HAL
open science

Non-parametric regression on the hyper-sphere with uniform design

Jean-Baptiste Monnier

To cite this version:

Jean-Baptiste Monnier. Non-parametric regression on the hyper-sphere with uniform design. Test Publication, 2011, 20 (2), pp.412-446. 10.1007/s11749-011-0233-7 . hal-00552982

HAL Id: hal-00552982

https://hal.science/hal-00552982

Submitted on 6 Jan 2011

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

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

Non-parametric regression on the hyper-sphere with uniform design

Jean-Baptiste Monnier[1]

Abstract

This paper deals with the estimation of a function f defined on the sphere $\mathbb{S}^d$ of $\mathbb{R}^{d+1}$ from a sample of noisy observation points. We introduce an estimation procedure based on wavelet-like functions on the sphere called needlets and study two estimators $f^{\circledast}$ and $f^{\bigstar}$ respectively made adaptive through the use of a stochastic and deterministic needlet-shrinkage method. We show hereafter that these estimators are nearly-optimal in the minimax framework, explain why $f^{\circledast}$ outperforms $f^{\bigstar}$ and run finite sample simulations with $f^{\circledast}$ to demonstrate that our estimation procedure is easy to implement and fares well in practice. We are motivated by applications in geophysical and atmospheric sciences.

Keywords: Non-parametric regression, uniform design, minimax rate, needlets, needlet-shrinkage, stochastic thresholding.
Mathematics Subject classification (2000)62G08, 62G05, 62C20.

1 Introduction

Many branches of applied sciences call upon simple and powerful statistical methods to efficiently analyze spherical data. This is precisely the case of geophysical and atmospheric sciences, where data are usually collected via satellite or ground stations around the globe. Although there exists a basis of spherical harmonics in $\mathbb{L}^2(\mathbb{S}^2)$, its elements are poorly localized, which makes them of little use to represent locally-supported or multi-scale functions on the sphere (see Freeden and Michel (2004, p. 32)). Furthermore, the direct or indirect transposition of Euclidean wavelets to the sphere inherently leads to artificial distortions. This problem has been addressed by a proliferating literature leading to the creation of a wide variety of wavelet frames intrinsic to the sphere (see Freeden et al. (1998); Freeden and Michel (2004) for example). These spherical wavelet frames have since then found many applications in modeling the Earth's magnetic field (Holschneider et al. (2003); Maier (2005); Panet et al. (2005)), atmospheric flows (Fengler (2005)), oceanographic flows (Freeden et al. (2005)) or ionospheric currents (Mayer (2004)). At the same time, Narcowich and Ward (1996) introduced spherical basis functions (SBFs) and used them to design multi-resolution analysis MRA of spherical signals. SBFs were successfully applied in modeling the regional gravity field (Schmidt et al. (2007, 2006)) or the global temperature field (Li (1999); Li and Oh (2004)). However, the issue was raised that SBFs are actually single-scale (see Li (1999) for example), which, from a practical perspective, makes it difficult for the MRA construct given by Narcowich and Ward (1996) to discriminate global from local phenomena. To address that problem, Li (1999) proposed a multi-scale statistical method built upon SBFs of varying bandwidths, which led to new questions of bandwidth selection.
More recently, Narcowich et al. (2006, 2007) have shown it is possible to construct well concentrated frames on the hyper-sphere called needlets, which outperform previous spherical wavelet frames and SBFs in many ways. These needlets are very natural building blocks on the sphere and although they are not exactly an orthogonal basis, they behave almost like one (Narcowich et al. (2006)). They are in fact semi-orthogonal in the sense that any two needlets that are at least two levels apart are orthogonal. Besides each needlet is localized around a center point of $\mathbb{S}^d$ and decaying almost exponentially away from this point. Needlets improve in fact considerably on

[1]Jean-Baptiste Monnier
Université Denis Diderot, Paris 7
Laboratoire de Probabilités et Modèles Aléatoires
175 rue du Chevaleret, Paris, France
Office: 5B01
monnier@math.jussieu.fr

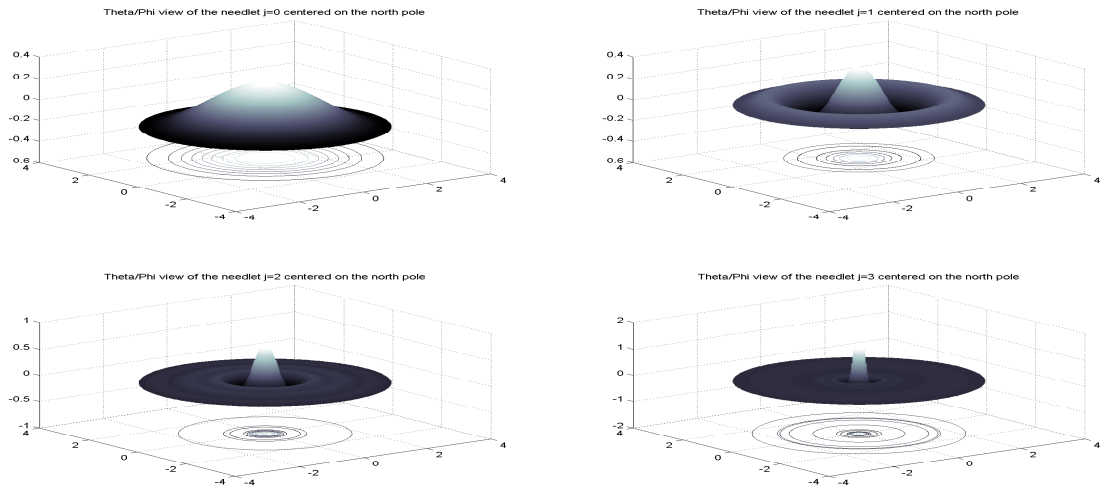

Figure 1: These graphs represent the needlets $\psi_{j,\eta}$ for $\eta = \eta_0 \triangleq (0,0,1)$ (Cartesian coordinates), and $j \in \{0,1,2,3\}$. We display a polar view of the needlets, that is $\psi_{j,\eta_0} = \psi_{j,\eta_0}(\varphi,\theta)$, where φ and θ stand respectively for the spherical coordinates colatitude and longitude. In this polar representation, we let φ play the role of the length of the radius vector, while θ stands for the angle between the latter radius and the x-axis

other spherical basis since, similarly to Euclidean wavelets, they characterize Besov spaces on the sphere and provide us with Jackson estimates of best approximations with needlets (Narcowich et al. (2006, Theorem 6.2)). In addition, needlets are truly multi-scale since they concentrate more and more around their center along increasing needlet levels (see Figure 1). As will be highlighted underneath, these fine features turn needlets into a powerful alternative to existing spherical wavelet frames or SBFs in performing sensible multi-scale modeling of spherical functions. They are indeed already extensively used for applications in astrophysics (e.g. Faÿ et al. (2008); Guilloux et al. (2009); Marinucci et al. (2008)).

In this paper, we consider the problem of recovering f from the observation of n independent and identically distributed (iid) realizations $(Y_i, T_i), i = 1, \ldots, n$ of the random vector (Y, T) generated by the model

$$Y = f(T) + \sigma V, \qquad T \sim \mathcal{U}(\mathbb{S}^d), \quad V \sim \mathcal{N}(0,1) \tag{1}$$

where T is uniformly distributed on the hyper-sphere: $T \sim \mathcal{U}(\mathbb{S}^d)$, V is a real-valued standard normal random variable: $V \sim \mathcal{N}(0,1)$, $\sigma \in \mathbb{R}^{+*}$ quantifies the magnitude of the error and f is a real-valued map of a wide Besov scale. In the sequel, we introduce an estimation procedure, which features multi-scale capabilities and is robust and easy to implement, since it rests in practice on the calibration of one single parameter (see Section 7). It is noteworthy that our results hold under the assumption that the data are uniformly scattered on the sphere, which is however not the case in most of the practical situations mentioned earlier. Further research is needed in order to circumvent that problem and show that our results eventually generalize to a warped needlets setting (Kerkyacharian and Picard (2004)).

We prove in this paper that two well chosen needlet estimators $f^{\circledast}$ and $f^{\bigstar}$ allow to reconstruct with near-minimax-optimality a very wide range of functions f defined on the sphere. In the sequel, we will write $f^{\diamond}$ to refer indifferently to $f^{\circledast}$ or $f^{\bigstar}$, unless stated otherwise. This will prove very convenient in statements that apply to both estimators. As is well-known, we mean by "near-optimality" that the approximation loss between $f^{\diamond}$ and f is within a logarithmic factor of the optimal minimax rate for a given *a priori* class. Optimal estimators tend to be overwhelmingly specific to the smoothness class and the loss for which they are optimal and become irrelevant

to any other setting. However, by relaxing the requirement of optimality to near-optimality, it becomes possible to choose $f^\diamond$ to be nearly-optimal over a wide range of losses and smoothness classes (Donoho et al. (1995)). In this paper, near-minimaxity is thus not only obtained for the celebrated $\mathbb{L}^2$-loss but also for the $\mathbb{L}^\infty$-loss, which means that $f^\diamond$ converges uniformly to f as the sample size grows. Besides, the near-optimality of $f^\diamond$ holds over a wide range of smoothness classes, which makes these results particularly interesting in practice, when there is only scarce information available on the objective function f.

The two needlet estimators $f^\circledast$ and $f^\bigstar$ are respectively made adaptive through the use of a stochastic and deterministic thresholding method. Although they verify a wide range of similar properties, the proofs of their near-optimality differ slightly. We run numerical simulations on $f^\circledast$ and show how the stochastic thresholding outperforms the deterministic one. As described in Section 7, this is due to the fact that the stochastic thresholding parameter adjusts to the magnitude of the sample noise at each sample coefficient $\hat{\beta}_{j,\eta}$ (see Section 5). Furthermore, we study the impact of the dimension d on both estimators. Despite the well-known deterioration of minimax rates as the dimension d increases, we show that working on a higher dimensional underlying unit sphere $\mathbb{S}^d$ sharpens the constants in our minimaxity results and makes the construction of both estimators easier.

This paper extends the optimality results of Baldi et al. (2009) to the regression setting, which is essentially made possible thanks to the sharp localization properties of needlets, as will be made precise in the proof of Lemma 10.3.

The plan of the paper is as follows. In Section 2 and Section 3 we review some background material on needlets and Besov spaces on the sphere. Readers familiar with these latter matters may jump directly to Section 4, where we introduce the model and set notations that will be used throughout the paper. Section 5 presents our thresholding estimators, whose minimax performances are stated in Section 6. Section 7 describes the performance of the estimators on some simulated data. Finally, Section 8–Section 10 contain the proofs.

In the sequel, the symbol $\triangleq$ stands for *equal by definition*. In addition, for two functions A, a of the variable γ, we write $A(\gamma) \approx a(\gamma)$ when there exist constants c, C independent of γ such that $ca(\gamma) \leq A(\gamma) \leq Ca(\gamma)$ for all values of γ. Furthermore, we will write $C(\beta)$ to mean that the constant C depends on parameter β.

2 Needlets and their properties

In this section, we give an overview of the needlets construction and describe some of their properties. It is inspired from Narcowich et al. (2006) where the needlets were originally introduced and presents similar material as in Baldi et al. (2009, Section 2). We first introduce spherical harmonics. We then detail the needlet construction. It is essentially divided into two steps: a Littlewood-Paley decomposition and a quadrature formula.

2.1 Spherical harmonics

In what follows, we write $\mathfrak{M}$ the surface measure of $\mathbb{S}^d$, that is the unique positive measure on $\mathbb{S}^d$ which is invariant under rotation and has the area ω_d of $\mathbb{S}^d$ as total mass, that is $\omega_d = 2\pi^{(d+1)/2}/\Gamma(\frac{d+1}{2})$. Denote by $\mathcal{H}_l$ the set of spherical harmonics of degree l (Stein and Weiss (1975, Chap. 4)). We can consider $\mathcal{H}_l$ as a subspace of $\mathbb{L}^2(\mathbb{S}^d)$ with inner product $(f, g) = \int_{\mathbb{S}^d} f(x)g(x)\mathfrak{M}(dx)$ and show that $\mathcal{H}_k \perp \mathcal{H}_l$ for $k \neq l$ and the collection of finite linear combinations of $\cup\mathcal{H}_l$ is dense in $\mathbb{L}^2(\mathbb{S}^d)$. Moreover we can compute $a_{l,d} \triangleq \dim \mathcal{H}_l = \mathcal{O}(l^{d-1})$ and exhibit an orthonormal basis $\{Y_{l,m}; m = 1, \ldots, a_{l,d}\}$ of $\mathcal{H}_l$ and, subsequently, an orthonormal basis $\cup_{l\geq 0}\{Y_{l,1}, \ldots, Y_{l,a_{l,d}}\}$ of $\mathbb{L}^2(\mathbb{S}^d)$. Thus, for any $f \in \mathbb{L}^2(\mathbb{S}^d)$, the orthogonal projector on $\mathcal{H}_l$ is given

by

$$P_{\mathcal{H}_l} f(\xi) = \sum_{m=1}^{a_{l,d}} (f, Y_{l,m}) Y_{l,m}(\xi) = \int_{\mathbb{S}^d} P_l(\xi, x) f(x) \mathfrak{M}(dx)$$

where we have written $P_l(\xi, x) = \sum_{m=1}^{a_{l,d}} Y_{l,m}(\xi) Y_{l,m}(x)$ the projection kernel on $\mathcal{H}_l$. Thanks to the addition theorem Müller (1966, Theorem 2), we have $P_l(\xi, x) = c_{l,d} L_l(\xi \centerdot x)$ where $c_{l,d} \triangleq a_{l,d}/\omega_d$, the operator "$\centerdot$" stands for the usual Euclidean scalar product of $\mathbb{R}^{d+1}$, and L_l is the Legendre polynomial of degree l and dimension $d+1$ (Müller (1966, p 16)). From now on, we will write $P_l(\xi \centerdot x)$ in place of $P_l(\xi, x)$.

Remark 2.1. *For practical implementation, let's recall that we have*

$$\int_{-1}^{1} L_l(t) L_k(t) (1-t^2)^{\frac{d-2}{2}} dt = \widetilde{c}_{l,d} \delta_{l,k}, \qquad \widetilde{c}_{l,d} = \frac{\omega_d}{\omega_{d-1}} \frac{1}{a_{l,d}} = \frac{\sqrt{\pi}\Gamma(\frac{d}{2})}{a_{l,d}\Gamma(\frac{d+1}{2})}$$

In Section 7, we will run simulations on the sphere of $\mathbb{R}^3$. In this case, $d = 2$, so that we have $\omega_2 = 4\pi$, $\widetilde{c}_{l,2} = \frac{2}{2l+1}$ and we can write $N_l = \sqrt{\frac{2l+1}{2}} L_l$ the normalized Legendre polynomials. Besides $c_{l,2} = \frac{2l+1}{4\pi}$ and we will therefore use kernels of the form $P_l(\xi \centerdot x) = \frac{2l+1}{4\pi} L_l(\xi \centerdot x) = \sqrt{\frac{2l+1}{8\pi^2}} N_l(\xi \centerdot x)$ for simulations.

A direct application of Parseval's formula in $\mathbb{L}^2(\mathbb{S}^d)$ leads to

$$\int_{\mathbb{S}^d} P_l(\xi \centerdot x) P_k(x \centerdot \tau) \mathfrak{M}(dx) = \delta_{l,k} P_l(\xi \centerdot \tau) \tag{2}$$

Let's write $\mathscr{P}_l$ the space of spherical harmonics of degree at most l. Obviously, the kernel $K_l = \sum_{k=0}^{l} P_k$ stands for the orthogonal projector on $\mathscr{P}_l$. Unfortunately, the poor localization properties of K_l are a major obstacle to its use for function decomposition outside the $\mathbb{L}^2$ framework. This problem is circumvented in Narcowich et al. (2007) by the introduction of a new kernel whose construction is based on the Littlewood-Paley decomposition.

2.2 Littlewood-Paley decomposition

Let φ be a C^∞ function on $\mathbb{R}$, symmetric and decreasing on $\mathbb{R}^+$ such that supp$\varphi \subset [-1, 1]$, $\varphi(z) = 1$ if $|z| \leq \frac{1}{2}$ and $0 \leq \varphi(z) \leq 1$ otherwise. We set $b^2(z) \triangleq \varphi(z/2) - \varphi(z)$ so that $b^2(z) \geq 0$ and $b(z) \neq 0$ only if $\frac{1}{2} \leq |z| \leq 2$. We now define the following kernels for $j \geq 0$,

$$\Theta_j \triangleq \sum_{l \geq 0} b^2(l/2^j) P_l = \sum_{2^{j-1} < l < 2^{j+1}} b^2(l/2^j) P_l$$
$$\Psi_j \triangleq \sum_{l \geq 0} b(l/2^j) P_l = \sum_{2^{j-1} < l < 2^{j+1}} b(l/2^j) P_l$$

For any $\xi, \tau \in \mathbb{S}^d$, we write

$$\{\Psi_j * f\}(\xi) \triangleq \int_{\mathbb{S}^d} \Psi_j(\xi \centerdot x) f(x) \mathfrak{M}(dx),$$
$$\{\Psi_j * \Psi_k\}(\xi \centerdot \tau) \triangleq \int_{\mathbb{S}^d} \Psi_j(\xi \centerdot x) \Psi_k(x \centerdot \tau) \mathfrak{M}(dx)$$

It appears as a direct consequence of eq. (2) that $\Theta_j \triangleq \Psi_j * \Psi_j$ for $j \geq 0$. As detailed in Narcowich et al. (2006, Theorem 2.2), these new kernels are in fact nearly exponentially localized in the sense that, for any $k > 0$, there exists a constant $c_k > 0$ such that, for any $\xi, \eta \in \mathbb{S}^d$,

$$|\Theta_j(\xi \centerdot \eta)|, |\Psi_j(\xi \centerdot \eta)| \leq \frac{c_k 2^{jd}}{\{1 + 2^j dist(\xi, \eta)\}^k}$$

where *dist* stands for the geodesic distance on the sphere. As proved in Narcowich et al. (2007, Theorem 3.1), we have the following result

Proposition 2.1. *For every $f \in \mathbb{L}^p(\mathbb{S}^d)$ and $1 \leq p < \infty$, or if $p = \infty$ and f is continuous, the following identity holds in $\mathbb{L}^p$,*

$$f = P_0 * f + \lim_{J \to \infty} \sum_{j=0}^{J} \Theta_j * f$$

2.3 Quadrature formula and needlets

The needlets arise as by-products of kernel Ψ by means of a quadrature formula (Narcowich et al. (2006, Corollary 2.9)). For all $l \in \mathbb{N}$, there indeed exists a finite subset $\mathcal{X}_l \subset \mathbb{S}^d$ and positive real numbers $\lambda_\eta > 0$, indexed by the elements $\eta \in \mathcal{X}_l$, such that for all $f \in \mathscr{P}_l$

$$\int_{\mathbb{S}^d} f(x)\mathfrak{M}(dx) = \sum_{\eta \in \mathcal{X}_l} \lambda_\eta f(\eta)$$

In particular, it is obvious from the above definition of the operator Ψ_j that $x \mapsto \Psi_j(\xi, x) \in \mathscr{P}_{[2^{j+1}]}$, so that $x \mapsto \Psi_j(\xi, x)\Psi_j(x, \tau) \in \mathscr{P}_{[2^{j+2}]}$. Thus we can apply the quadrature formula to write

$$\Theta_j(\xi \centerdot \tau) = \int_{\mathbb{S}^d} \Psi_j(\xi \centerdot x)\Psi_j(x \centerdot \tau)\mathfrak{M}(dx) = \sum_{\eta \in \mathcal{X}_{[2^{j+2}]}} \lambda_\eta \Psi_j(\xi \centerdot \eta)\Psi_j(\eta \centerdot \tau) \tag{3}$$

Now, write $\mathcal{X}_{[2^{j+2}]} = \mathcal{Z}_j$ and define the needlet of center $\eta \in \mathcal{Z}_j$ and level j by

$$\psi_{j,\eta}(\xi) \triangleq \sqrt{\lambda_\eta}\Psi_j(\xi \centerdot \eta)$$

With these notations, eq. (3) leads to

$$\Theta_j * f(\xi) = \sum_{\eta \in \mathcal{X}_{[2^{j+2}]}} \sqrt{\lambda_\eta}\Psi_j(\xi \centerdot \eta)\{\sqrt{\lambda_\eta}\Psi_j * f(\eta)\} = \sum_{\eta \in \mathcal{Z}_j} (\psi_{j,\eta}, f)\psi_{j,\eta}(\xi)$$

As described in Narcowich et al. (2006, Corollary 2.9), the choice of the sets of cubature points $\mathcal{Z}_j$ is not unique, but one can impose the conditions $\#\mathcal{Z}_j \approx 2^{jd}$ and $\lambda_\eta \approx 2^{-jd}$. These last results together with Proposition 2.1 lead to the following equality in $\mathbb{L}^p$, $p \geq 1$,

$$f = P_0 * f + \sum_{j \geq 0} \sum_{\eta \in \mathcal{Z}_j} (f, \psi_{j,\eta})\psi_{j,\eta} \tag{4}$$

The $\psi_{j,\eta}$'s appear therefore as building blocks on the sphere. They obviously inherit the fine localization properties of the Ψ_j's, which prompted Narcowich, Petrushev and Ward to call them *needlets*. Besides, they are more and more localized around their center as j increases and therefore capture sample phenomena occurring at finer and finer scales. These last features turn them into a very handy tool to tackle statistical problems on the sphere. From this localization property it follows that for $0 < p \leq \infty$,

$$c_p 2^{jd\{\frac{1}{2} - \frac{1}{p}\}} \leq \|\psi_{j,\eta}\|_p \leq C_p 2^{jd\{\frac{1}{2} - \frac{1}{p}\}} \tag{5}$$

Moreover, as shown in Baldi et al. (2009, Lemma 2), we have the following useful lemma

Lemma 2.1. *For any $j \geq 0$, we have,*

1) For every $0 < p \leq \infty$

$$\| \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta \psi_{j,\eta} \|_p \leq c 2^{jd\{\frac{1}{2}-\frac{1}{p}\}} \left(\sum_{\eta \in \mathcal{Z}_j} |\lambda_\eta|^p \right)^{\frac{1}{p}} \tag{6}$$

2) For every $1 \leq p \leq +\infty$

$$\left(\sum_{\eta \in \mathcal{Z}_j} |(f, \psi_{j,\eta})|^p \right)^{\frac{1}{p}} 2^{jd\{\frac{1}{2}-\frac{1}{p}\}} \leq c \|f\|_p \tag{7}$$

We refer the reader to the above-mentioned article for a detailed proof. Let's now turn to the case where f belongs to a Besov space on the sphere and describe how these latter spaces relate to needlets.

3 Besov spaces on the sphere and needlets

In this section we summarize the main properties of Besov spaces on the sphere following the presentations given in Baldi et al. (2009); Narcowich et al. (2006, Section 5). Besov spaces on the sphere can be defined as follows

Definition 3.1. *The Besov space* $B^s_{rq} \triangleq B^s_{rq}(\mathbb{S}^d)$*, where* $s \in \mathbb{R}$*,* $0 < r, q \leq \infty$*, is the set of all measurable functions on* $\mathbb{S}^d$ *such that*

$$\|f\|_{B^s_{rq}} \triangleq \left(\sum_{j=0}^{\infty} \{2^{js} \|\Theta_j * f\|_r\}^q \right)^{\frac{1}{q}} < \infty$$

where the ℓ^q*-norm is replaced by the* ℓ^∞*-norm when* $q = \infty$*. It is in fact possible to show that this definition is independent of the choice of* φ *used to build kernels* Θ_j*.*

Besides, the following noteworthy theorem (Narcowich et al. (2006, Theorem 5.5)) sheds some light on the tight intertwining between Besov spaces and needlet coefficients.

Theorem 3.1. *Let be given* s, r, q *such that* $1 \leq r \leq +\infty$*,* $s > 0$*,* $0 \leq q \leq +\infty$*. For any sequence* $\{g_{j,\eta}, \eta \in \mathcal{Z}_j, j \geq 0\}$*, we write*

$$\|g\|_{\boldsymbol{b}^s_{rq}} = \| \left\{ 2^{j\{s+\frac{d}{2}-\frac{d}{r}\}} \|\{g_{j,\eta}\}_{\eta \in \mathcal{Z}_j}\|_{\ell^r} \right\}_{j \geq 0} \|_{\ell^q}$$

In addition, for any measurable function f*, we define* $\beta_{j,\eta} \triangleq (f, \psi_{j,\eta})$ *provided it makes sense, and we consider the sequence* $\{\beta_{j,\eta}, \eta \in \mathcal{Z}_j, j \geq 0\}$*. Then we have* $\|\beta\|_{\boldsymbol{b}^s_{rq}} \approx \|f\|_{B^s_{rq}}$ *and, thus,* $f \in B^s_{r,q}$ *if and only if*

$$\|\beta\|_{\boldsymbol{b}^s_{rq}} < \infty \tag{8}$$

In the sequel we shall write $\|f\|_{B^s_{rq}}$ in place of $\|\beta\|_{\boldsymbol{b}^s_{rq}}$. Furthermore we will denote by $B^s_{r,q}(M)$ the ball of radius M of the Besov space $B^s_{r,q}$. Let's now recall the Besov embedding theorem on the sphere. We refer the reader to Baldi et al. (2009, Theorem 5) for a detailed proof.

Theorem 3.2. *(The Besov embedding)*

$$B^s_{r,q} \subseteq B^s_{p,q}, \qquad \text{if } p \leq r \leq \infty$$

$$B^s_{r,q} \subseteq B^{s+\frac{d}{p}-\frac{d}{r}}_{p,q}, \qquad \text{if } s > \frac{d}{r} - \frac{d}{p} \text{ and } r \leq p \leq \infty$$

4 Setting and notations

In this section, we describe the model and introduce notations that will be used throughout the paper. We start with a few notations. For $d \geq 1$, $s > \frac{d}{r}$, $0 < r \leq \infty$, we set

$$\vartheta \triangleq \vartheta_r^p \triangleq \frac{s + \frac{d}{p} - \frac{d}{r}}{2(s + \frac{d}{2} - \frac{d}{r})}, \qquad \varsigma \triangleq \frac{s}{2s + d}, \qquad \vartheta^\infty \triangleq \vartheta_r^\infty \triangleq \frac{s - \frac{d}{r}}{2(s + \frac{d}{2} - \frac{d}{r})}$$

We will write $X \sim \mathbb{P}$ to mean that the random variable X follows law $\mathbb{P}$, $X \sim Y$ to mean that X and Y have the same law and denote the standard Gaussian law by $\mathcal{N}(0,1)$. In the sequel C and c stand for absolute constants, which may vary from line to line or even inside a same equation. And in order to lighten the notations, we sometimes write $A \equiv \widetilde{A}$ to mean that A is a lighter typographical way to refer to $\widetilde{A}$. We write $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space on which Y, T and V are defined according to $Y = f(T) + V$, where T is uniformly distributed on $\mathbb{S}^d$, V is normal with mean zero and standard deviation σ and $f \in B^s_{rq}(M)$. In particular, we can take $(\Omega, \mathcal{F}, \mathbb{P})$ to be the canonical probability space associated to the vector (V, T), that is, $\Omega \equiv \mathbb{R} \times \mathbb{S}^d$, $\mathcal{F} = \mathscr{B}(\mathbb{R} \times \mathbb{S}^d)$, for all $w = (v, t) \in \mathbb{R} \times \mathbb{S}^d$, $(V, T)\{w\} = (V(v), T(t)) = (v, t) = w$ and $\mathbb{P} \equiv \mathbb{P}^{V,T} = \mathbb{P}^V \otimes \mathbb{P}^T$. Obviously, $\mathbb{P}^V(dv) = \varphi_\sigma(v)\lambda(dv)$, where $\varphi_\sigma(v)$ is the normal density with mean zero and standard deviation σ and λ the Lebesgue measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$; and $\mathbb{P}^T$ is the uniform law on the sphere $\mathbb{S}^d$, that is $\mathbb{P}^T(dt) = \mathfrak{M}(dt)/\omega_d$ where $\mathfrak{M}$ stands for the spherical surface measure introduced earlier. We write as well $\mathbb{P}_f \triangleq \mathbb{P}^{Y,T}$ the law of the vector (Y, T) and $\mathbb{E}_f$ the expectation with respect to $\mathbb{P}_f$. When there is no ambiguity, we denote $\mathbb{P} \equiv \mathbb{P}_f$ and $\mathbb{E} \equiv \mathbb{E}_f$. Alternatively and when appropriate, we will write the model $Y = f(T) + \sigma V$ with corresponding modifications.

5 Needlet estimation of f on the sphere

Given the set of n iid observations (Y_i, T_i), $i = 1, \ldots, n$, we can compute

$$\frac{1}{n}\sum_{i \leq n} Y_i \psi_{j,\eta}(T_i) = \frac{1}{n}\sum_{i \leq n} f(T_i)\psi_{j,\eta}(T_i) + \frac{1}{n}\sum_{i \leq n} \sigma V_i \psi_{j,\eta}(T_i)$$

In the sequel, we will adopt the following notations $y^*_{j,\eta} \triangleq \omega_d \sum_{i \leq n} Y_i \psi_{j,\eta}(T_i)/n$, $\zeta^*_{j,\eta} \triangleq \omega_d \sum_{i \leq n} f(T_i)\psi_{j,\eta}(T_i)/n$ and $\gamma^*_{j,\eta} \triangleq \omega_d \sum_{i \leq n} \sigma V_i \psi_{j,\eta}(T_i)/n$. In addition we write $\varrho^2_{j,\eta} \triangleq \sum_{i \leq n} \psi_{j,\eta}(T_i)^2/n$ and $\xi_{j,\eta} = \sqrt{n}\gamma^*_{j,\eta}/(\sigma\omega_d\varrho_{j,\eta})$. Since the V_i's are iid standard normal and independent from the T_i's, we know that $\mathbb{V}\mathrm{ar}(\gamma^*_{j,\eta}|T_1, \ldots T_n) = \sigma^2\omega_d^2\varrho^2_{j,\eta}/n$. Thus $\xi_{j,\eta} \sim \mathcal{N}(0,1)$ conditionally on the T_i's. We therefore observe the sequence $\{y^*_{j,\eta}, j \geq 0, \eta \in \mathcal{Z}_j\}$, such that $y^*_{j,\eta} = \zeta^*_{j,\eta} + \{\sigma\omega_d\varrho_{j,\eta}/\sqrt{n}\}\xi_{j,\eta}$, for all $j \geq 0$ and all $\eta \in \mathcal{Z}_j$.

Eq. (4) shows that the estimation of f by $f^\diamond \equiv f^\diamond(Y_i, T_i; i = 1, \ldots n)$ reduces to the estimation of its needlet coefficients $\beta_{j,\eta}$ and $P_0 * f$. It is easily proved that $\hat{\beta}_{j,\eta} \triangleq y^*_{j,\eta}$ and $\hat{P}_0 \triangleq \sum Y_i/n$ are respectively strongly consistent and unbiased estimator of $\beta_{j,\eta}$ and $P_0 * f$. We further want the estimator $f^\diamond$ to be adaptive to inhomogeneous smoothness. In that perspective we use a hard thresholding method, which aims at canceling out coefficient estimates $\hat{\beta}_{j,\eta}$ that result mainly from noise. In the sequel we study two estimators $f^\circledast$ and $f^\bigstar$ built respectively upon a stochastic and a deterministic thresholding method. We denote by $\beta^\circledast_{j,\eta}$ the needlet coefficients of $f^\circledast$ and set $\beta^\circledast_{j,\eta} \triangleq \hat{\beta}_{j,\eta}\mathbb{1}_{\{|\hat{\beta}_{j,\eta}| \geq \kappa_{j,\eta}t(n)\}}$, where $\kappa_{j,\eta} = \varpi\varrho_{j,\eta}$, ϖ is a constant and $t(n) = \sqrt{\log n/n}$. Similarly, we denote by $\beta^\bigstar_{j,\eta}$ the needlet coefficients of $f^\bigstar$ and set $\beta^\bigstar_{j,\eta} \triangleq \hat{\beta}_{j,\eta}\mathbb{1}_{\{|\hat{\beta}_{j,\eta}| \geq \kappa t(n)\}}$, where κ is a constant. In the sequel, we will write $f^\diamond$, $\beta^\diamond_{j,\eta}$ and $\varkappa$ to refer indifferently to $f^\circledast$, $\beta^\circledast_{j,\eta}$ and ϖ, or $f^\bigstar$, $\beta^\bigstar_{j,\eta}$ and κ.

Finally, we cut the series expansion of $f^\diamond$ at level J such that $2^{Jd} = n/\{C_0 \log n\}$. With these

notations, the needlet estimator of f can be written as

$$f_J^\diamond = \hat{P}_0 + \sum_{j=0}^{J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j,\eta}^\diamond \psi_{j,\eta} \tag{9}$$

Before we move on to the study of the minimax rates for the estimator $f^\diamond$, notice that in the above construction of $f^\diamond$, we remain free to choose the values of $\varkappa$ and C_0. We will see later that C_0 is in fact very much related to $\varkappa$ so that we are truly left with one tuning parameter $\varkappa$. We will give some hints on ways of evaluating it in Section 7.

6 Minimax rates for $\mathbb{L}^p$ norms and Besov spaces on the sphere

In the sequel, we will denote by $\mathfrak{C}_z^\circledast$ the set of conditions

$$\begin{aligned} \varpi &\geq 4 \max\left(4e^2\omega_d^2 \|f\|_\infty^2, \omega_d^2\sigma^2, z+1\right), \\ C_0 &\geq \varpi \max\left(\frac{2C_\infty^2}{c_2^2 e^2}, \frac{2}{m^-}\right), \\ 2^{Jd} &= \frac{n}{C_0 \log n} \end{aligned} \tag{$\mathfrak{C}_z^\circledast$}$$

and $\mathfrak{C}_z^\bigstar$ the set of conditions

$$\begin{aligned} \kappa &\geq 4 \max\left(2\omega_d C_2 \max(4\|f\|_\infty^2, 3\sigma^2), z+1\right), \\ C_0 &> \max(6\omega_d \|f\|_\infty C_\infty, \kappa/\{2m^-\}), \\ 2^{Jd} &= \frac{n}{C_0 \log n} \end{aligned} \tag{$\mathfrak{C}_z^\bigstar$}$$

where the constants C_∞, c_2 are defined in Proposition 8.2 and m^- is defined in Lemma 10.1. Once again, the couple $f^\diamond, \mathfrak{C}_z^\diamond$ will denote indifferently $f^\circledast, \mathfrak{C}_z^\circledast$ or $f^\bigstar, \mathfrak{C}_z^\bigstar$. We now present two theorems that describe the asymptotic properties of the estimator $f^\diamond$ of f. In a first theorem, we compute an upper-bound on the loss of our estimator over Besov balls and $\mathbb{L}^p$-norms. Its proof can be found in Section 8.

Theorem 6.1. *Let be given $f \in \mathbb{L}^\infty$. Consider the estimator $f^\diamond$ (see eq. (9)) of f built upon n iid observations (Y_i, T_i) drawn from the model stated in eq. (1). Then, for $d \geq 1$, $s > \frac{d}{r}$, $0 < r \leq \infty$, we have*

a) For any $z > 1$, there exists some constant c_∞ such that, as soon as the conditions $\mathfrak{C}_z^\diamond$ are verified,

$$\sup_{f \in B_{r,q}^s(M)} \mathbb{E}\|f^\diamond - f\|_\infty^z \leq c_\infty (\log n)^{2z} \left(\frac{n}{\log n}\right)^{-z\vartheta^\infty} \tag{10}$$

b) For any $1 \leq p < \infty$, there exists a constant c_p such that, as soon as the conditions $\mathfrak{C}_p^\diamond$ are verified,

$$\sup_{f \in B_{r,q}^s(M)} \mathbb{E}\|f^\diamond - f\|_p^p \leq c_p \{\log n\}^p \left(\frac{n}{\log n}\right)^{-p\vartheta}, \qquad \text{if } r \leq \frac{dp}{2s+d} \text{ and } p \geq 2 \tag{11}$$

$$\sup_{f \in B_{r,q}^s(M)} \mathbb{E}\|f^\diamond - f\|_p^p \leq c_p \{\log n\}^p \left(\frac{n}{\log n}\right)^{-p\varsigma}, \qquad \text{if } r > \frac{dp}{2s+d} \tag{12}$$

Notice that the above result places $f^\diamond$ within a larger $\log n$ factor of the $n/\log n$ term than in Baldi et al. (2009, Theorem 8). The proof we present here is marginally simpler than theirs and eventually more systematic since it introduces a function of a floating parameter l as an upper-bound and optimizes with respect to l. To be more specific, in contrary to Baldi et al. (2009, Proposition 15), Proposition 8.2 does not use the fact that there exists an index $J_1(s)$ beyond which $|\beta_{j,\eta}| \leq t(n)$, which makes our demonstration simpler albeit less precise.
In a second theorem, we compute a lower bound on the loss of f estimators over Besov balls and $\mathbb{L}^p$-norms. Its proof follows similar lines as the proof of the lower bound detailed in Baldi et al. (2009). It is therefore not reported here but made available at `www.math.jussieu.fr/~monnier`.

Theorem 6.2. *(Lower bound) We write* $\inf_\theta$ *the lower-bound over all estimators* θ *of* f*, that is all measurable functions of the* $Y_i, T_i, i = 1 \ldots, n$*. Then, for* $d \geq 1$*,* $s > \frac{d}{r}$*,* $0 < r \leq \infty$*, we have*

a) If $1 \leq p \leq 2$*,*

$$\inf_\theta \sup_{f \in B^s_{rq}(M)} \mathbb{E}_f \|\theta - f\|_p^p \geq cn^{-p\varsigma}$$

b) If $2 < p \leq +\infty$

$$\inf_\theta \sup_{f \in B^s_{rq}(M)} \mathbb{E}_f \|\theta - f\|_p^p \geq \begin{cases} cn^{-p\varsigma}, & \text{if } r > \frac{dp}{2s+d} \\ cn^{-p\vartheta}, & \text{if } r \leq \frac{dp}{2s+d} \end{cases}$$

These two theorems demonstrate that our estimator $f^\diamond$ is in fact nearly-optimal in all the above settings. Although these minimaxity results hold for a " big enough" sample size n, the estimator $f^\diamond$ fares well in practice over finite samples too, as will be shown through simulations in the next section.

7 Simulations

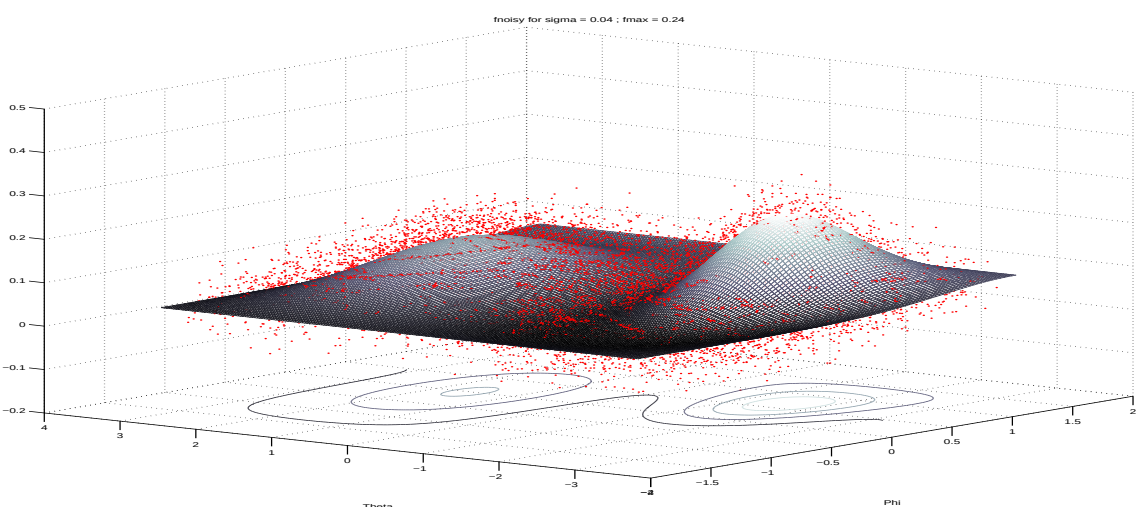

Figure 2: Display of the function $f : x \in \mathbb{S}^2 \mapsto 0.65 \exp(-k_1\|x - x_1\|_2)/b_1 + 0.35 \exp(-k_2\|x - x_2\|_2)/b_2$ on a grid of points of the unit sphere of $\mathbb{R}^3$ parametrized by their spherical coordinates colatitude φ and longitude θ. We choose $x_1 = (0, 1, 0)$, $x_2 = (0, -0.8, 0.6)$, $k_1 = 0.7$, $k_2 = 2$ and $b_i = \int_{\mathbb{S}^2} \exp(k_i\|x - x_i\|_2)\mathfrak{M}(dx)$, $i = 1, 2$. We set $\sigma = 0.04$ and represent $N = 10,000$ noisy observations Y_i at locations T_i simulated using the transformation $\theta = 2\pi(\text{rand}() - .5)$ and $\varphi = \sin^{-1}(2\text{rand}() - 1)$

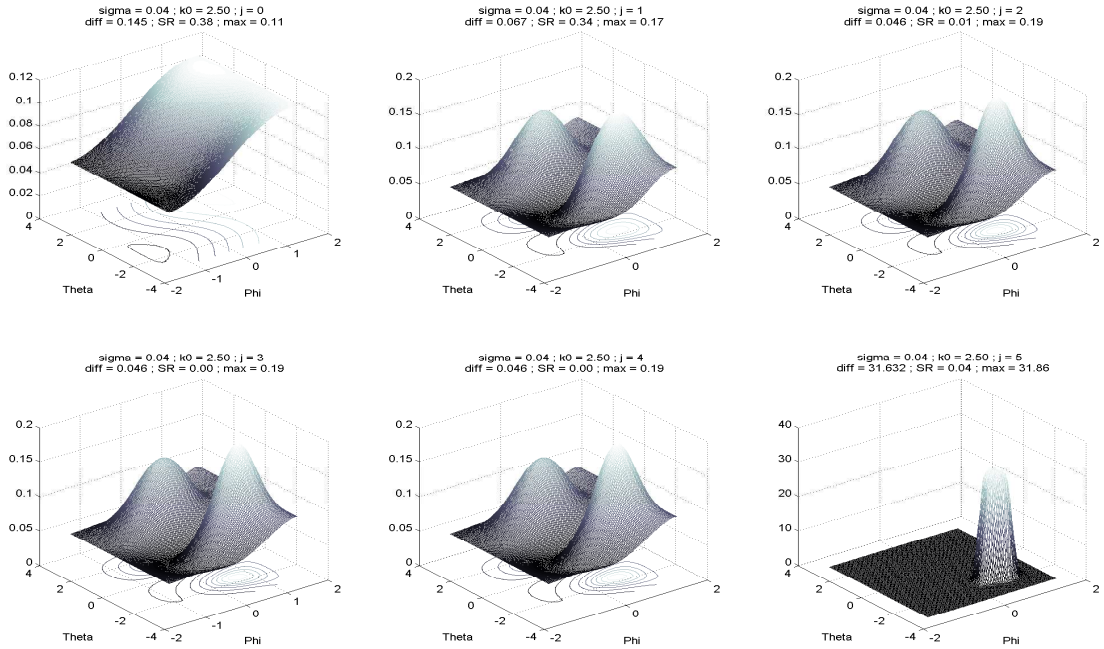

Figure 3: Computation of $f^{\circledast}$ from the observation of $20,000$ vectors (Y_i, T_i). To clearly picture the contribution of each level j to the value of the estimator at each point, we graph $f_J^{\circledast}$ for $J = 0, 1, 2, 3$ on the grid. In the title of each sub-figure we indicate the value of σ, ϖ (constant multiple of k_0), J (corresponds to j), *diff* (which stands for the value of $\|f^{\circledast} - f\|_\infty$ on the grid), *SR* (which stands for "survival rate" and displays the percentage of coefficients that survives thresholding at level J), and max (which gives the maximum value of the estimator $f^{\circledast}$ on the grid)

For the sake of brevity, we only report here the main results of our simulations. The interested reader is referred to the addendum at www.math.jussieu.fr/~monnier for a thorough discussion. Remark we expect the stochastic thresholding to outperform the deterministic one, since it adjusts the constant thresholding parameter ϖ by the sample noise standard deviation $\varrho_{j,\eta}$ at sample coefficient $\hat{\beta}_{j,\eta}$. The comparison between $f^{\bigstar}$ and $f^{\circledast}$ on simulated data shows that the two estimators perform similarly at needlet levels $0 \leq j \leq 3$, due to the fact that $\varrho_{j,\eta}$ remains almost constant across cubature points η at these resolution levels. However, $\varrho_{j,\eta}$ varies more widely from one cubature point to another at higher needlet levels causing $f^{\circledast}$ to adjust more efficiently to the noise than $f^{\bigstar}$ and outperform it.

We therefore run simulations with $f^{\circledast}$. Notice that the condition on ϖ depends on $\|f\|_\infty$, which is unknown in practice. However, if we have any prior insight into f, we can replace $\|f\|_\infty$ by any real constant that overshoots it, which gives us more flexibility. We compute conditions $\mathfrak{C}_z^{\circledast}$ numerically. With our test function f (see Figure 2), such that $\|f\|_\infty = 0.24$, we obtain $C_0 \geq \varpi 5 \cdot 10^5$. It thus appears that our proof of the near-optimality of $f^{\circledast}$ imposes drastic conditions on the parameter C_0. From a theoretical standpoint, $f_J^{\circledast}$ is the near optimal estimator of f as soon as $N/\log N$ is of order $C_0 2^{2J}$, which means that we would have to gather unrealistically large data samples in order to build an estimator $f_J^{\circledast}$ of f that would contain information up to degree of resolution J for large C_0. However, numerical simulations tend to demonstrate that our procedure fares well under much less drastic conditions.

The most obvious way of fixing the free thresholding parameter ϖ is to monitor the proportion of coefficients that are zeroed out at each resolution level as a function of ϖ. It is clear that the more coefficients thresholded at high levels, the smoother the estimator. The ultimate choice of ϖ

should therefore be related to an *a priori* knowledge of the smoothness of f. Numerical simulations show that $f^{\circledast}$ recovers the overall shape of f, even for values of σ that are large in front of $\|f\|_\infty$. Without over-fitting the data, we obtain an estimation error $\|f_2^{\circledast} - f\|_\infty = 0.046$ for $\sigma = 0.04$ (see Figure 3) and $\|f_2^{\circledast} - f\|_\infty = 0.062$ for $\sigma = 0.5$.
Finally, notice that, despite the "curse of dimensionality" on minimax rates, conditions $\mathfrak{C}_z^\diamond$ are loosened as d increases while the upper-bound constants that appear in Theorem 6.1 become smaller.

8 Proof of the minimax rate

Let us now move on to the proofs of the theorems presented in Section 6. We first introduce two propositions that we will need later on in the proof.

Proposition 8.1. *For all $z > 0$, we have*

$$\mathbb{E}|\hat{P}_0 - P_0 * f|^z \leq Cn^{-\frac{z}{2}}$$

In the sequel, we denote by $\mathfrak{D}^{\circledast}(\gamma, z)$ the set of conditions

$$\begin{aligned} \varpi &\geq \max\left(16e^2\omega_d^2\|f\|_\infty^2, 4\omega_d^2\sigma^2, z, 4\left\{\frac{2\gamma}{d}+1\right\}\right), \\ C_0 &\geq \max\left(\frac{1}{m^-}\max\left(\frac{z}{2}, \frac{\gamma}{d}\right), \varpi\max\left(\frac{2C_\infty^2}{c_2^2e^2}, \frac{2}{m^-}\right)\right), \\ 2^{Jd} &= \frac{n}{C_0\log n} \end{aligned} \qquad \mathfrak{D}^{\circledast}(\gamma, z)$$

and $\mathfrak{D}^{\bigstar}(\gamma, z)$ the set of conditions

$$\begin{aligned} \kappa &\geq \max\left(8\omega_d C_2\max(4\|f\|_\infty^2, 3\sigma^2), \max\left(z, 8(\frac{\gamma}{d}+\frac{1}{2})\right)\right), \\ C_0 &> \max(6\omega_d\|f\|_\infty C_\infty, \kappa/\{2m^-\}), \\ 2^{Jd} &= \frac{n}{C_0\log n} \end{aligned} \qquad \mathfrak{D}^{\bigstar}(\gamma, z)$$

As usual, the couple $f^\diamond, \mathfrak{D}^\diamond(\gamma, z)$ will denote indifferently $f^{\circledast}, \mathfrak{D}^{\circledast}(\gamma, z)$ or $f^{\bigstar}, \mathfrak{D}^{\bigstar}(\gamma, z)$.

Proposition 8.2. *Consider the constant m^- defined in Lemma 10.1, and c_2 and C_∞ such that, for all $j \geq 0, \eta \in \mathcal{Z}_j$, $c_2 \leq \|\psi_{j,\eta}\|_2$ and $\|\psi_{j,\eta}\|_\infty \leq C_\infty 2^{jd/2}$. For any $f \in \mathbb{L}^\infty$, $\gamma > 0$, $z > 1$, $l \in [0, z]$, and under the set of conditions $\mathfrak{D}^\diamond(\gamma, z)$ on $\varkappa$, C_0 and J, the following two inequalities hold true,*

$$\sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}^\diamond - \beta_{j,\eta}|^z \leq C\left\{t(n)^{z-l}(J+1)^z\sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l + n^{-\frac{z}{2}}\right\} \tag{13}$$

$$\sum_{j=0}^{J} 2^{j(\gamma-d)}\mathbb{E}\sum_{\eta}|\beta_{j,\eta}^\diamond - \beta_{j,\eta}|^z \leq C\left\{t(n)^{z-l}\sum_{j=0}^{J}2^{j(\gamma-d)}\left(\sum_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l\right) + n^{-\frac{z}{2}}\right\} \tag{14}$$

We delay their proofs to Section 9 and Section 10. The proof of Proposition 8.2 in the deterministic thresholding case follows very similar lines as with stochastic thresholding. For the sake of brevity, we therefore only detail the proof in the stochastic thresholding case. The interested reader is referred to `www.math.jussieu.fr/~monnier` for a full proof. Let us now prove that Proposition 8.1 and Proposition 8.2 yield the statements of Theorem 6.1.

8.1 Minimax rate for the $\mathbb{L}^\infty$-norm

We begin with eq. (10). Let's first assume that $r = q = \infty$, that is $f \in B^s_{\infty,\infty}(M)$. It is clear that

$$\begin{aligned}\mathbb{E}\|f^\diamond - f\|_\infty^z \leq C\Big\{&\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta}-\beta_{j,\eta})\psi_{j,\eta}\|_\infty^z \\ &+\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\|_\infty^z + \mathbb{E}|\hat{P}_0 - P_0 * f|^z\Big\} \triangleq I + II + III\end{aligned} \tag{15}$$

Let's now prove that each of the terms I, II and III are at most $\mathcal{O}(\{\log n\}^{2z}\{n/\log n\}^{-zs/(2s+d)})$. We indeed see immediately that III is of the good order thanks to Proposition 8.1 and $\frac{z}{2} \geq \frac{zs}{2s+d}$. For II we have

$$\begin{aligned}II^{\frac{1}{z}} &\leq C\sum_{j>J}\|\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\|_\infty \leq \sum_{j>J}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\|\psi_{j,\eta}\|_\infty \\ &\leq C\sum_{j>J}2^{-j(s+\frac{d}{2})}C_\infty 2^{\frac{jd}{2}} \leq C2^{-Js} \leq C\left\{\frac{\log n}{n}\right\}^{\frac{s}{d}}\end{aligned}$$

where the second inequality is a direct result from Lemma 2.1, eq. (6); and the third inequality comes from Theorem 3.1, eq. (8) together with eq. (5) and the fact that $f \in B^s_{\infty,\infty}(M)$. It is now enough to notice that $z\frac{s}{d} \geq \frac{zs}{2s+d}$ to conclude that II is of the good order.
For I, we apply successively the triangular inequality and Hölder with the pair of conjugate exponents z and $z/(z-1)$, $z > 1$,

$$\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta}-\beta_{j,\eta})\psi_{j,\eta}\|_\infty^z = (J+1)^{z-1}\mathbb{E}\sum_{j=0}^{J}\|\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta}-\beta_{j,\eta})\psi_{j,\eta}\|_\infty^z$$

And finally, we apply Lemma 2.1, eq. (6) to get

$$I \leq C(J+1)^{z-1}\sum_{j=0}^{J}2^{jd\frac{z}{2}}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta^\diamond_{j,\eta}-\beta_{j,\eta}|^z$$

Then we apply Proposition 8.2, eq. (13) with $\gamma = d\frac{z}{2}$ and thus under conditions $\mathfrak{D}^\diamond(dz/2, z)$. Notice in particular that this latter set of conditions is equivalent to $\mathfrak{C}^\diamond_z$. This leads us to,

$$\sum_{j=0}^{J}2^{jd\frac{z}{2}}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta^\diamond_{j,\eta}-\beta_{j,\eta}|^z \leq C\left\{t(n)^{z-l}(J+1)^z\sum_{j=0}^{J}2^{jd\frac{z}{2}}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j\eta}|^l + n^{-\frac{z}{2}}\right\}$$

where $l \in [0, z]$. We denote A, B the two terms in-between braces on the right-hand-side above. Notice first that the last term B is of the right order as $\frac{z}{2} > \frac{zs}{2s+d}$. Moreover, since $f \in B^s_{\infty\infty}(M)$, we have $\sup_{\eta\in\mathcal{Z}_j}|\beta_{j\eta}|^l \leq M^l 2^{-jl(s+\frac{d}{2})}$. The first term A can therefore be bounded by

$$A \leq Ct(n)^{z-l}(J+1)^z\sum_{j\leq J}2^{jd\frac{z}{2}}2^{-jl(s+\frac{d}{2})} = Ct(n)^{z-l}(J+1)^z\sum_{j\leq J}2^{j\alpha(l)}$$

where we have written $\alpha(l) \triangleq \frac{2s+d}{2}(l^* - l)$ and $l^* = \frac{dz}{2s+d}$. Obviously $\alpha(l)$ is a decreasing function of l. Moreover we have $z > l^*$. We can thus explore all the following cases,

i) $\boxed{l = l^*}$ implies $A \leq C(\log n)^{z+1}t(n)^{z-l^*} = C(\log n)^{z+1}[t(n)^2]^{\frac{sz}{2s+d}}$, as $z - l^* = 2\frac{sz}{2s+d}$, which is of the right order.

ii) $\boxed{l > l^*}$ implies $\alpha(l) < 0$ and $A \leq C(\log n)^z t(n)^{z-l}$. However $\frac{z-l}{2} \geq \frac{sz}{2s+d}$ is impossible as $\frac{z-l}{2} < \frac{z-l^*}{2} = \frac{sz}{2s+d}$.

iii) $\boxed{l < l^*}$ implies $\alpha(l) > 0$ and $A \leq C(\log n)^z t(n)^{z-l-\frac{2\alpha(l)}{d}}$. Notice that $z - l - \frac{2\alpha(l)}{d} = 2\frac{ls}{d}$ and $\frac{1}{2}(2\frac{ls}{d}) \geq \frac{zs}{2s+d}$ leads to $l \geq l^*$, which is impossible.

Therefore, we have $I \leq C(\log n)^{2z} \{\log n/n\}^{\frac{sz}{2s+d}}$, which finishes to prove eq. (10) of Theorem 6.1 in the case where $r = q = \infty$. The Besov embedding $B^s_{rq}(M) \subset B^{s-\frac{d}{r}}_{\infty\infty}(M)$ allows for a direct transposition of this result to the general case where r and q are chosen arbitrarily. Thus it appears that, in the general case, each of the terms I, II and III are at most $\mathcal{O}(\{\log n\}^{2z}\{n/\log n\}^{-z\vartheta^\infty})$. Which finishes the proof of eq. (10).

8.2 Minimax rate for the $\mathbb{L}^p$-norm in the regular case, $r > \frac{dp}{2s+d}$

Let's prove eq. (12), that is the regular case. Since $B^s_{r,q}(M) \subset B^s_{p,q}(M)$ for $r \geq p$, this case will be assimilated to the case $p = r$, and from now on, we only consider $r \leq p$. We have

$$\begin{aligned}\mathbb{E}\|f^\diamond - f\|_p^p \leq C\Big\{&\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta} - \beta_{j,\eta})\psi_{j,\eta}\|_p^p \\ &+ \|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\|_p^p + \mathbb{E}|\hat{P}_0 - P_0 * f|^p\Big\} \triangleq I + II + III\end{aligned} \tag{16}$$

Let's now prove that each of the terms I, II and III are at most $\mathcal{O}(\{\log n\}^p\{n/\log n\}^{-p\varsigma})$. We indeed see immediately that III is of the good order thanks to Proposition 8.1 and $\frac{p}{2} \geq \frac{sp}{2s+d}$. For II we have

$$\begin{aligned}II^{\frac{1}{p}} &\leq C\sum_{j>J}\|\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\|_p \leq C\sum_{j>J}2^{-j(s-\frac{d}{r}+\frac{d}{p})} \\ &\leq C2^{-J(s-\frac{d}{r}+\frac{d}{p})} = C\left\{\frac{\log n}{n}\right\}^{\frac{s}{d}-\frac{1}{r}+\frac{1}{p}}\end{aligned}$$

where the second inequality comes from Theorem 3.1, eq. (8), and uses the embedding $B^s_{r,q}(M) \subset B^{s-\frac{d}{r}+\frac{d}{p}}_{p,q}(M)$ for $r \leq p$. Thus, we have to find the conditions which lead to $\frac{s}{d} - \frac{1}{r} + \frac{1}{p} \geq \frac{s}{2s+d}$. Since we are in the regular case, we have $r > \frac{dp}{2s+d}$, that is $\frac{s}{2s+d} \leq \frac{rs}{dp}$. Thus, as $r \leq p$ and $s > \frac{d}{r}$, we have

$$0 \leq \frac{r}{d}\left(\frac{1}{r} - \frac{1}{p}\right)\left(s - \frac{d}{r}\right) = \left(\frac{s}{d} - \frac{1}{r} + \frac{1}{p}\right) - \frac{rs}{dp} \leq \left(\frac{s}{d} - \frac{1}{r} + \frac{1}{p}\right) - \frac{s}{2s+d}$$

which shows that II is of the good order.
For I, we use the triangular inequality together with Hölder inequality to get

$$\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta} - \beta_{j,\eta})\psi_{j,\eta}\|_p^p \leq C(J+1)^{p-1}\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}|\beta^\diamond_{j,\eta} - \beta_{j,\eta}|^p.$$

Then we apply Proposition 8.2, eq. (14) with $\gamma = d\frac{p}{2}, z = p$ and thus under conditions $\mathfrak{D}^\diamond(dp/2, p)$. This leads us to,

$$\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\mathbb{E}\sum_{\eta}|\beta^\diamond_{j,\eta} - \beta_{j,\eta}|^p \leq C\left\{t(n)^{p-l}\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\left(\sum_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l\right) + n^{-\frac{p}{2}}\right\}$$

where $l \in [0,p]$. We denote A, B the two terms in-between braces on the right-hand-side above. Notice first that the last term B is of the right order as $\frac{p}{2} > \frac{sp}{2s+d}$. Now choose $l \in [0,r]$. The embedding $B^s_{rq}(M) \subset B^s_{lq}(M)$ ensures that $\sum_{\eta\in\mathcal{Z}_j} |\beta_{j,\eta}|^l \le M^l 2^{-jl\{s+\frac{d}{2}-\frac{d}{l}\}}$. The first term A can therefore be bounded by

$$A \le Ct(n)^{p-l} \sum_{j\le J} 2^{jd(\frac{p}{2}-1)} 2^{-jl\{s+\frac{d}{2}-\frac{d}{l}\}} = Ct(n)^{p-l} \sum_{j\le J} 2^{j\alpha(l)}$$

where we have written $\alpha(l) \triangleq \frac{2s+d}{2}(l^* - l)$ and $l^* = \frac{dp}{2s+d}$. Obviously $\alpha(l)$ is a decreasing function of l. Moreover, as we are in the regular case, we have $r > l^*$. We can thus explore all cases,

i) $\boxed{l = l^*}$ implies $A \le C(\log n)t(n)^{p-l^*} = C\log n[t(n)^2]^{\frac{sp}{2s+d}}$, as $p - l^* = 2\frac{sp}{2s+d}$, which is of the right order.

ii) $\boxed{l > l^*}$ implies $\alpha(l) < 0$ and $A \le Ct(n)^{p-l}$. However $\frac{p-l}{2} \ge \frac{sp}{2s+d}$ is impossible as $\frac{p-l}{2} < \frac{p-l^*}{2} = \frac{sp}{2s+d}$.

iii) $\boxed{l < l^*}$ implies $\alpha(l) > 0$ and $A \le Ct(n)^{p-l-\frac{2\alpha(l)}{d}}$. Notice that $p - l - \frac{2\alpha(l)}{d} = 2\frac{ls}{d}$ and $\frac{1}{2}(2\frac{ls}{d}) \ge \frac{ps}{2s+d}$ leads to $l \ge l^*$, which is impossible.

Thus, we have $I \le C(\log n)^p \left\{\frac{\log n}{n}\right\}^{\frac{sp}{2s+d}}$, which finishes to prove eq. (12).

8.3 Minimax rate for the $\mathbb{L}^p$-norm in the sparse case, $r \le \frac{dp}{2s+d}$

Let's now turn to the proof of eq. (11). We proceed as above and observe first that in order to have $s > 0$ as well as $r \le \frac{pd}{2s+d} \Leftrightarrow s \le \frac{pd}{2}(\frac{1}{r} - \frac{1}{p})$, it is necessary that $p \ge r$. As above, we have

$$\begin{aligned}\mathbb{E}\|f^\diamond - f\|_p^p \le C\Big\{&\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta} - \beta_{j,\eta})\psi_{j,\eta}\|_p^p \\ &+ \|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\|_p^p + \mathbb{E}|\hat{P}_0 - P_0 * f|^p\Big\} \triangleq I + II + III\end{aligned} \tag{17}$$

Let's prove that each of the terms I, II and III are at most $\mathcal{O}(\{\log n\}^p\{n/\log n\}^{-p\vartheta})$. For $p \ge 2$, III is of the right order. For II, using again the embedding $B^s_{r,q}(M) \subset B^{s-\frac{d}{r}+\frac{d}{p}}_{p,q}(M)$ for $r \le p$ and Theorem 3.1, we have

$$II^{\frac{1}{p}} \le C\Big\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\beta_{j,\eta}\psi_{j,\eta}\Big\|_p \le C2^{-J(s-\frac{d}{r}+\frac{d}{p})} \le C\left\{\frac{\log n}{n}\right\}^{\frac{s}{d}-\frac{1}{r}+\frac{1}{p}}$$

And the constraint $p(\frac{s}{d} - \frac{1}{r} + \frac{1}{p}) \ge \frac{p}{2}\frac{s+\frac{d}{p}-\frac{d}{r}}{s+\frac{d}{2}-\frac{d}{r}}$ holds true since $s > \frac{d}{r}$.
For I, we use again the triangular inequality together with Hölder inequality to get

$$\mathbb{E}\|\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}(\beta^\diamond_{j,\eta} - \beta_{j,\eta})\psi_{j,\eta}\|_p^p \le C(J+1)^{p-1}\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}|\beta^\diamond_{j,\eta} - \beta_{j,\eta}|^p.$$

Then we apply Proposition 8.2, eq. (14) with $\gamma = d\frac{p}{2}, z = p$ and thus under conditions $\mathfrak{D}^\diamond(dp/2, p)$. This leads us to,

$$\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\mathbb{E}\sum_{\eta}|\beta^\diamond_{j,\eta} - \beta_{j,\eta}|^p \le C\left\{t(n)^{p-l}\sum_{j=0}^{J}2^{jd(\frac{p}{2}-1)}\left(\sum_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l\right) + n^{-\frac{p}{2}}\right\}$$

where $l \in [0,p]$. We denote A, B the two terms in-between braces on the right-hand-side above. Notice first that, similarly with III above, the last term B is of the right order. Now choose $l \in [r,p]$. The embedding $B^s_{rq}(M) \subset B^{s-\frac{d}{r}+\frac{d}{l}}_{lq}(M)$ ensures that $\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \leq M^l 2^{-jl\{s+\frac{d}{2}-\frac{d}{r}\}}$. The first term A can therefore be bounded by

$$A \leq Ct(n)^{p-l} \sum_{j \leq J} 2^{jd(\frac{p}{2}-1)} 2^{-jl\{s+\frac{d}{2}-\frac{d}{r}\}} = Ct(n)^{p-l} \sum_{j \leq J} 2^{j\alpha(l)}$$

where we have written $\alpha(l) \triangleq (s+\frac{d}{2}-\frac{d}{r})(l^*-l)$ and

$$l^* = \frac{d(\frac{p}{2}-1)}{s+\frac{d}{2}-\frac{d}{r}}$$

As $s > \frac{d}{r}$, $\alpha(l)$ is a decreasing function of l. Moreover

$$l^* - \frac{dp}{2s+d} = \frac{d}{r(s+\frac{d}{2}-\frac{d}{r})} \left\{ \frac{dp}{2s+d} - r \right\}$$

Since we are in the sparse case, that is $r \leq \frac{dp}{2s+d}$, and $s > d/r$, we have $r \leq \frac{dp}{2s+d} \leq l^*$. Notice now that $l^* < p$ since $l^* < p \Leftrightarrow 0 < p(s-\frac{d}{r})+d$. Thus we can choose $l \in [r,p]$ with $r \leq \frac{dp}{2s+d} \leq l^* < p$. We again optimize with respect to l,

i) $\boxed{l = l^*}$ implies $A \leq C(\log n)t(n)^{p-l^*}$ and $\frac{p-l^*}{2} = p\vartheta$, which is of the right order.

ii) $\boxed{l > l^*}$ leads to $A \leq Ct(n)^{p-l}$ and $\frac{p-l}{2} < \frac{p-l^*}{2} = p\vartheta$, which makes it impossible to have $\frac{p-l}{2} \geq p\vartheta$.

iii) $\boxed{l < l^*}$ leads to $A \leq Ct(n)^{p-l-\frac{2\alpha(l)}{d}}$ and $\frac{1}{2}(p-l-\frac{2\alpha(l)}{d}) \geq p\vartheta$ leads to $l \geq l^*$, which is impossible.

Thus, we have $I \leq C(\log n)^p \left\{ \frac{\log n}{n} \right\}^{p\vartheta}$, which finishes to prove eq. (11). □

9 Proof of Proposition 8.1

First notice that

$$\mathbb{E}\left(|\hat{P}_0 - P_0 * f|^z\right) \leq C \left\{ \mathbb{E}|\frac{1}{n}\sum_{i \leq n}\{f(T_i) - \mathbb{E}f(T)\}|^z + \mathbb{E}|\frac{1}{n}\sum_{i \leq n} V_i|^z \right\}$$

We write $X_i = f(T_i) - \mathbb{E}f(T)$. We have $|X_i| \leq 2\|f\|_\infty$, $\mathbb{E}X_i = 0$. And we define $t^2 \triangleq \mathbb{E}X_i^2$ and $s(z) \triangleq \mathbb{E}|X_i|^z < \infty$. We can thus apply Rosenthal inequality (see Petrov (1995, p. 54)) to obtain, for $z \geq 2$, $\mathbb{E}|\sum X_i/n|^z \leq C\{n^{-z+1}s(z) + n^{-\frac{z}{2}}t^z\} \leq Cn^{-\frac{z}{2}}$. For $0 < z \leq 2$, the well-known ordering of the $\mathbb{L}^p$ norms $\|.\|_{\mathbb{L}^1} \leq \|.\|_{\mathbb{L}^2}$ on probability spaces and the fact that the X_i's are mutually independent and centered lead to $\mathbb{E}|\sum X_i/n|^z \leq \left(\mathbb{E}|X_i|^2\right)^{\frac{z}{2}} n^{-\frac{z}{2}} \leq t^z n^{-\frac{z}{2}}$. Moreover notice that, since the V_i's are iid standard normal, $\sum V_i/n \sim n^{-\frac{1}{2}}\sigma Z$, where Z is standard normal. And thus $\mathbb{E}|\sum V_i/n|^z = n^{-\frac{z}{2}}\sigma^z \mathbb{E}|Z|^z \leq Cn^{-\frac{z}{2}}$, which concludes the proof. □

10 Proof of Proposition 8.2

In what follows, we will write $\|.\|_p$ to denote the usual norm of $\mathbb{L}^p(\mathbb{S}^d, \mathfrak{M}/\omega_d)$. This will make the transition between expectations over functions of uniform random variables and $\mathbb{L}^p$-norms easier. Besides, with this notation, eq. (5) transforms into

$$c_p/(\omega_d^{1/p}) \leq \|\psi_{j,\eta}\|_p 2^{-jd\{\frac{1}{2}-\frac{1}{p}\}} \leq C_p/(\omega_d^{1/p}) \tag{18}$$

10.1 Two useful Lemmas

In this paragraph we introduce two lemmas that will prove very helpful in the demonstration of Proposition 8.2. The first one is concerned with finding an upper-bound to the probability of the events $\{\varrho_{j,\eta}^2 \leq s\}$ and $\{\varrho_{j,\eta}^2 \geq t\}$. It goes as follows

Lemma 10.1. *For any $s \in (0, \|\psi_{j,\eta}\|_2^2)$ and any $t \in (\|\psi_{j,\eta}\|_2^2, +\infty)$, we have*

$$\mathbb{P}(\varrho_{j,\eta}^2 < s) \leq n^{-\nu_0(C_0,s)}, \qquad \mathbb{P}(\varrho_{j,\eta}^2 > t) \leq n^{-\nu_*(C_0,t)}$$

where we have written

$$\nu_0(C_0, s) \triangleq \frac{C_0(\|\psi_{j,\eta}\|_2^2 - s)^2}{(2C_4^4/\omega_d) + \frac{4}{3}C_\infty^2\{\|\psi_{j,\eta}\|_2^2 - s\}},$$
$$\nu_*(C_0, t) \triangleq \frac{C_0(t - \|\psi_{j,\eta}\|_2^2)^2}{(2C_4^4/\omega_d) + \frac{4}{3}C_\infty^2\{t - \|\psi_{j,\eta}\|_2^2\}}$$

Remember for later use that

$$m_{j,\eta} \triangleq \nu_0(1, \frac{\|\psi_{j,\eta}\|_2^2}{2}) = \nu_*(1, \frac{3\|\psi_{j,\eta}\|_2^2}{2}) = \frac{\|\psi_{j,\eta}\|_2^4}{(8C_4^4/\omega_d) + \frac{8}{3}C_\infty^2\|\psi_{j,\eta}\|_2^2}$$

Besides, since the map $g : x \in \mathbb{R}^+ \mapsto \frac{x^2}{8C_4^4 + \frac{8}{3}C_\infty^2 x}$ is non-decreasing and, given eq. (18), we have, for all $j \geq 0$, $\eta \in \mathcal{Z}_j$,

$$m^- \triangleq g(c_2^2/\omega_d) \leq m_{j,\eta} \leq m^+ \triangleq g(C_2^2/\omega_d)$$

Proof. This is in fact a direct application of Bernstein inequality. Let's start with the term $\mathbb{P}(\varrho_{j,\eta}^2 > t)$. Notice that

$$\begin{aligned}\mathbb{P}(\varrho_{j,\eta}^2 > t) &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \psi_{j,\eta}(T_i)^2 > t\right)\\ &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \{\psi_{j,\eta}(T_i)^2 - \|\psi_{j,\eta}\|_2^2\} > t - \|\psi_{j,\eta}\|_2^2\right)\end{aligned}$$

We write $X_i \triangleq \psi_{j,\eta}(T_i)^2 - \|\psi_{j,\eta}\|_2^2$. Now, given that $\mathbb{E}X_i = 0$, $\|X_i\|_\infty \leq 2\|\psi_{j,\eta}(.)^2\|_\infty \leq 2C_\infty^2 2^{jd} \leq 2C_\infty^2 2^{Jd}$ and $\mathbb{E}X_i^2 \leq \|\psi_{j,\eta}\|_4^4 \leq C_4^4\{2^{jd(\frac{1}{2}-\frac{1}{4})}\}^4/\omega_d \leq C_4^4 2^{Jd}/\omega_d$ and $2^{Jd} \leq n/\{C_0 \log n\}$, we can apply Bernstein inequality. For $t - \|\psi_{j,\eta}\|_2^2 > 0$, we get

$$\begin{aligned}\mathbb{P}(\varrho_{j,\eta}^2 > t) &\leq \exp\left(-\frac{n(t - \|\psi_{j,\eta}\|_2^2)^2}{2\mathbb{E}\{\psi_{j,\eta}(T_i)^2 - \|\psi_{j,\eta}\|_2^2\}^2 + \frac{2}{3}t\|\psi_{j,\eta}(.)^2 - \|\psi_{j,\eta}\|_2^2\|_\infty}\right)\\ &\leq \exp\left(-\frac{n(t - \|\psi_{j,\eta}\|_2^2)^2}{2^{Jd}(2C_4^4/\omega_d + \frac{4}{3}C_\infty^2\{t - \|\psi_{j,\eta}\|_2^2\})}\right)\\ &\leq \exp\left(-\frac{C_0(t - \|\psi_{j,\eta}\|_2^2)^2}{2C_4^4/\omega_d + \frac{4}{3}C_\infty^2\{t - \|\psi_{j,\eta}\|_2^2\}}\log n\right)\\ &= \exp\left(-\nu_*(C_0, t)\log n\right)\end{aligned}$$

As regards the term $\mathbb{P}(\varrho_{j,\eta}^2 < s)$, write

$$\begin{aligned}\mathbb{P}(\varrho_{j,\eta}^2 < s) &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \psi_{j,\eta}(T_i)^2 < s\right)\\ &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \{\psi_{j,\eta}(T_i)^2 - \|\psi_{j,\eta}\|_2^2\} < -(\|\psi_{j,\eta}\|_2^2 - s)\right)\\ &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (-X_i) > \|\psi_{j,\eta}\|_2^2 - s\right)\end{aligned}$$

Besides $(-X_i)$ verifies the same hypotheses as X_i above. Thus for $s \in (0, \|\psi_{j,\eta}\|_2^2)$, we have $\|\psi_{j,\eta}\|_2^2 - s > 0$ and we are brought back to the preceding case. Which concludes the proof. □

The second Lemma focuses on finding an upper bound on the probability that $\kappa_{j,\eta}t(n)$ be beyond or under a given constant value $g_{j,\eta}$.

Lemma 10.2. *Let $z > 0$ be given. Let $g_{j,\eta}$ be a strictly positive constant that eventually depends on j and η. For any $l \in [0, z]$, any $s \in (0, \|\psi_{j,\eta}\|_2^2)$ and any $t \in (\|\psi_{j,\eta}\|_2^2, +\infty)$, we have*

$$\mathbb{P}(g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n)) \leq C g_{j,\eta}^{2l} t(n)^{-2l} + n^{-\nu_0(C_0,s)} \tag{19}$$

$$\mathbb{P}(g_{j,\eta} < 2\kappa_{j,\eta}t(n)) \leq C g_{j,\eta}^{l-z} t(n)^{z-l} + n^{-\nu_*(C_0,t)} \tag{20}$$

In particular eq. (19) together with the obvious fact that, for any $x, y \geq 0$, $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$, lead to

$$\mathbb{P}(g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n))^{\frac{1}{2}} \leq C g_{j,\eta}^{l} t(n)^{-l} + n^{-\frac{\nu_0(C_0,s)}{2}} \tag{21}$$

Besides, if $g_{j,\eta} \leq C2^{-j\frac{d}{2}}$, we have obviously that

$$g_{j,\eta}^z \leq C g_{j,\eta}^l \tag{22}$$

Notice in particular that Lemma 2.1, eq. (7) and the fact that $f \in \mathbb{L}^\infty$ lead to $|\beta_{j,\eta}| \leq C2^{-j\frac{d}{2}}$, which in turn implies that both $|\beta_{j,\eta}|$ and $\sup_{\eta\in\mathcal{Z}_j} |\beta_{j,\eta}|$ verify eq. (22).

Proof. Let's start with the proof of eq. (19). For any $s \in (0, \|\psi_{j,\eta}\|_2^2)$, we have

$$\mathbb{E}\mathbb{1}_{\{g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n)\}} = \mathbb{E}\mathbb{1}_{\{g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n)\}}(\mathbb{1}_{\{\varrho_{j,\eta}^2 \geq s\}} + \mathbb{1}_{\{\varrho_{j,\eta}^2 \leq s\}})$$

For any $l \geq 0$, the left term can be bounded as follows

$$\begin{aligned}\mathbb{E}\mathbb{1}_{\{g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n)\}}\mathbb{1}_{\{\varrho_{j,\eta}^2 \geq s\}} &\leq \mathbb{E}\mathbb{1}_{\{g_{j,\eta} \geq \frac{\varpi\sqrt{s}}{2}t(n)\}}\\ &\leq \mathbb{1}_{\{g_{j,\eta} \geq \frac{\varpi\sqrt{s}}{2}t(n)\}} g_{j,\eta}^{2l}\left\{\frac{\varpi\sqrt{s}}{2}t(n)\right\}^{-2l}\\ &\leq C g_{j,\eta}^{2l} t(n)^{-2l}\end{aligned}$$

As regards the right one, we have

$$\mathbb{E}\mathbb{1}_{\{g_{j,\eta} \geq \frac{\kappa_{j,\eta}}{2}t(n)\}}\mathbb{1}_{\{\varrho_{j,\eta}^2 \leq s\}} \leq \mathbb{P}(\varrho_{j,\eta}^2 \leq s) \leq n^{-\nu_0(C_0,s)}$$

where the last inequality is a direct application of Lemma 10.1. This finishes to prove eq. (19). Let's now turn to eq. (20). Its proof follows the same lines. For any $t \in (\|\psi_{j,\eta}\|_2^2, +\infty)$, we have

$$\mathbb{E}\mathbb{1}_{\{g_{j,\eta} < 2\kappa_{j,\eta}t(n)\}} = \mathbb{E}\mathbb{1}_{\{g_{j,\eta} < 2\kappa_{j,\eta}t(n)\}}(\mathbb{1}_{\{\varrho_{j,\eta}^2 \leq t\}} + \mathbb{1}_{\{\varrho_{j,\eta}^2 > t\}})$$

The term on the right-hand-side can therefore be bounded by

$$\mathbb{E}\mathbb{1}_{\{g_{j,\eta}<2\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\varrho_{j,\eta}^2>t\}} \leq \mathbb{P}(\varrho_{j,\eta}^2 > t) \leq n^{-\nu_*(C_0,t)}$$

where the second inequality is a direct application of Lemma 10.1. Let l be chosen such that $l \in [0,z]$, that is $z-l \geq 0$ and turn to the other term. We have

$$\begin{aligned}\mathbb{E}\mathbb{1}_{\{g_{j,\eta}<2\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\varrho_{j,\eta}^2\leq t\}} &\leq \mathbb{E}\mathbb{1}_{\{g_{j,\eta}<2\sqrt{t}\varpi t(n)\}} \\ &\leq g_{j,\eta}^{l-z}(2\sqrt{t}\varpi t(n))^{z-l} = Cg_{j,\eta}^{l-z}t(n)^{z-l}\end{aligned}$$

which finishes to prove eq. (20). □

10.2 Proof of Proposition 8.2 in the stochastic thresholding case

The proof relies on the following Lemma, which gives upper-bounds on various measures of the deviation of $\hat{\beta}_{j,\eta}$ from the true coefficient $\beta_{j,\eta}$, the last one in eq. (25) being threshold-dependent.

Lemma 10.3. *Let's define the constant C_∞ by $\|\psi_{j,\eta}\|_\infty \leq C_\infty 2^{jd/2}$ as in eq. (5) and $C_0 > 0$. For any $q \geq 1$, there exist constants $C_1(q), C_2(q)$ such that, as soon as the following conditions are verified,*

$$\begin{aligned} &f \in \mathbb{L}^\infty, && \varpi \geq 4\omega_d^2 \max(4e^2\|f\|_\infty^2, \sigma^2), \\ &C_0 > \varpi \max(2C_\infty^2/\{c_2^2e^2\}, 2/m^-), && 2^J \triangleq n/\{C_0 \log n\}\end{aligned}$$

we can write,

$$\mathbb{E}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^q \leq C_1(q)n^{-\frac{q}{2}} \tag{23}$$

$$\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^q \leq C_2(q)(j+1)^q n^{-\frac{q}{2}} \tag{24}$$

$$\mathbb{P}\left(|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq \varpi\varrho_{j,\eta}t(n)\right) \leq 2n^{-\frac{\varpi}{2}} \tag{25}$$

Proof. **Proof of eq. (25).** A triangular inequality gives $|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \leq |\zeta_{j,\eta}^* - \beta_{j,\eta}| + |\gamma_{j,\eta}^*|$. We notice in particular that, for any $u > 0$, $\{|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq u\} \subset \{|\zeta_{j,\eta}^* - \beta_{j,\eta}| \geq \frac{u}{2}\} \bigcup \{|\gamma_{j,\eta}^*| \geq \frac{u}{2}\}$. And therefore

$$\mathbb{P}(|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq u) \leq \mathbb{P}(|\zeta_{j,\eta}^* - \beta_{j,\eta}| \geq \frac{u}{2}) + \mathbb{P}(|\gamma_{j,\eta}^*| \geq \frac{u}{2}) \triangleq I(u) + II(u) \tag{26}$$

Obviously, this result holds true for a stochastic u as well. Therefore we can take $u = \kappa_{j,\eta}t(n)$, where we have written $\kappa_{j,\eta} = \varpi\varrho_{j,\eta}$. Thanks to eq. (26), we only need to bound $I(\kappa_{j,\eta}t(n)/2)$ and $II(\kappa_{j,\eta}t(n)/2)$. We first deal with this latter term.

Upper bound for $II(\kappa_{j,\eta}t(n)/2)$. Recall from Section 5 that $|\gamma_{j,\eta}^*| \sim \omega_d\varrho_{j,\eta}\sigma|\xi|/\sqrt{n}$, where $\xi \sim \mathcal{N}(0,1)$ conditionally on the T_i's. We can thus write

$$\begin{aligned} II(\kappa_{j,\eta}t(n)/2) &= \mathbb{P}(|\gamma_{j,\eta}^*| \geq \frac{\kappa_{j,\eta}t(n)}{2}) = \mathbb{P}(|\xi| \geq \frac{\varpi\sqrt{\log n}}{2\sigma\omega_d}) \\ &\leq \frac{4\sigma\omega_d}{\varpi\sqrt{2\pi\log n}}\exp\left(-\frac{\varpi^2\log n}{8\sigma^2\omega_d^2}\right) \leq n^{-\varpi^2/8\sigma^2\omega_d^2}\end{aligned}$$

where the before last inequality results from a conditioning with respect to the T_i's and a regular Gaussian tail-inequality and the last inequality holds true for n big enough. So in order to get eq. (25), we just need to pick ϖ such that $\varpi \geq 4\sigma^2\omega_d^2$. We now turn to the other term.

Upper bound for $I(\kappa_{j,\eta}t(n)/2)$. We write $X_i \triangleq \omega_d f(T_i)\psi_{j,\eta}(T_i) - \beta_{j,\eta}$. Obviously $\zeta_{j,\eta}^* - \beta_{j,\eta} =$

$\sum X_i/n$, where the sum is over all the X_i's, $i = 1, \ldots, n$. Since $\kappa_{j,\eta} = \varpi \varrho_{j,\eta}$, the problem reduces to finding an upper bound to $I(\varrho_{j,\eta}\delta)$ where δ is a given constant. We will subsequently obtain an upper bound for $I(\kappa_{j,\eta}t(n)/2)$ by choosing $\delta \equiv \varpi t(n)/2$. With these notations, we have $I(\varrho_{j,\eta}\delta) = \mathbb{P}(|\sum X_i| \geq n\varrho_{j,\eta}\delta)$. In a first step, we remove the absolute values and focus on finding an upper bound to $\mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}\delta)$. Let h be a strictly positive non-decreasing function. We can write

$$h(\sum X_i) \geq h(\sum X_i)\mathbb{1}_{\{\sum X_i \geq n\varrho_{j,\eta}\delta\}} \geq h(n\varrho_{j,\eta}\delta)\mathbb{1}_{\{\sum X_i \geq n\varrho_{j,\eta}\delta\}}$$

Taking the expectation leads to

$$\mathbb{E}h(\sum X_i) \geq \mathbb{E}h(n\varrho_{j,\eta}\delta)\mathbb{1}_{\{\sum X_i \geq \varrho_{j,\eta}\delta\}} \tag{27}$$

Since h is a strictly positive function, we can apply a reverse Hölder inequality to the right-hand-side (Adams and Fournier (2003, Theorem 2.12)). That is, with $p = 1/2$ and $q = -1$ as conjugate exponents

$$\mathbb{E}h(n\varrho_{j,\eta}\delta)\mathbb{1}_{\{\sum X_i \geq n\varrho_{j,\eta}\delta\}} \geq \mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}\delta)^2 \left(\mathbb{E}\frac{1}{h(n\varrho_{j,\eta}\delta)}\right)^{-1} \tag{28}$$

Let's choose $h(x) \triangleq e^{\lambda x}$, where $\lambda > 0$. Combining eq. (27) and eq. (28), we get

$$\mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}n\delta)^2 \leq \mathbb{E}e^{-\lambda n\varrho_{j,\eta}\delta}\mathbb{E}e^{\lambda \sum X_i}$$

This together with the independence of the T_i's and thus of the X_i's leads to

$$\mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}\delta)^2 \leq \mathbb{E}e^{-\lambda n\varrho_{j,\eta}\delta}\left(\mathbb{E}e^{\lambda X_i}\right)^n \tag{29}$$

Let's first look for an upper bound to $(\mathbb{E}e^{\lambda X_i})^n = \exp(n \log \mathbb{E}e^{\lambda X_i})$. Obviously

$$e^{\lambda X_i} = 1 + \lambda X_i + (e^{\lambda X_i} - 1 - \lambda X_i) = 1 + \lambda X_i + \lambda^2 X_i^2 \theta(\lambda X_i) \tag{30}$$

where we have written $\theta(x) = (e^x - 1 - x)/(x^2)\mathbb{1}_{\{x \neq 0\}} + (1/2)\mathbb{1}_{\{x=0\}}$. As is easily verified, θ is a positive non-decreasing function. Moreover, notice that $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 \leq \omega_d^2 \|f\|_\infty^2 \|\psi_{j,\eta}\|_2^2$ and $\|X_i\|_\infty \leq 2\omega_d \|f\|_\infty \|\psi_{j,\eta}\|_\infty$ and $\|\psi_{j,\eta}\|_\infty \leq C_\infty 2^{Jd/2}$. Therefore, taking the expectation in eq. (30) leads to

$$\mathbb{E}e^{\lambda X_i} \leq 1 + \lambda^2 \omega_d^2 \|f\|_\infty^2 \|\psi_{j,\eta}\|_2^2 \theta(2\lambda\omega_d \|f\|_\infty C_\infty 2^{Jd/2})$$

Notice finally that $\log(1+u) \leq u$ for all $u \geq 0$, so that

$$\begin{aligned}(\mathbb{E}e^{\lambda X_i})^n &\leq \exp\{n \log(1 + \lambda^2 \omega_d^2 \|f\|_\infty^2 \|\psi_{j,\eta}\|_2^2 \theta(\lambda\omega_d \|f\|_\infty C_\infty 2^{Jd/2}))\} \\ &\leq \exp\{n\lambda^2 \omega_d^2 \|f\|_\infty^2 \|\psi_{j,\eta}\|_2^2 \theta(2\lambda\omega_d \|f\|_\infty C_\infty 2^{Jd/2})\}\end{aligned} \tag{31}$$

We now turn to the problem of bounding $\mathbb{E}e^{-\lambda n\varrho_{j,\eta}\delta}$. Notice that

$$\mathbb{E}e^{-\lambda n\varrho_{j,\eta}\delta} \leq \mathbb{E}e^{-\lambda n\varrho_{j,\eta}\delta}\mathbb{1}_{\{\varrho_{j,\eta}^2 \leq t\}} + \mathbb{E}e^{-\lambda n\sqrt{t}\delta}\mathbb{1}_{\{\varrho_{j,\eta}^2 > t\}} \triangleq A_0 + B_0 \tag{32}$$

We have obviously $B_0 \leq e^{-\lambda n\sqrt{t}\delta}$. As regards A_0, notice that for any $t \in (0, \|\psi_{j,\eta}\|_2^2)$, we have $A_0 \leq \mathbb{P}(\varrho_{j,\eta}^2 \leq t) \leq n^{-\nu_0(C_0,t)}$ where the last inequality is a direct application of Lemma 10.1. Combining the upper bounds on A_0 and B_0 together with eq. (32), eq. (31) and eq. (29), we obtain

$$\begin{aligned}\mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}\delta)^2 &\leq (e^{-\lambda n\sqrt{t}\delta} + e^{-\nu_0(C_0,t)\log n})e^{n\lambda^2\omega_d^2\|f\|_\infty^2\|\psi_{j,\eta}\|_2^2\theta(2\lambda\omega_d\|f\|_\infty C_\infty 2^{Jd/2})} \\ &\triangleq e^{-\nu_1(t,C_0,\delta,\lambda)\log n} + e^{-\nu_2(t,C_0,\delta,\lambda)\log n}\end{aligned}$$

where

$$\nu_1(t, C_0, \delta, \lambda) \triangleq \frac{n}{\log n}\sqrt{t}\lambda\delta - \frac{n}{\log n C_\infty^2 2^{Jd}}\frac{\|\psi_{j,\eta}\|_2^2}{4}\theta_1(2\lambda\omega_d\|f\|_\infty C_\infty 2^{Jd/2})$$
$$\nu_2(t, C_0, \delta, \lambda) \triangleq \nu_0(C_0, t) - \frac{n}{\log n C_\infty^2 2^{Jd}}\frac{\|\psi_{j,\eta}\|_2^2}{4}\theta_1(2\lambda\omega_d\|f\|_\infty C_\infty 2^{Jd/2})$$

and $\theta_1(x) = x^2\theta(x)$. Recall that $2^{Jd} = n/\{C_0 \log n\}$, that is $2^{-Jd/2} = \sqrt{C_0}t(n)$. As $C_\infty 2^{Jd/2}$ explodes, it is clear that we have to take $\lambda = \lambda_0 \triangleq a2^{-Jd/2} = a\sqrt{C_0}t(n),\ a > 0$ for $\nu_i,\ i = 1, 2$ to have a lower bound. Besides we take $\delta = \delta_0 \triangleq \varpi t(n)/2$. With these parameters values, we have $\theta_1(2\lambda_0\omega_d\|f\|_\infty C_\infty 2^{Jd/2}) = \theta_1(2\omega_d\|f\|_\infty C_\infty a) \leq 2\omega_d^2\|f\|_\infty^2 C_\infty^2 a^2 e^{2\omega_d\|f\|_\infty C_\infty a}$, where the last inequality follows from the mean value theorem. Hence, if we choose $C_0 \triangleq b^2\varpi,\ b > 0$, we obtain

$$\nu_1(t, b^2\varpi, \delta_0, \lambda_0) \geq \varpi\left(\sqrt{t}ab\frac{\sqrt{\varpi}}{2} - \frac{(ab)^2}{2}\omega_d^2\|f\|_\infty^2\|\psi_{j,\eta}\|_2^2 e^{2\omega_d\|f\|_\infty C_\infty a}\right) \tag{33}$$
$$\nu_2(t, b^2\varpi, \delta_0, \lambda_0) \geq \varpi\left(b^2\nu_0(1, t) - \frac{(ab)^2}{2}\omega_d^2\|f\|_\infty^2\|\psi_{j,\eta}\|_2^2 e^{2\omega_d\|f\|_\infty C_\infty a}\right) \tag{34}$$

In order to verify eq. (25), we just need to pick $t \in (0, \|\psi_{j,\eta}\|_2^2)$, $a > 0$ and $b > 0$ such that $\nu_i(t, b^2\varpi, \delta_0, \lambda_0) \geq \varpi,\ i = 1, 2$. Given eq. (33) and eq. (34), it leads to the following three constraints on the parameters ϖ, a, b, t

$$\sqrt{\varpi} \geq \frac{2}{ab\sqrt{t}}\left(1 + \frac{(ab)^2}{2}\omega_d^2\|f\|_\infty^2\|\psi_{j,\eta}\|_2^2 e^{2\omega_d\|f\|_\infty C_\infty a}\right)$$
$$\nu_0(1, t) \geq \frac{1}{b^2} + \frac{a^2}{2}\|f\|_\infty^2\omega_d^2\|\psi_{j,\eta}\|_2^2 e^{2\omega_d\|f\|_\infty C_\infty a}$$
$$t \in (0, \|\psi_{j,\eta}\|_2^2)$$

which are obviously always feasible. In particular, we can take $t = t_0 \triangleq \|\psi_{j,\eta}\|_2^2/2$, $a = c_b\sqrt{2}/(b\omega_d\|f\|_\infty)$, $c_b = 1/(e\|\psi_{j,\eta}\|_2)$, $b \geq \max(C_\infty\sqrt{2}/(c_2 e), \sqrt{2/m^-})$, where c_2 and m^- have been defined in Lemma 10.1 and $\varpi \geq \{4\omega_d\|f\|_\infty/(\sqrt{2}c_b\sqrt{t_0})\}^2 = 16e^2\omega_d^2\|f\|_\infty^2$. Under these conditions, we have thus proved that $\mathbb{P}(\sum X_i \geq n\varrho_{j,\eta}\varpi t(n)) \leq n^{-\varpi/2}$. Besides, it is clear that $\mathbb{P}(\sum X_i \leq -n\varrho_{j,\eta}\varpi t(n)) = \mathbb{P}(\sum(-X_i) \geq n\varrho_{j,\eta}\varpi t(n))$. And the $(-X_i)$'s verify the same properties as the X_i's, which prompts us to apply the same reasoning as above. This leads straightforwardly to $\mathbb{P}(|\sum X_i| \geq n\varrho_{j,\eta}\varpi t(n)) \leq 2n^{-\varpi/2}$ and finishes the proof of eq. (25).

Proof of eq. (24). First notice that with $\#\mathcal{Z}_j = 2^{jd}$ and for $q \geq 1$ we have

$$\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^q = \int_{\mathbb{R}^+} qu^{q-1}\mathbb{P}(\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq u)du$$
$$\leq \int_{\mathbb{R}^+} qu^{q-1}\left\{1 \wedge 2^{jd}\mathbb{P}(|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq u)\right\}du \tag{35}$$

In order to prove eq. (24), we need to find an upper bound to right-hand-side of eq. (35). However we have shown earlier in eq. (26) that

$$\mathbb{P}(|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq u) \leq I(u) + II(u)$$

Upper bound for $II(u)$. Recall from Section 5 that $|\gamma_{j,\eta}^*| \sim \omega_d\sigma\varrho_{j,\eta}|\xi|/\sqrt{n}$, where $\xi \sim \mathcal{N}(0, 1)$ conditionally on the T_i's. Therefore we can write, for any $t > 0$,

$$\mathbb{P}(|\gamma_{j,\eta}^*| \geq \frac{u}{2}) = \mathbb{P}(\omega_d\sigma\varrho_{j,\eta}|\xi| \geq \frac{u}{2}\sqrt{n}; \varrho_{j,\eta}^2 > t) + \mathbb{P}(\omega_d\sigma\varrho_{j,\eta}|\xi| \geq \frac{u}{2}\sqrt{n}; \varrho_{j,\eta}^2 \leq t)$$
$$\triangleq A_1(u) + A_2(u)$$

We use a regular Gaussian tail-inequality to bound $A_2(u)$,

$$\begin{aligned}
A_2(u) &= \mathbb{E}\mathbb{1}_{\{\omega_d\sigma\varrho_{j,\eta}|\xi|\geq\frac{u}{2}\sqrt{n}\}}\mathbb{1}_{\{\varrho^2_{j,\eta}\leq t\}} = \mathbb{E}\mathbb{1}_{\{\varrho^2_{j,\eta}\leq t\}}\mathbb{E}[\mathbb{1}_{\{\omega_d\sigma\varrho_{j,\eta}|\xi|\geq\frac{u}{2}\sqrt{n}\}}|T_1,\dots,T_n] \\
&\leq \mathbb{E}\mathbb{1}_{\{\varrho^2_{j,\eta}\leq t\}}\left\{1\wedge\frac{4\sigma\omega_d\varrho_{j,\eta}}{u\sqrt{2\pi n}}\exp\left(-\frac{u^2n}{8\omega_d^2\sigma^2\varrho^2_{j,\eta}}\right)\right\} \\
&\leq \frac{4\omega_d\sigma\sqrt{t}}{u\sqrt{2\pi n}}\exp\left(-\frac{u^2n}{8\omega_d^2\sigma^2t}\right)
\end{aligned}$$

In order to find an upper bound for $A_1(u)$, we use the independence of the T_i's from the V_i's together with the fact that $\|\psi_{j,\eta}\|_\infty \leq C_\infty 2^{jd/2}$, and thus $\varrho_{j,\eta}\leq C_\infty 2^{jd/2}$. We write indeed

$$\begin{aligned}
A_1(u) &= \mathbb{P}(\omega_d\sigma\varrho_{j,\eta}|\xi|\geq\frac{u}{2}\sqrt{n};\varrho^2_{j,\eta}>t) = \mathbb{E}\mathbb{E}\left[\mathbb{1}_{\{\omega_d\sigma\varrho_{j,\eta}|\xi|\geq\frac{u}{2}\sqrt{n}\}}|T_1,\dots,T_n\right]\mathbb{1}_{\{\varrho^2_{j,\eta}>t\}} \\
&\leq \mathbb{E}\mathbb{E}\left[\mathbb{1}_{\{|\xi|\geq\frac{u}{2\omega_d\sigma C_\infty}\sqrt{n}2^{-jd/2}\}}|T_1,\dots,T_n\right]\mathbb{1}_{\{\varrho^2_{j,\eta}>t\}} \\
&\leq \mathbb{E}\left\{1\wedge\frac{4C_\infty\omega_d\sigma 2^{jd/2}}{u\sqrt{2\pi n}}\exp\left(-\frac{u^2}{8\omega_d^2\sigma^2C_\infty^2}n2^{-jd}\right)\right\}\mathbb{1}_{\{\varrho^2_{j,\eta}>t\}} \\
&\leq \left\{1\wedge\frac{4C_\infty\omega_d\sigma 2^{Jd/2}}{u\sqrt{2\pi n}}\exp\left(-\frac{u^2}{8\omega_d^2\sigma^2C_\infty^2}n2^{-Jd}\right)\right\}\mathbb{P}(\varrho^2_{j,\eta}>t) \\
&\leq \left\{1\wedge\frac{4C_\infty\omega_d\sigma}{u\sqrt{2\pi C_0}}\exp\left(-\frac{u^2C_0}{8\omega_d^2\sigma^2C_\infty^2}\right)\right\}\exp\left(-\nu_*(C_0,t)\log n\right) \qquad (36)
\end{aligned}$$

where the last inequality is a direct application of Lemma 10.1 with $t\in(\|\psi_{j,\eta}\|_2^2,\infty)$.
Upper bound for $I(u)$. We can write

$$\zeta^*_{j,\eta}-\beta_{j,\eta} = \frac{\omega_d}{n}\sum_{i\leq n}\{f(T_i)\psi_{j,\eta}(T_i)-\mathbb{E}f(T_i)\psi_{j,\eta}(T_i)\}$$

Since $\|\psi_{j,\eta}\|_\infty\leq C_\infty 2^{\frac{jd}{2}}$, we have $\|f(.)\psi_{j,\eta}(.)-\mathbb{E}f(T_i)\psi_{j,\eta}(T_i)\|_\infty\leq 2\|f\|_\infty C_\infty 2^{\frac{jd}{2}}\leq 2\|f\|_\infty C_\infty 2^{\frac{Jd}{2}}$. Moreover we have $w^2\triangleq\mathbb{E}\{f(T_i)\psi_{j,\eta}(T_i)-\mathbb{E}f(T_i)\psi_{j,\eta}(T_i)\}^2\leq\|f\|_\infty^2\|\psi_{j,\eta}\|_2^2$. We can thus apply a Bernstein-type inequality to find

$$\begin{aligned}
\mathbb{P}\left(|\zeta^*_{j,\eta}-\beta_{j,\eta}|\geq\frac{u}{2}\right) &= \mathbb{P}\left(\frac{\omega_d}{n}|\sum_{i\leq n}\{f(T_i)\psi_{j,\eta}(T_i)-\mathbb{E}f(T_i)\psi_{j,\eta}(T_i)\}|\geq\frac{u}{2}\right) \\
&\leq 2\exp\left(-\frac{nu^2}{8\omega_d^2w^2+\frac{8}{3}\omega_d\|f\|_\infty C_\infty 2^{\frac{Jd}{2}}u}\right)
\end{aligned}$$

And, using the convexity of the exponential map, we get,

$$\begin{aligned}
\exp\left(-\frac{nu^2}{8\omega_d^2w^2+(8/3)\omega_d\|f\|_\infty C_\infty 2^{Jd/2}u}\right) &\leq \exp\left(-\frac{nu^2}{16\omega_d^2w^2}\right) \\
&\quad+\exp\left(-\frac{nu}{(16/3)\omega_d\|f\|_\infty C_\infty 2^{Jd/2}}\right)
\end{aligned}$$

Upper bound to $\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat\beta_{j,\eta}-\beta_{j,\eta}|^q$. We have shown above that we can write $\mathbb{P}(|\hat\beta_{j,\eta}-\beta_{j,\eta}|\geq u)\leq I(u)+II(u)\leq I(u)+\{A_1(u)+A_2(u)\}$ and we have computed upper bounds to the three terms on the right-hand-side. Let's now use these results to actually prove eq. (24). Notice that for $x,y\in\mathbb{R}^+$ we have $1\wedge(x+y)\leq 1\wedge x+1\wedge y$. As the three terms $A_1(u),A_2(u)$ and $I(u)$ are

positive, we deduce from eq. (35) that

$$\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^q \le \int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}I(u)\}du + \int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}A_1(u)\}du + \int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}A_2(u)\}du \tag{37}$$

We proceed to bound the three integrals that appear on the right-hand-side of eq. (37). Notice first that, since $2^{Jd}\le n/C_0$,

$$\begin{aligned}\int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}A_1(u)\}du &\le \int_{\mathbb{R}^+} qu^{q-1}2^{Jd}A_1(u)du\\ &\le ne^{-\nu_*(C_0,t)\log n}\int_{\mathbb{R}^+}\frac{q}{C_0}u^{q-1}\left\{1\wedge\frac{4\omega_d\sigma C_\infty}{u\sqrt{2\pi C_0}}\exp\left(-\frac{u^2C_0}{8\omega_d^2\sigma^2C_\infty^2}\right)\right\}du\\ &= C(q)n^{-\nu_*(C_0,t)+1}\end{aligned}$$

where we have denoted the integral on the right-hand-side of the before-last line by $C(q)$. Therefore, in order to get eq. (24), it is enough to choose t such that $\nu_*(C_0,t)-1\ge q/2$, which is always possible since $\nu_*(C_0,t)$ is a positive, strictly increasing and unbounded function of t for $t>\|\psi_{j,\eta}\|_2^2$. Thus we can pick any $t\ge t_q$, where $t_q\triangleq[\nu_*^{-1}(C_0,.)](1+q/2)$, with obvious notations. Let's now fix $t>t_q$ and turn to the integral of $A_2(u)$. Notice first that for $u^2\ge 4^2jd\omega_d^2\sigma^2t/n$, that is $u\ge u^*\triangleq 4\sqrt{jd\omega_d^2\sigma^2t/n}$, we have

$$2^{jd}\exp\left(-\frac{u^2n}{8\omega_d^2\sigma^2t}\right)\le\exp\left(-\frac{u^2n}{16\omega_d^2\sigma^2t}-\frac{u^2n}{16\omega_d^2\sigma^2t}+jd\right)\le\exp\left(-\frac{u^2n}{16\omega_d^2\sigma^2t}\right)$$

Thus we can write

$$\begin{aligned}&\int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}A_2(u)\}du\\ &\le\int_{0\le u\le u^*}qu^{q-1}du+\int_{u\ge u^*}qu^{q-1}\left\{1\wedge 2^{jd}A_2(u)\right\}du\\ &\le(4\sqrt{dt})^qj^{\frac{q}{2}}n^{-\frac{q}{2}}+\int_{u\ge u^*}qu^{q-1}\left\{1\wedge\frac{4\omega_d\sigma\sqrt{t}}{u\sqrt{2\pi n}}2^{jd}\exp\left(-\frac{u^2n}{8\omega_d^2\sigma^2t}\right)\right\}du\\ &\le(4\sqrt{dt})^qj^{\frac{q}{2}}n^{-\frac{q}{2}}+\int_{u\ge u^*}qu^{q-1}\left\{1\wedge\frac{4\omega_d\sigma\sqrt{t}}{u\sqrt{2\pi n}}\exp\left(-\frac{u^2n}{16\omega_d^2\sigma^2t}\right)\right\}du\\ &\le(4\sqrt{dt})^qj^{\frac{q}{2}}n^{-\frac{q}{2}}+n^{-\frac{q}{2}}\int_{v\ge 4\sqrt{jd\omega_d^2\sigma^2t}}qv^{q-1}\left\{1\wedge\frac{4\omega_d\sigma\sqrt{t}}{v\sqrt{2\pi}}\exp\left(-\frac{v^2}{16\omega_d^2\sigma^2t}\right)\right\}dv\\ &\le C'(q)(j+1)^{\frac{q}{2}}n^{-\frac{q}{2}}\end{aligned}$$

where the before last inequation results from the change of variable $v=u\sqrt{n}$ and the last inequation from the fact that, for $j\ge 0$, the integral on the domain $v\ge 4\sqrt{jd\omega_d^2\sigma^2t}$ is smaller than the integral on $v\in\mathbb{R}^+$. Finally, we deal with the integral of $I(u)$ in eq. (37). It can be bounded using the same technique as for the integral of $A_2(u)$. We have indeed, for $u\ge u^*\triangleq(j\omega_d/\sqrt{n})\max(4w\sqrt{2d},32d\|f\|_\infty C_\infty/3)$,

$$\begin{aligned}&2^{jd}\left\{\exp\left(-\frac{nu^2}{16\omega_d^2w^2}\right)+\exp\left(-\frac{u\sqrt{n}}{(16/3)\omega_d\|f\|_\infty C_\infty}\right)\right\}\\ &\le\exp\left(-\frac{nu^2}{32\omega_d^2w^2}\right)+\exp\left(-\frac{u\sqrt{n}}{(32/3)\omega_d\|f\|_\infty C_\infty}\right)\end{aligned}$$

Thus we can write

$$\begin{aligned}
&\int_{\mathbb{R}^+} qu^{q-1}\{1\wedge 2^{jd}I(u)\}du \\
&\leq \int_{0\leq u\leq u^*} qu^{q-1}du + \int_{u\geq u^*} qu^{q-1}\left\{1\wedge 2^{jd}I(u)\right\}du \\
&\leq a^q j^q n^{-\frac{q}{2}} + \int_{u\geq u^*} qu^{q-1}1\wedge\left\{\exp\left(-\frac{nu^2}{32\omega_d^2 w^2}\right)+\exp\left(-\frac{u\sqrt{n}}{\frac{32}{3}\omega_d\|f\|_\infty C_\infty}\right)\right\}du \\
&\leq C''(q)(j+1)^q n^{-\frac{q}{2}}
\end{aligned}$$

where in the last equation we have made the change of variable $v = u\sqrt{n}$ and used the fact that for $j \geq 0$ the resulting integral is bounded by a constant, since $\omega_d^2 w^2 \leq 2\omega_d\|f\|_\infty^2 C_2^2$. Which concludes the proof of eq. (24).

Proof of eq. (23). We could prove eq. (23) using the same arguments as for eq. (24). However, we can also see eq. (23) as a direct consequence of Rosenthal inequality (see Petrov (1995, p. 54)). We have indeed, for $q > 0$,

$$\mathbb{E}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^q \leq C(q)(\mathbb{E}|\zeta^*_{j,\eta} - \beta_{j,\eta}|^q + \mathbb{E}|\gamma^*_{j,\eta}|^q)$$

Recall that we have $\gamma^*_{j,\eta} = \omega_d \sum V_i\psi_{j,\eta}(T_i)/n$. The $\{\omega_d V_i\psi_{j,\eta}(T_i)\}_{i\leq n}$ are iid with $\mathbb{E}\omega_d V_i\psi_{j,\eta}(T_i) = 0$ and $\mathbb{E}\{\omega_d V_i\psi_{j,\eta}(T_i)\}^2 = \omega_d^2\sigma^2\|\psi_{j,\eta}\|_2^2 \leq C\omega_d\sigma^2$ and, for $q \geq 2$, $\mathbb{E}|\omega_d V_i\psi_{j,\eta}(T_i)|^q = \omega_d^q\mathbb{E}|V_i|^q\|\psi_{j,\eta}\|_q^q \leq C2^{jd(\frac{q}{2}-1)}\omega_d^{q-1} \leq C\omega_d^{q-1}2^{Jd(\frac{q}{2}-1)}$. Thus, for $q \geq 2$, we have

$$\begin{aligned}
\mathbb{E}|\gamma^*_{j,\eta}|^q &= \mathbb{E}|\sum \omega_d V_i\psi_{j,\eta}(T_i)/n|^q \\
&\leq C\left\{n^{1-q}\mathbb{E}|\omega_d V\psi_{j,\eta}(T)|^q + n^{-\frac{q}{2}}\left(\mathbb{E}\{\omega_d V\psi_{j,\eta}(T)\}^2\right)^{\frac{q}{2}}\right\} \\
&\leq C\left\{n^{1-q}C\omega_d^{q-1}2^{Jd(\frac{q}{2}-1)} + n^{-\frac{q}{2}}\omega_d^{q/2}\sigma^q\right\} \leq C\omega_d^{q-1}n^{-\frac{q}{2}}
\end{aligned}$$

where the last inequality comes from the fact that $2^{Jd} \leq n/C_0$. And for $0 < q \leq 2$,

$$\begin{aligned}
\mathbb{E}|\gamma^*_{j,\eta}|^q = \mathbb{E}|n^{-1}\sum \omega_d V_i\psi_{j,\eta}(T_i)|^q &\leq \left\{\mathbb{E}|n^{-1}\sum \omega_d V_i\psi_{j,\eta}(T_i)|^2\right\}^{\frac{q}{2}} \\
&= \left\{\sum \mathbb{E}|n^{-1}\omega_d V_i\psi_{j,\eta}(T_i)|^2\right\}^{\frac{q}{2}} \leq C\omega_d^{q/2}\sigma^q n^{-\frac{q}{2}}
\end{aligned}$$

In the same way, for $q \geq 2$, denote $X_i = \omega_d f(T_i)\psi_{j,\eta}(T_i) - \mathbb{E}\omega_d f(T_i)\psi_{j,\eta}(T_i)$. Notice that we have $\mathbb{E}X_i = 0$ and,

$$\begin{aligned}
\mathbb{E}X_i^2 &\leq \omega_d^2\mathbb{E}f(T)^2\psi_{j,\eta}(T)^2 \leq \|f\|_\infty^2 C_2^2\omega_d, \\
\mathbb{E}|X_i|^q &\leq C\omega_d^q\mathbb{E}f(T)^q\psi_{j,\eta}(T)^q \leq C\omega_d^{q-1}\|f\|_\infty^q C_q^q 2^{Jd(\frac{q}{2}-1)}
\end{aligned}$$

So that

$$\begin{aligned}
\mathbb{E}|\zeta^*_{j,\eta} - \beta_{j,\eta}|^q &= \mathbb{E}|n^{-1}\omega_d\sum\{f(T_i)\psi_{j,\eta}(T_i) - \mathbb{E}f(T_i)\psi_{j,\eta}(T_i)\}|^q \\
&\leq C(n^{1-q}\omega_d^{q-1}\|f\|_\infty^q n^{\frac{q}{2}-1} + n^{-\frac{q}{2}}\omega_d^{q/2}\|f\|_\infty^{q/2}) \leq C\omega_d^{q-1}n^{-\frac{q}{2}}
\end{aligned}$$

And for $0 < q \leq 2$, $\mathbb{E}|\zeta^*_{j,\eta} - \beta_{j,\eta}|^q \leq C\|f\|_\infty^q\omega_d^{q/2}n^{-\frac{q}{2}}$, which finishes to prove eq. (23). □

Let's turn to the proof of Proposition 8.2 in the stochastic thresholding case.

We start with the proof of eq. (13). We break down the problem into four cases,

$$\sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta^{\circledast}_{j,\eta} - \beta_{j,\eta}|^z$$

$$
\begin{aligned}
&\leq \sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\}}\\
&+\sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|<\frac{\kappa_{j,\eta}}{2}t(n)\}}\\
&+\sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|<\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\geq 2\kappa_{j,\eta}t(n)\}}\\
&+\sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|<\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|<2\kappa_{j,\eta}t(n)\}}\\
&\triangleq Bb+Bs+Sb+Ss
\end{aligned}
$$

We then bound separately each of the four terms Bb, Ss, Sb, Bs. Let's start with Bb,

$$
\begin{aligned}
Bb &\leq \sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\}}\\
&\leq C\sum_{j=0}^{J} 2^{j\gamma}(j+1)^z n^{-\frac{z}{2}}\left(\mathbb{E}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\}}\right)^{\frac{1}{2}}
\end{aligned}
$$

where the last inequality comes as a direct application of Cauchy-Schwarz inequality and eq. (24). We can now apply Lemma 10.2, eq. (21) to the term on the right-hand-side with $g_{j,\eta}=\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|$ and $s=s_0\triangleq\|\psi_{j,\eta}\|_2^2/2$ to get

$$
\begin{aligned}
Bb &\leq C\sum_{j=0}^{J} 2^{j\gamma}(j+1)^z n^{-\frac{z}{2}}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l t(n)^{-l}+C\sum_{j=0}^{J} 2^{j\gamma}(j+1)^z n^{-\frac{z}{2}}n^{-\frac{\nu_0(C_0,s_0)}{2}}\\
&\leq Cn^{-\frac{z}{2}}t(n)^{-l}(J+1)^z\sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l+Cn^{-\frac{z}{2}}(J+1)^z 2^{J\gamma}n^{-C_0\frac{m^-}{2}}\\
&\leq Ct(n)^{z-l}(J+1)^z\sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l+Cn^{-\frac{z}{2}}(J+1)^z n^{\frac{\gamma}{d}-\frac{C_0m^-}{2}}
\end{aligned}
$$

where m^- has been defined in Lemma 10.1 and the last inequality uses the fact that $2^{Jd}\leq n/C_0$. The term on the far right is of the good order as soon as $\gamma/d-C_0m^-/2<0$ since in that case $(J+1)^z n^{\gamma/d-C_0m^-/2}\to 0$ as $n\to\infty$. Hence the term on the far right is of the good order as soon as $C_0>2\gamma/dm^-$.
Let's now turn to Ss. We can apply Lemma 10.2, eq. (20) with $g_{j,\eta}=\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|$ and $t=t_*=3\|\psi_{j,\eta}\|_2^2/2$. This leads to

$$
\begin{aligned}
Ss &\leq \sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\left(C\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^{l-z}t(n)^{z-l}+n^{-\nu_*(C_0,t_*)}\right)\\
&\leq C\left(t(n)^{z-l}+n^{-C_0m^-}\right)\sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l\leq Ct(n)^{z-l}\sum_{j=0}^{J}2^{j\gamma}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^l
\end{aligned}
$$

where the before last inequality uses Lemma 10.2, eq. (22) and the last inequality is valid as soon as $C_0m^-\geq(z-l)/2$. In particular, it is enough to take $C_0\geq z/2m^-$. Let's now turn to Bs. We

have, using eq. (24), eq. (25) and Cauchy-Schwarz inequality,

$$
\begin{aligned}
Bs &\leq \sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|<2\kappa_{j,\eta}t(n)\}}\\
&\leq \sum_{j=0}^{J} 2^{j\gamma}\left\{\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^{2z}\right\}^{\frac{1}{2}}\left\{2^{jd}\mathbb{P}\left(|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\right)\right\}^{\frac{1}{2}}\\
&\leq \sum_{j=0}^{J} 2^{j\gamma}\left\{C(j+1)^{2z}n^{-z}\right\}^{\frac{1}{2}}\left\{2^{jd}2n^{-\frac{\varpi}{4}}\right\}^{\frac{1}{2}}\leq C2^{J\{\gamma+\frac{d}{2}\}}(J+1)^z n^{-\frac{z}{2}}n^{-\frac{\varpi}{8}}\\
&\leq C(J+1)^z n^{\frac{\gamma}{d}+\frac{1}{2}-\frac{\varpi}{8}}n^{-\frac{z}{2}}\leq Cn^{-\frac{z}{2}}
\end{aligned}
$$

since $2^{Jd}\leq n/C_0$ and for $\varpi>4(\frac{2\gamma}{d}+1)$, $(J+1)^z n^{\frac{\gamma}{d}+\frac{1}{2}-\frac{\varpi}{8}}\to 0$ as $n\to\infty$. Finally, notice that Lemma 2.1, eq. (7) and the fact that $f\in\mathbb{L}^\infty$ lead to $|\beta_{j,\eta}|\leq C2^{-j\frac{d}{2}}$. And this together with eq. (25) allows us to write

$$
\begin{aligned}
Sb &\leq \sum_{j=0}^{J} 2^{j\gamma}\mathbb{E}\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}-\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{\sup_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|\geq 2\kappa_{j,\eta}t(n)\}}\\
&\leq \sum_{j=0}^{J} 2^{j\gamma}C2^{-jz\frac{d}{2}}2^{jd}\mathbb{P}\{|\beta_{j,\eta}-\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}\leq Cn^{-\frac{\varpi}{2}}\sum_{j=0}^{J}2^{j\{\gamma+d-d\frac{z}{2}\}}
\end{aligned}
$$

Here two cases arise. On the one hand, if $\gamma+d-d\frac{z}{2}\leq 0$, that is $z\geq 2(\frac{\gamma}{d}+1)$, we have asymptotically $Sb\leq Cn^{-\frac{\varpi}{2}}\log n\leq Cn^{-\frac{z}{2}}$ for $\varpi>z$. On the other hand, if $\gamma+d-d\frac{z}{2}>0$, that is $z<2(\frac{\gamma}{d}+1)$, we have $Sb\leq Cn^{-\frac{\varpi}{2}}2^{J\{\gamma+d-d\frac{z}{2}\}}\leq Cn^{-\frac{\varpi}{2}+\frac{\gamma}{d}+1-\frac{z}{2}}\leq Cn^{-\frac{z}{2}}$ for $\varpi\geq 2(\frac{\gamma}{d}+1)$. Therefore, it is enough to pick $\varpi>\max(z,2\{\frac{\gamma}{d}+1\})$ in order to get $Sb\leq Cn^{-z/2}$.

Let's now turn to the proof of eq. (14). We proceed in exactly the same way as in the previous paragraph. We start by breaking down the problem into four cases,

$$
\begin{aligned}
&\sum_{j=0}^{J}2^{j(\gamma-d)}\mathbb{E}\sum_{\eta\in\mathcal{Z}_j}|\beta^{\circledast}_{j,\eta}-\beta_{j,\eta}|^z\\
&\leq \sum_{j=0}^{J}2^{j(\gamma-d)}\mathbb{E}\sum_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{|\beta_{j,\eta}|\geq\frac{\kappa_{j,\eta}}{2}t(n)\}}\\
&+\sum_{j=0}^{J}2^{j(\gamma-d)}\mathbb{E}\sum_{\eta\in\mathcal{Z}_j}|\hat{\beta}_{j,\eta}-\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|\geq\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{|\beta_{j,\eta}|<\frac{\kappa_{j,\eta}}{2}t(n)\}}\\
&+\sum_{j=0}^{J}2^{j(\gamma-d)}\mathbb{E}\sum_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|<\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{|\beta_{j,\eta}|\geq 2\kappa_{j,\eta}t(n)\}}\\
&+\sum_{j=0}^{J}2^{j(\gamma-d)}\mathbb{E}\sum_{\eta\in\mathcal{Z}_j}|\beta_{j,\eta}|^z\mathbb{1}_{\{|\hat{\beta}_{j,\eta}|<\kappa_{j,\eta}t(n)\}}\mathbb{1}_{\{|\beta_{j,\eta}|<2\kappa_{j,\eta}t(n)\}}\\
&\triangleq Bb+Bs+Sb+Ss\ .
\end{aligned}
$$

We then bound separately each of the four terms Bb, Ss, Sb, Bs. Using eq. (23), we get, for any

$l \geq 0$,

$$
\begin{aligned}
Bb &\leq \sum_{j=0}^{J} \sum_{\eta \in \mathcal{Z}_j} 2^{j(\gamma-d)} \mathbb{E}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^z \mathbb{1}_{\{|\beta_{j,\eta}| \geq \frac{\kappa_{j,\eta}}{2} t(n)\}} \\
&\leq \sum_{j=0}^{J} \sum_{\eta \in \mathcal{Z}_j} 2^{j(\gamma-d)} \left\{ \mathbb{E}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^{2z} \right\}^{\frac{1}{2}} \left\{ \mathbb{E}\mathbb{1}_{\{|\beta_{j,\eta}| \geq \frac{\kappa_{j,\eta}}{2} t(n)\}} \right\}^{\frac{1}{2}}
\end{aligned}
$$

We can now apply Lemma 10.3, eq. (23) to the left expectation and Lemma 10.2, eq. (21) to the right expectation with $g_{j,\eta} = |\beta_{j,\eta}|$ and $s = s_0$ to get

$$
\begin{aligned}
Bb &\leq C n^{-\frac{z}{2}} t(n)^{-l} \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \right) + C \sum_{j=0}^{J} \sum_{\eta \in \mathcal{Z}_j} 2^{j(\gamma-d)} n^{-\frac{z}{2} - \frac{C_0 m^-}{2}} \\
&\leq C t(n)^{z-l} \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \right) + C n^{-\frac{z}{2} + \frac{\gamma}{d} - \frac{C_0 m^-}{2}}
\end{aligned}
$$

where the last inequality uses the fact that $\#\mathcal{Z}_j \approx 2^{jd}$ and $2^{Jd} \leq n/C_0$. In particular, the far right term is of the good order as soon as $C_0 \geq 2\gamma/dm^-$. Let's now take $l \in [0, z]$ and turn to Ss. If we apply Lemma 10.2, eq. (20) with $g_{j,\eta} = |\beta_{j,\eta}|$ and $t = t_*$, we get

$$
\begin{aligned}
Ss &\leq \sum_{j=0}^{J} 2^{j(\gamma-d)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^z \mathbb{E}\mathbb{1}_{\{|\beta_{j,\eta}| < 2\kappa_{j,\eta} t(n)\}} \\
&\leq C t(n)^{z-l} \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \right) + n^{-C_0 m^-} \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^z \right) \\
&\leq C \left(t(n)^{z-l} + n^{-C_0 m^-} \right) \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \right) \\
&\leq C t(n)^{z-l} \sum_{j=0}^{J} 2^{j(\gamma-d)} \left(\sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^l \right)
\end{aligned}
$$

where the before last equation uses Lemma 10.2, eq. (22) with $g_{j,\eta} = |\beta_{j,\eta}|$ and the last equation is valid as soon as $C_0 m^- \geq \frac{z-l}{2}$, so that it is enough to take $C_0 \geq z/2m^-$. Let's now turn to Bs. We use Lemma 10.3, eq. (23) and eq. (25), and the fact that $\#\mathcal{Z}_j \approx 2^{jd}$ to show that

$$
\begin{aligned}
Bs &\leq \sum_{j=0}^{J} 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^z \mathbb{1}_{\{|\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq \frac{\kappa_{j,\eta}}{2} t(n)\}} \\
&\leq \sum_{j=0}^{J} 2^{j(\gamma-d)} \sum_{\eta \in \mathcal{Z}_j} \left(\mathbb{E}|\hat{\beta}_{j,\eta} - \beta_{j,\eta}|^{2z} \right)^{\frac{1}{2}} \mathbb{P}\left\{ |\hat{\beta}_{j,\eta} - \beta_{j,\eta}| \geq \frac{\kappa_{j,\eta}}{2} t(n) \right\}^{\frac{1}{2}} \\
&\leq C \sum_{j=0}^{J} 2^{j\gamma} n^{-\frac{z}{2}} n^{-\frac{\varpi}{4}} \leq C n^{-\frac{z}{2} + \frac{\gamma}{d} - \frac{\varpi}{4}} \leq C n^{-\frac{z}{2}}
\end{aligned}
$$

where, as usual, the before last inequality uses the fact that $2^{Jd} \leq n/C_0$ and the last inequality is valid as soon as $\varpi \geq 4\frac{\gamma}{d}$. Let's now turn to Sb. Notice again that f bounded implies that

$|\beta_{j,\eta}| \leq C2^{-\frac{jd}{2}}$. This together with eq. (25) and $\#Z_j \approx 2^{jd}$ leads us to

$$\begin{aligned} Sb &\leq \sum_{j=0}^{J} 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j,\eta}|^z \mathbb{1}_{\{|\beta_{j,\eta}-\hat{\beta}_{j,\eta}| \geq \kappa_{j,\eta} t(n)\}} \\ &\leq C \sum_{j=0}^{J} 2^{j(\gamma-d)} 2^{jd} 2^{-jd\frac{z}{2}} \mathbb{P}\{|\beta_{j,\eta} - \hat{\beta}_{j,\eta}| \geq \kappa_{j,\eta} t(n)\} \\ &\leq C \sum_{j=0}^{J} 2^{j\{-d\frac{z}{2}+\gamma\}} n^{-\frac{\varpi}{2}} \leq C n^{-\frac{\varpi}{2}} \sum_{j=0}^{J} 2^{j\{\gamma+d-d\frac{z}{2}\}} \end{aligned}$$

Here we face the same dichotomy as at the end of the previous paragraph. This concludes the proof of eq. (13). □

Acknowledgements. The author would like to thank Dominique Picard for serving as his adviser and for many fruitful discussions and suggestions. He would also like to thank anonymous referees for their suggestions, which helped improving the overall presentation of the paper.

References

Adams, R., Fournier, J.: Sobolev spaces, second edition. Elsevier Science, Oxford (2003)

Baldi, P., Kerkyacharian, G., Marinucci, D., Picard, D.: Adaptative density estimation for directional data using needlets. Ann. Stat. **37(6)**, 3362–3395 (2009)

Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D.: Wavelet shrinkage: Asymptotia? J. Roy. Stat. Soc. B **57(2)**, 301–369 (1995)

Faÿ, G., Guilloux, F., Cardoso, J.F., Delabrouille, J., Le Jeune, M.: CMB power spectrum estimation using wavelets. Phys. Rev. D **78(8)** (2008)

Fengler, M.: Vector spherical harmonic and vector wavelet based non-linear Garlekin schemes for solving the incompressible Navier-Stokes equation on the sphere. PhD thesis, University of Kaiserslautern, Mathematics Department, Geomathematics Group (2005)

Freeden, W., Gervens, T., Schreiner, M.: Constructive Approximation on the Sphere (With Applications to Geomathematics). Oxford Sciences Publication. Clarendon Press, Oxford. (1998)

Freeden, W., Michel, D., Michel, V.: Local multiscale approximations of geostrophic oceanic flow: theoretical background and aspects of scientific computing. Mar. Geod. **28**, 313–329 (2005)

Freeden, W., Michel, V.: Multiscale Potential Theory, with applications to Geoscience. Brikhäuser, Boston (2004)

Guilloux, F., Faÿ, G., Cardoso, J.F.: Practical wavelet design on the sphere. Appl. Comput. Harmon. Anal. **26(2)**, 143–160 (2009)

Holschneider, M., Chambodut, A., Mandea, M.: From global to regional analysis of the magnetic field on the sphere using wavelet frames. Phys. Earth Planet. In. **135**, 107–124 (2003)

Kerkyacharian, G., Picard, D.: Regression in random design and warped wavelets. Bernoulli **10(6)**, 1053–1105 (2004)

Li, T.H.: Multiscale representation and analysis of spherical data by spherical wavelets. SIAM J. Sci. Comput. **21**, 924–953 (1999)

Li, T.H., Oh, H.S.: Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptative method. J. Roy. Stat. Soc. B **66**, 221–238 (2004)

Maier, T.: Wavelet-mie-representations for solenoidal vector fields with applications to ionospheric geomagnetic data. SIAM J. Appl. Math. **65**, 1888–1912 (2005)

Marinucci, D., Pietrobon, D., Balbi, A., Baldi, P., Cabella, P., Kerkyacharian, G., Natoli, P., Picard, D., Vittorio, N.: Spherical needlets for CMB data analysis. Mon. Not. R. Astron. Soc. **383(2)**, 539–545 (2008)

Mayer, C.: Wavelet modelling of the spherical inverse source problem with application to geomagnetism. Inverse Probl. **20**, 1713–1728 (2004)

Müller, C.: Spherical Harmonics. Springer-Verlag, Berlin (1966)

Narcowich, F., Petruchev, P., Ward, J.: Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. J. Funct. Anal. **238**, 530–564 (2006)

Narcowich, F., Petruchev, P., Ward, J.: Localized tight frames on spheres. SIAM J. Math. Anal. **38(2)**, 574–594 (2007)

Narcowich, F., Ward, J.: Nonstationary wavelets on the m-sphere for scattered data. Appl. Comput. Harmon. Anal. **3**, 324–326 (1996)

Panet, I., Jamet, O., Diament, M., Chambodut, A.: Modelling the Earth's gravity field using wavelet frames. Springer (2005)

Petrov, V.V.: Limit theorems of probability theory: sequences of independent random variables. Oxford University Press (1995)

Schmidt, M., Fengler, M., Mayer-Gürr, T., Eicker, A., Kusche, J., Sánchez, L., Han, S.C.: Regional gravity modeling in terms of spherical base functions. J. Geodesy **81**, 17–38 (2007)

Schmidt, M., Han, S., Kusche, J., Sánchez, L., Shum, C.: Regional high-resolution spatiotemporal gravity modeling from GRACE data using spherical wavelets. Geophys. Res. Lett. **33(8)** (2006)

Stein, E., Weiss, G.: Fourier Analysis on Euclidean Spaces. Princeton University Press, Princeton, N.J. (1975)