \tag{30}$$

and $M^P_{j,\xi}$ is some upper bound on the sup-norm over of $\{\pm 1\} \times H^+$ of $G^P_{j,\xi}(x,y) - \mathbb{E}\left[G^P_{j,\xi}(X_i,Y_i)\right] = G^P_{j,\xi}(x,y) - \beta^{P,a}_{j,\xi}$, where the expectation is conditional on the sample used to estimate $\widehat{f_X}$, and $\beta^{P,a}_{j,\xi}$ the expectation of $\widehat{\beta}^{P,a}_{j,\xi}$, again conditional on the sample used to estimate $\widehat{f_X}$,

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

The following uniform upper bound could be used

$$M^P_{j,\xi} \le 2C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{\widehat{f_X}}\right\|_{\mathrm{L}^\infty(H^+)} \triangleq M^P_j. \tag{31}$$

5.2 Upper bounds

Below we denote, for $\pi \ge 1$, by

$$M^{P,a,J,r,\pi} = M + 2C_{r,\pi}2^{J(s+\nu+(d-1)(1/\pi-1/r)_+)}\left\|\frac{f_X}{\widehat{f_X}} - 1\right\|_{\mathrm{L}^\pi(H^+)}$$

where $C_{r,\pi} = 2C'''_p C_{\text{proj}} B(d,p)|\mathbb{S}^{d-1}|^{(1/r-1/\pi)_+}$ and C_{proj} is the constant of the L^p continuity of the smoothed projections (wee Lemma 2.4 (c) of [26]). The expectations in the theorem below are conditional on the sample that is used to estimate $\widehat{f_X}$.

Theorem 10 *Take J such that $2^{J(\nu+(d-1)/2)}\left\|\frac{1}{\widehat{f_X}}\right\|_{\mathrm{L}^\infty(H^+)} \le t_n^{-1}$. If $M > 0$, $r \ge 1$, $s > (d-1)/r$ and $q \ge 1$ we have*

(i) For any $z > 1$, there exists a constant $\tilde{c}_\infty = \tilde{c}_\infty(s,r)$ such that $\gamma > z/2 + 1$,

$$\sup_{f_\beta^- \in B^s_{r,q}(M)} \mathbb{E}\left\|\widehat{f_\beta}^{P,a,\rho} - f_\beta\right\|_\infty^z \le 3^{z-1}\inf_{\pi\ge 1}\left\{\tilde{c}_\infty(\log n)^{z-1}\left(M^{P,a,J,r,\pi}\right)^r\left(\left\|\frac{1}{\widehat{f_X}}\right\|_{\mathrm{L}^\infty(H^+)} t_n\right)^{\mu_{s,\infty}z}\right.$$
$$\left. + 2^{-Js}\left(C'''_\infty\right)^{-1}\left(M^{P,a,J,r,\pi} - M\right)\right\} \tag{32}$$

where

$$\mu_{s,\infty} = \frac{s-(d-1)/r}{s+\nu-(d-1)(1/r-1/2)}.$$

(ii) *For $p \in [1,\infty)$, $q \geq 1$ (with the restriction $q \leq r$ is $s = p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$), there exists some constant $\tilde{c}_p = \tilde{c}_p(s,r,p)$ such that if $\gamma > p/2$,*

$$\sup_{f_\beta^- \in B^s_{r,q}(M)} \mathbb{E}\left\|\widehat{f_\beta}^{P,a,\rho} - f_\beta\right\|_p^p \leq 3^{p-1} \inf_{\pi\geq 1}\left\{\tilde{c}_p (\log n)^{p-1}\left(M^{P,a,J,r,\pi}\right)^{\varpi}\left(\left\|\frac{1}{\widehat{f_X}}\right\|_{\mathrm{L}^\infty(H^+)} t_n\right)^{\mu p}\right.$$
$$\left. + 2^{-Js}\left(C_p'''\right)^{-1}\left(M^{P,a,J,r,\pi} - M\right)\right\} \tag{33}$$

where $\mu = \mu_d$ with

$$\mu_d = \frac{s}{s + \nu + (d-1)/2}$$

and $\varpi = r$ in the dense zone

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

and $\mu = \mu_s$ with

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

and ϖ is arbitrary such that $\varpi > p\frac{\nu + (d-1)(1/2 - 1/(p\vee z))}{s + \nu - (d-1)(1/r - 1/2)}$ in the sparse zone

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

Two quantities appear in the upper bound that account for the design and the estimation of the density of the design: $\left\|\frac{1}{\widehat{f_X}}\right\|_{\mathrm{L}^\infty(H^+)}$ and $\left(M^{P,a,J,r,\pi}\right)^r$. Since in most design distributions of interest in the original scale $\mathbb{R}^{d-1}$, the corresponding density on the sphere f_X is bounded from below, it is useful to work with estimators $\widehat{f_X}$ which are trimmed estimators of an original estimator $\widehat{f_X}^t = \max(\widehat{f_X}, t)$ for a properly chosen t. For such trimmed preliminary estimators we obtain $\left\|\frac{1}{\widehat{f_X}^t}\right\|_{\mathrm{L}^\infty(H^+)} = t^{-1}$.

Now, the quantity $\left\|\frac{f_X}{\widehat{f_X}^t} - 1\right\|_{\mathrm{L}^\pi(H^+)}$ appears in the term $M^{P,a,J,r,\pi}$. It is possible to use the upper bound

$$\left\|\frac{f_X}{\widehat{f_X}^t} - 1\right\|_{\mathrm{L}^\pi(H^+)} \leq t^{-1}\left\|\widehat{f_X} - f_X\right\|_{\mathrm{L}^\pi(H^+)}.$$

For a trimmed estimator, this yields, for example,

$$\begin{aligned}\left\|\frac{f_X}{\widehat{f_X}^t} - 1\right\|_{\mathrm{L}^\pi(H^+)} &\leq \sigma\left(0 < \widehat{f_X} < t\right)^{1/\pi} + t^{-1}\left\|f_X \mathbf{1}_{\widehat{f_X}<t}\right\|_{\mathrm{L}^\pi(H^+)} + t^{-1}\left\|\widehat{f_X} - f_X\right\|_{\mathrm{L}^\pi(H^+)} \\ &\leq \left(1 + t^{-1}\|f_X\|_{\mathrm{L}^\infty(H^+)}\right)\sigma\left(0 < \widehat{f_X} < t\right)^{1/\pi} + t^{-1}\left\|\widehat{f_X} - f_X\right\|_{\mathrm{L}^\pi(H^+)}.\end{aligned}$$

Moreover, for $u > 1$,

$$\sigma\left(0 < \widehat{f_X} < t\right) \leq \sigma\left(0 < f_X < ut\right) + \sigma\left(f_X - \widehat{f_X} > (u-1)t\right).$$

Note that on the one hand $\pi = 1$ is good for $\sigma\left(0 < \widehat{f_X} < t\right)^{1/\pi}$ to be as small as possible, but a multiplicative factor $2^{J(d-1)(1-1/p)}$ is paid. On the other hand choosing $\pi = p$ implies a multiplicative factor equal to 1. Thus, based on the upper bound, the best choice for π and t depends on the smoothness of $\widehat{f_X}$ and the sample size of the first sample, as well as the function $u \mapsto \sigma(0 < f_X < u)$.

6 Proof of Theorem 7

Let us prove two lower bounds. They yield the lower bounds in the dense and sparse zone. We conclude by checking for which value of the parameters one rate is larger than the other one.

6.1 Proof of the lower bound in the dense zone

Consider a set of measures $(P_m)_{m=0}^M$ indexed by a finite family of densities $(f_m)_{m=0}^M$ which are the distributions of an n i.i.d. sample of (Y, X) when $f_\beta = f_m$ and for a given f_X. The tower property of the conditional expectation yield that the Kullback-Leibler divergence between two measures P_m and P_0 is given by

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

It is easy to check that

$$\begin{aligned}
K(P_m, P_0) &\leq n\mathbb{E}_{f_X}\left[\mathcal{H}(f_m)(X)\left(\frac{\mathcal{H}(f_m)(X) - \mathcal{H}(f_0)(X)}{\mathcal{H}(f_0)(X)}\right) + (1 - \mathcal{H}(f_m)(X))\left(\frac{\mathcal{H}(f_0)(X) - \mathcal{H}(f_m)(X)}{1-\mathcal{H}(f_0)(X)}\right)\right] \\
&= n\mathbb{E}_{f_X}\left[\frac{\mathcal{H}(f_m - f_0)(X)^2}{\mathcal{H}(f_0)(X)} + \frac{\mathcal{H}(f_m - f_0)(X)^2}{1-\mathcal{H}(f_0)(X)}\right] \\
&= n\mathbb{E}_{f_X}\left[\frac{\mathcal{H}(f_m - f_0)(X)^2}{\mathcal{H}(f_0)(X)\left(1-\mathcal{H}(f_0)(X)\right)}\right].
\end{aligned}$$

The general reduction scheme together with the Corollary 2.6 of the Fano lemma from [28] yield:

Lemma 11 *If, for $\alpha \in (0,1)$, some positive integer $\mathcal{M}$*

(i) $f_m \in B^s_{r,q}(M)$ *for* $m = 1, \dots, \mathcal{M}$,

(ii) $\forall m \neq l,\ \|f_m - f_l\|_p \geq 2h > 0$,

(iii) $\frac{1}{\mathcal{M}+1}\sum_{m=1}^{\mathcal{M}} K(P_m, P_0) \leq \alpha \log \mathcal{M}$

then

$$\forall z \geq 1,\ \inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)} \mathbb{E}\left\|\widehat{f_\beta} - f_\beta\right\|_p^z \geq h^{-z}\left(\frac{\log(\mathcal{M}+1) - \log 2}{\log \mathcal{M}} - \alpha\right).$$

We take the indices m that correspond to vectors of 0 and 1 of size $|A_j|$. We consider the family

$$f_m = \frac{1}{|\mathbb{S}^{d-1}|} + \gamma \sum_{\xi \in A_j} m_\xi \psi_{j,\xi}$$

where γ is small enough to guarantee the positivity of f_m for all m and the fact that for one of them, f_0, corresponding to a vector m_0, $\forall x \in H^+$, $\left|\mathcal{H}(f_0^-)(x)\right| \leq c_b$ for some $c_b \in (0, \frac{1}{2})$. Because $\mathcal{H}(f_0) = \frac{1}{2} + \mathcal{H}(f_0^-)$, the last condition implies that $\mathcal{H}(f_0)(x)\,(1 - \mathcal{H}(f_0)(x)) \geq (\frac{1}{2} - c_b)^2 > 0$. This yields a family $\mathcal{A}_j$ of functions of cardinality $2^{\lfloor c_A 2^{j(d-1)} \rfloor}$. The Varshamov-Guilbert bound (see, *e.g.*, [28]) yields that there exists a subset $A'_j \subset A_j$ such that $\forall (m_1, m_2) \in \left(\{0,1\}^{A'_j}\right)^2$, $\sum_{\xi \in A'_j} |m_{1,\xi} - m_{2,\xi}| > \frac{c_A}{8} 2^{j(d-1)}$. We denote the corresponding family of functions

$$f_m = \frac{1}{|\mathbb{S}^{d-1}|} + \gamma \sum_{\xi \in A'_j} m_\xi \psi_{j,\xi}$$

by $\mathcal{A}'_j$, it is of cardinality $\mathcal{M} \geq 2^{\lfloor c_A 2^{j(d-1)} \rfloor / 8}$. When $p = \infty$, we work with the whole family $\mathcal{A}_j$.
$|\gamma| \leq 2^{-j(s+(d-1)/2)} M$ implies that $\forall f_m \in \mathcal{A}_j$, $f_m \in B^s_{r,q}(M)$. Take $|\gamma| \gtrsim 2^{-j(s-(d-1)/2)}$ as well. Indeed, when $r < \infty$,

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

It is straightforward to check that the same condition is also sufficient when $r = \infty$.
Lemma 1 (ii) yields that for $p \in [1, \infty)$, m_1 and m_2 in $\{0,1\}^{A'_j}$,

$$\|f_{m_1} - f_{m_2}\|_p \geq |\gamma| c''_{p,A} 2^{j(d-1)(1/2-1/p)} \left(\frac{c_A}{8} 2^{j(d-1)}\right)^{1/p} = c''_{p,A} \left(\frac{c_A}{8}\right)^{1/p} M 2^{-js} \triangleq 2h$$

while for $p = \infty$, m_1 and m_2 in $\{0,1\}^{A_j}$,

$$\|f_{m_1} - f_{m_2}\|_\infty \geq |\gamma| c''_{p,A} 2^{j(d-1)(1/2-1/p)} \left(\frac{c_A}{8} 2^{j(d-1)}\right)^{1/p} = c''_{\infty,A} M 2^{-js} \triangleq 2h_{\mathcal{A}}.$$

For m_0 and every m in $\{0,1\}^{A'_j}$, we get

$$\begin{aligned}
K(P_m, P_0) &\leq \left(\frac{1}{2} - c_b\right)^{-2} \|f_X\|_\infty n \, \|\mathcal{H}(f_m - f_0)\|_2^2 \\
&\leq \left(\frac{1}{2} - c_b\right)^{-2} \|f_X\|_\infty n 2^{-2(j-2)\nu} \|f_m - f_0\|_2^2 \quad \text{(from (10), writing the squared L}^2\text{-norm} \\
&\qquad \text{as the sum of the squared L}^2\text{-norm on the spaces } H^{k,d} \text{ for } k = 2^{j-2}+1, \ldots, 2^j - 1) \\
&\leq (C''_2)^2 \left(\frac{1}{2} - c_b\right)^{-2} \|f_X\|_{\mathrm{L}^\infty(H^+)} n 2^{-2(j-2)\nu} \gamma^2 \left\| (m_\xi - m_{0,\xi})_{\xi \in A'_j} \right\|_{\ell^2}^2 \quad \text{(Lemma 1 (i))} \\
&\leq (C''_2)^2 \left(\frac{1}{2} - c_b\right)^{-2} \|f_X\|_{\mathrm{L}^\infty(H^+)} n 2^{-2(j-2)\nu} \gamma^2 \left\| (m_\xi - m_{0,\xi})_{\xi \in A'_j} \right\|_{\ell^1} \\
&\leq (C''_2)^2 \left(\frac{1}{2} - c_b\right)^{-2} 2^{4\nu} \|f_X\|_{\mathrm{L}^\infty(H^+)} C_{\Xi}^2 n 2^{j(d-1-2\nu)} \gamma^2
\end{aligned}$$

$$\leq (C_2'')^2 \left(\frac{1}{2} - c_b\right)^{-2} 2^{4\nu} \|f_X\|_{\mathrm{L}^\infty(H^+)} C_\Xi^2 n 2^{-2j(s+\nu)}$$

In the $p = \infty$ case we replace A_j' by A_j above. C_Ξ is an upper bound and we can replace by the constants for A_j and A_j'. Condition (iii) of Lemma 11 is satisfied once

$$n\|f_X\|_{\mathrm{L}^\infty(H^+)} 2^{-2j(s+\nu+(d-1)/2)} \leq \frac{\alpha c_A (\log 2) \left(\frac{1}{2} - c_b\right)^2}{2^{4\nu+3} C_\Xi^2}.$$

The larger h or h_∞ above is obtained for the smaller j in the above condition thus $2^j \simeq (n\|f_X\|_\infty)^{1/2(s+\nu+(d-1)/2)}$. Lemma 11 now yields for every $p \in [1, \infty]$ and $z \geq 1$,

$$\forall z \geq 1, \inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)} \mathbb{E} \left\|\widehat{f_\beta} - f_\beta\right\|_p^z \gtrsim \left(\frac{1}{\sqrt{n\|f_X\|_{\mathrm{L}^\infty(H^+)}}}\right)^{\frac{sz}{s+\nu+(d-1)/2}}.$$

6.2 Proof of the lower bound in the sparse zone

Consider now the hypotheses

$$f_\xi = \frac{1}{|\mathbb{S}^{d-1}|} + \gamma \psi_{j,\xi}$$

where ξ belongs to 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_ξ that we denote f_0, $\forall x \in H^+$, $\left|\mathcal{H}(f_0^-)(x)\right| \leq c_b$ with $c_b \in (0, \frac{1}{2})$.

We denote by P_ξ the distributions of an n i.i.d. sample of (Y, X) when $f_\beta = f_\xi$ and for a given f_X. Here $\mathcal{M}$ is the cardinality of A_j thus $\mathcal{M} \simeq 2^{j(d-1)}$. Like in [29], we make use of the following lemma from [22]. We denote by $\Lambda(P_\xi, P_0)$ the likelihood ratio. Recall that $K(P_\xi, P_0) = \mathbb{E}_{P_\xi}[\Lambda(P_\xi, P_0)]$.

Lemma 12 *If, for $\pi_0 > 0$ and some positive integer $\mathcal{M}$*

- *(i)* $f_m \in B^s_{r,q}(M)$ *for* $m = 1, \ldots, \mathcal{M}$,
- *(ii)* $\forall m \neq l$, $\|f_m - f_l\|_p \geq 2h > 0$,
- *(iii)* $\forall m = 1, \ldots, \mathcal{M}$, $\Lambda(f_0, f_m) = \exp(z_n^m - v_n^m)$, *where* z_n^m *are random variables and* v_n^m *constants such that* $\mathbb{P}(z_n^m > 0) \geq \pi_0$ *and* $\exp\left(\sup_{m=1,\ldots,\mathcal{M}} v_n^m\right) \leq \mathcal{M}$,

then

$$\forall z \geq 1, \inf_{\widehat{f_\beta}} \sup_{f_\beta \in B^s_{r,q}(M)} \mathbb{E} \left\|\widehat{f_\beta} - f_\beta\right\|_p^z \geq \frac{h^{-z}\pi_0}{2}.$$

(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 $|\gamma| \lesssim 2^{-j(s-(d-1)(1/r-1/2))}$. h in (ii) is obtained as follows, if ξ and ξ' belong to A_i,

$$\|f_\xi - f_{\xi'}\|_p = |\gamma| \|\psi_{j,\xi} - \psi_{j,\xi'}\|_p$$

$$
\begin{aligned}
&\geq |\gamma| c''_{p,A} 2^{j(d-1)(1/2-1/p)} \\
&\gtrsim 2^{-j(s-(d-1)(1/r-1/p))}.
\end{aligned}
$$

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

Thus, condition (iii) is satisfied when

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

for $\alpha \in (0,1)$. The same computations as in the beginning of Section 4.1 yield that we need to impose $n 2^{-2j\nu}\gamma^2 \lesssim j$, thus

$$
\|f_X\|_{\mathrm{L}^\infty(H^+)} n 2^{-2j(s+\nu-(d-1)(1/r-1/2))} \lesssim j.
$$

We can check that it is possible to take

$$
2^j \simeq \left(\frac{n\|f_X\|_{\mathrm{L}^\infty(H^+)}}{\log\left(n\|f_X\|_{\mathrm{L}^\infty(H^+)}\right)}\right)^{\frac{1}{2(s+\nu-(d-1)(1/r-1/2))}}
$$

which yields the desired rate.

7 Proof of Theorem 8

7.1 A preliminary decomposition

We know from [8] that for all p in $[1,\infty]$

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

We also use that for $z \in [1,\infty)$,

$$
\left\|\widehat{f_\beta^-}^{I,a,\rho} - f_\beta^-\right\|_p^z \leq 2^{z-1}\left(\left\|\widehat{f_\beta^-}^{I,a,\rho} - f_\beta^{-\,I,a,J}\right\|_p^z + \left\|f_\beta^{-\,I,a,J} - f_\beta^-\right\|_p^z\right). \tag{34}
$$

The second term is the approximation error. Let us focus on the first term which corresponds to the error in the high dimensional space.

Lemma 1 (i) yields

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

When $p = \infty$, we thus have

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

while for $p < \infty$, we obtain

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

the last inequality is obtained using the fact that when $p \ge z$,

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

while by the Hölder inequality when $p \le z$,

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

Note that for the case $p < \infty$ the inequality is sharp if and only if $z = p$.

7.2 Coefficientwise analysis

For the simplicity of the notations we will sometimes drop the dependence on γ in the sets of indices. We first focus on

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

By construction,

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

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 will also use $T_{j,\xi,\gamma}^{b,-}$, $T_{j,\xi,\gamma}^{s,+}$ and $\Delta_{j,\xi,\gamma}^+$ that are respectively, with high probability, deterministic lower bound, upper bound and upper bound of the previous quantities. This yields

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

$$
\begin{aligned}
&\leq \max\Bigg(\left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right),\\
&\qquad \left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^b_{j,\xi,\gamma}}\right)\Bigg)\\
&\leq \max\Bigg(\left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}}, \mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right),\\
&\qquad \left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}, \mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}\right)\Bigg)
\end{aligned}
$$

and sorting them according to the number of random terms

$$
\begin{aligned}
\delta_{j,\xi,z} \leq \max\Bigg(&\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}}, \left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right),\\
&\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}, \left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}, \mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}\right)\Bigg)
\end{aligned}
$$

7.3 Scalewise analysis

Defining $M_{j,z}$ as

$$
M_{j,z} = \sup_{\xi\in\Xi_j}\left|\rho_{T_{j,\xi,\gamma}}\left(\widehat{\beta}^{I,a}_{j,\xi}\right) - \beta^a_{j,\xi}\right|^z = \sup_{\xi\in\Xi_j}\delta_{j,\xi,z}
$$

and $S_{j,z}$ as

$$
S_{j,z} = \sum_{\xi\in\Xi_j}\left|\rho_{T_{j,\xi,\gamma}}\left(\widehat{\beta}^{I,a}_{j,\xi}\right) - \beta^a_{j,\xi}\right|^z = \sum_{\xi\in\Xi_j}\delta_{j,\xi,z}
$$

we obtain

$$
\begin{aligned}
M_{j,z} &\leq \max\Bigg(\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}}, \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right),\\
&\qquad \sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}, \sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\Bigg)\\
&\triangleq \max(M^{S0}_{j,z}, M^{S1}_{j,z}, M^{B1}_{j,z}, M^{B2}_{j,z}) \leq M^{S0}_{j,z} + M^{S1}_{j,z} + M^{B1}_{j,z} + M^{B2}_{j,z}\\
S_{j,z} &\leq \sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}} + \sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\\
&\quad + \sum_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}} + \sum_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\\
&\triangleq S^{S0}_{j,z} + S^{S1}_{j,z} + S^{B1}_{j,z} + S^{B2}_{j,z}.
\end{aligned}
$$

We can bound the expectations of each term as follows

$$
\mathbb{E}\left[M^{S0}_{j,z}\right] = \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\leq T^{s,+}_{j,\xi,\gamma}}
$$

$$\mathbb{E}\left[M^{S1}_{j,z}\right] \le \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbb{E}\left[\sup_{\xi\in\Xi_j}\max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\right]$$

$$\le \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \left(\mathbb{P}\left\{\exists\xi\in\Xi_j,\ T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}\right\}+\mathbb{P}\left\{\exists\xi\in\Xi_j,\ \left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\right)$$

$$\mathbb{E}\left[M^{B1}_{j,z}\right] \le \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]$$

Using the Hölder inequality with $\tau > 1$ to be specified later

$$\mathbb{E}\left[M^{B2}_{j,z}\right] \le \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^{z\tau}\right]^{1/\tau} \mathbb{E}\left[\sup_{\xi\in\Xi_j}\max\left(\mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\right]^{1-1/\tau}$$

$$\le \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^{z\tau}\right]^{1/\tau} \left(\mathbb{P}\left\{\exists\xi\in\Xi_j,\ T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}\right\}+\mathbb{P}\left\{\exists\xi\in\Xi_j,\ \left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\right)^{1-1/\tau}$$

$$\mathbb{E}\left[S^{S0}_{j,z}\right] = \sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}$$

$$\mathbb{E}\left[S^{S1}_{j,z}\right] = \sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbb{E}\left[\max\left(\mathbf{1}_{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\right]$$

$$= \sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \left(\mathbb{P}\left\{T^{s,+}_{j,\xi,\gamma}<T^s_{j,\xi,\gamma}\right\}+\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\right)$$

$$\mathbb{E}\left[S^{B1}_{j,z}\right] = \sum_{\xi\in\Xi_j}\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}$$

$$\le \sum_{\xi\in\Xi_j}\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}$$

$$\mathbb{E}\left[S^{B2}_{j,z}\right] = \sum_{\xi\in\Xi_j}\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z \max\left(\mathbf{1}_{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}}, \mathbf{1}_{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}}\right)\right]$$

using the Hölder inequality with $\tau > 1$ to be specified later

$$\le \sum_{\xi\in\Xi_j}\left(\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^{z\tau}\right]\right)^{1/\tau}\left(\mathbb{P}\left\{T^{b,-}_{j,\xi,\gamma}>T^b_{j,\xi,\gamma}\right\}+\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|>\Delta_{j,\xi,\gamma}\right\}\right)^{1-1/\tau}.$$

The following concentration inequalities allow to control the stochastic terms appearing in those bounds.

7.4 Concentration inequalities

7.4.1 Bernstein inequality and the $\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z$ terms

We denote by

$$\left(\sigma^I_{j,\xi}\right)^2 = \mathbb{E}\left[\left(G^I_{j,\xi}(X_i,Y_i)-\beta^a_{j,\xi}\right)^2\right]$$

the variance of $G^I_{j,\xi}(X_i,Y_i)$, if $\sigma^I_{j,\xi} > 0$.

Lemma 13 *For any c_σ and c_M positive*

$$\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi}}\right)^z\right] \leq 2\left(c_{2,z}\left(\frac{2}{\sqrt{n}}\frac{1}{c_\sigma+c_M\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}}\right)^z + c_{1,z}\left(\frac{4}{3n}\frac{1}{c_\sigma\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M}\right)^z\right) \tag{35}$$

Proof. The Bernstein inequality yields

$$\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|\geq u\right\}\leq 2e^{-\frac{nu^2}{2\left(\left(\sigma^I_{j,\xi}\right)^2+M^I_{j,\xi}u/3\right)}}$$

setting $u=\tau(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi}$ yields

$$\begin{aligned}
\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|\geq \tau(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi})\right\} &\leq 2e^{-\frac{n\tau^2(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi})^2}{2\left(\left(\sigma^I_{j,\xi}\right)^2+M^I_{j,\xi}\tau(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi})/3\right)}}\\
&\leq 2\left(e^{-\frac{\tau^2 n(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi})^2}{4\left(\sigma^I_{j,\xi}\right)^2}}+e^{-\frac{\tau 3n(c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi})}{4M^I_{j,\xi}}}\right)\\
&\leq 2\left(e^{-\frac{1}{4}n\left(c_\sigma+c_M\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2}+e^{-\frac{3}{4}n\left(c_\sigma\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau}\right).
\end{aligned}$$

We use now

$$\mathbb{E}\left[|X|^z\right]=\int_{\mathbb{R}+}zu^{z-1}\mathbb{P}\{|X|>u\}du$$

and the upper bounds (25) and (26) to derive

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi}}\right)^z\right] &\leq \int_{\mathbb{R}+}z\tau^{z-1}\mathbb{P}\left\{\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi}}\geq\tau\right\}d\tau\\
&\leq \int_{\mathbb{R}+}z\tau^{z-1}2\left(e^{-\frac{1}{4}n\left(c_\sigma+c_M\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2}+e^{-\frac{3}{4}n\left(c_\sigma\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau}\right)d\tau
\end{aligned}$$

this yields (35) □

Taking $c_\sigma=1$ and $c_M=0$ yields

$$\mathbb{E}\left[\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\right]\leq 2\left(\left(2c^{1/z}_{2,z}\frac{\sigma^I_{j,\xi}}{\sqrt{n}}\right)^z+\left(\frac{4}{3}c^{1/z}_{1,z}\frac{M^I_{j,\xi}}{n}\right)^z\right) \tag{36}$$

while taking $c_\sigma = c'_\sigma \sqrt{\log n/n}$ and $c_M = c'_M \log n/(n-1)$ we obtain

$$\mathbb{E}\left[\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|}{c'_\sigma\sqrt{\log n}\frac{\sigma^I_{j,\xi}}{\sqrt{n}} + c'_M \log n \frac{M^I_{j,\xi}}{n-1}}\right)^z\right] \le 2\left(c_{2,z}\left(2\frac{1}{c'_\sigma\sqrt{\log n} + c'_M \frac{\sqrt{n}\log n}{n-1}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}}\right)^z + c_{1,z}\left(\frac{4}{3}\frac{1}{c'_\sigma\sqrt{n}\sqrt{\log n}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}} + c'_M \log n \frac{n}{n-1}}\right)^z\right)$$

$$\le 2\left(c_{2,z}\left(2\frac{1}{c'_\sigma\sqrt{\log n}}\right)^z + c_{1,z}\left(\frac{4}{3}\frac{1}{c'_M \log n}\right)^z\right) \tag{37}$$

The following lemma is useful for the $p=\infty$ case.

Lemma 14 *For any* $\Xi'_j \subset \Xi_j$,

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi} + c_M M^I_{j,\xi}}\right)^z\right] \le \left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{1}{c_\sigma + c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}}\right)^z\left(2 + \left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)\right)^{z/2}\right) + \left(\frac{8}{3n}\frac{1}{c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}} + c_M}\right)^z\left(2 + \left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right) \tag{38}$$

Proof. A uniform union bound yields

$$\mathbb{P}\left\{\sup_{\xi\in\Xi'_j}\frac{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi} + c_M M^I_{j,\xi}} \ge \tau\right\} \le \min\left(1, \left|\Xi'_j\right| 2\left(e^{-\frac{1}{4}n\left(c_\sigma + c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2} + e^{-\frac{3}{4}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}} + c_M\right)\tau}\right)\right)$$

$$\le \min\left(1, \left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma + c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2}\right) + \min\left(1, \left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}} + c_M\right)\tau}\right).$$

We thus derive, for any τ_1 and τ_2 positive,

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi} + c_M M^I_{j,\xi}}\right)^z\right] \le \int_{\mathbb{R}^+} z\tau^{z-1}\min\left(1, \left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma + c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2}\right)d\tau$$

$$+\int_{\mathbb{R}^+} z\tau^{z-1}\min\left(1,\left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau}\right)d\tau$$

$$\leq \tau_2^z + \int_{\tau\geq\tau_2} z\tau^{z-1}\left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2} d\tau$$

$$+\tau_1^z + \int_{\tau\geq\tau_1} z\tau^{z-1}\left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau} d\tau.$$

Let

$$\tau_1 = \frac{8}{3n}\frac{\log\left(c_{1,z}\left|\Xi'_j\right|\right)}{c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M} \quad \text{and} \quad \tau_2 = \frac{2\sqrt{2}}{\sqrt{n}}\frac{\sqrt{\log\left(c_{2,z}\left|\Xi'_j\right|\right)}}{c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}},$$

by construction

$$\forall\tau\geq\tau_1, \left|\Xi'_j\right| 2e^{-\frac{3}{4}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau} \leq \frac{2}{c_{1,z}}e^{-\frac{3}{8}n\left(c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M\right)\tau}$$

$$\forall\tau\geq\tau_2, \left|\Xi'_j\right| 2e^{-\frac{1}{4}n\left(c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2} \leq \frac{2}{c_{2,z}}e^{-\frac{1}{8}n\left(c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^2\tau^2}$$

This implies

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{c_\sigma\sigma^I_{j,\xi}+c_M M^I_{j,\xi}}\right)^z\right] \leq \left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{\sqrt{\log\left(c_{2,z}\left|\Xi'_j\right|\right)}}{c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}}\right)^z + 2\left(\frac{2\sqrt{2}}{\sqrt{n}}\frac{1}{c_\sigma+c_M \inf_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}}\right)^z$$
$$+\left(\frac{8}{3n}\frac{\log\left(c_{1,z}\left|\Xi'_j\right|\right)}{c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M}\right)^z + 2\left(\frac{8}{3n}\frac{1}{c_\sigma \inf_{\xi\in\Xi'_j}\frac{\sigma^I_{j,\xi}}{M^I_{j,\xi}}+c_M}\right)^z$$

which allows to conclude □

If $c_\sigma = 1$ and $c_M = 0$ (38) reduces to

$$\mathbb{E}\left[\sup_{\xi\in\Xi'_j}\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{\sigma^I_{j,\xi}}\right)^z\right] \leq \left(\frac{2\sqrt{2}}{\sqrt{n}}\right)^z\left(2+\left(\log\left(c_{2,z}\left|\Xi'_j\right|\right)\right)^{z/2}\right)$$
$$+\left(\frac{8}{3n}\sup_{\xi\in\Xi'_j}\frac{M^I_{j,\xi}}{\sigma^I_{j,\xi}}\right)^z\left(2+\left(\log\left(c_{1,z}\left|\Xi'_j\right|\right)\right)^z\right)$$

Note that one could have used the uniform bounds M_j^I (see (24)) and

$$\sigma_{j,\xi}^I \le C_2 B(d,2) \left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} 2^{j\nu} \triangleq \sigma_j^I \tag{39}$$

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

$$\mathbb{E}\left[\sup_{\xi\in\Xi_j'} \left|\widehat{\beta}_{j,\xi}^{I,a} - \beta_{j,\xi}^a\right|^z\right] \le \left(\frac{2\sqrt{2}}{\sqrt{n}}\sigma_j^I\right)^z \left(2 + \left(\log\left(c_{2,z}\left|\Xi_j'\right|\right)\right)^{z/2}\right) + \left(\frac{8}{3n}M_j^I\right)^z \left(2 + \left(\log\left(c_{1,z}\left|\Xi_j'\right|\right)\right)^z\right) \tag{40}$$

Along the same lines, with $c_\sigma = c_\sigma' t_n$ and $c_M = c_M' \log n/(n-1)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\xi\in\Xi_j'} \left(\frac{\left|\widehat{\beta}_{j,\xi}^{I,a} - \beta_{j,\xi}^a\right|}{c_\sigma'\sqrt{\log n}\frac{\sigma_{j,\xi}^I}{\sqrt{n}} + c_M' \log n \frac{M_{j,\xi}^I}{n-1}}\right)^z\right] \\
&\quad\le \left(\frac{2\sqrt{2}}{c_\sigma'\sqrt{\log n} + c_M' \log n/\sqrt{n-1}\inf_{\xi\in\Xi_j'}\frac{M_{j,\xi}^I}{\sigma_{j,\xi}^I}}\right)^z \left(2 + \left(\log\left(c_{2,z}\left|\Xi_j'\right|\right)c_{1,z}\right)^{z/2}\right) \\
&\qquad + \left(\frac{8/3}{c_\sigma'\sqrt{n\log n}\inf_{\xi\in\Xi_j'}\frac{\sigma_{j,\xi}^I}{M_{j,\xi}^I} + c_M'\log n}\right)^z \left(2 + \left(\log\left(c_{1,z}\left|\Xi_j'\right|\right)\right)^z\right) \\
&\quad\le \left(\frac{2\sqrt{2}}{c_\sigma'\sqrt{\log n}}\right)^z \left(2 + \left(\log\left(c_{2,z}\left|\Xi_j'\right|\right)\right)^{z/2}\right) + \left(\frac{8}{3c_M'\log n}\right)^z \left(2 + \left(\log\left(c_{1,z}\left|\Xi_j'\right|\right)\right)^z\right)
\end{aligned} \tag{41}$$

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

7.4.2 Empirical Bernstein and the probabilities

We define

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

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

Lemma 15 *The following upper bounds hold*

$$\mathbb{P}\left\{T^{b,-}_{j,\xi,\gamma} > T^{b}_{j,\xi,\gamma}\right\} \leq \frac{1}{n^\gamma}$$
$$\mathbb{P}\left\{T^{s,+}_{j,\xi,\gamma} < T^{s}_{j,\xi,\gamma}\right\} \leq \frac{1}{n^\gamma}$$
$$\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^{a}_{j,\xi}\right| > \Delta_{j,\xi,\gamma}\right\} \leq \frac{3}{n^\gamma},$$

$$\mathbb{P}\left\{\exists \xi \in \Xi_j :\ T^{s,+}_{j,\xi,\gamma} < T^{s}_{j,\xi,\gamma}\right\} \leq \sum_{\xi\in\xi_j} \mathbb{P}\left\{\ T^{s,+}_{j,\xi,\gamma} < T^{s}_{j,\xi,\gamma}\right\} \leq C_\Xi 2^{j(d-1)}\frac{1}{n^\gamma}$$
$$\mathbb{P}\left\{\exists \xi \in \Xi_j :\ T^{b,-}_{j,\xi,\gamma} > T^{b}_{j,\xi,\gamma}\right\} \leq \sum_{\xi\in\xi_j} \mathbb{P}\left\{T^{b,-}_{j,\xi,\gamma} > T^{b}_{j,\xi,\gamma}\right\} \leq C_\Xi 2^{j(d-1)}\frac{1}{n^\gamma}$$
$$\mathbb{P}\left\{\exists \xi \in \Xi_j :\ \left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^{a}_{j,\xi}\right| > \Delta_{j,\xi,\gamma}\right\} \leq \sum_{\xi\in\xi_j} \mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^{a}_{j,\xi}\right| > \Delta_{j,\xi,\gamma}\right\} \leq C_\Xi 2^{j(d-1)}\frac{3}{n^\gamma}.$$

Proof. Using the results of [25] we get

$$\mathbb{P}\left\{\sigma^I_{j,\xi} > \widehat{\sigma}^I_{j,\xi} + 2\sqrt{2u}\frac{M^I_{j,\xi}}{\sqrt{n-1}}\right\} \leq e^{-u}$$
$$\mathbb{P}\left\{\sigma^I_{j,\xi} < \widehat{\sigma}^I_{j,\xi} - 2\sqrt{2u}\frac{M^I_{j,\xi}}{\sqrt{n-1}}\right\} \leq e^{-u}$$
$$\mathbb{P}\left\{\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^{a}_{j,\xi}\right| > \sqrt{2u}\frac{\widehat{\sigma}_{j,\xi}}{\sqrt{n}} + \frac{14}{3}M^I_{j,\xi}\frac{u}{n-1}\right\} \leq 3e^{-u}$$

which yield the first inequalities. The second set of inequalities are obtained using a union bound ☐

7.5 The $p = \infty$ case

7.5.1 Error in the high dimensional space

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

with

$$\mathbb{E}\left[M^{S0}_{j,z}\right] = \sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right| \leq T^{s,+}_{j,\xi,\gamma}}$$
$$\mathbb{E}\left[M^{S1}_{j,z}\right] \leq C_\Xi 2^{j(d-1)}\frac{4}{n^\gamma}\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z$$
$$\mathbb{E}\left[M^{B1}_{j,z}\right] \leq \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi} - \beta^{a}_{j,\xi}\right|^z \mathbf{1}_{\left|\beta^a_{j,\xi}\right| > T^{b,-}_{j,\xi,\gamma}}\right]$$

$$\mathbb{E}\left[M_{j,z}^{B2}\right] \leq \left(C_\Xi 2^{j(d-1)}\frac{4}{n^\gamma}\right)^{1-1/\tau} \left(\left(\frac{2\sqrt{2}}{\sqrt{n}}\sigma_j^I\right)^z \left(2^{1/\tau} + \left(\sqrt{\log\left(|\Xi_j|\, c_{1,z\tau}\right)}\right)^z\right) + \left(\frac{8}{3n}M_j^I\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_j|\, c_{1,z\tau}\right)\right)^z\right)\right)$$

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

This yields

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

7.5.2 The $R'_{1,\infty,z}$ and $R'_{2,\infty,z}$ terms

The $R'_{1,\infty,z}$ is exactly the term appearing in Theorem 8 and thus we only need to bound $R'_{2,\infty,z}$.

As in the $p < \infty$ case, one can plug the uniform bounds on $\sigma_{j,\xi}^I$ and $M_{j,\xi}^I$ as well as the bounds $|\Xi_j| \leq |\Xi_J|$ to obtain

$$\begin{aligned}
R'_{2,\infty,z} &\leq \left(C_\Xi\frac{4}{n^\gamma}\right)^{1-1/\tau} \sum_{j=0}^J 2^{j(d-1)(z/2+1-1/\tau)} \\
&\quad\times \left(\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2B(d,2)2^{j\nu}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(c_{2,z\tau}\,|\Xi_J|\right)\right)^{z/2}\right)\right. \\
&\qquad\left. + \left(\frac{8}{3n}C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\right)^z \left(2^{1/\tau} + \left(\log\left(c_{1,z\tau}\,|\Xi_J|\right)\right)^z\right)\right) \\
&\leq \left(C_\Xi\frac{4}{n^\gamma}\right)^{1-1/\tau}\left[\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2B(d,2)\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_J|\, c_{2,z\tau}\right)\right)^{z/2}\right)\sum_{j=0}^J 2^{j(\nu z+(d-1)(z/2+1-1/\tau))}\right. \\
&\qquad\left. + \left(\frac{8}{3n}C_\infty B(d,\infty)\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\right)^z \left(2^{1/\tau} + \left(\log\left(|\Xi_J|\, c_{1,z\tau}\right)\right)^z\right)\sum_{j=0}^J 2^{j(\nu z+(d-1)(z+1-1/\tau))}\right] \\
&\leq \left(C_\Xi\frac{4}{n^\gamma}\right)^{1-1/\tau}\left[\left(\frac{2\sqrt{2}}{\sqrt{n}}C_2B(d,2)\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\right)^z \left(2^{1/\tau} + \left(\log\left(c_{2,z\tau}\,|\Xi_J|\right)\right)^{z/2}\right)\frac{2^{J(\nu z+(d-1)(z/2+1-1/\tau))}}{1-2^{-(\nu z+(d-1)(z/2+1-1/\tau))}}\right.
\end{aligned}$$

$$+\left(\frac{8}{3n}C_\infty B(d,\infty)\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\right)^z\left(2^{1/\tau}+\left(\log\left(c_{1,z\tau}\left|\Xi_J\right|\right)\right)^z\right)\frac{2^{J(\nu z+(d-1)(z+1-1/\tau))}}{1-2^{-(\nu z+(d-1)(z+1-1/\tau))}}\Bigg]$$

7.5.3 The $O'_{\infty,z}$ term

Denote by

$$O'_{z,j}=\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}}\right]$$

Because $T^{s,++}_{j,\xi,\gamma}\ge T^{s,+}_{j,\xi,\gamma}$, we get

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

thus

$$O'_{z,j}\le\left(1+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left(\frac{\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|}{T^{b,-}_{j,\xi,\gamma}}\right)^z\right]\right)\sup_{\xi\in\Xi_j}\left\{\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,++}_{j,\xi,\gamma}}\right\}+\mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|>T^{s,++}_{j,\xi,\gamma}}\right].$$

Using now (41) with $c'_\sigma=\sqrt{2\gamma}$ and $c'_M=\frac{2}{3}\gamma$, we get as $|\Xi_j|\le C_\Xi 2^{j(d-1)}$ gives the upper bound in Theorem 8.

Remark that using (41) with $\Xi'_j=\Xi_j$ is rough since the sup could be taken on the much smaller subset $\Xi'_j=\left\{\xi\in\Xi_j:\ T^{s,++}_{j,\xi,\gamma}\ge\left|\beta^a_{j,\xi}\right|>T^{b,-}_{j,\xi,\gamma}\right\}$.

7.6 The $p<\infty$ case

7.6.1 Error in the high dimensional space

We obtain

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

with

$$\mathbb{E}\left[S^{S0}_{j,z}\right]=\sum_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z\mathbf{1}_{\left|\beta^a_{j,\xi}\right|\le T^{s,+}_{j,\xi,\gamma}}$$

$$\mathbb{E}\left[S_{j,z}^{S1}\right] \leq \frac{4}{n^\gamma}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z$$

$$\mathbb{E}\left[S_{j,z}^{B1}\right] \leq \sum_{\xi\in\Xi_j}\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^{I,a}-\beta_{j,\xi}^a\right|^z\right]\mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{b,-}}$$

$$\mathbb{E}\left[S_{j,z}^{B2}\right] \leq \frac{4^{1-1/\tau}}{n^{\gamma(1-1/\tau)}}\sum_{\xi\in\Xi_j}2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_{j,\xi}^I}{\sqrt{n}}\right)^z+\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_{j,\xi}^I}{n}\right)^z\right)$$

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

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

7.6.2 The $R_{1,p,z}$ and $R_{2,p,z}$ terms

The $R_{1,p,z}$ term appears as is in Theorem 8.

To bound the term $R_{2,p,z}$, we rely on the uniform bounds M_j^I in (24) and σ_j^I in (39). We obtain

$$\begin{aligned}&\sum_{\xi\in\Xi_j}2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_j^I}{\sqrt{n}}\right)^z+\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_j^I}{n}\right)^z\right)\\
&\quad\leq\sum_{\xi\in\Xi_j}2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)2^{j\nu}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\frac{1}{\sqrt{n}}\right)^z\\
&\qquad+\sum_{\xi\in\Xi_j}2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\frac{1}{n}\right)^z\\
&\quad\leq C_\Xi 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z/2}\frac{1}{n^{z/2}}2^{j((d-1)+z\nu)}\\
&\qquad+C_\Xi 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^z\frac{1}{n^z}2^{j((d-1)+z(\nu+(d-1)/2))}\end{aligned}$$

thus

$$\sum_{j=0}^{J}2^{j(d-1)z(1/2-1/(p\vee z))}\sum_{\xi\in\Xi_j}2^{1/\tau}\left(\left(2c_{2,z\tau}^{1/(z\tau)}\frac{\sigma_j^I}{\sqrt{n}}\right)^z+\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}\frac{M_j^I}{n}\right)^z\right)$$

$$\le\sum_{j=0}^{J}2^{j(d-1)z(1/2-1/(p\vee z))}C_\Xi 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z/2}\frac{1}{n^{z/2}}2^{j((d-1)+z\nu)}$$

$$+\sum_{j=0}^{J}2^{j(d-1)z(1/2-1/(p\vee z))}C_\Xi 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z}\frac{1}{n^{z}}2^{j((d-1)+z(\nu+(d-1)/2))}$$

$$\le C_\Xi 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z/2}\frac{1}{n^{z/2}}\sum_{j=0}^{J}2^{jz(\nu+(d-1)/z+(d-1)(1/2-1/(p\vee z)))}$$

$$+C_\Xi 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z}\frac{1}{n^{z}}\sum_{j=0}^{J}2^{jz(\nu+(d-1)/z+(d-1)(1-1/(p\vee z)))}$$

$$\le\frac{C_\Xi 2^{1/\tau}\left(2c_{2,z\tau}^{1/(z\tau)}C_2B(d,2)\right)^z}{1-2^{-z(\nu+(d-1)/z+(d-1)(1/2-1/(p\vee z)))}}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z/2}\frac{1}{n^{z/2}}2^{Jz(\nu+(d-1)/z+(d-1)(1/2-1/(p\vee z)))}$$

$$+\frac{C_\Xi 2^{1/\tau}\left(\frac{4}{3}c_{1,z\tau}^{1/(z\tau)}C_\infty B(d,\infty)\right)^z}{1-2^{-z(\nu+(d-1)/z+(d-1)(1-1/(p\vee z)))}}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{z}\frac{1}{n^{z}}2^{Jz(\nu+(d-1)/z+(d-1)(1-1/(p\vee z)))}\quad(\text{as } \nu=d/2).$$

7.6.3 The $O_{p,z}$ term

Denote by

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

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

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

$$\le\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^{I,a}-\beta_{j,\xi}^a\right|^z\right]\mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{s,++}}+\frac{\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^{I,a}-\beta_{j,\xi}^a\right|^z\right]}{\left(T_{j,\xi,\gamma}^{b,-}\right)^z}\left|\beta_{j,\xi}^a\right|^z\mathbf{1}_{T_{j,\xi,\gamma}^{s,++}\ge\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{b,-}},$$

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

Now using the results of Section 7.4.1 with $T_{j,\xi,\gamma}^{b,-}=\sqrt{2\gamma}t_n\sigma_{j,\xi}^I+\frac{2}{3}\gamma\frac{\log n}{n-1}M_{j,\xi}^I$

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

$$\leq 2\left(\left(\frac{\sqrt{2}c_{2,z}^{1/z}}{\sqrt{\gamma\log n}}\right)^z+\left(\frac{2c_{1,z}^{1/z}}{\gamma\log n}\right)^z\right)$$

thus

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

8 Proof of Theorem 9

The proof of this result requires to upper bound the approximation error, the $R_{1,p,z}$ and $O_{p,z}$ terms in the upper bound of Theorem 8 when $z=p$ using the prior knowledge that the unknown f_β^- belongs to the ellipsoid $B_{r,q}^s(M)$.

8.1 The $p\in[1,\infty)$ case

8.1.1 The approximation error

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

From Lemma 1 (i) and the definition of the Besov spaces as a sequence space,

$$2^{Js}\left\|\sum_{j>J}\sum_{\xi\in\Xi_j}\beta_{j,\xi}^a\psi_{j,\xi}\right\|_p\leq C_p''\sum_{j>J}2^{j(s+(d-1)(1/2-1/p))}\left\|\left(\beta_{j,\xi}^a\right)_{\xi\in\Xi_j}\right\|_{\ell^p}\leq\left\|f_\beta^-\right\|_{B_{I,q}^s}$$

which yields that

$$\left\|\sum_{j>J}\sum_{\xi\in\Xi_j}\beta_{j,\xi}^a\psi_{j,\xi}\right\|_p\leq C_p''2^{-Js}\left\|f_\beta^-\right\|_{B_{I,q}^s}\leq\begin{cases}MC_\Xi^{1/p-1/r} & \text{if } r\geq p\\ M2^{-J(s-(d-1)(1/r-1/p))} & \text{if } r\leq p\end{cases}.$$

It is enough to consider the worst case where $r\leq p$ and to check that $\frac{s-(d-1)(1/r-1/p)}{\nu+(d-1)/2}\geq\mu$ in the two zones. On the first zone $s\geq\left(\nu+\frac{d-1}{2}\right)\left(\frac{p}{r}-1\right)$ thus $s+\nu+\frac{d-1}{2}\geq\left(\nu+\frac{d-1}{2}\right)\frac{p}{r}$ which yields $\frac{s}{s+\nu+\frac{d-1}{2}}\leq\frac{s}{\left(\nu+\frac{d-1}{2}\right)\frac{p}{r}}$. Because $s>(d-1)/r$ and $p\geq r$, $s-\frac{d-1}{r}+\frac{d-1}{p}-\frac{sr}{p}=(d-1)\left(\frac{sr}{d-1}-1\right)\left(\frac{1}{r}-\frac{1}{p}\right)\geq 0$, which yields $s-(d-1)(1/r-1/p)\geq\frac{sr}{p}$ and gives the result.
On the second zone, it is straightforward, because $s>(d-1)/r$, that $\frac{s-(d-1)(1/r-1/p)}{\nu+(d-1)/2}\geq\frac{s-(d-1)(1/r-1/p)}{s+\nu-(d-1)(1/r-1/2)}$.

8.1.2 The $R_{1,p,p}$ and $R_{2,p,p}$ terms

Using Lemma 2 (iii) we obtain that

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

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

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

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, $\frac{1}{\sqrt{n}}2^{J(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2} \leq 1$ and $\frac{2^{J(d-1)}}{n}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)} \leq 1$ (it even decays to 0). 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 τ large enough such that $\gamma(1-1/\tau) > p/2$.

8.1.3 The $O_{p,p}$ term

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

$$\begin{aligned}
T_{j,\xi,\gamma}^{s,++} &= 3\sqrt{2\gamma}t_n C_2 B(d,2)2^{j\nu}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2} + 52C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\frac{\gamma\log n}{n-1} \\
&\leq 2^{j\nu}\sqrt{\gamma}t_n\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\left(3\sqrt{2}C_2B(d,2) + 52C_\infty B(d,\infty)\frac{n\sqrt{\gamma}}{n-1}\right) \quad \text{(because of the upper bound on } J) \\
&\leq 2^{j\nu}\sqrt{\gamma}t_n\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\left(3\sqrt{2}C_2B(d,2) + 104C_\infty B(d,\infty)\right) \quad (\text{for } n \geq 2) \\
&\triangleq T_{j,\gamma}^{s,++}
\end{aligned}$$

as well as the following consequence of (36)

$$\begin{aligned}
\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^{I,a} - \beta_{j,\xi}^{a}\right|^p\right] &\leq 2\left(\left(2c_{2,p}^{1/p}\frac{\sigma_j^I}{\sqrt{n}}\right)^p + \left(\frac{4}{3}c_{1,p}^{1/p}\frac{M_j^I}{n}\right)^p\right) \\
&\leq 2\left(2c_{2,p}^{1/p}C_2B(d,2)2^{j\nu}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}\frac{1}{\sqrt{n}}\right)^p \\
&\qquad + 2\left(\frac{8}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)2^{j(\nu+(d-1)/2)}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}\frac{1}{n}\right)^p
\end{aligned}$$

$$\leq 2^{jp\nu}\frac{1}{n^{p/2}}\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{p/2} 2^{p+1}\left(c_{2,p}^{1/p}C_2B(d,2)+\frac{4}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)\right)^p$$
$$\leq \frac{\left(T_{j,\gamma}^{s,++}\right)^p}{(\gamma\log n)^{p/2}}2\left(2\frac{c_{2,p}^{1/p}C_2B(d,2)+\frac{4}{3}c_{1,p}^{1/p}C_\infty B(d,\infty)}{3\sqrt{2}C_2B(d,2)+104C_\infty B(d,\infty)\left(\sqrt{\gamma}\right)}\right)^p$$
$$\leq \frac{\left(T_{j,\gamma}^{s,++}\right)^p}{(\gamma\log n)^{p/2}}2\left(\frac{\sqrt{2}}{3}c_{2,p}^{1/p}+\frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}}\right)^p$$

Let

$$C_\gamma = 3\sqrt{2}C_2B(d,2)+104C_\infty B(d,\infty)\sqrt{\gamma}$$
$$C_{\sigma,p} = 2^{1/p}\left(\frac{\sqrt{2}}{3}c_{2,p}^{1/p}+\frac{c_{1,p}^{1/p}}{78\sqrt{\gamma}}\right).$$

Now, for any $0<z<p$,

$$\sum_{\xi\in\Xi_j}\left(\left|\beta_{j,\xi}^a\right|^p\mathbf{1}_{\left|\beta_{j,\xi}^a\right|\leq T_{j,\gamma}^{s,++}}+\mathbb{E}\left[\left|\widehat{\beta}_{j,\xi}^{I,a}-\beta_{j,\xi}^a\right|^p\right]\mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\gamma}^{s,++}}\right)$$
$$\leq \sum_{\xi\in\Xi_j}\left(\left|\beta_{j,\xi}^a\right|^p\mathbf{1}_{\left|\beta_{j,\xi}^a\right|\leq T_{j,\gamma}^{s,++}}+\frac{\left(T_{j,\gamma}^{s,++}\right)^p}{(\gamma\log n)^{p/2}}C_{\sigma,p}^p\mathbf{1}_{\left|\beta_{j,\xi}^a\right|>T_{j,\xi,\gamma}^{s,++}}\right)$$
$$\leq \left(1+\frac{C_{\sigma,p}^p}{(\gamma\log n)^{p/2}}\right)\left(T_{j,\gamma}^{s,++}\right)^{p-z}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z$$
$$\leq \left(1+\frac{C_{\sigma,p}^p}{(\gamma\log n)^{p/2}}\right)\left(\sqrt{\gamma}t_n\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}C_\gamma\right)^{p-z}2^{j\nu(p-z)}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^a\right|^z.$$

We then need to sum over j and will 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 $j_0<j\leq J$. 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

$$A+B = \left(\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}t_n\right)^{p-z_1}\sum_{j=0}^{j_0}2^{j[\nu(p-z_1)+(d-1)(p/2-1)]}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^{I,a}\right|^{z_1}$$
$$+\left(\left\|\frac{1}{f_X}\right\|_{\mathrm{L}^\infty(H^+)}^{1/2}t_n\right)^{p-z_2}\sum_{j=j_0+1}^{J}2^{j[\nu(p-z_2)+(d-1)(p/2-1)]}\sum_{\xi\in\Xi_j}\left|\beta_{j,\xi}^{I,a}\right|^{z_2}$$

where we choose adequately z_1, z_2 and j_0 in the two zones. Because of Lemma 2 (i) we only consider $p\geq r$. Let us first consider the dense zone. We define

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

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

$$s = \left(\nu + \frac{d-1}{2}\right)\left(\frac{p}{\tilde{r}} - 1\right). \tag{42}$$

With $z_2 = r$, we get

$$B \le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} \sum_{j=j_0+1}^{J} 2^{j[\nu(p-r)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} \left| \beta_{j,\xi}^{I,a} \right|^r .$$

Lemma 2 (iii) gives that

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

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

$$s + (d-1)\left(\frac{1}{2} - \frac{1}{r}\right) = \frac{(d-1)p}{2\tilde{r}} - \frac{d-1}{r} + \nu \left(\frac{p}{\tilde{r}} - 1\right), \tag{43}$$

thus

$$\begin{aligned} B &\le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} \sum_{j=j_0+1}^{J} 2^{jp\left(1-\frac{r}{\tilde{r}}\right)\left(\nu+\frac{d-1}{2}\right)} D_j^r \\ &\lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} 2^{j_0 p\left(1-\frac{r}{\tilde{r}}\right)\left(\nu+\frac{d-1}{2}\right)} \end{aligned}$$

for $q \ge 1$ if $r > \tilde{r}$ and for $q \le r$ if $r = \tilde{r}$ (i.e. $s = p\left(\nu + \frac{d-1}{2}\right)\left(\frac{1}{r} - \frac{1}{p}\right)$). Taking $2^{j_0 \frac{p}{\tilde{r}}\left(\nu+\frac{d-1}{2}\right)} \simeq \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{-1}$ we get

$$B \lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\tilde{r}}$$

which is the rate that we expect in that zone.
As for A, we take $z_1 = \overline{r} < \tilde{r} \le r$, this yields, using Lemma 2 (iii),

$$\begin{aligned} A &\le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{j[\nu(p-\overline{r})+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} \left| \beta_{j,\xi}^{I,a} \right|^{\overline{r}} \\ &\lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{j[\nu(p-\overline{r})+(d-1)(p/2-1)-\overline{r}(s+(d-1)(1/2-1/\overline{r}))]} \\ &\lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} \sum_{j=0}^{j_0} 2^{jp(\nu+(d-1)/2)(1-\overline{r}/\tilde{r})} \quad \text{(using (42))} \\ &\lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} 2^{j_0 p(\nu+(d-1)/2)(1-\overline{r}/\tilde{r})} \end{aligned}$$

$$\lesssim M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\tilde{r}} \quad \text{(from the definition of } j_0\text{)}.$$

Let us now consider the sparse zone. We define by

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

in a such a way that

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

$$\tilde{r} - r = \frac{(p-r)((d-1)/2 + \nu) - rs}{s + \nu - (d-1)(1/r - 1/2)} > 0$$

and

$$s + (d-1)\left(\frac{1}{2} - \frac{1}{r}\right) = \frac{(d-1)p}{2\tilde{r}} - \frac{d-1}{\tilde{r}} + \nu\left(\frac{p}{\tilde{r}} - 1\right). \tag{44}$$

Take $z_1 = r$.

$$\begin{aligned} A &\le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} \sum_{j=0}^{j_0} 2^{j[\nu(p-r)+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} \left| \beta_{j,\xi}^{I,a} \right|^r \\ &\le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} \sum_{j=0}^{j_0} 2^{j\left[(\nu+(d-1)/2-(d-1)/p)\frac{p}{\tilde{r}}(\tilde{r}-r)\right]} D_j^r \quad \text{(using (44))} \\ &\lesssim \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-r} 2^{j_0\left[(\nu+(d-1)/2-(d-1)/p)\frac{p}{\tilde{r}}(\tilde{r}-r)\right]} M^r, \end{aligned}$$

the last inequality holds because $\nu + (d-1)/2 - (d-1)/p > 0$, indeed because we are in the sparse zone $\nu + (d-1)/2 \ge s/(p/r - 1) = sr/(p-r) \ge 2/(p-r) \ge (d-1)/p$. Taking $2^{j_0(\nu+(d-1)(1/2-1/p))\frac{p}{\tilde{r}}} \simeq \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{-1}$ yields the upper bound of the order of

$$M^r \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\tilde{r}}$$

for A.
For B we take $z_2 = \overline{r} > \tilde{r} > r$,

$$\begin{aligned} B &\le \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} \sum_{j=j_0+1}^{J} 2^{j[\nu(p-\overline{r})+(d-1)(p/2-1)]} \sum_{\xi \in \Xi_j} \left| \beta_{j,\xi}^{I,a} \right|^{\overline{r}} \\ &\lesssim \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\overline{r}} \sum_{j=j_0+1}^{J} 2^{j(\nu+(d-1)(1/2-1/p))p(r-\overline{r})/\tilde{r}} D_j^{\overline{r}} \quad \text{(using (44))} \end{aligned}$$

$$\lesssim \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\bar{r}} 2^{j_0(\nu+(d-1)(1/2-1/p))p(r-\bar{r})/\tilde{r}} M^{\bar{r}}$$
$$\lesssim \left(\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)}^{1/2} t_n \right)^{p-\tilde{r}} M^{\bar{r}}.$$

8.2 The $p = \infty$ case

We simply consider the case where $r = q = \infty$ and deduce the general case using Lemma 2 (ii).

8.2.1 The approximation error

As f belongs to $B^s_{\infty,\infty}(M)$,

$$\begin{aligned}\left\| \sum_{j>J} \sum_{\xi\in\Xi_j} \beta^a_{j,\xi}\psi_{j,\xi} \right\|_\infty &\le \sum_{j>J} \left\| \sum_{\xi\in\Xi_j} \beta^a_{j,\xi}\psi_{j,\xi} \right\|_\infty \\ &\le MC_\infty \sum_{j>J} 2^{j(d-1)/2} 2^{-j(s+(d-1)/2)} \\ &\le MC_\infty 2^{-Js}.\end{aligned}$$

From the choice of J and the fact that $\left\| \frac{1}{f_X} \right\|_{\mathrm{L}^\infty(H^+)} \ge 1$ (because supp $f_X = H^+$) we get

$$\left\| \sum_{j>J} \sum_{\xi\in\Xi_j} \beta^a_{j,\xi}\psi_{j,\xi} \right\|_\infty \le t_n^{s/(\nu+(d-1)/2)},$$

this term is negligible because $\frac{s}{\nu+(d-1)/2} \ge \frac{s}{s\nu+(d-1)/2}$.

8.2.2 The $R'_{1,\infty,z}$ and $R'_{2,\infty,z}$ terms

Using the definition of the Besov norm, we obtain that

$$R'_{1,\infty,z} \le \frac{4}{n^\gamma}(J+1)^{z-1} M^z C''^z_\infty C_\Xi \sum_{j=0}^{J} 2^{-jzs} 2^{j(d-1)}$$

thus

$$R'_{1,\infty,z} \le \frac{4}{n^\gamma} 2^{J(d-1)} (J+1)^{z-1} M^z C''^z_\infty C_\Xi.$$

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

Due to the choice of J the bracket term in the expression of $R'_{2,\infty,z}$ in Theorem 8 is less than 1, as well 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 $(\log n)^{z/2}$. Thus

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

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

8.2.3 The $O'_{\infty,z}$ term

Note that here $a_{n,\infty,z,J}$ is of the order of a constant. We shall proceed like for the $O_{p,p}$ term in Section 8.1.3. Using (40) we obtain that up to another constant (previously of the order of $1+o(1)$),

$$\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^z \mathbf{1}_{|\beta^a_{j,\xi}|\le T^{s,++}_{j,\xi,\gamma}} + \mathbb{E}\left[\sup_{\xi\in\Xi_j}\left|\widehat{\beta}^{I,a}_{j,\xi}-\beta^a_{j,\xi}\right|^z \mathbf{1}_{|\beta^a_{j,\xi}|>T^{s,++}_{j,\xi,\gamma}}\right] \lesssim \left(\sqrt{\gamma}t_n\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)}\right)^{z-\overline{z}} 2^{j\nu(z-\overline{z})}\sup_{\xi\in\Xi_j}\left|\beta^a_{j,\xi}\right|^{\overline{z}}$$

for arbitrary $\overline{z}\in[0,z]$. We thus need to upper bound

$$\begin{aligned} A+B &= \left(\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)} t_n\right)^{z-z_1}\sum_{j=0}^{j_0} 2^{j[\nu(z-z_1)+(d-1)z/2]}\sup_{\xi\in\Xi_j}\left|\beta^{I,a}_{j,\xi}\right|^{z_1}\\ &+\left(\left\|\frac{1}{f_X}\right\|^{1/2}_{\mathrm{L}^\infty(H^+)} t_n\right)^{z-z_2}\sum_{j=j_0+1}^{J} 2^{j[\nu(z-z_2)+(d-1)z/2]}\sup_{\xi\in\Xi_j}\left|\beta^{I,a}_{j,\xi}\right|^{z_2}\end{aligned}$$

for some well chosen $0\le j_0\le J$, z_1 and z_2. Because f belongs to $B^s_{\infty,\infty}(M)$, for all $\overline{z}\ge 1$,

$$\sup_{\xi\in\Xi_j}\left|\beta^{I,a}_{j,\xi}\right|^{\overline{z}} \le M^{\overline{z}}2^{-j(s+(d-1)/2)\overline{z}}.$$

The result follows using this upper bound in A and B and computing $A+B$ with $z_1=0$, j_0 such that $2^{j_0}\simeq t_n^{-1/(s+\nu+(d-1)/2)}$ and $z_2=z$.

9 Proof of Theorem 10

The proof consists in a slight modification of the proof of Theorem 9 using the decomposition

$$\left\|\widehat{f_\beta^-}^{P,a,\rho}-f_\beta^-\right\|_p^z \le 3^{z-1}\left(\left\|\widehat{f_\beta^-}^{P,a,\rho}-f_\beta^{-\,P,a,J}\right\|_p^z+\left\|f_\beta^{-\,P,a,J}-f_\beta^{-\,I,a,J}\right\|_p^z+\left\|f_\beta^{-\,I,a,J}-f_\beta^-\right\|_p^z\right),$$

and the two following lemmas.

Lemma 16

$$\forall\pi\ge1,\ \|f_\beta^{-\,P,a,J}-f_\beta^{-\,I,a,J}\|_p\le 2C_{proj}B(d,p)|\mathbb{S}^{d-1}|^{(1/p-1/\pi)_+}2^{J(\nu+(d-1)(1/\pi-1/p)_+)}\left\|\frac{f_X}{\widehat{f_X}}-1\right\|_{\mathrm{L}^\pi(H^+)} \tag{45}$$

where $a_+\triangleq\max(a,0)$.

Proof.

$$\begin{aligned}\|f_\beta^{-\,P,a,J}-f_\beta^{-\,I,a,J}\|_p &= \left\|\mathcal{H}^{-1}\left(R^{P,a,J}-R^{I,a,J}\right)\right\|_p\\ &\le B(d,p)2^{J\nu}\left\|R^{P,a,J}-R^{I,a,J}\right\|_p \quad \text{(Proposition 4)}\end{aligned}$$

$$\leq 2C_{\text{proj}}B(d,p)|\mathbb{S}^{d-1}|^{(1/p-1/\pi)_+}2^{J(\nu+(d-1)(1/\pi-1/p)_+)}\left\|R\left(\frac{f_X}{\widehat{f_X}}-1\right)\right\|_{\mathrm{L}^\pi(H^+)}$$

Conclusion follows from the L^p continuity of the smoothed projections (Lemma 2.4 (c)) and the Nikolski inequality (Proposition 2.5) of [26], the Hölder inequality and since $R^{P,a,J}$ and $R^{I,a,J}$ are odd □
The constant C_{proj} could be taken independent of p, it is enough to take the uniform upper bound on the L^1 norm of the smoothed projection kernels with respect to one of its argument according to the Young inequality (see [15]).

The following lemma is used in the analysis to relate the smoothness of the true function with that of the function with a plugged-in preliminary estimator of the density of the design.

Lemma 17 *If $f_\beta^- \in B^s_{r,q}(M)$ then, for any $\pi \geq 1$, $f_\beta^{-P,a,J} \in B^s_{r,q}(M^{P,a,J,r,\pi})$.*

A maximal resolution J should be imposed to obtain an additive term of the order of a constant, it depends on the quality of the estimation of f_X and its smallness at certain points through $\left\|\frac{f_X}{\widehat{f_X}}-1\right\|_{\mathrm{L}^\pi(H^+)}$.

Proof. As long as $j \leq J$, $\langle f_\beta^- - f_\beta^{-P,a,J}, \psi_{j,\xi}\rangle = \langle f_\beta^{-I,a,J} - f_\beta^{-P,a,J}, \psi_{j,\xi}\rangle$, thus we get, with $J=j$, using Lemma 1 (iii),

$$\begin{aligned}\left\|2^{j(s+(d-1)(1/2-1/r))}\left\|\beta^a_{j,\xi}-\beta^{P,a}_{j,\xi}\right\|_{\ell^r}\right\|_{\ell^q(\{0,\dots,J\})} &\leq C_p'''\left\|2^{js}\left\|f_\beta^{-P,a,j}-f_\beta^{-I,a,j}\right\|_r\right\|_{\ell^q(\{0,\dots,J\})}\\ &\leq C_{r,\pi}\left\|\frac{f_X}{\widehat{f_X}}-1\right\|_{\mathrm{L}^\pi(H^+)}\left\|2^{j(s+\nu+(d-1)(1/\pi-1/r)_+)}\right\|_{\ell^q(\{0,\dots,J\})}\end{aligned}$$

using Lemma 16. □

References

[1] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395. MR2549563

[2] Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regression. *Ann. Statist.* **20** 1970–1984. MR1193321

[3] Bertin, K., Le Pennec, E. and Rivoirard, V. (2011). Adaptive Dantzig density estimation. *Ann. Inst. H. Poincaré Probab. Statist.* **47** 43–74.

[4] Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard models. *Rev. Econom. Stud.* **49** 403–410. MR0667501

[5] Funk, P. (1916). Uber eine geometrische anwendung der abelschen integralgleichung. *Math. Ann.* **77** 129–135. MR1511851

[6] Gaïffas, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statist. Sinica* **19** 427–447. MR2514170

[7] Gautier, E., and Hoderlein, S. (2011). Estimating the distribution of treatment effects. Working paper.

[8] Gautier, E., and Kitamura, Y. (2009). Nonparametric estimation in random coefficients binary choice models. Preprint arXiv:0907.2451.

[9] Groemer, H. (1996). *Geometric Applications of Fourier Series and Spherical Harmonics*. Cambridge University Press, Cambridge, *Encyclopedia of Mathematics and its Applications*. MR1412143

[10] 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. MR1389889

[11] Heckman, J. J., and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52** 271–320. MR0735309

[12] Hoderlein, S., Klemelä, J., and Mammen, E. (2010). Reconsidering the random coefficient model. *Econometric Theory* **26** 804–837. MR2646480

[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. MR1649224

[14] Johnstone, I. M., and Raimondo, M. (2004). Periodic boxcar deconvolution and diophantine approximation. *Ann. Statist.* **32** 1781–1804. MR2102493

[15] 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. MR0677597

[16] Kim, P. T., and Koo, J. Y. (2000). Directional mixture models and optimal esimation of the mixing density. *Canad. J. Statist.* **28** 383–398. MR1792056

[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. MR2563258

[18] Kerkyacharian, G., and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli* **10** 1053-1105. MR2108043

[19] Kerkyacharian, G., Petrushev, P., Picard D., and Willer, T. (2007). Needlet algorithms for estimation in inverse problems. *Electron. J. Stat.* **1** 30–76. MR2312145

[20] Kerkyacharian, G., Picard, D., and Raimondo, M. (2007). Adaptive boxcar deconvolution on full Lebesgue measure sets. *Statist. Sinica* **17** 317—340. MR2352512

[21] Kerkyacharian, G., Phan Ngoc, T. M., and Picard, D. (2011). Localized deconvolution on the sphere. *Ann. Statist.* **39** 1042–1068 .

[22] Korostelev, V., and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction*. Springer, New York. MR1226450

[23] Kulik, R., Raimondo, R. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *Ann. Statist.* **37** 3396–3430. MR2549564

[24] Mc Fadden, D. (2001). Economic choices - Nobel Lecture, December 2000. *American Economic Review* **91** 351–378.

[25] Maurer, A., and Pontil, M. (2009). Empirical Bernstein bounds and sample variance penalization. *COLT 2009*.

[26] Narcowich, F., Petrushev, P., and Ward, J. (2006). Decomposition of Besov and Triebel–Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564. MR2253732

[27] Rubin, B. (1999). Inversion and characterization of the hemispherical transform. *J. Anal. Math.* **77** 105–128. MR1753484

[28] Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York. MR2724359

[29] Willer, T. (2005). Deconvolution in white noise with a random blurring function. Preprint arXiv:math/0505142.

[30] Zalcman, L. (1992). A bibliographic survey of the Pompeiu problem, in *Approximation by Solutions of Partial Differential Equations* (B. Fuglede et al., eds.). Kluwer Academic Publ., Amsterdam, 185–194. MR1168719