

# Radon needlet thresholding

GÉRARD KERKYACHARIAN<sup>\*</sup>, ERWAN LE PENNEC<sup>\*\*</sup> and  
DOMINIQUE PICARD<sup>†</sup>

<sup>1</sup>*LPMA, CNRS, Université Paris Diderot, 175 rue du Chevaleret, 75013 Paris, France.*  
*E-mail: <sup>\*</sup>kerk@math.jussieu.fr; <sup>\*\*</sup>lepenec@math.jussieu.fr; <sup>†</sup>picard@math.jussieu.fr*

We provide a new algorithm for the treatment of the noisy inversion of the Radon transform using an appropriate thresholding technique adapted to a well-chosen new localized basis. We establish minimax results and prove their optimality. In particular, we prove that the procedures provided here are able to attain minimax bounds for any  $\mathbb{L}_p$  loss. It is important to notice that most of the minimax bounds obtained here are new to our knowledge. It is also important to emphasize the adaptation properties of our procedures with respect to the regularity (sparsity) of the object to recover and to inhomogeneous smoothness. We perform a numerical study that is of importance since we especially have to discuss the cubature problems and propose an averaging procedure that is mostly in the spirit of the cycle spinning performed for periodic signals.

*Keywords:* minimax estimation; second-generation wavelets; statistical inverse problems

## 1. Introduction

We consider the problem of inverting noisy observations of the  $d$ -dimensional Radon transform. Obviously, the most immediate examples occur for  $d = 2$  or  $3$ . However, no major differences arise from considering the general case.

There is considerable literature on the problem of reconstructing structures from their Radon transforms, which is a fundamental problem in medical imaging and, more generally, in tomography. In our approach, we focus on several important points. We produce a procedure that is efficient from an  $\mathbb{L}_2$  point of view, since this loss function mimics quite well in many situations the preferences of the human eye. On the other hand, we have at the same time the requirement of clearly identifying the local bumps, of being able to estimate the different level sets well. We also want the procedure to enjoy good adaptation properties. In addition, we require the procedure to be simple to implement.

At the heart of such a problem, there is a notable conflict between the inversion part – which, in the presence of noise, creates an instability reasonably handled by a singular value decomposition (SVD) approach – and the fact that the SVD basis is very rarely localized and, as a consequence, capable of representing local features of images, which are especially important to recover.

Our strategy is to follow the approach started in [11], which utilizes a construction borrowed from [21] (see also [13]) of localized frames based on orthogonal polynomials on the ball, which are closely related to the Radon transform SVD basis.

To achieve the goals presented above, and especially adaptation to different regularities and local inhomogeneous smoothness, we add a fine-tuned subsequent thresholding process to the estimation performed in [11].

This improves considerably the performances of the algorithm, both from a theoretical point of view and a numerical point of view. In effect, the new algorithm provides a much better spatial adaptation, as well as adaptation to the classes of regularity. We prove here that the bounds obtained by the procedure are minimax over a large class of Besov spaces and any  $\mathbb{L}_p$  loss: we provide upper bounds for the performance of our algorithm and lower bounds for the associated minimax rate.

It is important to notice that because we consider different  $\mathbb{L}_p$  losses, we provide rates of convergence of new types attained by our procedure. Those rates are minimax since they are confirmed by lower-bound inequalities.

The problem of choosing appropriate spaces of regularity on the ball reflecting the standard objects analyzed in tomography is a highly non-trivial problem. We decided to consider the spaces that seem to stay closest to our natural intuition, that is, those that generalize to the ball the approximation properties by polynomials.

The procedure gives very promising results in the simulation study. We show that the estimates obtained by thresholding the needlets outperform those obtained either by thresholding the SVD or by the linear needlet estimate proposed in [11]. An important issue in the needlet scheme is the choice of the quadrature in the needlet construction. We discuss the possibilities proposed in the literature and consider a cubature formula based on the full tensorial grid on the sphere, introducing an averaging close to the cycle-spinning method.

Among others, one amazing result is that, to attain minimax rates in the  $\mathbb{L}_\infty$  norm, we need to modify the estimator. This result is also corroborated by the numerical results: see Theorem 2 and Figures 4 and 5.

In the first section, we introduce the Radon transform and the associated SVD basis. The following section summarizes the construction of the localized basis, the needlets. The procedure is introduced in Section 4, where the main theoretical results are stated for upper bounds and lower bounds. Section 5 details the simulation study. Section 6 details important properties of the needlet basis. The proof of the two main results stated in Section 4 is postponed in the two last sections.

## 2. Radon transform and white noise model

### 2.1. Radon transform

Here we recall the definitions and some basic facts about the Radon transform (cf. [10,14,19]). Denote by  $B^d$  the unit ball in  $\mathbb{R}^d$ , that is,  $B^d = \{x = (x_1, \dots, x_d) \in \mathbb{R}^d: |x| \leq 1\}$  with  $|x| = (\sum_{i=1}^d x_i^2)^{1/2}$  and, by  $\mathbb{S}^{d-1}$ , the unit sphere in  $\mathbb{R}^d$ . The Lebesgue measure on  $B^d$  will be denoted by  $dx$  and the usual surface measure on  $\mathbb{S}^{d-1}$  by  $d\sigma(x)$  (sometimes we will also deal with the surface measure on  $\mathbb{S}^d$ , which will be denoted by  $d\sigma_d$ ). We let  $|A|$  denote the measure  $|A| = \int_A dx$  if  $A \subset B^d$  and  $|A| = \int_A d\sigma(x)$  if  $A \subset \mathbb{S}^{d-1}$ .

The Radon transform of a function  $f$  is defined by

$$Rf(\theta, s) = \int_{\substack{y \in \theta^\perp \\ s\theta + y \in B^d}} f(s\theta + y) \, dy, \quad \theta \in \mathbb{S}^{d-1}, s \in [-1, 1],$$

where  $dy$  is the Lebesgue measure of dimension  $d - 1$  and  $\theta^\perp = \{x \in \mathbb{R}^d: \langle x, \theta \rangle = 0\}$ . With a slight abuse of notation, we will rewrite this integral as

$$Rf(\theta, s) = \int_{\langle y, \theta \rangle = s} f(y) \, dy.$$

By Fubini’s theorem, we have

$$\int_{-1}^1 Rf(\theta, s) \, ds = \int_{B^d} f(x) \, dx.$$

It is easy to see (cf. [19]) that the Radon transform is a bounded linear operator mapping  $\mathbb{L}_2(B^d, dx)$  into  $\mathbb{L}_2(\mathbb{S}^{d-1} \times [-1, 1], d\mu(\theta, s))$ , where

$$d\mu(\theta, s) = d\sigma(\theta) \frac{ds}{(1 - s^2)^{(d-1)/2}}.$$

## 2.2. Noisy observation of the Radon transform

We consider observations of the form

$$dY(\theta, s) = Rf(\theta, s) \, d\mu(\theta, s) + \varepsilon \, dW(\theta, s),$$

where the unknown function  $f$  belongs to  $\mathbb{L}_2(B^d, dx)$ . The meaning of this equation is that, for any  $\varphi(\theta, s)$  in  $\mathbb{L}_2(\mathbb{S}^{d-1} \times [-1, 1], d\mu(\theta, s))$ , one can observe

$$\begin{aligned} Y_\varphi &= \int \varphi(\theta, s) \, dY(\theta, s) = \int_{\mathbb{S}^{d-1} \times [-1, 1]} Rf(\theta, s) \varphi(\theta, s) \, d\mu(\theta, s) + \varepsilon \int \varphi(\theta, s) \, dW(\theta, s) \\ &= \langle Rf, \varphi \rangle_\mu + \varepsilon W_\varphi. \end{aligned}$$

Here  $W_\varphi = \int \varphi(\theta, s) \, dW(\theta, s)$  is a Gaussian field of zero mean and covariance

$$\mathbb{E}(W_\varphi, W_\psi) = \int_{\mathbb{S}^{d-1} \times [-1, 1]} \varphi(\theta, s) \psi(\theta, s) \, d\sigma(\theta) \frac{ds}{(1 - s^2)^{(d-1)/2}} = \langle \varphi, \psi \rangle_\mu.$$

The goal is to recover the unknown function  $f$  from the observation of  $Y$ . Our idea is to refine the algorithms proposed in [11] using thresholding methods. In [11], estimation schemes are derived that combine the stability and computability of SVD decompositions with the localization and multiscale structure of wavelets. To this end, a specific frame (essentially following the construction from [13]) is used. It comprises elements of nearly exponential localization and is, in addition, compatible with the SVD basis of the Radon transform.

### 2.3. Singular value decomposition of the Radon transform

The SVD of the Radon transform was first established in [5,15]. In this regard, we also refer the reader to [19,28].

#### 2.3.1. Jacobi and Gegenbauer polynomials

The Radon SVD bases are defined in terms of Jacobi and Gegenbauer polynomials. The Jacobi polynomials  $P_n^{(\alpha,\beta)}$ ,  $n \geq 0$ , constitute an orthogonal basis for the space  $\mathbb{L}_2([-1, 1], w_{\alpha,\beta}(t) dt)$  with weight  $w_{\alpha,\beta}(t) = (1 - t)^\alpha (1 + t)^\beta$ ,  $\alpha, \beta > -1$ . They are standardly normalized by  $P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n}$  and then [1,7,25]

$$\int_{-1}^1 P_n^{(\alpha,\beta)}(t) P_{n'}^{(\alpha,\beta)}(t) w_{\alpha,\beta}(t) dt = \delta_{n,n'} h_n^{(\alpha,\beta)},$$

where

$$h_n^{(\alpha,\beta)} = \frac{2^{\alpha+\beta+1}}{(2n + \alpha + \beta + 1)} \frac{\Gamma(n + \alpha + 1)\Gamma(n + \beta + 1)}{\Gamma(n + 1)\Gamma(n + \alpha + \beta + 1)}.$$

The Gegenbauer polynomials  $C_n^\lambda$  are a particular case of Jacobi polynomials and are traditionally defined by

$$C_n^\lambda(t) = \frac{(2\lambda)_n}{(\lambda + 1/2)_n} P_n^{(\lambda-1/2,\lambda-1/2)}(t), \quad \lambda > -1/2,$$

where, by definition,  $(a)_n = a(a + 1) \cdots (a + n - 1) = \frac{\Gamma(a+n)}{\Gamma(a)}$  (note that in [25] the Gegenbauer polynomial  $C_n^\lambda$  is denoted by  $P_n^\lambda$ ). It is readily seen that  $C_n^\lambda(1) = \binom{n+2\lambda-1}{n} = \frac{\Gamma(n+2\lambda)}{n!\Gamma(2\lambda)}$  and

$$\int_{-1}^1 C_n^\lambda(t) C_m^\lambda(t) (1 - t^2)^{\lambda-1/2} dt = \delta_{n,m} h_n^{(\lambda)} \quad \text{with } h_n^{(\lambda)} = \frac{2^{1-2\lambda} \pi}{\Gamma(\lambda)^2} \frac{\Gamma(n + 2\lambda)}{(n + \lambda)\Gamma(n + 1)}.$$

#### 2.3.2. Polynomials on $B^d$ and $\mathbb{S}^{d-1}$

Let  $\Pi_n(\mathbb{R}^d)$  be the space of all polynomials in  $d$  variables of degree  $\leq n$ . We denote by  $\mathcal{P}_n(\mathbb{R}^d)$  the space of all homogeneous polynomials of degree  $n$  and by  $\mathcal{V}_n(\mathbb{R}^d)$  the space of all polynomials of degree  $n$  that are orthogonal to lower-degree polynomials with respect to the Lebesgue measure on  $B^d$ . Of course,  $\mathcal{V}_0(\mathbb{R}^d)$  will be the set of all constants. We have the following orthogonal decomposition:

$$\Pi_n(\mathbb{R}^d) = \bigoplus_{k=0}^n \mathcal{V}_k(\mathbb{R}^d).$$

Also, denote by  $\mathbb{H}_n(\mathbb{R}^d)$  the subspace of all harmonic homogeneous polynomials of degree  $n$  (i.e.,  $Q \in \mathbb{H}_n(\mathbb{R}^d)$  if  $Q \in \mathcal{P}_n(\mathbb{R}^d)$  and  $\Delta Q = 0$ ) and, by  $\mathbb{H}_n(\mathbb{S}^{d-1})$ , the (injective) restriction of

the polynomials from  $\mathbb{H}_n(\mathbb{R}^d)$  to  $\mathbb{S}^{d-1}$ . It is well known that

$$N_{d-1}(n) = \dim(\mathbb{H}_n(\mathbb{S}^{d-1})) = \binom{n+d-1}{d-1} - \binom{n+d-3}{d-1} \sim n^{d-2}.$$

Let  $\Pi_n(\mathbb{S}^{d-1})$  be the space of restrictions to  $\mathbb{S}^{d-1}$  of polynomials of degree  $\leq n$  on  $\mathbb{R}^d$ . It is also well known that

$$\Pi_n(\mathbb{S}^{d-1}) = \bigoplus_{m=0}^n \mathbb{H}_m(\mathbb{S}^{d-1})$$

(the orthogonality is, with respect to the surface measure,  $d\sigma$  on  $\mathbb{S}^{d-1}$ ).  $\mathbb{H}_l(\mathbb{S}^{d-1})$  is called the space of spherical harmonics of degree  $l$  on the sphere  $\mathbb{S}^{d-1}$ .

Let  $Y_{l,m}$ ,  $1 \leq m \leq N_{d-1}(l)$ , be an orthonormal basis of  $\mathbb{H}_l(\mathbb{S}^{d-1})$ , that is,

$$\int_{\mathbb{S}^{d-1}} Y_{l,m}(\xi) \overline{Y_{l,m'}(\xi)} d\sigma(\xi) = \delta_{m,m'}.$$

Then the natural extensions of  $Y_{l,m}$  on  $B^d$  are defined by  $Y_{l,m}(x) = |x|^l Y_{l,m}(\frac{x}{|x|})$  and satisfy

$$\begin{aligned} \int_{B^d} Y_{l,m}(x) \overline{Y_{l,m'}(x)} dx &= \int_0^1 r^{d-1} \int_{\mathbb{S}^{d-1}} Y_{l,m}(r\xi) \overline{Y_{l,m'}(r\xi)} d\sigma(\xi) dr \\ &= \int_0^1 r^{d+2l-1} \int_{\mathbb{S}^{d-1}} Y_{l,m}(\xi) \overline{Y_{l,m'}(\xi)} d\sigma(\xi) dr = \delta_{m,m'} \frac{1}{2l+d}. \end{aligned}$$

For more details, we refer the reader to [6].

The spherical harmonics on  $\mathbb{S}^{d-1}$  and orthogonal polynomials on  $B^d$  are naturally related to Gegenbauer polynomials. Thus, the kernel of the orthogonal projector onto  $\mathbb{H}_n(\mathbb{S}^{d-1})$  can be written as (see [24]):

$$\sum_{m=1}^{N_{d-1}(n)} Y_{l,m}(\xi) \overline{Y_{l,m}(\theta)} = \frac{2n+d-2}{(d-2)|\mathbb{S}^{d-1}|} C_n^{(d-2)/2}(\langle \xi, \theta \rangle). \tag{1}$$

The ‘‘ridge’’ Gegenbauer polynomials  $C_n^{d/2}(\langle x, \xi \rangle)$  are orthogonal to  $\Pi_{n-1}(B^d)$  in  $\mathbb{L}_2(B^d)$  and the kernel  $L_n(x, y)$  of the orthogonal projector onto  $\mathcal{V}_n(B^d)$  can be written in the form (see [22, 28])

$$\begin{aligned} L_n(x, y) &= \frac{2n+d}{|\mathbb{S}^{d-1}|^2} \int_{\mathbb{S}^{d-1}} C_n^{d/2}(\langle x, \xi \rangle) C_n^{d/2}(\langle y, \xi \rangle) d\sigma(\xi) \\ &= \frac{(n+1)_{d-1}}{2^d \pi^{d-1}} \int_{\mathbb{S}^{d-1}} \frac{C_n^{d/2}(\langle x, \xi \rangle) C_n^{d/2}(\langle y, \xi \rangle)}{\|C_n^{d/2}\|^2} d\sigma(\xi). \end{aligned} \tag{2}$$

The following important identities are valid for “ridge” Gegenbauer polynomials:

$$\int_{B^d} C_n^{d/2}(\langle \xi, x \rangle) C_n^{d/2}(\langle \eta, x \rangle) dx = \frac{h_n^{(d/2)}}{C_n^{d/2}(1)} C_n^{d/2}(\langle \xi, \eta \rangle), \quad \xi, \eta \in \mathbb{S}^{d-1}, \tag{3}$$

and, for  $x \in B^d, \eta \in \mathbb{S}^{d-1}$ ,

$$\int_{\mathbb{S}^{d-1}} C_n^{d/2}(\langle \xi, x \rangle) C_n^{d/2}(\langle \xi, \eta \rangle) d\sigma(\xi) = |\mathbb{S}^{d-1}| C_n^{d/2}(\langle \eta, x \rangle); \tag{4}$$

see [22]. By (2) and (4)

$$L_n(x, \xi) = \frac{(2n + d)}{|\mathbb{S}^{d-1}|} C_n^{d/2}(\langle x, \xi \rangle), \quad \xi \in \mathbb{S}^{d-1},$$

and again by (2)

$$\int_{\mathbb{S}^{d-1}} L_n(x, \xi) L_n(y, \xi) d\sigma(\xi) = (2n + d) L_n(x, y).$$

### 2.3.3. The SVD of the Radon transform

Assume that  $\{Y_{l,m}: 1 \leq m \leq N_{d-1}(l)\}$  is an orthonormal basis for  $\mathbb{H}_l(\mathbb{S}^{d-1})$ . Then it is standard and easy to see that the family of polynomials,

$$f_{k,l,m}(x) = (2k + d)^{1/2} P_j^{(0,l+d/2-1)}(2|x|^2 - 1) Y_{l,m}(x), \\ 0 \leq l \leq k, k - l = 2j, 1 \leq i \leq N_{d-1}(l),$$

form an orthonormal basis of  $\mathcal{V}_k(B^d)$ ; see [6]. Here, as before,  $Y_{l,m}(x) = |x|^l Y_{l,m}(x/|x|)$ . On the other hand, the collection

$$g_{k,l,m}(\theta, s) = [h_k^{(d/2)}]^{-1/2} (1 - s^2)^{(d-1)/2} C_k^{d/2}(s) Y_{l,m}(\theta), \quad k \geq 0, l \geq 0, 1 \leq m \leq N_{d-1}(l),$$

is an orthonormal basis of  $\mathbb{L}_2(\mathbb{S}^{d-1} \times [-1, 1], d\mu(\theta, s))$ . Most important, the Radon transform  $R: \mathbb{L}_2(B^d) \mapsto \mathbb{L}_2(\mathbb{S}^{d-1} \times [-1, 1], d\mu(\theta, s))$  is a one-to-one mapping and

$$Rf_{k,l,m} = \lambda_k g_{k,l,m}, \quad R^* g_{k,l,m} = \lambda_k f_{k,l,m},$$

where

$$\lambda_k^2 = \frac{2^d \pi^{d-1}}{(k + 1)(k + 2) \cdots (k + d - 1)} = \frac{2^d \pi^{d-1}}{(k + 1)_{d-1}} \sim k^{-d+1}.$$

More precisely, we have for any  $f \in \mathbb{L}_2(B^d)$

$$Rf = \sum_{k \geq 0} \lambda_k \sum_{0 \leq l \leq k, k-l \equiv 0 \pmod{2}} \sum_{1 \leq m \leq N_{d-1}(l)} \langle f, f_{k,l,m} \rangle g_{k,l,m}.$$

Furthermore, for  $f \in \mathbb{L}_2(B^d)$ ,

$$f = \sum_{k \geq 0} \lambda_k^{-1} \sum_{0 \leq l \leq k, k-l \equiv 0 \pmod{2}} \sum_{1 \leq m \leq N_{d-1}(l)} \langle Rf, g_{k,l,m} \rangle_{\mu} f_{k,l,m}.$$

In the above identities, the convergence is in  $\mathbb{L}_2$ .

For the Radon SVD, we refer the reader to [11,19,28].

### 3. Construction of needlets on the ball

In this section, we briefly recall the construction of the needlets on the ball. This construction is due to [21]. Its aim is to build a very well-localized tight frame constructed using the eigenvectors of the Radon transform. For more precision, we refer the reader to [11,12,21]

Let  $\{f_{k,l,m}\}$  be the orthonormal basis of  $\mathcal{V}_k(B^d)$  defined in Section 2.3.3. Denote by  $T_k$  the index set of this basis, that is,  $T_k = \{(l, m): 0 \leq l \leq k, l \equiv k \pmod{2}, 0 \leq m \leq N_{d-1}(l)\}$ . Then the orthogonal projector of  $\mathbb{L}_2(B^d)$  onto  $\mathcal{V}_k(B^d)$  can be written in the form

$$L_k f = \int_{B^d} f(y) L_k(x, y) dy \quad \text{with } L_k(x, y) = \sum_{l,m \in T_k} f_{k,l,m}(x) f_{k,l,m}(y).$$

Using (1),  $L_k(x, y)$  can be written in the form

$$\begin{aligned} L_k(x, y) &= (2k + d) \sum_{l \leq k, k-l \equiv 0 \pmod{2}} P_j^{(0,l+d/2-1)}(2|x|^2 - 1) |x|^l P_j^{(0,l+d/2-1)}(2|y|^2 - 1) |y|^l \\ &\quad \times \sum_m Y_{l,m} \left( \frac{x}{|x|} \right) Y_{l,m} \left( \frac{y}{|y|} \right) \\ &= \frac{(2k + d)}{|\mathbb{S}^{d-1}|} \sum_{l \leq k, k-l \equiv 0 \pmod{2}} P_j^{(0,l+d/2-1)}(2|x|^2 - 1) |x|^l P_j^{(0,l+d/2-1)}(2|y|^2 - 1) |y|^l \\ &\quad \times \left( 1 + \frac{l}{d/2 - 1} \right) C_l^{d/2-1} \left( \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle \right). \end{aligned}$$

Another representation of  $L_k(x, y)$  has already been given in (2). Clearly,

$$\int_{B^d} L_k(x, z) L_{k'}(z, y) dz = \delta_{k,k'} L_k(x, y) \tag{5}$$

and, for  $f \in \mathbb{L}_2(B^d)$ ,

$$f = \sum_{k \geq 0} L_k f \quad \text{and} \quad \|f\|_2^2 = \sum_k \|L_k f\|_2^2 = \sum_k \langle L_k f, f \rangle. \tag{6}$$

The construction of the needlets is based on the classical Littlewood–Paley decomposition and a subsequent discretization.

Let  $a \in C^\infty[0, \infty)$  be a cut-off function such that  $0 \leq a \leq 1$ ,  $a(t) = 1$  for  $t \in [0, 1/2]$  and  $\text{supp } a \subset [0, 1]$ . We next use this function to introduce a sequence of operators on  $\mathbb{L}_2(B^d)$ . For  $j \geq 0$ , write

$$A_j f(x) = \sum_{k \geq 0} a\left(\frac{k}{2^j}\right) L_k f(x) = \int_{B^d} A_j(x, y) f(y) \, dy$$

$$\text{with } A_j(x, y) = \sum_k a\left(\frac{k}{2^j}\right) L_k(x, y).$$

Also, we define  $B_j f = A_{j+1} f - A_j f$ . Then, setting  $b(t) = a(t/2) - a(t)$ , we have

$$B_j f(x) = \sum_k b\left(\frac{k}{2^j}\right) L_k f(x) = \int_{B^d} B_j(x, y) f(y) \, dy$$

$$\text{with } B_j(x, y) = \sum_k b\left(\frac{k}{2^j}\right) L_k(x, y).$$

Obviously, for  $f \in \mathbb{L}_2(B^d)$ ,

$$\langle A_j f, f \rangle = \sum_k a\left(\frac{k}{2^j}\right) \langle L_k f, f \rangle \leq \|f\|_2^2.$$

An important result from [21] (cf. [13]) asserts that the kernels  $A_j(x, y)$ ,  $B_j(x, y)$  have nearly exponential localization. Namely, for any  $M > 0$  there exists a constant  $C_M > 0$  such that

$$|A_j(x, y)|, |B_j(x, y)| \leq C_M \frac{2^{jd}}{(1 + 2^j d(x, y))^M \sqrt{W_j(x)} \sqrt{W_j(y)}}, \quad x, y \in B^d, \quad (7)$$

where

$$W_j(x) = 2^{-j} + \sqrt{1 - |x|^2}, \quad |x|^2 = |x|_d^2 = \sum_{i=1}^d x_i^2 \quad (8)$$

and

$$d(x, y) = \text{Arccos}(\langle x, y \rangle + \sqrt{1 - |x|^2} \sqrt{1 - |y|^2}), \quad \langle x, y \rangle = \sum_{i=1}^d x_i y_i.$$

Let us define

$$C_j(x, y) = \sum_k \sqrt{a\left(\frac{k}{2^j}\right)} L_k(x, z) \quad \text{and} \quad D_j(x, y) = \sum_k \sqrt{b\left(\frac{k}{2^j}\right)} L_k(x, z).$$

Note that  $C_j$  and  $D_j$  have the same localization as the localization of  $A_j, B_j$  in (7) (cf. [21]). Using (5), we get,

$$A_j(x, y) = \int_{B^d} C_j(x, z)C_j(z, y) dz, \quad B_j(x, y) = \int_{B^d} D_j(x, z)D_j(z, y) dz. \quad (9)$$

And, obviously,  $z \mapsto C_j(x, z)C_j(z, y)$  (resp.,  $D_j(x, z)D_j(z, y)$ ) are polynomial of degrees  $< 2^{j+1}$ .

The following proposition follows from results in [21] and [27] and establishes a cubature formula.

**Proposition 1.** *Let  $\{B(\tilde{\xi}_i, \rho) : i \in I\}$  be a maximal family of disjoint spherical caps of radius  $\rho = \tau 2^{-j}$  with centers on the hemisphere  $\mathbb{S}_+^d$ . Then for sufficiently small  $0 < \tau \leq 1$  the set of points  $\chi_j = \{\xi_i : i \in I\}$  obtained by projecting the set  $\{\tilde{\xi}_i : i \in I\}$  on  $B^d$  is a set of nodes of a cubature formula that is exact for  $\Pi_{2^{j+2}}(B^d)$ : for any  $P \in \Pi_{2^{j+2}}(B^d)$ ,*

$$\int_{B^d} P(u) du = \sum_{\xi \in \chi_j} \omega_{j,\xi} P(\xi),$$

where, moreover, the coefficients  $\omega_{j,\xi}$  of this cubature are positive and satisfy  $\omega_{j,\xi} \sim W_j(\xi)2^{-jd}$ , and the cardinality of the set  $\chi_j$  is of order  $2^{jd}$ .

### 3.1. Needlets

Going back to identities (9) and applying the cubature formula described in Proposition 1, we get

$$A_j(x, y) = \int_{B^d} C_j(x, z)C_j(z, y) dz = \sum_{\xi \in \chi_j} \omega_{j,\xi} C_j(x, \xi)C_j(y, \xi) \quad \text{and}$$

$$B_j(x, y) = \int_{B^d} D_j(x, z)D_j(z, y) dz = \sum_{\xi \in \chi_j} \omega_{j,\xi} D_j(x, \xi)D_j(y, \xi).$$

We define the *father needlets*,  $\varphi_{j,\xi}$ , and the *mother needlets*,  $\psi_{j,\xi}$ , by

$$\varphi_{j,\xi}(x) = \sqrt{\omega_{j,\xi}}C_j(x, \xi) \quad \text{and} \quad \psi_{j,\xi}(x) = \sqrt{\omega_{j,\xi}}D_j(x, \xi), \quad \xi \in \chi_j, j \geq 0.$$

We also set  $\psi_{-1,0} = \frac{1}{|B^d|}$  and  $\chi_{-1} = \{0\}$ . From above, it follows that

$$A_j(x, y) = \sum_{\xi \in \chi_j} \varphi_{j,\xi}(x)\varphi_{j,\xi}(y), \quad B_j(x, y) = \sum_{\xi \in \chi_j} \psi_{j,\xi}(x)\psi_{j,\xi}(y).$$

Therefore,

$$A_j f(x) = \int_{B^d} A_j(x, y) f(y) dy = \sum_{\xi \in \chi_j} \langle f, \varphi_{j,\xi} \rangle \varphi_{j,\xi} = \sum_{\xi \in \chi_j} \alpha_{j,\xi} \varphi_{j,\xi}, \quad \alpha_{j,\xi} = \langle f, \varphi_{j,\xi} \rangle$$

and

$$B_j f(x) = \int_{B^d} B_j(x, y) f(y) dy = \sum_{\xi \in \chi_j} \langle f, \psi_{j,\xi} \rangle \psi_{j,\xi} = \sum_{\xi \in \chi_j} \beta_{j,\xi} \psi_{j,\xi}, \quad \beta_{j,\xi} = \langle f, \psi_{j,\xi} \rangle.$$

It is easy to prove (see [21]) that

$$\|\varphi_{j,\xi}\|_2 \leq 1.$$

From (6) and the fact that  $\sum_{j \geq 0} b(t2^{-j}) = 1$  for  $t \in [1, \infty)$ , it readily follows that

$$f = \sum_{j \geq -1} \sum_{\xi \in \chi_j} \langle f, \psi_{j,\xi} \rangle \psi_{j,\xi}, \quad f \in \mathbb{L}_2(B^d),$$

and, taking the inner product with  $f$ , it leads to

$$\|f\|_2^2 = \sum_j \sum_{\xi \in \chi_j} |\langle f, \psi_{j,\xi} \rangle|^2.$$

In turn, this shows that the family  $\{\psi_{j,\xi}\}$  is a tight frame for  $\mathbb{L}_2(B^d)$ .

### 4. Needlet inversion of a noisy Radon transform and minimax performances

Our estimator is based on an appropriate thresholding of a needlet expansion as follows.  $f$  can be decomposed using the frame above:

$$f = \sum_{j \geq -1} \sum_{\xi \in \chi_j} \langle f, \psi_{j,\xi} \rangle \psi_{j,\xi}.$$

Our estimation procedure will be defined by the following steps

$$\widehat{\alpha}_{k,l,m} = \frac{1}{\lambda_k} \int g_{k,l,m} dY, \tag{10}$$

$$\widehat{\beta}_{j,\xi} = \sum_{k,l,m} \gamma_{k,l,m}^{j,\xi} \widehat{\alpha}_{k,l,m} \tag{11}$$

with

$$\gamma_{k,l,m}^{j,\xi} = \langle g_{k,l,m}, \psi_{j,\xi} \rangle$$

and

$$\hat{f} = \sum_{j=-1}^{J_\varepsilon} \sum_{\xi \in \chi_j} \hat{\beta}_{j,\xi} 1_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} c_\varepsilon\}} \psi_{j,\xi} \tag{12}$$

with

$$\nu = (d - 1)/2. \tag{13}$$

Hence, our procedure has three steps: the first one (10) corresponds to the inversion of the operator in the SVD basis, the second one (11) projects on the needlet basis and the third one (12) ends up the procedure with a final thresholding. The tuning parameters of this estimator are

- The range  $J_\varepsilon$  of resolution levels will be taken such that

$$2^{J_\varepsilon(d-1/2)} \leq (\varepsilon \sqrt{\log 1/\varepsilon})^{-1} < 2^{(J_\varepsilon+1)(d-1/2)}.$$

- The threshold constant  $\kappa$  is an important tuning of our method. The theoretical point of view asserts that, for  $\kappa$  above a constant (for which our evaluation is probably not optimal), the minimax properties hold. Evaluations of  $\kappa$  from the simulation point of view are also given.
- $c_\varepsilon$  is a constant depending on the noise level. We shall see that the following choice is appropriate:

$$c_\varepsilon = \varepsilon \sqrt{\log 1/\varepsilon}.$$

- Notice that the threshold function for each coefficient contains  $2^{j\nu}$ . This is due to the inversion of the Radon operator and the concentration relative to the  $g_{k,l,m}$ 's of the needlets.
- It is important to remark here that, unlike the (linear) procedures proposed in [11], this one does not require the knowledge of the regularity while, as will be seen in the sequel, it attains bounds that are as good as the linear ones and even better since they are handling much wider ranges for the parameters of the Besov spaces.

We will consider the minimax properties of this estimator on the Besov bodies constructed on the needlet basis. In [13], it is proved that these spaces can also be described as approximation spaces, so they have a genuine meaning and can be compared to standard Sobolev spaces.

We define here the Besov body,  $B_{\pi,r}^s$ , as the space of functions  $f = \sum_{j \geq -1} \sum_{\xi \in \chi_j} \beta_{j,\xi} \psi_{j,\xi}$  such that

$$\sum_j 2^{j s r} \left( \sum_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\pi)^\pi \right)^{r/\pi} < \infty$$

(with the obvious modifications for the cases  $\pi$  or  $r = \infty$ ) and  $B_{\pi,r}^s(M)$ , the ball of radius  $M$  of this space.

**Theorem 1.** *For  $0 < r \leq \infty$ ,  $\pi \geq 1$ ,  $1 \leq p < \infty$ , there exist some constant  $c_p = c_p(s, \pi, r, M)$ ,  $\kappa_0$  such that if  $\kappa \geq \kappa_0$ ,  $s > (d + 1)(\frac{1}{\pi} - \frac{1}{p})_+$  and, in addition, if  $\pi < p$ ,  $s > \frac{d+1}{\pi} - \frac{1}{2}$ :*

(1) If  $\frac{1}{p} < \frac{d}{d+1}$ ,

$$\begin{aligned} & \sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|\hat{f} - f\|_p^p)^{1/p} \\ & \leq c_p (\log 1/\varepsilon)^{p/2} \\ & \quad \times (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)(1/\pi-1/p))/(s+d-(d+1)/\pi) \wedge s/(s+d-1/2) \wedge (s-2(1/\pi-1/p))/(s+d-2/\pi)}. \end{aligned}$$

(2) If  $\frac{d}{d+1} \leq \frac{1}{p}$  and  $d > 2$  or  $p > 1$ ,

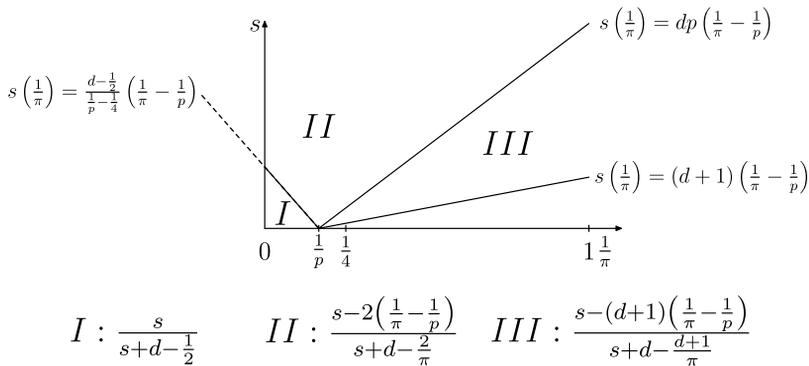
$$\sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|\hat{f} - f\|_p^p)^{1/p} \leq c_p (\log 1/\varepsilon)^{p/2} (\varepsilon \sqrt{\log 1/\varepsilon})^{s/(s+d-1/2) \wedge (s-2(1/\pi-1/p))/(s+d-2/\pi)}.$$

(3) If  $d = 2$  and  $p = 1$ ,

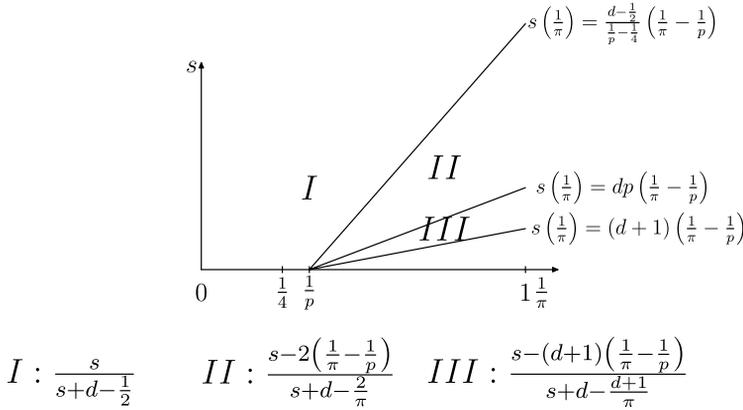
$$\sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|\hat{f} - f\|_1) \leq c_1 (\log 1/\varepsilon)^{1/2} (\varepsilon \sqrt{\log 1/\varepsilon})^{s/(s+2-1/2)}.$$

**Remark 1.** Up to logarithmic terms, the rates observed here are minimax, as will appear in the following theorem. It is known that in this kind of estimation, full adaptation yields unavoidable extra logarithmic terms. The rates of the logarithmic terms obtained in these theorems are, most of the time, suboptimal (for instance, for obvious reasons, the case  $p = 2$  yields fewer logarithmic terms). A more detailed study could lead to optimized rates, which we decided not to include here for the sake of simplicity.

The cumbersome comparisons of the different rates of convergence are summarized in Figures 1 and 2 for the case  $0 < \frac{1}{p} < \frac{d}{d+1}$ . These figures illustrate and highlight the differences between the cases  $p > 4$  and  $p < 4$ . We put  $\frac{1}{p}$  as the horizontal axis and the regularity  $s$  as the vertical axis. As explained later, after the lower-bound results, zones I and II correspond to



**Figure 1.** The three different minimax rate type zones are shown with respect to the Besov space parameters  $s$  and  $\pi$  for a fixed loss norm  $L^p$  with  $0 < \frac{1}{p} < \frac{1}{4}$ .



**Figure 2.** The three different minimax rate type zones are shown with respect to the Besov space parameters  $s$  and  $\pi$  for a fixed loss norm  $p$  with  $\frac{1}{4} < \frac{1}{p} < \frac{d}{d+1}$ .

two different types of the so called “dense” case, whereas zone III corresponds to the “sparse” case.

For the case of an  $\mathbb{L}_\infty$  loss function, we have a slightly different result since the thresholding depends on the  $\mathbb{L}_\infty$  norm of the local needlet. Let us consider the following estimate:

$$\hat{f}_\varepsilon = \sum_{j=-1}^{J_\varepsilon} \sum_{\xi \in \chi_j} \hat{\beta}_{j,\xi} \mathbf{1}_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} c_\varepsilon\}} \psi_{j,\xi},$$

$$2^{J_\varepsilon d} = (\varepsilon \sqrt{\log 1/\varepsilon})^{-1}.$$

Then, for this estimate, we have the following results:

**Theorem 2.** For  $0 < r \leq \infty$ ,  $\pi \geq 1$ ,  $s > \frac{d+1}{\pi}$ , there exist some constants  $c_\infty = c_\infty(s, \pi, r, M)$  such that if  $\kappa^2 \geq 4\tau_\infty$ , where  $\tau_\infty := \sup_{j,\xi} 2^{-j(d+1)/2} \|\psi_{j,\xi}\|_\infty$ ,

$$\sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|\hat{f}_\infty - f\|_\infty \leq c_\infty (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)/\pi)/(s+d-(d+1)/\pi)}.$$

The following theorem states lower bounds for the minimax rates over Besov spaces in this model.

**Theorem 3.** Let  $\mathcal{E}$  be the set of all estimators, for  $0 < r \leq \infty$ ,  $\pi \geq 1$ ,  $s > \frac{d+1}{\pi}$ .

(a) There exists some constant  $C_\infty = C_\infty(s, \pi, r, M)$  such that,

$$\inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|f^* - f\|_\infty \geq C_\infty (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)/\pi)/(s+d-(d+1)/\pi)}.$$

(b) For  $1 \leq p < \infty$ , there exists some constant  $C_p = C_p(s, \pi, r, M)$  such that if  $s > (\frac{d+1}{\pi} - \frac{d+1}{p})_+$ ,

(1) If  $\frac{1}{p} < \frac{d}{d+1}$

$$\inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|f^* - f\|_p^p)^{1/p} \geq C_p \varepsilon^{(s-(d+1)(1/\pi-1/p))/(s+d-(d+1)/\pi) \wedge s/(s+d-1/2) \wedge (s-2(1/\pi-1/p))/(s+d-2/\pi)}.$$

(2) If  $\frac{d}{d+1} \leq \frac{1}{p}$  and  $d > 2$  or  $p > 1$

$$\inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|f^* - f\|_p^p)^{1/p} \geq C_p \varepsilon^{s/(s+d-1/2) \wedge (s-2(1/\pi-1/p))/(s+d-2/\pi)}.$$

(3) If  $d = 2$  and  $p = 1$

$$\inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} (\mathbb{E} \|f^* - f\|_1) \geq C_p \varepsilon^{s/(s+2-1/2)}.$$

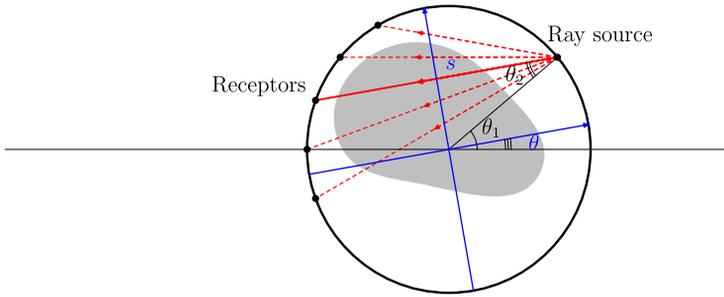
**Remark 2.** A careful look at the proof shows that the different rates observed in the two preceding theorems can be “explained” by geometrical considerations. In fact, depending on the cubature points around which they are centered, the needlets do not behave the same way. In particular, their  $\mathbb{L}_p$  norms differ. This leads us to consider two different regions on the sphere, one near the pole and one closer to the equator. In these two regions, we considered dense and sparse cases in the usual way. This yielded four rates. Then it appeared that one of them (sparse) is always dominated by the others.

## 5. Applications to the fan beam tomography

### 5.1. The 2D case: Fan beam tomography

When  $d = 2$ , the Radon transform studied in this paper is the fan beam Radon transform used in a computed axial tomography (CAT) scan. The geometry of such a device is illustrated in Figure 3. An object is placed in the middle of the scanner and X-rays are sent from a pointwise source,  $S(\theta_1)$ , making an angle  $\theta_1$  with a reference direction. Rays go through the object and the energy decay between the source and an array of receptors is measured. As the log decay along the ray is proportional to the integral of the density  $f$  of the object along the same ray, the measurements are

$$\tilde{R}f(\theta_1, \theta_2) = \int_{e_{\theta_1} + \lambda e_{(\theta_1 - \theta_2)}} \in B^2 f(x) d\lambda$$



**Figure 3.** Simplified CAT device.

with  $e_\theta = (\cos \theta, \sin \theta)$ . This is equivalent to the classical Radon transform

$$Rf(\theta, s) = \int_{\substack{y \in \theta^\perp, \\ s\theta + y \in B^2}} f(s\theta + y) \, dy, \quad \theta \in \mathbf{S}^1, s \in [-1, 1],$$

for  $\theta = \theta_1 - \theta_2$  and  $s = \sin \theta_2$ . The ray source is then rotated to a different angle and the measurement process is repeated. In our Gaussian white noise model, we measure the continuous function  $Rf(\theta, s)$  through the process  $dY = Rf(\theta, s) \, d\theta \frac{ds}{(1-s^2)} + \varepsilon \, dW(\theta, s)$ . The measure  $d\theta \frac{ds}{(1-s^2)}$  corresponds to the uniform measure  $d\theta_1 \, d\theta_2$  by the change of the variable that maps  $(\theta_1, \theta_2)$  into  $(\theta, s)$ . Our goal is to recover the unknown function,  $f$ , from the observation of  $Y$  using the needlet thresholding mechanism described in the previous sections.

In our implementation, we exploit the tensorial structure of the SVD basis of the disk in polar coordinates:

$$f_{k,l,m}(r, \theta) = (2k + 2)^{1/2} P_j^{(0,l)}(2|r|^2 - 1) |r|^l Y_{l,m}(\theta), \quad 0 \leq l \leq k, k - l = 2j, 1 \leq m \leq 2,$$

where  $P_j^{0,l}$  is the corresponding Jacobi polynomial, and  $Y_{l,1}(\theta) = c_l \cos(l\theta)$  and  $Y_{l,2}(\theta) = c_l \sin(l\theta)$  with  $c_0 = \frac{1}{\sqrt{2\pi}}$  and  $c_l = \frac{1}{\sqrt{\pi}}$ , otherwise. The basis of  $S^2 \times [-1, 1]$  has a similar tensorial structure as it is given by

$$g_{k,l,m}(\theta, s) = [h_k]^{-1/2} (1 - s^2)^{1/2} C_k^1(s) Y_{l,m}(\theta), \quad k \geq 0, l \geq 0, 1 \leq m \leq 2,$$

where  $C_k^1$  is the Gegenbauer of parameter 1 and degree  $k$ . We recall that the corresponding eigenvalues are

$$\lambda_k = \frac{2\sqrt{\pi}}{\sqrt{k+1}}.$$

### 5.2. SVD, needlet and cubature

In our numerical studies, we compare four different type of estimators: linear SVD estimators, thresholded SVD estimators, linear needlet estimators and thresholded needlet estimators. They are defined from the measurement of the values of the Gaussian field on the SVD basis function

$Y_{g_{k,l,m}}$  and the following linear estimates of, respectively, the SVD basis coefficients  $\langle f, f_{k,l,m} \rangle$  and the needlet coefficients  $\langle f, \psi_{j,\xi} \rangle$ ,

$$\widehat{\alpha}_{k,l,m} = \frac{1}{\lambda_k} Y_{g_{k,l,m}} = \frac{1}{\lambda_k} \int g_{k,l,m} dY$$

and

$$\widehat{\beta}_{j,\xi} = \sqrt{\omega_{j,\xi}} \sum_k \sqrt{b(k/2^j)} \sum_{l,m} g_{k,l,m}(\xi) \widehat{\alpha}_{k,l,m}.$$

The estimators we consider are respectively defined as:

linear SVD estimates	$\widehat{f}_J^{\text{LS}} = \sum_{k < 2^J} \sum_{l,m} \widehat{\alpha}_{k,l,m} f_{k,l,m},$
linear needlet estimates	$\widehat{f}_J^{\text{LN}} = \sum_{j < J} \sum_{\xi} \widehat{\beta}_{j,\xi} \psi_{j,\xi},$
thresholded SVD estimates	$\widehat{f}_T^{\text{TS}} = \sum_{k < 2^J} \sum_{l,m} \rho_{T_k}(\widehat{\alpha}_{k,l,m}) f_{k,l,m},$
thresholded needlet estimates	$\widehat{f}_T^{\text{TN}} = \sum_{j < J} \sum_{\xi} \rho_{T_{j,\xi}}(\widehat{\beta}_{j,\xi}) \psi_{j,\xi},$

where  $\rho_T(\cdot)$  is the hard threshold function with threshold  $T$ :

$$\rho_T(x) = \begin{cases} x, & \text{if } |x| \geq T, \\ 0, & \text{otherwise.} \end{cases}$$

A more precise description is given in Table 1. In our experiments, the values of  $Y_{g_{k,l,m}}$  have been obtained from an initial approximation of  $\langle f, f_{k,l,m} \rangle$  computed with a very fine cubature to which a Gaussian i.i.d. sequence is added.

We have used, in our numerical experiments, thresholds of the form

$$T_k = \frac{\kappa}{\lambda_k} \varepsilon \sqrt{\log 1/\varepsilon} \quad \text{and} \quad T_{j,\xi} = \kappa \sigma_{j,\xi} \varepsilon \sqrt{\log 1/\varepsilon},$$

where  $\sigma_{j,\xi}$  is the standard deviation of the noisy needlet coefficients when  $f = 0$  and  $\varepsilon = 1$ :

$$\sigma_{j,\xi}^2 = \omega_{j,\xi} \sum_k b(k/2^j) \sum_{l,m} g_{k,l,m}(\xi)^2.$$

Note that, while the needlet threshold is different than in Theorem 1, as  $\sigma_{j,\xi}$  is of order  $2^{j\nu}$ , its conclusions remain valid.

An important issue in the needlet scheme is the choice of the cubature in the needlet construction. Proposition 1 ensures the existence of a suitable cubature  $\xi_j$  for every level  $j$  based on a cubature  $\widetilde{\xi}_j$  on the sphere but gives neither an explicit construction of the points on the sphere nor an explicit formula for the weights  $\omega_{j,\xi}$ . Those ingredients are, nevertheless, central in the numerical scheme and should be specified.

**Table 1.** Algorithmic description of the considered estimators

	Linear SVD $\hat{f}^{LS}$	Thresh. SVD $\hat{f}^{TS}$	Linear needlet $\hat{f}^{LN}$	Thresh. needlet $\hat{f}^{TN}$
Observation	$dY = Rf d\theta \frac{ds}{\sqrt{1-s^2}} + \varepsilon dW$			
SVD Dec	$Y_{h_{k,l}} = \langle Y, g_{k,l,m} \rangle = \langle Rf, g_{k,l,m} \rangle + \varepsilon_{k,l,m} = \mu_k \langle f, f_{k,l,m} \rangle + \varepsilon_{k,l,m}$			
Inv. Radon	$\hat{\alpha}_{k,l,m} = \frac{1}{\mu_k} \langle Y, g_{k,l,m} \rangle = \langle f, f_{k,l,m} \rangle + \frac{1}{\mu_k} \varepsilon_{k,l,m}$			
Needlet transf.	$\hat{\beta}_{j,\xi} = \sqrt{\omega_{j,\xi}} \sum_k \sqrt{b(\frac{k}{2^j})} \times \sum_{l,m} f_{k,l,m}(\xi) \hat{\alpha}_{k,l,m}$			
Coeff. mod.	$\hat{\alpha}_{k,l,m}^{SL} = \mathbf{1}_{k \leq k_{\max}} \hat{\alpha}_{k,l,m}$	$\hat{\alpha}_{k,l,m}^{ST} = \rho_{T_k}(\hat{\alpha}_{k,l,m})$	$\hat{\beta}_{j,\xi}^{NL} = \mathbf{1}_{j < j_{\max}} \hat{\beta}_{j,\xi}$	$\hat{\beta}_{j,\xi}^{NT} = \rho_{T_{j,\xi}}(\hat{\beta}_{j,\xi})$
Needlet inv.	$\hat{\alpha}_{k,l,m}^* = \sum_j \sqrt{b(\frac{k}{2^j})} \times \sum_{\xi \in \chi_j} \sqrt{\omega_{j,\xi}} f_{k,l,m}(\xi) \hat{\beta}_{j,\xi}^*$			
SVD rec.	$\hat{f}^* = \sum_{k,l,m} \hat{\alpha}_{k,l,m}^* f_{k,l,m}$			

Three possibilities have been considered: a numerical cubature deduced from an almost uniform cubature of the half sphere available, an approximate cubature deduced from the Healpix cubature on the sphere and a cubature obtained by subsampling a tensorial cubature associated to the latitude and longitude coordinates on the sphere. The first strategy has been considered by Baldi *et al.* [2] in a slightly different context; there is, however, a strong limitation on the maximum degree of the cubature available and, thus, this solution has been abandoned. The Healpix strategy, also considered by Baldi *et al.* in another paper [3], is easily implementable but, as it is based on an approximate cubature, fails to be precise enough in our numerical experiments. The last strategy relies on the subsampling on a tensorial grid on the sphere. While such a strategy provides a simple way to construct an admissible cubature, the computation of the cubature weights is becoming an issue as no closed form is available.

To overcome those issues, we have considered a cubature formula based on the full tensorial grid appearing as proposed by [16]. This cubature does not satisfy the condition of Proposition 1, but its weights can be computed explicitly. Furthermore, we argue that, using our modified threshold, we can still control the risk of the estimator. Indeed, note first that the modified threshold is such that the thresholding of a needlet depends only on its scale parameter  $j$  and on its center  $\xi$  and not on the corresponding cubature weight  $\omega_{j,\xi}$ . Assume now that we have a collection of  $K$  cubature, each satisfying conditions of Proposition 1 and thus defining a suitable estimate,  $\hat{f}_k$ . We can use the “average” cubature obtained by adding all the cubature points and averaging the cubature weights. This new cubature defines a new estimate,  $\hat{f}$ , satisfying

$$\hat{f} = \frac{1}{K} \sum_{k=1}^K \hat{f}_k.$$

By convexity, for any  $p \geq 1$ ,

$$\|f - \hat{f}\|_p^p = \left\| f - \frac{1}{K} \sum_{k=1}^K \hat{f}_k \right\|_p^p \leq \frac{1}{K} \sum_{k=1}^K \|f - \hat{f}_k\|_p^p$$

and, thus, this average estimator is as efficient as the worst estimator in the family  $\hat{f}_k$ . The full tensorial cubature is an average of suitable cubatures. The corresponding estimator satisfies the error bounds of Theorems 1 and 2. Note that this principle is quite close to the cycle-spinning method introduced by Donoho *et al.* Indeed, the same kind of numerical gain is obtained with this method. The numerical comparison of the Healpix cubature and our tensorial cubature is largely in favor of our scheme. Furthermore, as already noticed by [16], the tensorial structure of the cubature leads to some simplification in the numerical implementation of the needlet estimator. The resulting scheme is almost as fast as the Healpix-based one.

### 5.3. Numerical results

In this section, we compare five “estimators” (linear SVD with best scale, linear needlet with best scale, thresholded SVD with best  $\kappa$ , thresholded needlet with best  $\kappa$  and thresholded needlet with  $\kappa = 3$ ) for seven different norms ( $\mathbb{L}_1, \mathbb{L}_2, \mathbb{L}_4, \mathbb{L}_6, \mathbb{L}_7, \mathbb{L}_8, \mathbb{L}_{10}$  and  $\mathbb{L}_\infty$ ) and seven noise levels  $\varepsilon$  ( $2^k/1000$  for  $k$  in  $0, 1, \dots, 6$ ). Each subfigure of Figures 4 and 5 plots the logarithm of the estimation error for a specific norm against the opposite of the logarithm of the noise level. Note that the subfigure overall aspect is explained by the error decay when the noise level diminishes. The good theoretical behavior of the thresholded needlet estimator is confirmed numerically: the thresholded needlet estimator with an optimized  $\kappa$  appears as the best estimator for every norm while a fixed  $\kappa$  yields a very good estimator, except for the  $\mathbb{L}_\infty$  case, as expected by our theoretical results. These results are confirmed visually by the reconstructions of Figure 6. In the needlet ones, errors are smaller and much more localized than in their SVD counterparts. Observe also how the fine structures are much more preserved with the thresholded needlet estimate than with any other methods.

We conclude this paper with some sections devoted to the proofs of our results.

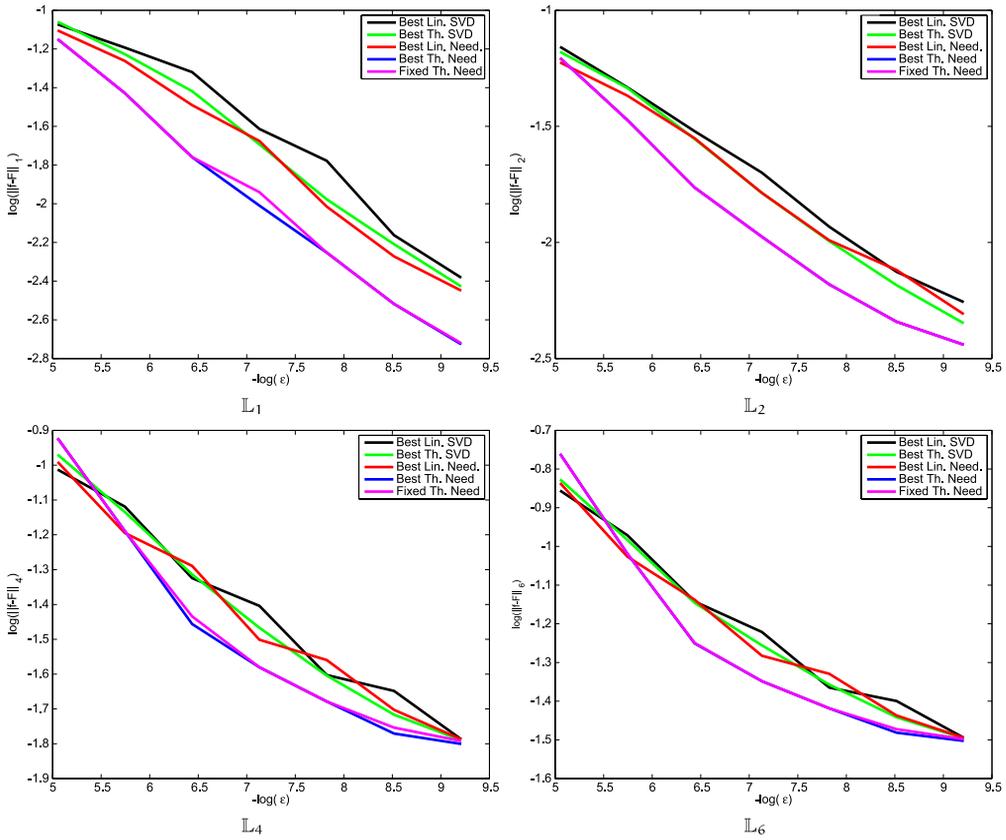
## 6. Needlet properties

### 6.1. Key inequalities

The following inequalities are true (and proved in [13,17,18,20]) and will be fundamental in the sequel. In the following lines,  $g_{j,\xi}$  will stand either for  $\varphi_{j,\xi}$  or  $\psi_{j,\xi}$ :

$$\forall j \in \mathbb{N}, \forall \xi \in \chi_j \quad 0 < c \leq \|g_{j,\xi}\|_2^2 \leq 1, \tag{14}$$

$$\forall j \in \mathbb{N}, \xi \in \chi_j, \forall x \in \mathcal{X} \quad \sum_{\xi \in \chi_j} \|g_{j,\xi}\|_1 |g_{j,\xi}(x)| \leq C < \infty, \tag{15}$$



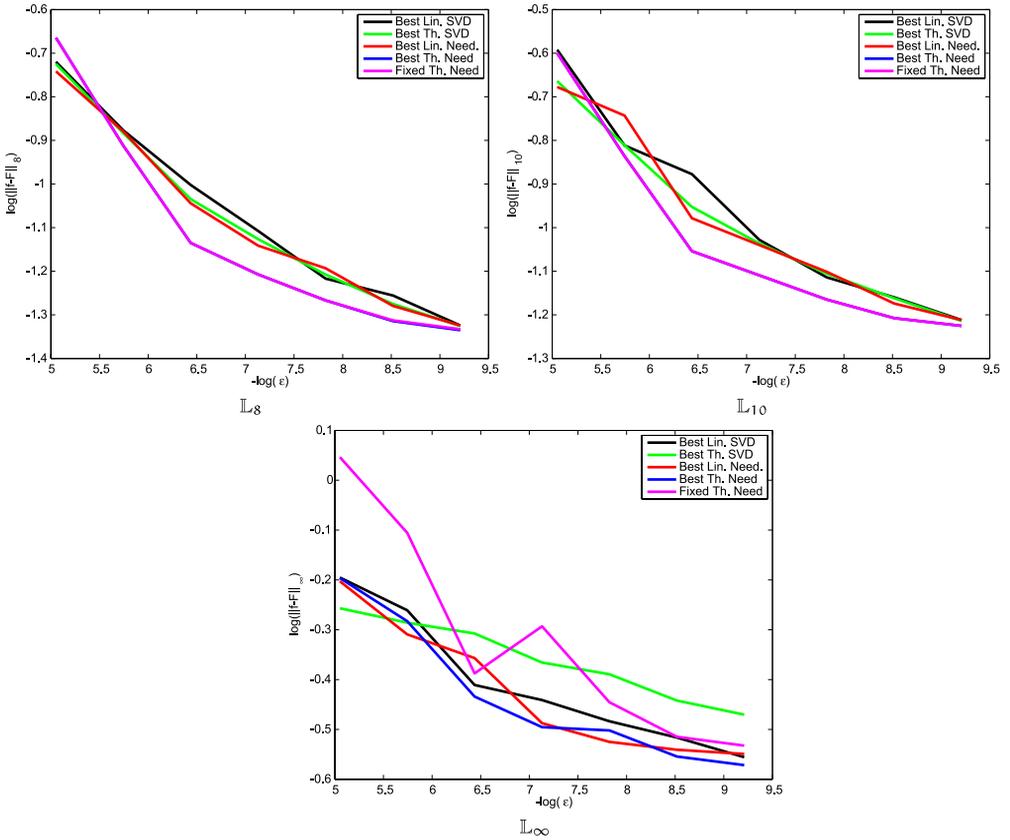
**Figure 4.** Estimation results in  $\log \mathbb{L}_p$  norm. Each figure shows the decay of the logarithm of the error against the logarithm of the noise parameter for the specified norm.

$$|g_{j,\xi}(x)| \leq C_M \frac{2^{jd/2}}{\sqrt{W_j(x)}(1 + 2^j d(x, \xi))^M} \tag{16}$$

(recall that  $W_j(x)$  has been defined in (8)). From these inequalities, one can deduce the following ones (see [12]): for all  $1 \leq p \leq \infty$ ,

$$\left( \sum_{\xi \in \mathcal{X}_j} |\langle f, g_{j,\xi} \rangle|^p \|g_{j,\xi}\|_p^p \right)^{1/p} \leq C \|f\|_p, \tag{17}$$

$$\left\| \sum_{\xi \in \mathcal{X}_j} \lambda_\xi g_{j,\xi}(x) \right\|_p \leq \left( \frac{C}{c} \right)^2 \left( \sum_{\xi \in \mathcal{X}_j} \|\lambda_\xi g_{j,\xi}\|_p^p \right)^{1/p}. \tag{18}$$



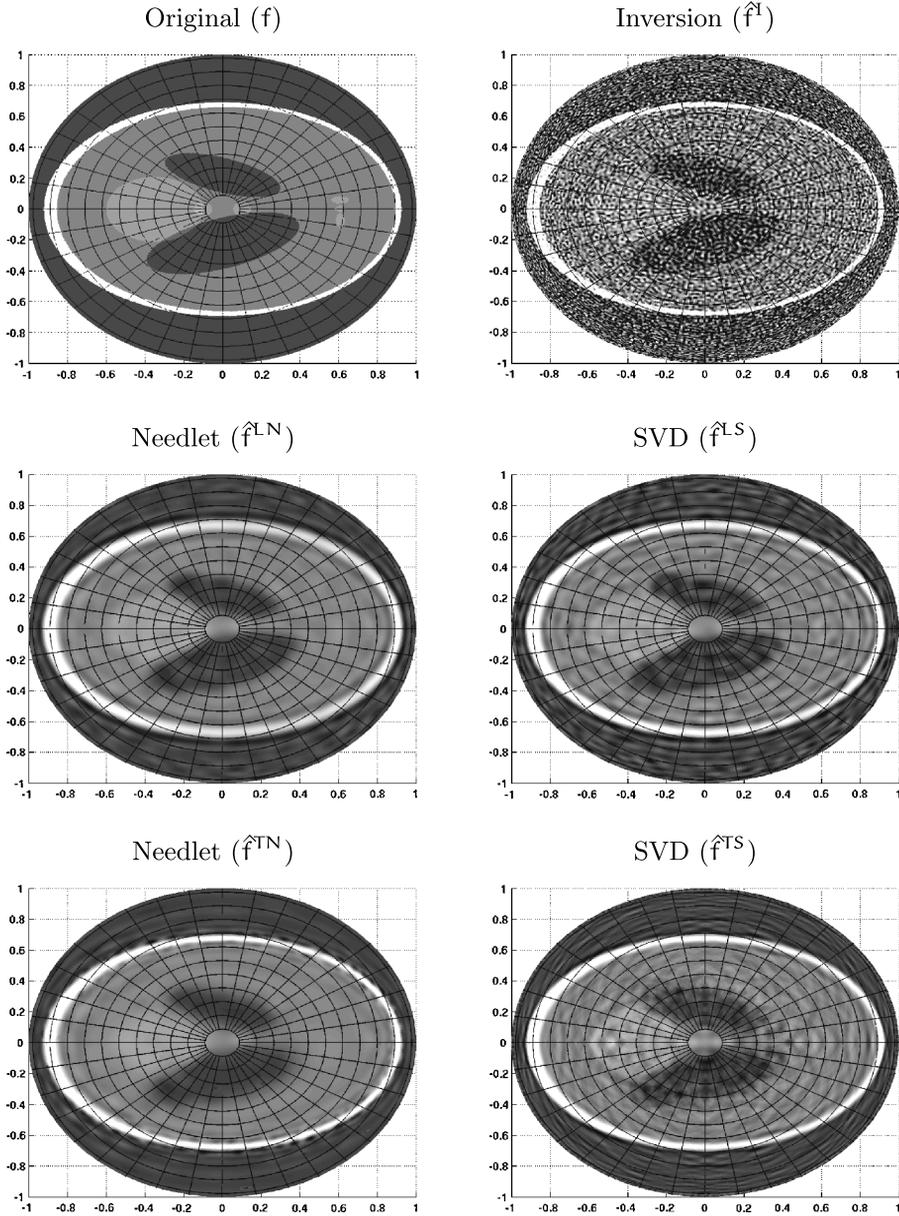
**Figure 5.** Estimation results in  $\log \mathbb{L}_p$  norm. Each figure shows the decay of the logarithm of the error against the logarithm of the noise parameter for the specified norm.

### 6.2. Besov embeddings

It is a key point to clarify how the Besov spaces defined above may be included in each of the others. As will be seen, the embeddings will parallel the standard embeddings of usual Besov spaces, but with important differences, which, in particular, yield new minimax rates of convergence as detailed above.

We begin with an evaluation of the different  $\mathbb{L}_p$  norms of the needlets. More precisely, in [13] it is shown that, for  $0 < p \leq \infty$ ,

$$\|\psi_{j,\xi}\|_p \sim \|\varphi_{j,\xi}\|_p \sim \left(\frac{2^{jd}}{W_j(\xi)}\right)^{1/2-1/p}, \quad \xi \in \chi_j. \tag{19}$$



**Figure 6.** Visual comparison for the original Logan–Shepp phantom with  $\varepsilon = 8/1000$ . Errors are much more localized in the needlet-based estimates compared to the fully delocalized errors of the SVD based estimates. Fine structures are much more restored in the thresholded needlet estimate than in the other estimates.

The following inequalities are proved in [11]:

$$\sum_{\xi \in \chi_j} \|g_{j,\xi}\|_p^p \leq c2^{j(d p/2 + (p/2 - 2)_+)} \quad \text{if } p \neq 4, \tag{20}$$

$$\sum_{\xi \in \chi_j} \|g_{j,\xi}\|_p^p \leq c j 2^{j d p/2} \quad \text{if } p = 4. \tag{21}$$

We are now able to state the embedding results (see [12]).

**Theorem 4.** (1)  $1 \leq p \leq \pi \leq \infty \Rightarrow B_{\pi,r}^s \subseteq B_{p,r}^s$ .

(2)  $\infty \geq p \geq \pi > 0, s > (d + 1)(1/\pi - 1/p), \Rightarrow B_{\pi,r}^s \subseteq B_{p,r}^{s - (d+1)(1/\pi - 1/p)}$ .

## 7. Proof of the upper bounds

A important tool for the proof of the upper bounds that clarifies the thresholding procedure is the following lemma.

**Lemma 1.** For all  $j \geq -1, \xi \in \chi_j, \widehat{\beta}_{j,\xi}$  has a Gaussian distribution with mean  $\beta_{j,\xi}$  and variance  $\sigma_{j,\xi}^2$ , with

$$\sigma_{j,\xi}^2 \leq c 2^{j(d-1)} \varepsilon^2.$$

**Proof.** As we can write

$$\begin{aligned} \widehat{\beta}_{j,\xi} &= \sum_{k,l,m} \gamma_{k,l,m}^{j,\xi} \int_{B^d} f f_{k,l,m} \, dx + \sum_{k,l,m} \gamma_{k,l,m}^{j,\xi} \frac{\varepsilon}{\lambda_k} Z_{k,l,m} \\ &= \beta_{j,\xi} + Z_{j,\xi}. \end{aligned}$$

Here the summation is over  $\{(k, l, m): 0 \leq k < 2^j, 0 \leq l \leq k, l \equiv k \pmod{2}, 1 \leq m \leq N_{d-1}(l)\}$ . Since the  $Z_{k,l,m}$ 's are independent  $N(0, 1)$  random variables,  $Z_{j,\xi} \sim N(0, \sigma_{j,\xi}^2)$ , we have

$$\sigma_{j,\xi}^2 = \varepsilon^2 \sum_{k,l,m} |\gamma_{k,l,m}^{j,\xi}|^2 \frac{\binom{k}{d}}{\pi^{d-1} 2^d k} \leq \frac{(2^j)^{d-1}}{\pi^{d-1} 2^d} \leq c 2^{j(d-1)} \varepsilon^2 \tag{22}$$

with  $c = (d/2\pi)^{d-1}$ . Here, we used that  $\{f_{k,l,m}\}$  is an orthonormal basis for  $\mathbb{L}_2$  and hence  $\sum_{k,l,m} |\gamma_{k,l,m}^{j,\xi}|^2 = \|\psi_{j,\xi}\|_2^2 \leq 1$ . □

Let us now begin with the second theorem, the proof of which is slightly simpler.

## 7.1. Proof of Theorem 2

In this proof, as well as in the other one,  $C$  will denote any *constant* in the sense of Theorems 1–3; we assume that  $f$  belongs to the Besov ball  $B_{\pi,r}^s(M)$  and measure the loss in  $\mathbb{L}_p$  norm (here,  $p = \infty$ ). Then  $C$  denotes a generic quantity that depends only on  $s, r, p$  and  $M$ . Note that the exact value denoted by  $C$  may vary from one line to another.

We have, if we denote

$$\begin{aligned} A_J(f) &:= \sum_{j>J} \sum_{\xi \in \chi_j} \beta_{j,\xi} \psi_{j,\xi}, \\ \|\hat{f}_\infty - f\|_\infty &\leq \|\hat{f}_\infty - A_J(f)\|_\infty + \|A_J(f) - f\|_\infty \\ &\leq \|\hat{f}_\infty - A_J(f)\|_\infty + C \|f\|_{B_{\pi,r}^s} 2^{-J(s-(d+1)/\pi)}. \end{aligned}$$

We have used, as  $B_{\pi,r}^s \subset B_{\infty,r}^{s-(d+1)/\pi}$ ,

$$\|A_J f - f\|_\infty \leq C \|f\|_{B_{\pi,r}^s} 2^{-J(s-(d+1)/\pi)}.$$

Moreover,

$$2^{-J(s-(d+1)/\pi)} \leq (\varepsilon \sqrt{\log 1/\varepsilon})^{-(s-(d+1)/\pi)/d} \leq (\varepsilon \sqrt{\log 1/\varepsilon})^{-(s-(d+1)/\pi)/(s-(d+1)/\pi+d)}$$

as  $s > \frac{d+1}{\pi}$ .

We have, using (18),

$$\begin{aligned} \|\hat{f}_\infty - A_J(f)\|_\infty &\leq \sum_{j<J} \left\| \sum_{\xi \in \chi_j} (\hat{\beta}_{j,\xi} 1_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} - \beta_{j,\xi}) \psi_{j,\xi} \right\|_\infty \\ &\leq \sum_{j<J} \left( \left\| \sum_{\xi \in \chi_j} ((\hat{\beta}_{j,\xi} - \beta_{j,\xi}) \psi_{j,\xi} 1_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right\|_\infty \right. \\ &\quad \left. + \left\| \sum_{\xi \in \chi_j} (\beta_{j,\xi} \psi_{j,\xi} 1_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty < \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right\|_\infty \right) \\ &\leq \frac{C}{c} \sum_{j<J} \left( \sup_{\xi \in \chi_j} (|\hat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty 1_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right. \\ &\quad \left. + \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty 1_{\{\|\hat{\beta}_{j,\xi}\| \|\psi_{j,\xi}\|_\infty < \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right). \end{aligned}$$

We decompose the first term of the last inequality

$$\begin{aligned}
& \sup_{\xi \in \mathcal{X}_j} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&= \sup_{\xi \in \mathcal{X}_j} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\quad \times \left( \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} + \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\leq \sup_{\xi \in \mathcal{X}_j} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\quad + \sup_{\xi \in \mathcal{X}_j} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right)
\end{aligned}$$

and the second one

$$\begin{aligned}
& \sup_{\xi \in \mathcal{X}_j} \left( |\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&= \sup_{\xi \in \mathcal{X}_j} \left( |\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\quad \times \left( \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq 2\kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} + \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < 2\kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\leq \sup_{\xi \in \mathcal{X}_j} \left( |\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\
&\quad + \sup_{\xi \in \mathcal{X}_j} \left( |\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right).
\end{aligned}$$

Now we will bound each of the four terms coming from the last two inequalities. Since for  $X \sim N(0, \sigma^2)$ , we have

$$\mathbb{E}(|Y| \mathbf{1}_{\{|Y| > \lambda \sigma\}}) = \sigma \frac{2}{\sqrt{2\pi}} \int_{\lambda}^{\infty} y e^{-y^2/2} dy = e^{-\lambda^2/2} \frac{2}{\sqrt{2\pi}} \leq e^{-\lambda^2/2}.$$

Noticing that the standard deviation of  $(\widehat{\beta}_{j,\xi} - \beta_{j,\xi}) \|\psi_{j,\xi}\|_\infty$  is smaller than  $\tau_\infty 2^{jd} \varepsilon$  (using Lemma 1 and (16)), we have

$$\begin{aligned}
& \sum_{j \leq J} \mathbb{E} \left( \sup_{\xi \in \mathcal{X}_j} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \right) \\
&\leq \sum_{j \leq J} \sum_{\xi \in \mathcal{X}_j} \mathbb{E} \left( |\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \geq (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \leq c 2^{Jd} \varepsilon^{\kappa^2/2} \tau_\infty^2 \\
&\leq C \varepsilon^{\kappa^2/2} \tau_\infty^2 - 1 \sqrt{\log 1/\varepsilon}^{-1} \leq C \varepsilon \sqrt{\log 1/\varepsilon}
\end{aligned}$$

if  $\kappa^2 \geq 4\tau_\infty^2$ , where we have used  $\text{Card } \chi_j \leq c2^{jd}$ . This proves that this term will be of the right order.

For the second term, let us observe that we have, using Theorem 4,

$$\|\beta_{j,\xi}\| \|\psi_{j,\xi}\|_\infty \leq C \|f\|_{B_{\pi,r}^s} 2^{-j(s-(d+1)/\pi)},$$

so only the  $j$ 's index such that  $j \leq j_1$  will verify this inequality:

$$2^{j_1} \sim \left( \frac{2C \|f\|_{B_{\pi,r}^s}}{\kappa} (\varepsilon \sqrt{\log 1/\varepsilon}) \right)^{-1/(s+d-(d+1)\pi)}.$$

On the other side, using the Pisier lemma [23],

$$\mathbb{E} \left( \sup_{\xi \in \chi_j} (|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty) \right) \leq \tau_\infty 2^{jd} \varepsilon \sqrt{2 \log 2c2^{jd}}.$$

So

$$\begin{aligned} & \sum_{j \leq j_1} \mathbb{E} \left( \sup_{\xi \in \chi_j} (|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{\|\beta_{j,\xi}\| \|\psi_{j,\xi}\|_\infty > (\kappa/2) 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right) \\ & \leq \tau_\infty \sum_{j \leq j_1} 2^{jd} \varepsilon \sqrt{2 \log 2c2^{jd}} \leq C \varepsilon j_1^{1/2} 2^{j_1 d} \\ & \lesssim C (\|f\|_{B_{\pi,r}^s})^{(s-(d+1)\pi)/(s+d-(d+1)\pi)} (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)\pi)/(s+d-(d+1)\pi)}. \end{aligned}$$

This proves that this term will be of the right order. Concerning the first term of the second inequality,

$$\begin{aligned} & \mathbb{E} \left( \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right) \\ & \leq C \|f\|_{B_{\pi,r}^s} \sum_{\xi \in \chi_j} P(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}) \end{aligned}$$

but

$$P(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}) \leq e^{-(\kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon})^2 / (2(\varepsilon 2^{j(d-1)/2} \|\psi_{j,\xi}\|_\infty)^2)} \leq \varepsilon^{\kappa^2/2\tau_\infty^2}.$$

So

$$\begin{aligned} & \sum_{j \leq J} \mathbb{E} \left( \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty > \kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right) \\ & \leq C \|f\|_{B_{\pi,r}^s} J \varepsilon^{\kappa^2/2\tau_\infty^2} \leq C \|f\|_{B_{\pi,r}^s} (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)\pi)/(s+d-(d+1)\pi)} \end{aligned}$$

if  $\kappa^2/2 \geq \tau_\infty^2$ . This proves that this term will be of the right order. Concerning the second term of the second inequality,

$$\begin{aligned} & \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < 2\kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\ & \leq 2\kappa 2^{jd} \varepsilon \sqrt{\log 1/\varepsilon} \wedge C \|f\|_{B_{\pi,r}^s} 2^{-j(s-(d+1)/\pi)}, \end{aligned}$$

let us again take

$$2^{j_1} \sim (\varepsilon \sqrt{\log 1/\varepsilon})^{-1/(s+d-(d+1)\pi)}$$

$$\begin{aligned} & \sum_{j \leq J} \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty \mathbf{1}_{\{|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\infty < 2\kappa 2^{jd} \varepsilon \log 1/\varepsilon\}}) \\ & \leq C \|f\|_{B_{\pi,r}^s} \left( \varepsilon \log 1/\varepsilon \sum_{j \leq j_1} 2^{jd} + \sum_{j_1 < j \leq J} 2^{-j(s-(d+1)/\pi)} \right) \\ & \lesssim C \|f\|_{B_{\pi,r}^s} (\varepsilon \sqrt{\log 1/\varepsilon})^{(s-(d+1)/\pi)/(s+d-(d+1)/\pi)}. \end{aligned}$$

This ends the proof of Theorem 2.

## 7.2. Proof of Theorem 1

As in the previous proof, we begin with the decomposition,

$$\|\hat{f}_\varepsilon - f\|_p^p \leq 2^{p-1} (\|\hat{f}_\varepsilon - A_J(f)\|_p^p + \|A_J(f) - f\|_p^p).$$

Using the fact that, for  $\pi \geq p$ ,  $B_{\pi,r}^s \subset B_{p,\infty}^s$ , and for  $\pi \leq p$ ,  $B_{\pi,r}^s \subset B_{p,\infty}^{s-(d+1)(1/\pi-1/p)}$ , we obtain

$$\|A_J(f) - f\|_p^p \leq C \|f\|_{B_{\pi,r}^s}^p 2^{-Jsp} \quad \text{if } \pi \geq p$$

and

$$\|A_J(f) - f\|_p^p \leq C \|f\|_{B_{\pi,r}^s}^p 2^{-J(s-(d+1)(1/\pi-1/p))p} \quad \text{if } \pi \leq p.$$

We have  $2^{-Jsp} \leq (\varepsilon \sqrt{\log 1/\varepsilon})^{sp/(d-1/2)} \leq (\varepsilon \sqrt{\log 1/\varepsilon})^{sp/(s+d-1/2)}$ . Obviously, this term has the right rate for  $\pi \geq p$ . For  $\pi < p$ ,

$$\begin{aligned} 2^{-J(s-(d+1)(1/\pi-1/p))p} & \leq (\varepsilon \log 1/\varepsilon)^{(s-(d+1)(1/\pi-1/p))/(d-1/2)} \\ & \leq (\varepsilon \log 1/\varepsilon)^{(s-(d+1)(1/\pi-1/p))/(s+d-(d+1)/\pi)}, \end{aligned}$$

thanks to  $s \geq (d+1)/\pi - 1/2$ . This gives the right rate for  $dp > d+1$ . For  $dp \leq d+1$ , we have (again as  $s \geq (d+1)/\pi - 1/2$ ),  $s - (d+1)(\frac{1}{\pi} - \frac{1}{p}) \geq d - \frac{1}{2}$ , so  $2^{-J(s-(d+1)(1/\pi-1/p))p} \leq$

$(\varepsilon \log 1/\varepsilon)^{(s-(d+1)(1/\pi-1/p))/(d-1/2)} \leq (\varepsilon \log 1/\varepsilon)^{s/(s+d-1/2)}$ . Finally, this proves that the bias term above always has the right rate.

Let us now investigate the stochastic term

$$\mathbb{E} \|\hat{f} - A_J(f)\|_p^p \leq C J^{p-1} \sum_{j < J} \mathbb{E} \left\| \sum_{\xi \in \mathcal{X}_j} (\hat{\beta}_{j,\xi} \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} - \beta_{j,\xi}) \psi_{j,\xi} \right\|_p^p.$$

But

$$\begin{aligned} & \left\| \sum_{\xi \in \mathcal{X}_j} (\hat{\beta}_{j,\xi} \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} - \beta_{j,\xi}) \psi_{j,\xi} \right\|_p^p \\ & \leq C 2^{p-1} \left( \left\| \sum_{\xi \in \mathcal{X}_j} ((\hat{\beta}_{j,\xi} - \beta_{j,\xi}) \psi_{j,\xi} \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right\|_p^p \right. \\ & \quad \left. + \left\| \sum_{\xi \in \mathcal{X}_j} (\beta_{j,\xi} \psi_{j,\xi} \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \right\|_p^p \right) \\ & \leq C \left( \sum_{\xi \in \mathcal{X}_j} |\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right. \\ & \quad \left. + \sum_{\xi \in \mathcal{X}_j} |\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right). \end{aligned}$$

In turn,

$$\begin{aligned} & |\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \\ & = |\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \left( \mathbf{1}_{\{|\beta_{j,\xi}| \geq (\kappa/2) 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} + \mathbf{1}_{\{|\beta_{j,\xi}| < (\kappa/2) 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\ & \leq |\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi} - \beta_{j,\xi}| \geq (\kappa/2) 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} + |\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \mathbf{1}_{\{|\beta_{j,\xi}| > (\kappa/2) 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \end{aligned}$$

and

$$\begin{aligned} |\beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} & = |\beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \\ & \quad \times \left( \mathbf{1}_{\{|\beta_{j,\xi}| \geq \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} + \mathbf{1}_{\{|\beta_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\ & \leq |\beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi} - \beta_{j,\xi}| > \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} + |\beta_{j,\xi}|^p \mathbf{1}_{\{|\beta_{j,\xi}| < \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}. \end{aligned}$$

We now have the following bound using direct computation:

$$\mathbb{E} (|\hat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \mathbf{1}_{\{|\hat{\beta}_{j,\xi} - \beta_{j,\xi}| \geq (\kappa/2) 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \leq C 2^{j\nu} \varepsilon^p (\kappa \sqrt{\log 1/\varepsilon})^{p-1} \varepsilon^{\kappa^2/2}.$$

Hence,

$$\begin{aligned}
& J^{p-1} \sum_{j \leq J} \sum_{\xi \in \chi_j} \mathbb{E}(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| \geq (\kappa/2)2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\
& \leq C J^{p-1} \sum_{j \leq J} \varepsilon^p (\kappa \sqrt{\log 1/\varepsilon})^{p-1} \varepsilon^{\kappa^2/2} \sum_{\xi \in \chi_j} \|\psi_{j,\xi}\|_p^p \\
& \leq C \sqrt{\log 1/\varepsilon}^{p-1} \varepsilon^p (\kappa \sqrt{\log 1/\varepsilon})^{p-1} \varepsilon^{\kappa^2/2} \sum_{j \leq J} 2^{j p \nu} 2^{j(d p/2 + (p/2 - 2)_+)} \\
& \leq C \varepsilon^p
\end{aligned}$$

if  $\kappa \geq \sqrt{2p}$  is large enough (we have used (19)).

Using the forthcoming inequality (24),

$$\begin{aligned}
& \sum_{\xi \in \chi_j} \mathbb{E}(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\beta_{j,\xi}| > (\kappa/2)2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\
& \leq C \varepsilon^p \sum_{\xi \in \chi_j} 2^{j\nu p} \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\beta_{j,\xi}| > (\kappa/2)2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}} \\
& \leq C \varepsilon^p \left( \frac{\kappa}{2} \varepsilon \sqrt{\log 1/\varepsilon} \right)^{-q}.
\end{aligned}$$

Hence,

$$\begin{aligned}
& J^{p-1} \sum_{j \leq J} \sum_{\xi \in \chi_j} \mathbb{E}(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\beta_{j,\xi}| > (\kappa/2)2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\
& \leq C J^p \sqrt{\log 1/\varepsilon}^{-p} \left( \frac{\kappa}{2} \right)^{-q} (\varepsilon \sqrt{\log 1/\varepsilon})^{p-q}
\end{aligned}$$

and, as  $J \leq \frac{1}{d-1/2} \log 1/\varepsilon + 1$ ,

$$\leq C (\varepsilon \sqrt{\log 1/\varepsilon})^{p-q} (\log 1/\varepsilon)^{p/2}.$$

This term is thus of the right rate. Let us now turn to

$$\begin{aligned}
& \sum_{\xi \in \chi_j} \mathbb{E}(|\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| > \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\
& = \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p P(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| > \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}).
\end{aligned}$$

As the standard deviation of  $\widehat{\beta}_{j,\xi} - \beta_{j,\xi}$  is smaller than  $\varepsilon 2^{j(d-1)/2}$ ,

$$P(|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| > \kappa 2^{j\nu} \varepsilon \sqrt{\log 1/\varepsilon}) \leq \varepsilon^{\kappa^2/2}.$$

So

$$\begin{aligned} & J^{p-1} \sum_{j \leq J} \mathbb{E} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\widehat{\beta}_{j,\xi} - \beta_{j,\xi}| > \kappa 2^{jv} \varepsilon \sqrt{\log 1/\varepsilon}\}} \right) \\ & \leq C \|f\|_{B_{\pi,r}^s}^p J^{p-1} \varepsilon^{\kappa^2/2} \leq C (\varepsilon \sqrt{\log 1/\varepsilon})^p \end{aligned}$$

if  $\kappa^2$  is large enough (where we have used that  $B_{\pi,r}^s \subset B_{p,r}^{s'} \subset \mathbb{L}_p$  with  $s' = s - (d+1)(1/\pi - 1/p)$ ). Hence this term also is of the right order.

Let us turn now to the last one, using (24)

$$\sum_{\xi \in \chi_j} |\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\beta_{j,\xi}| < 2\kappa 2^{jv} \varepsilon \sqrt{\log 1/\varepsilon}\}} \leq (2\kappa \varepsilon \sqrt{\log 1/\varepsilon})^{p-q}.$$

Hence,

$$\begin{aligned} & J^{p-1} \sum_{j \leq J} \sup_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_p^p \mathbf{1}_{\{|\beta_{j,\xi}| < 2\kappa 2^{jv} \varepsilon \sqrt{\log 1/\varepsilon}\}}) \\ & \leq C \|f\|_{B_{\pi,r}^s} J^p (\varepsilon \sqrt{\log 1/\varepsilon})^{p-q} \leq C \|f\|_{B_{\pi,r}^s} \sqrt{\log 1/\varepsilon}^p (\varepsilon \sqrt{\log 1/\varepsilon})^{p-q}. \end{aligned}$$

This proves that all the terms have the proper rate. It remains now to state and prove the following lemma.

**Lemma 2.** *Let  $A = \{(s, \pi), s > (d+1)(\frac{1}{\pi} - \frac{1}{p}) \cap (s > 0)\}$ , and  $f \in B_{\pi,r}^s, 1 \leq \pi \leq \infty, 1 \leq p < \infty$ . If  $\sum_{\xi \in \chi_j} (|\beta_{j,\xi}| \|\psi_{j,\xi}\|_\pi)^\pi = \rho_j^\pi 2^{-js\pi}$  with  $\rho \in l_r(\mathbb{N})$ , then, with  $v = \frac{d-1}{2}$ ,*

$$\sum_{\xi \in \chi_j} \left( \frac{|\beta_{j,\xi}|}{2^{jv}} \right)^q (2^{jv} \|\psi_{j,\xi}\|_p)^p \leq C \rho_j^q,$$

where  $q < p$  is as follows:

(1)  $p - q = \frac{sp}{s+d-1/2} (q = \frac{(d-1/2)p}{s+d-1/2})$  in the following domain  $I$ :

$$\{(s, \pi), (s(1/p - 1/4) \geq (d - 1/2)(1/\pi - 1/p)) \cap A\}.$$

Moreover, we have the following slight modification at the frontier: the domain becomes

$$\{(s, \pi), (s(1/p - 1/4) = (d - 1/2)(1/\pi - 1/p)) \cap A\}$$

and the inequality

$$\sum_{\xi \in \chi_j} \left( \frac{|\beta_{j,\xi}|}{2^{jv}} \right)^q (2^{jv} \|\psi_{j,\xi}\|_p)^p \leq C \rho_j^q j^{1-q/\pi}.$$

(2)  $p - q = \frac{(s-2(1/\pi-1/p))p}{s+d-2/\pi}$  ( $q = \frac{dp+2}{s+d-2/\pi}$ ) in the following domain II:

$$\{(s, \pi)(s > dp(1/\pi - 1/p)) \cap (s(1/p - 1/4) < (d - 1/2)(1/\pi - 1/p)) \cap A\}.$$

(3)  $p - q = \frac{(s-(d+1)(1/\pi-1/p))p}{s+d-(d+1)/\pi}$  ( $q = \frac{dp-(d+1)}{s+d-(d+1)/\pi}$ ) in the following domain III:

$$\left\{ (s, \pi), \left( dp \left( \frac{1}{\pi} - \frac{1}{p} \right) \geq s \right) \cap A, \text{ for } \frac{1}{p} < \frac{d}{d+1} \right\}.$$

This lemma is to be used essentially through the following corollary.

**Corollary 1.** *Respectively, in the domains I, II, III, we have, for  $q$  described in the lemma and  $f \in B_{\pi,r}^s$ ,*

$$\sum_{\xi \in \chi_j} 1_{\{|\beta_{j,\xi}|/2^{jv} \geq \lambda\}} (2^{jv} \|\psi_{j,\xi}\|_p)^p \leq C \rho_j^q \lambda^{-q}, \tag{23}$$

$$\sum_{\xi \in \chi_j} 1_{\{|\beta_{j,\xi}|/2^{jv} \leq 2^{jv}\lambda\}} |\beta_{j,\xi}|^p \|\psi_{j,\xi}\|_p^p \leq C \rho_j^q \lambda^{p-q} \tag{24}$$

with an obvious modification for

$$\{(s, \pi), (s(1/p - 1/4) = (d - 1/2)(1/\pi - 1/p)) \cap A\}.$$

**Proof.** Let us recall that on a measure space  $(X, \mu)$  we have, if  $h \in \mathbb{L}_q(\mu)$  then  $\mu(|h| \geq \lambda) \leq \frac{\|h\|_q^q}{\lambda^q}$  and, as  $q < p$ ,

$$\begin{aligned} \int_{|h| \leq \lambda} |h|^p \, d\mu &\leq \int (|h| \wedge \lambda)^p \, d\mu = \int_0^\lambda p x^{p-1} \mu(|h| \geq x) \, dx \\ &\leq \int_0^\lambda p x^{p-1} \frac{\|h\|_q^q}{x^q} \, dx = \frac{p \|h\|_q^q}{p - q} \lambda^{p-q}. \end{aligned}$$

For the corollary, we take  $X = \chi_j$ ,  $\mu(\xi) = (2^{jv} \|\psi_{j,\xi}\|_p)^p$  and  $h(\xi) = \frac{|\beta_{j,\xi}|}{2^{jv}}$ . □

**Proof of Lemma 2.** Let us fix  $q$  (chosen later) and investigate separately the two cases  $q \geq \pi$  and  $q < \pi$ .

For  $q \geq \pi$ , we have, using (19),

$$\begin{aligned} I_j(f, q, p) &= \sum_{\xi \in \chi_j} \left| \frac{\beta_{j,\xi}}{2^{jv}} \right|^q \|2^{jv} \psi_{j,\xi}\|_p^p \sim 2^{jv(p-q)} \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^q \left( \frac{2^{jd}}{W_j(\xi)} \right)^{p/2-1} \\ &\leq 2^{jv(p-q)} \left( \sum_{\xi \in \chi_j} \left( |\beta_{j,\xi}|^q \left( \frac{2^{jd}}{W_j(\xi)} \right)^{p/2-1} \right)^{\pi/q} \right)^{q/\pi} \end{aligned}$$

$$\begin{aligned}
 &= 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \left( \frac{2^{jd}}{W_j(\xi)} \right)^{(p/2-1)\pi/q} \right)^{q/\pi} \\
 &= 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \left( \frac{2^{jd}}{W_j(\xi)} \right)^{\pi/2-1} \left( \frac{2^{jd}}{W_j(\xi)} \right)^{(\pi/q)(p/2-1)-(\pi/2-1)} \right)^{q/\pi} \\
 &\leq 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \left( \frac{2^{jd}}{W_j(\xi)} \right)^{\pi/2-1} \right)^{q/\pi} 2^{j(d+1)((p-q)/2+q(1/\pi-1/q))}
 \end{aligned}$$

if we choose  $q$  such that  $(sq + d(p - q) + (d + 1)q(\frac{1}{\pi} - \frac{1}{q})) = 0$  gives  $q = \frac{pd-(d+1)}{s+d-(d+1)\pi}$ . Hence,  $p - q = \frac{s-(d+1)(1/\pi-1/p)}{s+d-(d+1)\pi} p$ ;  $q - \pi = -\pi \frac{s-pd(1/\pi-1/p)}{s+d-(d+1)\pi}$ . Thus, in domain III,

$$\left\{ \left( \frac{1}{p} < \frac{d}{d+1} \right) \cap (s - (d + 1)(1/\pi - 1/p) > 0) \cap (s - pd(1/\pi - 1/p) \leq 0) \right\},$$

we have  $0 < q < p$ ,  $\pi \leq q$ ,  $\sum_{\xi \in \chi_j} |\frac{\beta_{j,\xi}}{2^{j\nu}}|^q \|2^{j\nu} \psi_{j,\xi}\|_p^p \leq \rho_j^q$ .

For  $q < \pi$ , we have, using (19),

$$\begin{aligned}
 I_j(f, q, p) &= \sum_{\xi \in \chi_j} \left| \frac{\beta_{j,\xi}}{2^{j\nu}} \right|^q \|2^{j\nu} \psi_{j,\xi}\|_p^p \sim 2^{j\nu(p-q)} \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^q \left( \frac{2^{jd}}{W_j(\xi)} \right)^{p/2-1}, \\
 &2^{j\nu(p-q)} \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^q \left( \frac{2^{jd}}{W_j(\xi)} \right)^{(\pi/2-1)q/\pi} \left( \frac{2^{jd}}{W_j(\xi)} \right)^{(p/2-1)-(\pi/2-1)q/\pi} \\
 &\leq 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \left( \frac{2^{jd}}{W_j(\xi)} \right)^{\pi/2-1} \right)^{q/\pi} \\
 &\quad \times \left( \sum_{\xi \in \chi_j} \left( \frac{2^{jd}}{W_j(\xi)} \right)^{\pi/(\pi-q)((p/2-1)-(\pi/2-1)q/\pi)} \right)^{1-q/\pi} \\
 &\sim 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \|\psi_{j,\xi}\|_\pi^\pi \right)^{q/\pi} \left( \sum_{\xi \in \chi_j} \left( \frac{2^{jd}}{W_j(\xi)} \right)^{(\pi(p-q)/(2(\pi-q)))-1} \right)^{1-q/\pi} \\
 &\sim 2^{j\nu(p-q)} \left( \sum_{\xi \in \chi_j} |\beta_{j,\xi}|^\pi \|\psi_{j,\xi}\|_\pi^\pi \right)^{q/\pi} \left( \sum_{\xi \in \chi_j} \|\psi_{j,\xi}\|_\pi^{\pi(p-q)/(\pi-q)} \right)^{1-q/\pi}.
 \end{aligned}$$

Now let us investigate separately the cases  $\frac{\pi(p-q)}{(\pi-q)}$  smaller, greater or equal to 4.

Case  $\frac{\pi(p-q)}{(\pi-q)} < 4$ . Using (19)–(21), we have

$$I_j(f, q, p) \leq C2^{j\nu(p-q)} \rho_j^q 2^{-jsq} 2^{jd(p-q)/2} \leq C\rho_j^q.$$

If we define  $q$  such that  $-sq + (p - q)(d - 1/2) = 0$ , that is,  $q = \frac{p(d-1/2)}{s+d-1/2}$ , then  $p - q = \frac{sp}{s+d-1/2} > 0$ . So  $\pi - q = \pi \frac{s-(d-1/2)p(1/\pi-1/p)}{s+d-1/2} > 0 \Leftrightarrow \frac{s}{p} > (d-1/2)(1/\pi - 1/p)$  and  $\frac{\pi(p-q)}{(\pi-q)} = \frac{sp}{s-(d-1/2)p(1/\pi-1/p)} < 4 \Leftrightarrow s(1/p - 1/4) > (d-1/2)(1/\pi - 1/p)$ . Hence we only need to impose  $s(1/p - 1/4) > (d-1/2)(1/\pi - 1/p)$  and domain I is given by

$$\{(s - (d + 1)(1/\pi - 1/p) > 0) \cap (s > 0)\} \cap \{s(1/p - 1/4) > (d - 1/2)(1/\pi - 1/p)\}$$

on which  $I_j(f, q, p) \leq C\rho_j^{p(d-1/2)/(s+d-1/2)}$ .

Case  $\frac{\pi(p-q)}{(\pi-q)} > 4$ . Using (19)–(21), we have

$$I_j(f, q, p) \leq C2^{j\nu(p-q)} \rho_j^q 2^{-jsq} 2^{jd(p-q)/2} 2^{j((p-q)/2-2(\pi-q)/\pi)}.$$

If we put  $(p - q)d - sq - 2\frac{\pi-q}{\pi} = 0 \Leftrightarrow q = \frac{pd-2}{s+d-2/\pi}$ , we have  $p - q = \frac{s-2(1/\pi-1/p)}{s+d-2/\pi} p > 0 \Leftrightarrow s - 2(1/\pi - 1/p) > 0$  and  $\pi - q = \frac{s-dp(1/\pi-1/p)}{s+d-2/\pi} \pi > 0 \Leftrightarrow s - dp(1/\pi - 1/p) > 0$ .

Moreover,  $\frac{\pi(p-q)}{(\pi-q)} = \frac{s-2(1/\pi-1/p)}{s-dp(1/\pi-1/p)} p > 4 \Leftrightarrow s(1/p - 1/4) < (d - 1/2)(1/\pi - 1/p)$ . Hence, on the domain

$$\{(0 < s) \cap (s > (d + 1)(1/\pi - 1/p)) \cap (s > dp(1/\pi - 1/p)) \cap (s(1/p - 1/4) < (d - 1/2)(1/\pi - 1/p))\},$$

we have  $I_j(f, q, p) \leq C\rho_j^{(pd-2)/(s+d-2/\pi)}$ .

Case  $\frac{\pi(p-q)}{(\pi-q)} = 4$ . Using (19)–(21), we have

$$I_j(f, q, p) \leq C2^{j\nu(p-q)} \rho_j^q 2^{-jsq} j^{1-q/\pi} 2^{jd(p-q)/2} \leq C\rho_j^q j^{1-q/\pi}$$

if  $(p - q)(d - 1/2) - sq = 0 \Leftrightarrow q = p \frac{d-1/2}{s+d-1/2}$ . This is realized either if  $p = 4 = \pi$  and for  $s > 0$  or if  $p \neq 4, \pi \neq 4$  and  $0 < q = \pi \frac{p-4}{\pi-4} = p \frac{d-1/2}{s+d-1/2}$ . Moreover,  $q < \pi$  and  $q < p \Leftrightarrow 4 < p < \pi$ ; or  $4 < \pi < p$  and  $\frac{\pi(p-q)}{(\pi-q)} = 4 \Leftrightarrow s(1/p - 1/4) = (d - 1/2)(1/\pi - 1/p)$ . Hence, on the domain

$$\{(s, \pi), (s(1/p - 1/4) = (d - 1/2)(1/\pi - 1/p)) \cap (s > 0) \cap (s > (d + 1)(1/\pi - 1/p))\},$$

we have  $I_j(f, q, p) \leq C\rho_j^{(p(d-1/2))/(s+d-1/2)} j^{s/(s+d-1/2)}$ . □

### 8. Proof of the lower bounds

In this section, we prove the lower bounds: for  $0 < s < \infty, 1 \leq \pi \leq \infty, 0 < r \leq \infty, 0 < M < \infty$ , denoting by  $B_{\pi,r}^s(M)$  the ball of radius  $M$  of the space  $B_{\pi,r}^s$  and, by  $\mathcal{E}$ , the set of all estimators, we consider

$$\begin{aligned} \omega_p(s, \pi, r, M, \varepsilon) &= \inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|f^* - f\|_p^p, \\ \omega_\infty(s, \pi, r, M, \varepsilon) &= \inf_{f^* \in \mathcal{E}} \sup_{f \in B_{\pi,r}^s(M)} \mathbb{E} \|f^* - f\|_\infty. \end{aligned}$$

The main tool will be the classical lemma introduced by Fano in 1952 [8]. We will use the version of Fano’s lemma introduced in [4]. For details on general lower bound results, see also [26]. Let us recall that  $K(P, Q)$  denotes the Kullback information “distance” between  $P$  and  $Q$ .

**Lemma 3.** *Let  $\mathcal{A}$  be a sigma algebra on the space  $\Omega$ , and  $A_i \in \mathcal{A}, i \in \{0, 1, \dots, m\}$ , such that  $\forall i \neq j, A_i \cap A_j = \emptyset, P_i, i \in \{0, 1, \dots, m\}$  are  $m + 1$  probability measures on  $(\Omega, \mathcal{A})$ . Define*

$$p := \sup_{i=0}^m P_i(A_i^c) \quad \text{and} \quad \kappa := \inf_{j \in \{0,1,\dots,m\}} \frac{1}{m} \sum_{i \neq j} K(P_i, P_j),$$

then

$$p \geq \frac{1}{2} \wedge (C \sqrt{m} \exp(-\kappa)) \quad \text{with } C = \exp\left(-\frac{3}{e}\right). \tag{25}$$

This inequality will be used in the following way: Let  $H_\varepsilon$  be the Hilbert space of measurable functions on  $Z = \mathbb{S}^{d-1} \times [-1, 1]$  with the scalar product

$$\langle \varphi, \psi \rangle_\varepsilon = \varepsilon^2 \int_{\mathbb{S}^{d-1}} \int_{-1}^1 \varphi(\theta, s) \psi(\theta, s) \, d\sigma(\theta) \frac{ds}{(1 - s^2)^{(d-1)/2}}.$$

It is well known that there exists a (unique) probability measure on  $(\Omega, \mathcal{A})$ :  $Q_f$  the density of which, with respect to  $P$ , is

$$\frac{dQ_f}{dP} = \exp\left(W^\varepsilon(f) - \frac{1}{2} \|f\|_{H_\varepsilon}^2\right).$$

Let us now choose  $f_0, f_1, \dots, f_m$  in  $B_{\pi,r}^s(M)$  such that  $i \neq j \implies \|f_i - f_j\|_p \geq \delta$  and denote  $P_i = Q_{R(f_i/\varepsilon^2)}$ . Let  $f^*$  be an arbitrary estimator of  $f$ . Obviously, the sets  $A_i = (\|f^* - f\|_p < \frac{\delta}{2})$  are disjoint sets and we have, for  $i \neq j$ ,

$$K(P_i, P_j) = \frac{1}{2\varepsilon^2} \int_Z |R(f_i - f_j)|^2 \, d\mu.$$

Now

$$\begin{aligned}\omega_p(s, \pi, q, M, \varepsilon) &\geq \inf_{f^* \in \mathcal{E}} \sup_{f_i, i=0,1,\dots,m} \mathbb{E} \|f^* - f_i\|_p^p \\ &\geq \left(\frac{\delta}{2}\right)^p \inf_{f^* \in \mathcal{E}} \sup_{f_i, i=0,1,\dots,m} P\left(\|f^* - f_i\|_p \geq \frac{\delta}{2}\right).\end{aligned}$$

Likewise,

$$\omega_\infty(s, \pi, q, M, \varepsilon) \geq \left(\frac{\delta}{2}\right) \inf_{f^* \in \mathcal{E}} \sup_{f_i, i=0,1,\dots,m} P\left(\|f^* - f_i\|_\infty \geq \frac{\delta}{2}\right).$$

Using Fano's lemma,

$$\sup_{f_i, i=0,1,\dots,m} P\left(\|f^* - f_i\|_p \geq \frac{\delta}{2}\right) \geq \frac{1}{2} \wedge (C\sqrt{m} \exp(-\kappa))$$

with

$$\kappa = \inf_{j=0,\dots,M} \frac{1}{m} \sum_{i \neq j} \frac{1}{2\varepsilon^2} \int_Z |R(f_i - f_j)|^2 d\mu.$$

So if, for a given  $\varepsilon$ , we can find  $f_0, f_1, \dots, f_m$  in  $B_{\pi,r}^{s,0}(M)$  such that  $i \neq j \implies \|f_i - f_j\|_p \geq \delta(\varepsilon)$  and  $C\sqrt{m} \exp(-\kappa) \geq 1/2$ , then we have

$$\text{for } p < \infty, \quad \omega_p(s, \pi, q, M, \varepsilon) \geq \frac{1}{2} \delta(\varepsilon)^p \quad \text{and} \quad \omega_\infty(s, \pi, q, M, \varepsilon) \geq \frac{1}{2} \delta(\varepsilon).$$

In the sequel, we will choose, as usual, sets of functions containing either two items (sparse case) or a number of order  $2^{jd}$  or  $2^{j(d-1)}$  (dense cases). We will consider sets of functions that are basically linear combinations of needlets at a fixed level  $f = \sum_{\xi \in \chi_j} \beta_{j,\xi} \psi_{j,\xi}$ . Because the needlets have different orders of norms, depending on whether they are around the north pole or closer to the equator, we will have to investigate different cases. These differences will precisely yield the different minimax rates.

## 8.1. Reverse inequality

Because the needlets are not forming an orthonormal system, we cannot pretend that inequality (18) is an equivalence. Since, precisely in the lower-bound evaluations, we need to bound both sides of the  $\mathbb{L}_p$  norm for terms of the form  $\sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi}$  with  $A_j \subset \chi_j$ . The following section is devoted to this problem.

**Proposition 2.** For  $A_j \subset \chi_j$ ,

$$\frac{1}{C} \left( \sum_{\xi' \in A_j} \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi}, \psi_{j,\xi'} \right\|_p^p \right)^{1/p} \leq \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p \leq C \left( \sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}.$$

**Proof.** Let  $f = \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi}$ . Clearly, by (18),

$$\left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p \leq C \left( \sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p},$$

and by (17),

$$\left( \sum_{\xi' \in \chi_j} \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi}, \psi_{j,\xi'} \right\|_p^p \right)^{1/p} \leq C \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p,$$

so obviously,

$$\begin{aligned} & \frac{1}{C} \left( \sum_{\xi' \in A_j} \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi}, \psi_{j,\xi'} \right\|_p^p \right)^{1/p} \\ & \leq \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p \leq C \left( \sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}. \end{aligned} \quad \square$$

In the sequel, we will look for subset  $A_j$  with equilibrated  $\mathbb{L}_p$  norms, that is, such that there exists

$$0 < D_j, \quad \text{such that } \forall \xi \in A_j, \|\psi_{j,\xi}\|_p \sim D_j.$$

(Here and in the rest of this section,  $a_{j,\xi} \sim b_j$  will mean that there exist two absolute constants  $c_1$  and  $c_2$  – which will not be precised for the sake of simplicity – such that  $c_1 b_j \leq a_{j,\xi} \leq c_2 b_j$ , for all considered  $\xi$ .) As specified above,  $D_j$  may have different forms depending on the regions. Using (19), we have

$$\|\psi_{j,\xi}\|_p \sim \left( \frac{2^{jd}}{2^{-j} + \sqrt{1 - |\xi|^2}} \right)^{1/2-1/p}.$$

For our purpose, let us precise Proposition 1 by choosing the cubature points in the following way: We choose in the hemisphere  $\mathbb{S}_+^d$  strips  $S_k = B(A, (2k+1)\eta) \setminus B(A, 2k\eta)$  with  $\eta \sim \frac{\pi}{22^{j+1}}$ ,  $k \in \{0, \dots, 2^j - 1\}$  ( $A$  is the north pole). In each of these strips, we choose a maximal  $\eta$ -net of points  $\tilde{\xi}$ , whose cardinality is of order  $k^{d-1}$ . Projecting these points on the ball, we obtain cubature points  $\xi$  on the ball with coefficients  $\omega_{j,\xi} \sim 2^{-jd} W_j(\xi)$ . As a consequence, we have in the set  $\{x \in \mathbb{R}^d, |x| \leq \frac{1}{\sqrt{2}}\}$ , about  $2^{jd}$  points of cubature for which

$$D_j \sim \|\psi_{j,\xi}\|_p \sim 2^{jd(1/2-1/p)}.$$

And in the corona  $\{(1 - 2^{-2j} \leq |x| \leq 1)\}$ , we have about  $2^{j(d-1)}$  points of cubature for which

$$D_j \sim \|\psi_{j,\xi}\|_p \sim 2^{j(d+1)(1/2-1/p)}.$$

Now let us consider a set  $A_j$  of cubature points included in one of the two sets considered just above (either  $\{x \in \mathbb{R}^d, |x| \leq \frac{1}{\sqrt{2}}\}$  or  $\{(1 - 2^{-2j} \leq |x| \leq 1)\}$ ). Consider also the matrix (parametrized by  $A_j$ )

$$\mathbb{M}(A_j) = (\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle)_{\xi, \xi' \in A_j \times A_j}.$$

We have, for any  $\lambda \in l_p(A_j)$ , using Proposition 2,

$$\|\mathbb{M}(A_j)(\lambda)\|_{l_p(A_j)} \leq C' \|\lambda\|_{l_p(A_j)}.$$

On the other hand, let us observe that, using (14),

$$0 < c \leq \|\psi_{j,\xi}\|_2^2 = \langle \psi_{j,\xi}, \psi_{j,\xi} \rangle \leq 1.$$

Thus,

$$\mathbb{M}(A_j) = \text{Diag}(\mathbb{M}(A_j)) + \mathbb{M}'(A_j) = \text{Diag}(\mathbb{M}(A_j))(\text{Id} + [\text{Diag}(\mathbb{M}(A_j))]^{-1}\mathbb{M}'(A_j)),$$

where  $\text{Diag}(\mathbb{M}(A_j))$  is the diagonal matrix parametrized by  $A_j$  extracted from  $\mathbb{M}(A_j)$ . Clearly, each of the terms of  $[\text{Diag}(\mathbb{M}(A_j))]^{-1}$  is bounded by  $c^{-1}$ .

So if  $\|[\text{Diag}(\mathbb{M}(A_j))]^{-1}\mathbb{M}'(A_j)\|_{\mathcal{L}(l_p(A_j))} \leq \alpha < 1$ , we have

$$\|\mathbb{M}(A_j)^{-1}\|_{\mathcal{L}(l_p(A_j))} \leq c^{-1} \frac{1}{1 - \alpha}.$$

Let us prove that we can choose  $A_j$  large enough and such that such an  $\alpha$  exists. Using the Schur lemma (see [9], Appendix 29),

$$\|[\text{Diag}(\mathbb{M}(A_j))]^{-1}\mathbb{M}'(A_j)\|_{\mathcal{L}(l_p(A_j))} \leq c^{-1} \sup_{\xi \in A_j} \sum_{\xi' \neq \xi, \xi' \in A_j} |\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle|.$$

Now, using (16),

$$|\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle| \leq C_M^2 \int_{B^d} \frac{1}{\sqrt{W_j(x)}} \frac{1}{(1 + 2^j d(x, \xi))^M} \frac{1}{\sqrt{W_j(x)}} \frac{1}{(1 + 2^j d(x, \xi'))^M} dx$$

and thus, by triangular inequality,

$$\begin{aligned} |\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle| &\leq \frac{C_M^2}{(1 + 2^j d(\xi, \xi'))^M} \int_{B^d} \frac{1}{2^{-j} + \sqrt{1 - |x|^2}} dx \\ &\leq C_M^2 \frac{1}{(1 + 2^j d(\xi, \xi'))^M} |\mathbb{S}^{d-1}| \int_0^1 r^{d-2} dr. \end{aligned}$$

So

$$\forall M \quad |\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle| \leq C'_M \frac{1}{(1 + 2^j d(\xi, \xi'))^M}. \tag{26}$$

Now, let us choose  $A_j$  as a maximal  $K\eta$  net in the set  $\chi_j \cap \{x \in \mathbb{R}^d, |x| \leq \frac{1}{\sqrt{2}}\}$  (case 1) or as a maximal  $K\eta$  net in the set  $\chi_j \cap \{(1 - 2^{-2j} \leq |x| \leq 1)\}$  (case 2). Recall that  $\eta \sim \frac{\pi}{22^{j+1}}$  and  $K$  will be chosen later.

As, in case 1,

$$\text{Card}\{\xi', d(\xi', \xi) \sim Kl2^{-j}\} \lesssim (Kl)^d,$$

$$\sum_{\xi' \neq \xi, \xi' \in A_j} |\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle| \leq \sum_{l=1}^{2^{j/K}} (Kl)^d C_M \frac{1}{(1+Kl)^M} \leq C_M \sum_{l=1}^{2^{j/K}} (Kl)^d \frac{1}{(Kl)^M} \leq \frac{2C_M}{K^{M-d}} \leq \alpha$$

if  $M - d \geq 2$  and  $K$  is large enough. In case 2, again

$$\text{Card}\{\xi', d(\xi', \xi) \sim Kl2^{-j}\} \lesssim (Kl)^{d-1}$$

so

$$\begin{aligned} \sum_{\xi' \neq \xi, \xi' \in A_j} |\langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle| &\leq \sum_{l=1}^{2^{j/K}} (Kl)^{d-1} C_M \frac{1}{(1+Kl)^M} \\ &\leq C_M \sum_{l=1}^{2^{j/K}} (Kl)^{d-1} \frac{1}{(Kl)^M} \leq \frac{2C_M}{K^{M-d+1}} \leq \alpha \end{aligned}$$

if  $M - d \geq 1$  and  $K$  is large enough.

Hence,  $\mathbb{M}(A_j)$  is invertible in both cases and we have

$$c^{-1} \frac{1}{1-\alpha} \left( \sum_{\xi \in A_j} |\lambda_\xi|^p \right)^{1/p} \leq \left( \sum_{\xi' \in A_j} \left| \left\langle \sum_{\xi \in A_j} \lambda_\xi \langle \psi_{j,\xi}, \psi_{j,\xi'} \rangle \right\rangle^p \right)^{1/p}$$

and

$$\left( \sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p} \lesssim \left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p \lesssim \left( \sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}.$$

## 8.2. Lower bounds associated sparse/dense cases and different choices of $A_j$

Let  $j$  be fixed and choose

$$f = \sum_{\xi \in A_j} \beta_{j,\xi} \psi_{j,\xi}.$$

We have

$$f = \sum_{2^{j-1} < k < 2^{j+1}} P_k(f),$$

where  $P_k$  is the orthogonal projector on  $\mathcal{V}_k(B^d)$ . So

$$\begin{aligned} \|R(f)\|^2 &= \langle R^*R(f), f \rangle = \sum_{2^{j-1} < k < 2^{j+1}} \langle \lambda_k^2 P_k(f), f \rangle \\ &\leq \left( \sup_{2^{j-1} < k < 2^{j+1}} \lambda_k^2 \right) \sum_k \|P_k(f)\|^2 \leq C 2^{-j(d-1)} \sum_{\xi \in A_j} |\beta_{j,\xi}|^2. \end{aligned}$$

8.2.1. Sparse choice, case 1

Let  $f_i = \gamma \varepsilon_i \psi_{j,\xi_i}$ ,  $i \in \{1, 2\}$ ,  $\varepsilon_i$  is +1 or -1, in such a way that

$$\|f_1 - f_2\|_p = \|\gamma \psi_{j,\xi_1} - \gamma \psi_{j,\xi_2}\|_p = \gamma \|\psi_{j,\xi_1} - \psi_{j,\xi_2}\|_p \sim \gamma (\|\psi_{j,\xi_1}\| + \|\psi_{j,\xi_2}\|_p).$$

In case 1,  $\|\psi_{j,\xi}\|_r \sim 2^{jd(1/2-1/r)}$ . So

$$\begin{aligned} f_i \in B_{\pi,r}^s(1) &\iff \gamma 2^{jd(1/2-1/\pi)} \sim 2^{-js} \iff \gamma \sim 2^{-j(s+d(1/2-1/\pi))}, \\ \delta = \|f_1 - f_2\|_p &\sim \gamma 2^{jd(1/2-1/p)} \sim 2^{-j(s+d(1/2-1/\pi)-d(1/2-1/p))} = 2^{-j(s-d(1/\pi-1/p))}. \end{aligned}$$

On the other hand,

$$K(P_1, P_2) = \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-j(d-1)} \gamma^2 \sim \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-j(d-1)} 2^{-2j(s+d(1/2-1/\pi))} = \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-2j(s+d-1/2-d/\pi)}.$$

Now, by the Fano inequality, if  $j$  is chosen so that  $\varepsilon \sim 2^{-j(s+d-1/2-d/\pi)}$  (under the constraint  $s > d(1/\pi - (1 - 1/2d))$ ),

$$\left(\frac{2}{\delta}\right)^p \mathbb{E} \|f^* - f_i\|_p^p \geq P(\|f^* - f_i\|_p > \delta/2) \geq c.$$

So, necessarily,

$$\mathbb{E} \|f^* - f_i\|_p^p \geq c \delta^p \sim \varepsilon^{(s-d(1/\pi-1/p))/(s+d-1/2-d/\pi)}.$$

**Remark 3.** If

$$d \left( 1/\pi - \left( 1 - \frac{1}{2d} \right) \right) < s \leq d(1/\pi - 1/p)$$

(so, necessarily,  $\frac{1}{p} \leq 1 - \frac{1}{2d}$ ), then

$$\lim_{\varepsilon \rightarrow 0} \omega_p(s, \pi, q, M, \varepsilon) \geq C > 0.$$

8.2.2. Sparse choice, case 2

In case 2,  $\|\psi_{j,\xi}\|_r \sim 2^{j(d+1)(1/2-1/r)}$ , so

$$\begin{aligned} f_i \in B_{\pi,r}^s(1) &\iff \gamma 2^{j(d+1)(1/2-1/\pi)} \sim 2^{-js} \iff \gamma \sim 2^{-j(s+(d+1)(1/2-1/\pi))}, \\ \delta = \|f_1 - f_2\|_p &\sim \gamma 2^{j(d+1)(1/2-1/p)} \sim 2^{-j(s+(d+1)(1/2-1/\pi)-(d+1)(1/2-1/p)} \\ &= 2^{-j(s-(d+1)(1/\pi-1/p))}. \end{aligned}$$

On the other hand,

$$\begin{aligned} K(P_1, P_2) &= \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-j(d-1)} \gamma^2 = \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-j(d-1)} 2^{-2j(s+(d+1)(1/2-1/\pi))} \\ &\sim \frac{1}{2} \frac{1}{\varepsilon^2} 2^{-2j(s+d-(d+1)/\pi)}. \end{aligned}$$

Now, by the Fano inequality, if  $\varepsilon \sim 2^{-j(s+d-(d+1)/\pi)}$  (under the constraint  $s > (d+1)(\frac{1}{\pi} - \frac{d}{d+1})$ ),

$$\left(\frac{2}{\delta}\right)^p \mathbb{E} \|f^* - f_i\|_p^p \geq P(\|f^* - f_i\|_p > \delta/2) \geq c.$$

So, necessarily,

$$\mathbb{E} \|f^* - f_i\|_p^p \geq c\delta^p \sim \varepsilon^{(s-(d+1)(1/\pi-1/p))p/(s+d-(d+1)/\pi)}.$$

**Remark 4.** If

$$(d+1)\left(1/\pi - \frac{d}{d+1}\right) < s \leq (d+1)(1/\pi - 1/p)$$

(so, necessarily,  $\frac{1}{p} < \frac{d}{d+1}$ ),

$$\lim_{\varepsilon \rightarrow 0} \omega_p(s, \pi, q, M, \varepsilon) \geq C > 0.$$

8.2.3. Dense choice, case 1

In this case, we take

$$f_\rho = \gamma \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi}, \quad \varepsilon_\xi = \pm 1, \rho = (\varepsilon_\xi)_{\xi \in A_j}.$$

As we are in case 1, we have

$$\gamma^r \left\| \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi} \right\|_r^r \sim \gamma^r 2^{jd(r/2-1)} \sum_{\xi \in A_j} |\varepsilon_\xi|^r \sim \gamma^r 2^{jd(r/2-1)} \text{Card}(A_j) \sim \gamma^r 2^{jdr/2}.$$

Using the Varshamov–Gilbert theorem (see [26], Chapter 2), we consider a subset  $\mathcal{A}$  of  $\{-1, +1\}^{A_j}$  such that  $\text{Card}(\mathcal{A}) \sim 2^{(1/8)\text{Card}(A_j)}$  and for  $\rho \neq \rho'$ ,  $\rho, \rho' \in \mathcal{A}$ ,  $\|\rho - \rho'\|_1 \geq \frac{1}{2} \text{Card}(A_j)$ . Let us now restrict our set to

$$f_\rho = \gamma \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi}, \quad \varepsilon_\xi = \pm 1, \rho = (\varepsilon_\xi)_{\xi \in A_j}, \rho \in \mathcal{A},$$

$$f_\rho \in B_{\pi,r}^s(1) \iff \gamma \left( \sum_{\xi \in A_j} \|\psi_{j,\xi}\|_\pi^\pi \right)^{1/\pi} \sim \gamma 2^{jd/2} \sim 2^{-js}.$$

So we choose

$$\gamma \sim 2^{-j(s+d/2)}.$$

Moreover,

$$\delta = \|f_\rho - f_{\rho'}\|_p = \gamma \left\| \sum_{\xi \in A_j} (\varepsilon_\xi - \varepsilon'_\xi) \psi_{j,\xi} \right\|_p \sim \gamma \left( \sum_{\xi \in A_j} |\varepsilon_\xi - \varepsilon'_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}$$

$$\sim \gamma 2^{jd(1/2-1/p)} \|\rho - \rho'\|_1^{1/p} \sim 2^{-j(s+d/2)} 2^{jd/2} = 2^{-js}.$$

Let us compute the Kullback distance,

$$K(P_\rho, P_{\rho'}) = \frac{1}{2\varepsilon^2} 2^{-j(d-1)} \|f_\rho - f_{\rho'}\|_2^2 \sim \frac{1}{2\varepsilon^2} 2^{-j(d-1)} 2^{-2j(s+d/2)} 2^{jd} = \frac{1}{2\varepsilon^2} 2^{-2j(s+d/2-1/2)},$$

so, by the Fano inequality,

$$\frac{\mathbb{E} \|\hat{f} - f\|_p^p}{\delta^p} \geq 1/2 \wedge c 2^{(1/8)2^{jd}} e^{-(1/(2\varepsilon^2))2^{-2j(s+d/2-1/2)}} \geq 1/2$$

if

$$\varepsilon \sim 2^{-j(s+d-1/2)}.$$

This implies

$$\inf_{f \in B_{\pi,r}^s} \mathbb{E} \|\hat{f} - f\|_p^p \geq c \varepsilon^{sp/(s+d-1/2)}.$$

#### 8.2.4. Dense choice, case 2

Similar to the previous case, we take now (with a slight abuse of notation, since the subset  $A$  obtained using the Varshamov–Gilbert theorem is not the same  $A$ , as  $A_j$  has also changed)

$$f_\rho = \gamma \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi}, \quad \varepsilon_\xi = \pm 1, \quad \rho = (\varepsilon_\xi)_{\xi \in A_j}, \rho \in \mathcal{A}.$$

As we are in case 2, we have

$$\begin{aligned} \gamma^r \left\| \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi} \right\|_r &\sim \gamma^r 2^{j(d+1)(r/2-1)} \sum_{\xi \in A_j} |\varepsilon_\xi|^r \\ &\sim \gamma^r 2^{j(d+1)(r/2-1)} \text{Card}(A_j) \sim \gamma^r 2^{j[(d+1)r/2-2]}, \\ f_\rho \in B_{\pi,r}^s(1) &\iff \gamma \left( \sum_{\xi \in A_j} \|\psi_{j,\xi}\|_\pi^\pi \right)^{1/\pi} \sim \gamma (2^{j(d-1)} 2^{j(d+1)(\pi/2-1)})^{1/\pi} \\ &\sim \gamma 2^{-j((d+1)/2-2/\pi)} \sim 2^{-js}. \end{aligned}$$

So we choose

$$\gamma \sim 2^{-j(s+(d+1)/2-2/\pi)}.$$

Moreover,

$$\begin{aligned} \delta = \|f_\rho - f_{\rho'}\|_p &= \gamma \left\| \sum_{\xi \in A_j} (\varepsilon_\xi - \varepsilon'_\xi) \psi_{j,\xi} \right\|_p \sim \gamma \left( \sum_{\xi \in A_j} |\varepsilon_\xi - \varepsilon'_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p} \\ &\sim \gamma 2^{j(d+1)(1/2-1/p)} \|\rho - \rho'\|_1^{1/p} \sim \gamma 2^{j(d+1)(1/2-1/p)} 2^{j(d-1)1/p} \\ &\sim 2^{-j(s+(d+1)/2-2/\pi)} 2^{j((d+1)/2-2/p)} = 2^{-j(s-2(1/\pi-1/p))}. \end{aligned}$$

Let us compute the Kullback distance:

$$\begin{aligned} K(P_\rho, P_{\rho'}) &= \frac{1}{2\varepsilon^2} 2^{-j(d-1)} \|f_\rho - f_{\rho'}\|_2^2 \\ &\sim \frac{1}{2\varepsilon^2} 2^{-j(d-1)} 2^{-2j(s+(d+1)/2-2/\pi)} 2^{j(d-1)} = \frac{1}{2\varepsilon^2} 2^{-2j(s+(d+1)/2-2/\pi)}, \end{aligned}$$

so, by the Fano inequality,

$$\frac{\mathbb{E}\|\hat{f} - f\|_p^p}{\delta^p} \geq 1/2 \wedge c 2^{(1/8)2^j(d-1)} e^{-1/(2\varepsilon^2)2^{-2j(s+(d+1)/2-2/\pi)}} \geq 1/2$$

if

$$\varepsilon \sim 2^{-j(s+d-2/\pi)}.$$

This implies

$$\inf_{f \in B_{\pi,r}^s(1)} \mathbb{E}\|\hat{f} - f\|_p^p \geq c\varepsilon^{p(s-2(1/\pi-1/p))/(s+d-2/\pi)}.$$

**Remark 5.** The case  $p = \infty$  can be handled using the same arguments without difficulties.

## References

- [1] Andrews, G.E., Askey, R. and Roy, R. (1999). *Special Functions. Encyclopedia of Mathematics and Its Applications* **71**. Cambridge: Cambridge Univ. Press. [MR1688958](#)
- [2] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395. [MR2549563](#)
- [3] Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Asymptotics for spherical needlets. *Ann. Statist.* **37** 1150–1171. [MR2509070](#)
- [4] Birge, L. (2001). A new look at an old result: Fano’s lemma. Prepublication 632, LPMA.
- [5] Davison, M.E. (1981). A singular value decomposition for the Radon transform in  $n$ -dimensional Euclidean space. *Numer. Funct. Anal. Optim.* **3** 321–340. [MR0629949](#)
- [6] Dunkl, C.F. and Xu, Y. (2001). *Orthogonal Polynomials of Several Variables. Encyclopedia of Mathematics and Its Applications* **81**. Cambridge: Cambridge Univ. Press. [MR1827871](#)
- [7] Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F.G. (1981). *Higher Transcendental Functions. Vol. II*. Melbourne, FL: Robert E. Krieger Publishing Co. Inc. [MR0698780](#)
- [8] Fano, R. (1952). Class notes for transmission of information, course. 6.574. MIT, Cambridge, MA.
- [9] Grafakos, L. (2004). *Classical and Modern Fourier Analysis*. Upper Saddle River, NJ: Pearson Education, Inc. [MR2449250](#)
- [10] Helgason, S. (1999). *The Radon Transform*, 2nd ed. *Progress in Mathematics* **5**. Boston, MA: Birkhäuser. [MR1723736](#)
- [11] Kerkyacharian, G., Kyriazis, G., Le Pennec, E., Petrushev, P. and Picard, D. (2010). Inversion of noisy Radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.* **28** 24–45. [MR2563258](#)
- [12] Kerkyacharian, G. and Picard, D. (2009). New generation wavelets associated with statistical problems. In *The 8th Workshop on Stochastic Numerics* 119–146. Kyoto Univ.: Research Institute for Mathematical Sciences.
- [13] Kyriazis, G., Petrushev, P. and Xu, Y. (2008). Decomposition of weighted Triebel–Lizorkin and Besov spaces on the ball. *Proc. Lond. Math. Soc.* (3) **97** 477–513. [MR2439670](#)
- [14] Logan, B.F. and Shepp, L.A. (1975). Optimal reconstruction of a function from its projections. *Duke Math. J.* **42** 645–659. [MR0397240](#)
- [15] Louis, A.K. (1984). Orthogonal function series expansions and the null space of the Radon transform. *SIAM J. Math. Anal.* **15** 621–633. [MR0740700](#)
- [16] Muciaccia, P.F., Natoli, P. and Vittorio, N. (1997). Fast spherical harmonic analysis: A quick algorithm for generating and/or inverting full-sky high-resolution cosmic microwave background anisotropy maps. *The Astrophysical J. Lett.* **488** 63–66.
- [17] Narcowich, F., Petrushev, P. and Ward, J. (2006). Decomposition of Besov and Triebel–Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564. [MR2253732](#)
- [18] Narcowich, F.J., Petrushev, P. and Ward, J.D. (2006). Localized tight frames on spheres. *SIAM J. Math. Anal.* **38** 574–594 (electronic). [MR2237162](#)
- [19] Natterer, F. (2001). *The Mathematics of Computerized Tomography. Classics in Applied Mathematics* **32**. Philadelphia, PA: SIAM. Reprint of the 1986 original. [MR1847845](#)
- [20] Petrushev, P. and Xu, Y. (2005). Localized polynomial frames on the interval with Jacobi weights. *J. Fourier Anal. Appl.* **11** 557–575. [MR2182635](#)
- [21] Petrushev, P. and Xu, Y. (2008). Localized polynomial frames on the ball. *Constr. Approx.* **27** 121–148. [MR2336420](#)
- [22] Petrushev, P.P. (1999). Approximation by ridge functions and neural networks. *SIAM J. Math. Anal.* **30** 155–189 (electronic). [MR1646689](#)
- [23] Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. In *Banach Spaces, Harmonic Analysis, and Probability Theory (Storrs, Conn., 1980/1981). Lecture Notes in Math.* **995** 123–154. Berlin: Springer. [MR0717231](#)

- [24] Stein, E.M. and Weiss, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton Mathematical Series **32**. Princeton, NJ: Princeton Univ. Press. [MR0304972](#)
- [25] Szegő, G. (1975). *Orthogonal Polynomials*, 4th ed. American Mathematical Society, Colloquium Publications **XXIII**. Providence, RI: Amer. Math. Soc. [MR0372517](#)
- [26] Tsybakov, A.B. (2008). *Introduction to Nonparametric Estimation*. Berlin: Springer.
- [27] Xu, Y. (1998). Orthogonal polynomials and cubature formulae on spheres and on balls. *SIAM J. Math. Anal.* **29** 779–793 (electronic). [MR1617720](#)
- [28] Xu, Y. (2007). Reconstruction from Radon projections and orthogonal expansion on a ball. *J. Phys. A* **40** 7239–7253. [MR2344454](#)

*Received August 2009 and revised July 2010*