

ADAPTIVE DENSITY ESTIMATION FOR DIRECTIONAL DATA USING NEEDLETS

BY P. BALDI,¹ G. KERKYACHARIAN, D. MARINUCCI AND D. PICARD

Università di Roma Tor Vergata, Laboratoire de Probabilités et Modèles Aléatoires, Università di Roma Tor Vergata and Université Paris Diderot & Laboratoire de Probabilités et Modèles Aléatoires

This paper is concerned with density estimation of directional data on the sphere. We introduce a procedure based on thresholding on a new type of spherical wavelets called *needlets*. We establish a minimax result and prove its optimality. We are motivated by astrophysical applications, in particular in connection with the analysis of ultra high-energy cosmic rays.

1. Introduction. We consider the problem of estimating the density f of an independent sample of points X_1, \dots, X_n observed on the d -dimensional sphere \mathbb{S}^d of \mathbb{R}^{d+1} . Obviously, the most immediate examples of applications appear in the case $d = 2$. However, no major differences arise from considering the general case.

There is an abundant literature about this type of problems. In particular, minimax L^2 results have been obtained (see [14, 15]). These procedures are generally obtained using either orthogonal series methods associated with spherical harmonics (providing an estimator having poor local performances, as the spherical harmonics are spread all over the sphere) or kernel methods (which do not take advantage of the Fourier structure of the space of square integrable functions on the manifold).

In our approach we focus on two important points. We aim at a procedure of estimation which is efficient from a L^2 point of view (as it is a tradition in statistics to evaluate the procedure with the mean square error). On the other hand, we would like it to perform satisfactorily also from a local point of view (in infinity norm, for instance). To have these two requirements together seems to us a warrant to have good results in practice. In effect, it is very difficult to produce a loss function which reflects at the same time the requirement of clearly seeing the bumps of the density, of being able to well estimate different level sets, of testing whether there is a difference between the northern and southern hemispheres and so on.

In addition, we require this procedure to be simple to implement, as well as adaptive to inhomogeneous smoothness. This type of requirements is generally

Received July 2008; revised December 2008.

¹Supported by PRIN project Probabilità e Finanza 2009.

AMS 2000 subject classifications. Primary 62G07; secondary 62G20, 65T60.

Key words and phrases. Density estimation, spherical and directional data, thresholding, needlets.

well handled using thresholding estimates associated to wavelets. The problem requires a special construction adapted to the sphere, since usual tensorized wavelets never reflect the manifold structure of the sphere and necessarily create unwanted artifacts. Recently in [18, 19] a tight frame (i.e., a redundant family) was produced which enjoys enough properties to be successfully used for density estimation.

The fundamental properties of wavelets are their concentration in the Fourier domain as well as in the space domain. Here, obviously the “space” domain is the sphere itself whereas the Fourier domain is now obtained by replacing the “Fourier” basis by the basis of spherical harmonics which plays an analogous role on the sphere.

The construction [18, 19] produces a family of functions which very much resemble wavelets, the needlets, and in particular have very good concentration properties.

We use these needlets to construct an estimation procedure, and prove that this procedure attains optimal rates over various spaces of regularity.

Again, the problem of choosing appropriate spaces of regularity on the sphere is a serious question, and we decided to consider the spaces which may be the closest to our natural intuition: those which generalize to the sphere case the classical Hölder spaces.

Of course the estimator can produce a function that takes negative values. It is a well-known fact that, except for small regularity cases, it is impossible to find minimax estimators satisfying the positivity constraint. The rates of convergence ensure that this kind of artifacts rarely occur and of course in practice one replaces the negative value by 0.

In the first section we present [19] needlets, and describe spaces of regularity on the sphere. In the second one we define our estimation procedure, and describe its properties.

The novelties of this paper lie in the application of thresholding to the needlet coefficients, which gives a very simple and adaptive procedure which works on the sphere. We also focus here on giving the results in \mathbb{L}_∞ norm, and obtain the rates of convergence for many other loss functions as a consequence of the previous ones.

Our results are motivated by many recent developments in the area of observational astrophysics. As an example, we refer to experiments measuring incoming directions of Ultra High Energy Cosmic Rays, such as the AUGER Observatory (<http://www.auger.org>). Here, efficient estimation of the density function of these directional data may yield crucial insights into the physical mechanisms generating the observations. More precisely, a uniform density would suggest the high energy cosmic rays are generated by cosmological effects, such as the decay of massive particles generated during the Big Bang; on the other hand, if these cosmic rays are generated by astrophysical phenomena (such as acceleration into active galactic nuclei), then we should observe a density function which is highly nonuniform and tightly correlated with the local distribution of nearby galaxies.

Massive amount of data in this area are expected to be available in the next few years. The Auger Observatory will be based on two arrays of detectors; the first one covers an area larger than 3000 km² in Pampa Amarilla (Argentina), and has already started to collect observations. Some preliminary evidence was provided in [4], and a nonuniform distribution seems to be favored. The whole celestial sphere will actually be covered only when the construction of the northern hemisphere array, due to be built in eastern Colorado, will be completed a few years from now. Hence, in the immediate future efficient statistical techniques will be eagerly requested for the analysis of the forthcoming datasets.

A survey of statistical methodologies dealing with directional data on the sphere may be found in [10, 16, 17]. The generalization of estimation using orthogonal series methods to the case of compact Riemannian manifold can be found in [8]. See related works in [7, 9, 20, 21] and [11]. Kernel methods on the sphere have been investigated in [6]. Minimax rates for the equivalent of Sobolev spaces on the sphere associated can be found in [13, 15] and [14].

The plan of the paper is as follows. In Sections 2 and 3 we review some background material on needlets and Besov spaces. Section 4 introduces our thresholding estimator, whose minimax performances are stated in Section 5. Section 6 shows the performance of the estimators on some simulated data. Sections 7–9 contain the proofs.

2. Needlets. This construction is due to Narcowich, Petrushev and Ward [19]. Its aim is essentially to build a very well-localized tight frame constructed using spherical harmonics, as discussed below. It was recently extended with fruitful statistical applications to more general Euclidean settings (see [12]) and already exploited for estimation and testing problems in [1] and [2].

Let us denote by \mathbb{S}^d , the unit sphere of \mathbb{R}^{d+1} . We denote dx the surface measure of \mathbb{S}^d , that is the unique positive measure on \mathbb{S}^d which is invariant by rotation and has total mass $\omega_d = 2\pi^{(d+1)/2} / \Gamma(\frac{d+1}{2})$. The following decomposition is well known:

$$(1) \quad \mathbb{L}^2(dx) = \bigoplus_{l=0}^{\infty} \mathcal{H}_l,$$

where \mathcal{H}_l is the restriction to \mathbb{S}^d of the homogeneous polynomials on \mathbb{R}^{d+1} of degree l which are harmonic (i.e., $\Delta P = 0$, where Δ is the Laplacian on \mathbb{R}^{d+1}). This space is called the space of spherical harmonics of degree l (see [22], Chapter 4 and [24], Chapter 5). Its dimension is equal to $g_{l,d} = \binom{l+d}{d} - \binom{l+d-2}{d}$ and is therefore of order l^{d-1} . The orthogonal projector on \mathcal{H}_l is given by the kernel operator

$$(2) \quad \forall f \in \mathbb{L}^2(dx), \quad P_{\mathcal{H}_l} f(x) = \int_{\mathbb{S}^d} L_l(\langle x, y \rangle) f(y) dy,$$

where $\langle x, y \rangle$ is the standard scalar product of \mathbb{R}^{d+1} , and L_l is the Gegenbauer polynomial with parameter $\frac{d-1}{2}$ of degree l , defined on $[-1, +1]$ and normalized so that

$$(3) \quad \int_{-1}^1 L_l(t)L_k(t)(1-t^2)^{d/2-1} dt = \frac{g_{l,d}2^{d/2}\Gamma(\frac{d}{2})^2}{\Gamma(d)\omega_d^2}\delta_{l,k} \\ = \frac{g_{l,d}2^{d/2}\Gamma(\frac{d+1}{2})^2\Gamma(\frac{d}{2})^2}{4\Gamma(d)\pi^{d+1}}\delta_{l,k}.$$

For the main situation of interest, $d = 2$, the right-hand side above is equal to $\frac{2l+1}{8\pi^2}$. Recall that if $d = 2$, the usual normalization of the Legendre polynomial ($L_l(1) = 1$) gives the square of their L^2 norm equal to $\frac{2}{2l+1}$. Therefore these must be multiplied by $(2l + 1)/(4\pi)$, in order to satisfy (3).

Let us point out the following reproducing property of the projection operators:

$$(4) \quad \int_{\mathbb{S}^d} L_l(\langle x, y \rangle)L_k(\langle y, z \rangle) dy = \delta_{l,k}L_l(\langle x, z \rangle).$$

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

Let φ be a C^∞ function on \mathbb{R} , symmetric and decreasing on \mathbb{R}^+ supported in $|\xi| \leq 1$, such that $1 \geq \varphi(\xi) \geq 0$ and $\varphi(\xi) = 1$ if $|\xi| \leq \frac{1}{2}$. We set

$$b^2(\xi) = \varphi\left(\frac{\xi}{2}\right) - \varphi(\xi) \geq 0$$

so that

$$(5) \quad \forall |\xi| \geq 1, \quad \sum_{j \geq 0} b^2\left(\frac{\xi}{2^j}\right) = 1.$$

Remark that $b(\xi) \neq 0$ only if $\frac{1}{2} \leq |\xi| \leq 2$. Let us now define the operator $\Lambda_j = \sum_{l \geq 0} b^2(\frac{l}{2^j})L_l$ and the associated kernel

$$\Lambda_j(x, y) = \sum_{l \geq 0} b^2\left(\frac{l}{2^j}\right)L_l(\langle x, y \rangle) = \sum_{2^{j-1} < l < 2^{j+1}} b^2\left(\frac{l}{2^j}\right)L_l(\langle x, y \rangle).$$

The following proposition is obvious.

PROPOSITION 1. For every $f \in L^2$

$$(6) \quad f = \lim_{J \rightarrow \infty} L_0(f) + \sum_{j=0}^J \Lambda_j(f).$$

Moreover, if $M_j(x, y) = \sum_{l \geq 0} b(\frac{l}{2^j})L_l(\langle x, y \rangle)$, then

$$(7) \quad \Lambda_j(x, y) = \int M_j(x, z)M_j(z, y) dz.$$

Let

$$\mathcal{P}_l = \bigoplus_{m=0}^l \mathcal{H}_m$$

the space of the restrictions to \mathbb{S}^d of the polynomials of degree $\leq l$. The following quadrature formula is true: for all $l \in \mathbb{N}$ there exists a finite subset $\mathcal{X}_l \subset \mathbb{S}^d$ and positive real numbers $\lambda_\eta > 0$, indexed by the elements $\eta \in \mathcal{X}_l$, such that

$$(8) \quad \forall f \in \mathcal{P}_l, \quad \int_{\mathbb{S}^d} f(x) dx = \sum_{\eta \in \mathcal{X}_l} \lambda_\eta f(\eta).$$

Then the operator M_j defined in the subsection above is such that $z \mapsto M_j(x, z) \in \mathcal{P}_{[2^j+1]}$, so that

$$z \mapsto M_j(x, z)M_j(z, y) \in \mathcal{P}_{[2^{j+2}]},$$

and we can write

$$\Lambda_j(x, y) = \int M_j(x, z)M_j(z, y) dz = \sum_{\eta \in \mathcal{X}_{[2^{j+2}]}} \lambda_\eta M_j(x, \eta)M_j(\eta, y).$$

This implies

$$\begin{aligned} \Lambda_j f(x) &= \int \Lambda_j(x, y) f(y) dy = \int \sum_{\eta \in \mathcal{X}_{[2^{j+2}]}} \lambda_\eta M_j(x, \eta)M_j(\eta, y) f(y) dy \\ &= \sum_{\eta \in \mathcal{X}_{[2^{j+2}]}} \sqrt{\lambda_\eta} M_j(x, \eta) \int \sqrt{\lambda_\eta} M_j(\eta, y) f(y) dy. \end{aligned}$$

We denote

$$\mathcal{X}_{[2^{j+2}]} = \mathcal{Z}_j, \quad \psi_{j,\eta}(x) := \sqrt{\lambda_\eta} M_j(x, \eta) \quad \text{for } \eta \in \mathcal{Z}_j.$$

The choice of the sets \mathcal{Z}_j of cubature points is not unique, but one can impose the conditions

$$(9) \quad \frac{1}{c} 2^{dj} \leq \#\mathcal{Z}_j \leq c 2^{dj}, \quad \frac{1}{c} 2^{-dj} \leq \lambda_\eta \leq c 2^{-dj}$$

for some $c > 0$. Actually in the simulations of Section 6 we make use of some sets of cubature points for $d = 2$ such that $\#\mathcal{Z}_j = 2^{2j+4}$ exactly (the corresponding weights being however not identical). We have, using (6),

$$(10) \quad f = L_0(f) + \sum_j \sum_{\eta \in \mathcal{Z}_j} \langle f, \psi_{j,\eta} \rangle_{\mathbb{L}^2(\mathbb{S}^d)} \psi_{j,\eta}.$$

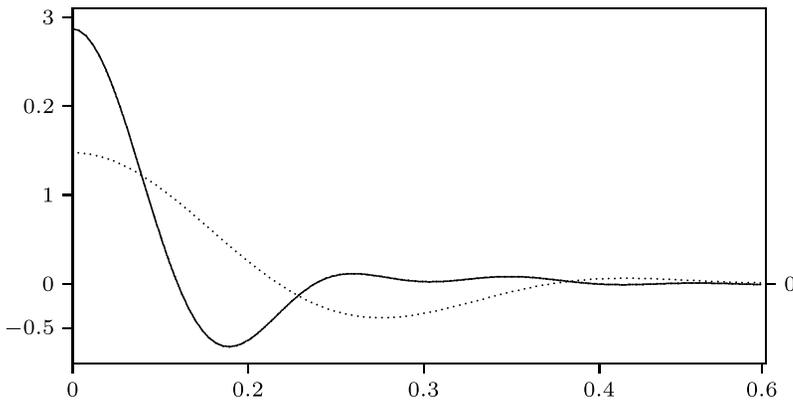


FIG. 1. The value of the needlet $\psi_{j,\xi}$ as a function of the distance from the cubature point ξ for $j = 3$ (dots) and $j = 4$ (solid). This emphasizes the localization properties of needlets as j increases.

The main result of Narcowich, Petrushev and Ward [19] is the following localization property of the $\psi_{j,\eta}$, called needlets: for any k there exists a constant c_k such that, for every $\xi \in \mathbb{S}^d$:

$$(11) \quad |\psi_{j,\eta}(\xi)| \leq \frac{c_k 2^{jd/2}}{(1 + 2^{jd/2} \text{dist}(\eta, \xi))^k},$$

where d is the natural geodesic distance on the sphere [for $d = 2$, $d(\xi, \eta) = \arccos\langle \eta, \xi \rangle$] (see Figure 1). In other words needlets are almost exponentially localized around any cubature point, which motivates their name. From this localization property it follows (see [19]) that for $1 \leq p \leq +\infty$ there exist a positive constant C_p such that

$$(12) \quad \frac{1}{C_p} 2^{jd(1/2-1/p)} \leq \|\psi_{j\eta}\|_p \leq C_p 2^{jd(1/2-1/p)}.$$

Also, the following holds:

LEMMA 2. (1) For every $0 < p \leq +\infty$

$$(13) \quad \left\| \sum_{\xi \in \mathcal{Z}_j} \lambda_\xi \psi_{j,\xi} \right\|_p \leq c 2^{jd(1/2-1/p)} \left(\sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p \right)^{1/p}.$$

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

$$(14) \quad \left(\sum_{\xi \in \mathcal{Z}_j} |\langle f, \psi_{j\xi} \rangle|^p \right)^{1/p} 2^{jd(1/2-1/p)} \leq c \|f\|_p.$$

PROOF. Let us prove (13) for $p = +\infty$. Using (11) and Lemma 6 of [1]

$$\begin{aligned} \sup_{x \in \mathbb{S}^d} \left| \sum_{\xi \in \mathcal{Z}_j} \lambda_\xi \psi_{j,\xi}(x) \right| &\leq \sup_{\xi \in \mathcal{Z}_j} |\lambda_\xi| \sup_{x \in \mathbb{S}^d} \sum_{\xi \in \mathcal{Z}_j} |\psi_{j,\xi}(x)| \\ &\leq \sup_{\xi \in \mathcal{Z}_j} |\lambda_\xi| c_3 \sup_{x \in \mathbb{S}^d} \sum_{\xi \in \mathcal{Z}_j} \frac{2^{jd/2}}{(1 + 2^{jd/2} d(\xi, x))^3} \\ &\leq \tilde{c}_3 2^{jd/2} \sup_{\xi \in \mathcal{Z}_j} |\lambda_\xi|. \end{aligned}$$

If $1 \leq p < +\infty$, by Hölder inequality, if $\frac{1}{p} + \frac{1}{p'} = 1$ so that $\frac{p}{p'} = p - 1$,

$$\begin{aligned} \left(\sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi \psi_{j,\xi}(x)| \right)^p &= \left(\sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi| |\psi_{j,\xi}(x)|^{1/p} |\psi_{j,\xi}(x)|^{1/p'} \right)^p \\ &\leq \left(\sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p |\psi_{j,\xi}(x)| \right) \left(\sum_{\xi \in \mathcal{Z}_j} |\psi_{j,\xi}(x)| \right)^{p-1} \\ &\leq \tilde{c}_3^{p-1} 2^{jd/2(p-1)} \sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p |\psi_{j,\xi}(x)|, \end{aligned}$$

where the last inequality comes again from (11) and Lemma 6 of [1]. Now integrating and using (12) for $p = 1$,

$$\left\| \sum_{\xi \in \mathcal{Z}_j} \lambda_\xi \psi_{j,\xi}(x) \right\|_p^p \leq 2^{jd/2(p-1)} \sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_1 \leq 2^{jd/2(p-2)} \sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p$$

from which (13) follows. The remaining case $0 < p \leq 1$ follows immediately by subadditivity, as

$$\left\| \sum_{\xi \in \mathcal{Z}_j} \lambda_\xi \psi_{j,\xi}(x) \right\|_p^p \leq \sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p \|\psi_{j,\xi}(x)\|_p^p.$$

As for (2) clearly if $p = +\infty$

$$\begin{aligned} C 2^{jd/2} \sup_{\xi \in \mathcal{Z}_j} |\langle f, \varphi_{j,\xi} \rangle| &\leq C 2^{jd/2} \sup_{\xi \in \mathcal{Z}_j} \int |f(x)| |\varphi_{j,\xi}(x)| dx \\ &\leq C 2^{jd/2} \|f\|_\infty \sup_{\xi \in \mathcal{Z}_j} \|\varphi_{j,\xi}\|_1 \leq C' \|f\|_\infty \end{aligned}$$

and if $p = 1$

$$\begin{aligned} \sum_{\xi \in \mathcal{Z}_j} |\langle f, \varphi_{j,\xi} \rangle| 2^{-jd/2} &\leq 2^{-jd/2} \sum_{\xi \in \mathcal{Z}_j} \int |f(x)| |\varphi_{j,\xi}(x)| dx \\ &= 2^{-jd/2} \int |f(x)| \sum_{\xi \in \mathcal{Z}_j} |\varphi_{j,\xi}(x)| dx \leq C \|f\|_1. \end{aligned}$$

Let now $1 < p < \infty$

$$\sum_{\xi \in \mathcal{Z}_j} |\langle f, \varphi_{j,\xi} \rangle|^p 2^{jd(p/2-1)} \leq 2^{jd(p/2-1)} \sum_{\xi \in \mathcal{Z}_j} \left(\int |f(x)| |\varphi_{j,\xi}(x)| dx \right)^p.$$

But, by Hölder’s inequality, for p' such that $\frac{1}{p} + \frac{1}{p'} = 1$, then $\frac{p}{p'} = p - 1$ and

$$\begin{aligned} \left(\int |f(x)| |\varphi_{j,\xi}(x)| dx \right)^p &= \left(\int |f(x)| |\varphi_{j,\xi}(x)|^{1/p} |\varphi_{j,\xi}(x)|^{1/p'} dx \right)^p \\ &\leq \int |f(x)|^p |\varphi_{j,\xi}(x)| dx \left(\int |\varphi_{j,\xi}(x)| dx \right)^{p-1} \\ &= \int |f(x)|^p |\varphi_{j,\xi}(x)| dx \|\varphi_{j,\xi}\|_1^{p-1}. \end{aligned}$$

So

$$\begin{aligned} \sum_{\xi \in \mathcal{Z}_j} |\langle f, \varphi_{j,\xi} \rangle|^p 2^{jd(p/2-1)} &\leq 2^{-jd/2(p-1)} 2^{jd(p/2-1)} \sum_{\xi \in \mathcal{Z}_j} \int |f(x)|^p |\varphi_{j,\xi}(x)| dx \\ &= 2^{-jd/2} \int |f(x)|^p \sum_{\xi \in \mathcal{Z}_j} |\varphi_{j,\xi}(x)| dx \leq C \|f\|_p^p. \quad \square \end{aligned}$$

EXAMPLE 3. Relation (12) for $p = 2$ states that the L^2 norm of $\psi_{j\xi}$ is bounded with respect to j and also bounded away from 0 from below. Assume $d = 2$. Then using (4) it is actually easy to see that, keeping in mind that $L_l(1) = \frac{2^{l+1}}{4\pi}$,

$$\|\psi_{j\xi}\|_2^2 = \lambda_\xi \sum_{l \geq 0} b^2 \left(\frac{l}{2^j} \right) L_l(1) = \frac{\lambda_\eta}{4\pi} \sum_{l \geq 0} b^2 \left(\frac{l}{2^j} \right) (2l + 1).$$

Assuming that the cubature points are of cardinality 2^{2j+4} and that they sum up to 4π , $\lambda_\eta \sim 4\pi \cdot 2^{-2j-4}$ as $j \rightarrow \infty$. If the previous relation were an equality we could recognize in the right-hand term the Riemann sum

$$\frac{1}{8} \left(\frac{1}{2^j} \sum_{l \geq 0} b^2 \left(\frac{l}{2^j} \right) \frac{l}{2^j} + \frac{1}{2^{2j}} \sum_{l \geq 0} b^2 \left(\frac{l}{2^j} \right) \right)$$

that converges, as $j \rightarrow \infty$, to the integral

$$I = \frac{1}{8} \int_{1/2}^2 t b^2(t) dt$$

which depends on the choice of the function b . This L^2 norm shall appear frequently in the sequel. For instance, in the development (10) of the function $f = \psi_{j_0\xi_0}$ the coefficient $\beta_{j_0\xi_0} = \langle f, \psi_{j,\eta} \rangle$ would be exactly equal to $\|\psi_{j\xi}\|_2^2$. As it

is clear that it would be desirable for this coefficient to be as large as possible, the value of the integral above can be seen as a measure of the localization properties of the system of needlets and can be used as a criterion of goodness of the choice of the function b . With the choice we made (see Section 6) the quantity I above is $\simeq 0.107$.

3. Besov spaces on the sphere and needlets. In this section we summarize the main properties of Besov spaces and needlets, as established in [19].

Let $f : \mathbb{S}^d \rightarrow \mathbb{R}$ a measurable function. We define

$$E_k(f, r) = \inf_{P \in \mathcal{P}_k} \|f - P\|_r$$

the infimum of the distances in L^r of f from the polynomials of degree k . Then the Besov space $B_{r,q}^s$ is defined as the space of functions such that

$$f \in L^r \quad \text{and} \quad \left(\sum_{k=0}^{\infty} (k^s E_k(f, r))^q \frac{1}{k} \right)^{1/q} < +\infty.$$

Remarking that $k \rightarrow E_k(f, r)$ is decreasing, by a standard condensation argument, as

$$(2^{js} E_{2^{j+1}}(f, r))^q \leq \sum_{k=2^j}^{2^{j+1}-1} (k^s E_k(f, r))^q \frac{1}{k} \leq 2(2^{(j+1)s} E_{2^j}(f, r))^q$$

this is equivalent to

$$f \in L^r \quad \text{and} \quad \left(\sum_{j=0}^{\infty} (2^{js} E_{2^j}(f, r))^q \right)^{1/q} < +\infty.$$

THEOREM 4. *Let $1 \leq r \leq +\infty$, $s > 0$, $0 \leq q \leq +\infty$. Let f a measurable function and define*

$$\langle f, \psi_{j,\xi} \rangle = \int_{\mathbb{S}^d} f(x) \psi_{j,\xi}(x) dx \stackrel{\text{def}}{=} \beta_{j,\xi}$$

provided the integrals exists. Then $f \in B_{r,q}^s$ if and only if, for every $j = 1, 2, \dots$,

$$\left(\sum_{\xi \in \mathcal{X}_j} (\beta_{j,\xi} \|\psi_{j,\xi}\|_r)^r \right)^{1/r} = 2^{-js} \delta_j,$$

where $(\delta_j)_j \in \ell_q$.

As

$$c2^{jd(1/2-1/r)} \leq \|\psi_{j,\xi}\|_r \leq C2^{jd(1/2-1/r)}$$

for some positive constants c, C , the Besov space $B_{r,q}^s$ turns out to be a Banach space associate to the norm

$$(15) \quad \|f\|_{B_{r,q}^s} := \|(2^{j[s+d(1/2-1/r)]} \|(\beta_{j\eta})_{\eta \in \mathcal{L}_j}\|_{\ell_r})_{j \geq 0}\|_{\ell_q} < \infty.$$

In the sequel we shall denote by $B_{r,q}^s(M)$ the ball of radius M of the Besov space $B_{r,q}^s$.

THEOREM 5 (The Besov embedding). *If $p \leq r \leq \infty$ then $B_{r,q}^s \subseteq B_{p,q}^s$. If $s > d(\frac{1}{r} - \frac{1}{p})$,*

$$r \leq p \leq \infty \quad \Rightarrow \quad B_{r,q}^s \subseteq B_{p,q}^{s-d(1/r-1/p)}.$$

PROOF. By hypothesis

$$2^{jd(1/2-1/r)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \leq \delta_j 2^{-js}, \quad (\delta_j)_j \in \ell_q.$$

Let $p \leq r \leq \infty$, then

$$\begin{aligned} & 2^{jd(1/2-1/p)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^p \right)^{1/p} \\ &= 2^{jd/2} (2^{-jd} \text{card}(\mathcal{L}_j))^{1/p} \left(\frac{1}{\text{card}(\mathcal{L}_j)} \sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^p \right)^{1/p} \\ &\leq 2^{jd/2} (2^{-jd} \text{card}(\mathcal{L}_j))^{1/p} \left(\frac{1}{\text{card}(\mathcal{L}_j)} \sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \\ &= 2^{jd(1/2-1/r)} (2^{-jd} \text{card}(\mathcal{L}_j))^{1/p-1/r} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \\ &\leq C 2^{jd(1/2-1/r)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \leq C \delta_j 2^{-js}. \end{aligned}$$

On the other hand, if $r \leq p \leq \infty$,

$$\begin{aligned} 2^{jd(1/2-1/p)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^p \right)^{1/p} &\leq 2^{jd(1/2-1/p)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \\ &= 2^{jd(1/r-1/p)} 2^{jd(1/2-1/r)} \left(\sum_{\xi \in \mathcal{L}_j} |\beta_{j,\xi}|^r \right)^{1/r} \\ &\leq 2^{jd(1/r-1/p)} \delta_j 2^{-js} = \delta_j 2^{-j(s-d(1/r-1/p))}. \end{aligned}$$

It is easy to show (see [19]) that the Sobolev space H_s , defined through the norm

$$\|f\|_s = \|f\|_2 + \|\Delta^{s/2} f\|_2,$$

coincides with the Besov space B_{22}^s . □

4. Needlet estimation of a density on the sphere. Let us suppose that we observe X_1, \dots, X_n , i.i.d. random variables taking values on the sphere having common density f with respect to dx . f can be decomposed using the frame of needlets described above.

$$f = \frac{1}{|\mathbb{S}^d|} + \sum_{j \geq 0} \sum_{\eta} \beta_{j\eta} \psi_{j\eta}.$$

The needlet estimator is based on hard thresholding of a needlet expansion as follows. We start by letting

$$(16) \quad \hat{\beta}_{j\eta} := \frac{1}{n} \sum_{i=1}^n \psi_{j\eta}(X_i),$$

$$(17) \quad \hat{f} := \frac{1}{|\mathbb{S}^d|} + \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} \hat{\beta}_{j\eta} \psi_{j\eta} 1_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}}.$$

The tuning parameters of the needlet estimator are:

- The range $J = J(n)$ of resolution levels (frequencies) where the approximation (17) is used:

$$\Lambda_n = \{(j, \eta), 0 \leq j \leq J, \eta \in \mathcal{Z}_j\},$$

We shall see that the choice $2^J = (\frac{n}{\log n})^{1/d}$ is appropriate.

- The threshold constant κ . Evaluations of κ are given in the following section and also discussed in Section 6.
- t_n : is a sample size-dependent scaling factor. We shall see that an appropriate choice is

$$t_n = \left(\frac{\log n}{n}\right)^{1/2}.$$

EXAMPLE 6. In order to give a better intuition about the localization and near absence of correlation of the needlet coefficients, let us consider the case of a sample X_1, \dots, X_n of i.i.d. r.v.'s uniform on the sphere \mathbb{S}^2 of \mathbb{R}^3 . Then the distribution of the r.v. $\langle x, X_i \rangle$ is uniform on the interval $[-1, 1]$ and, if $\hat{\beta}_{\eta j}, \hat{\beta}_{\xi j}$ are the corresponding needlet coefficients associated to the cubature points η, ξ then of course they are centered r.v.'s and, thanks to (4), their covariance is equal to

$$E(\hat{\beta}_{\eta j} \hat{\beta}_{\xi j}) = \sqrt{\lambda_{\eta} \lambda_{\xi}} \sum_{l \geq 0} b^2 \left(\frac{l}{2j}\right) L_l(\langle \xi, \eta \rangle).$$

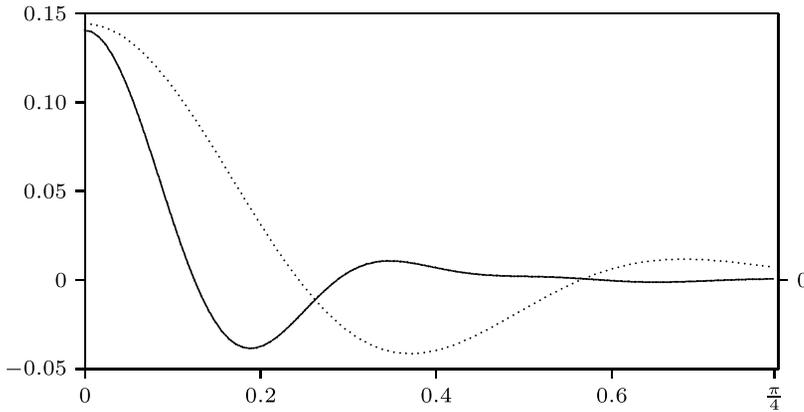


FIG. 2. The decay of the covariance of $\widehat{\beta}_{j,\xi}$ and $\widehat{\beta}_{j,\eta}$ as a function of the distance between the cubature points ξ and η for $j = 3$ (dots) and $j = 4$ (solid) (case of a uniformly distributed sample).

Setting $\eta = \xi$ we find

$$\text{Var}(\widehat{\beta}_{\eta j}) = \|\psi_{j\xi}\|_2^2$$

which is a quantity already discussed in Example 3. As for the correlation between coefficients, it is given by the function $\theta \rightarrow \lambda_\eta \sum_{l \geq 0} b^2(\frac{l}{2j}) L_l(\cos \theta)$, whose graph, for some values of j is plotted in Figure 2.

REMARK 7. Whereas coefficients associated to cubature points that are not too close are only slightly correlated, the random needlet coefficients $\widehat{\beta}_{j,\eta}$, $\eta \in \mathcal{X}_j$ are not independent and they even satisfy the linear relation

$$\sum_{\eta \in \mathcal{X}_j} \sqrt{\lambda_\eta} \widehat{\beta}_{j,\eta} = 0.$$

This comes from the fact that, as $y \rightarrow L_l(\langle y, x \rangle)$ for $l \leq 2^j$ is a polynomial of degree $\leq 2^{2j}$, one has

$$(18) \quad \sum_{\eta \in \mathcal{X}_j} \lambda_\eta L_l(\langle \eta, x \rangle) = \int_{\mathbb{S}^d} L_l(\langle y, x \rangle) dy = 0.$$

Therefore

$$\sum_{\eta \in \mathcal{X}_j} \sqrt{\lambda_\eta} \widehat{\beta}_{j,\eta} = \frac{1}{n} \sum_{i=1}^n \sum_{l>0} b\left(\frac{l}{2j}\right) \underbrace{\sum_{\eta \in \mathcal{X}_j} \lambda_\eta L_l(\langle \eta, X_i \rangle)}_{=0} = 0.$$

Relation (18) also implies that, for a given square integrable function f on \mathbb{S}^d , $\sum_{\eta \in \mathcal{Z}_j} \sqrt{\lambda_\eta} \beta_{j,\eta} = 0$. Actually

$$\sum_{\eta \in \mathcal{Z}_j} \sqrt{\lambda_\eta} \beta_{j,\eta} = \sum_{l>0} b\left(\frac{l}{2^j}\right) \underbrace{\int_{\mathbb{S}^d} \sum_{\eta \in \mathcal{Z}_j} \lambda_\eta L_l(\langle \eta, x \rangle) f(x) dx}_{=0}$$

5. Minimax rates for \mathbb{L}^p norms and Besov spaces on the sphere. We describe the performances of the procedure by the following theorem. Remark that the condition $s > \frac{d}{r}$ implies $f \in B_{r,q}^s \subset B_{\infty,q}^{s-d/r}$ so that f is continuous. By \mathbb{E}_f we denote the expectation taken with respect to a probability with respect to which the r.v.'s $(X_n)_n$ are i.i.d. with common density f .

THEOREM 8. For $0 < r \leq \infty, p \geq 1, s > \frac{d}{r}$ we have:

(a) For any $z > 1$, there exist some constants $c_\infty = c_\infty(s, p, r, M)$ such that if $\kappa > \frac{z+1}{6}$,

$$(19) \quad \sup_{f \in B_{r,q}^s(M)} \mathbb{E}_f \|\hat{f} - f\|_\infty^z \leq c_\infty (\log n)^{z-1} \left[\frac{n}{\log n} \right]^{-(s-d/r)/(2(s-d(1/r-1/2)))}$$

(b) For $1 \leq p < \infty$ there exist some constant $c_p = c_p(s, r, p, M)$ such that if $\kappa > \frac{p}{12}$,

$$(20) \quad \begin{aligned} & \sup_{f \in B_{r,q}^s(M)} \mathbb{E}_f \|\hat{f} - f\|_p^p \\ & \leq c_p (\log n)^{\alpha_p} \left[\frac{n}{\log n} \right]^{-(s-d(1/r-1/p))p/(2(s-d(1/r-1/2)))} \end{aligned}$$

where $\alpha_p = p - 1 + 1_{\{r=d p/2s+d\}}$, if $r \leq \frac{d p}{2s+d}$, whereas

$$(21) \quad \sup_{f \in B_{r,q}^s(M)} \mathbb{E}_f \|\hat{f} - f\|_p^p \leq c_p (\log n)^{p-1} \left[\frac{n}{\log n} \right]^{-s p/2s+d} \quad \text{if } r > \frac{d p}{2s+d}.$$

REMARK 9. Compared to the results obtained in [5] for instance, we see that the influence of the sphere do not appear in the minimax rates (although it does appear in the statistical procedures). As in [5], we find rates of convergence which compare to those obtained in [13, 14] in the homogeneous case (i.e., when $p = r$). It is the multiresolution and localization properties of the needlets that allow to obtain optimal rates in the nonhomogeneous case. Also note here that, contrary to the kernel methods, our procedure do not need an a priori knowledge of the regularity. It is fully adaptive.

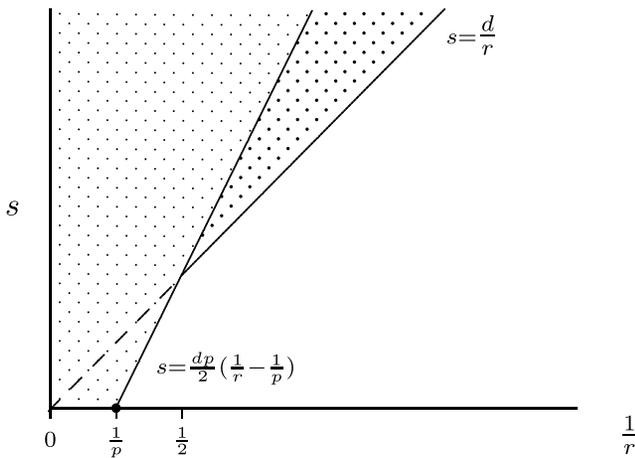


FIG. 3. Type of behavior of the minimax estimator as a function of s, r and p . The region with the large dots is the one corresponding to the sparse case of (20). The region with the small dots corresponds to the regular case so that (21) holds—see also Remark 10. It should be noticed that if $p \leq 2$ then the slope of the straight line starting at $\frac{1}{p}$ is smaller than the other one, so that the sparse region is empty.

Usually the case (20) is referred to as the *sparse case*, whereas (21) is the *regular case*. Remark that if $p \leq 2$, then we are always in the regular case (see also Figure 3).

REMARK 10. A closer look to the proof shows that, in the regular case, if we assume $\|f\|_\infty \leq M$, then we can drop the restriction $s > \frac{d}{r}$ without any modification if $1 \leq p \leq 2$. In the case $p > 2$, using an additional modification allowing J to depend also in p , $2^J = (\frac{n}{\log n})^{p/p-2}$ to be precise, we obtain the same rate under the same conditions as in the lower bound (up to logarithmic terms).

THEOREM 11 (Lower bound). (a) If $1 \leq p \leq 2$,

$$\sup_{f \in B_{r,q}^s(M)} \mathbb{E}_f(\|\hat{f} - f\|_p^p) \geq cn^{-sp/(2s+d)}.$$

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

$$\sup_{f \in B_{r,q}^s(M)} \mathbb{E}_f(\|\hat{f} - f\|_p^p) \geq \begin{cases} cn^{-sp/(2s+d)}, & \text{if } s > p \frac{d}{2} \left(\frac{1}{r} - \frac{1}{p} \right)_+, \\ cn^{-p(s+d(1/p-1/r))/(2(s+d(1/2-1/r)))}, & \text{if } \frac{d}{r} < s \leq p \frac{d}{2} \left(\frac{1}{r} - \frac{1}{p} \right). \end{cases}$$

REMARK 12. As already remarked, up to logarithmic terms, the rates observed are minimax. It is known that in this kind of estimation, full adaptation yields unavoidable extra logarithmic terms. The rates of the logarithmic terms obtained in Theorem 8 are suboptimal (for instance, for obvious reason the case $p = 2$ yields much less logarithmic terms). We have focused on a simple proof giving all the results in a rather clear and readable way. However, using a more intricate proof, the rates could be improved up to be comparable with those in [5].

6. Simulations. In this section we produce the result of numerical experiments on the sphere \mathbb{S}^2 . In both of them the major question concerns the choice of the values of J and κ . Actually in practical (finite sample) situations the values given in Theorem 8 should be considered just as a reasonable hint. The sets of cubature points in the simulations that follow have been taken from the web site of R. Womersley <http://web.maths.unsw.edu.au/~rsw>.

We realized the function φ of Section 2 by connecting the levels 0 and 1 with a function that is the primitive, suitably rescaled, of the function $x \rightarrow e^{-(1-x^2)^{-1}}$, set to be equal to 0 outside $[-1, 1]$. The shape of the resulting function b is given in Figure 4.

For this choice of b , we have

$$\frac{1}{8} \int_{1/2}^2 t b^2(t) dt \simeq 0.107$$

which, as remarked above, gives an indication about the square of the value of the L^2 norm of a needlet $\psi_{j\xi}$. In the first two examples below we considered samples of cardinality $n = 2000$ and $n = 8000$. The hint for the value of J of Theorem 8 is $J = \frac{1}{2} \log_2(\frac{n}{\log n})$, which gives the values $J \sim 4.02$ and $J \sim 4.9$, respectively. One should keep in mind that at a given level j it is necessary to have enough cubature points in order to integrate exactly all polynomials up to the degree $2(2^{j+1} - 1) = 2^{j+2} - 2$, which means $\sim 2^{2j+4}$ cubature points with

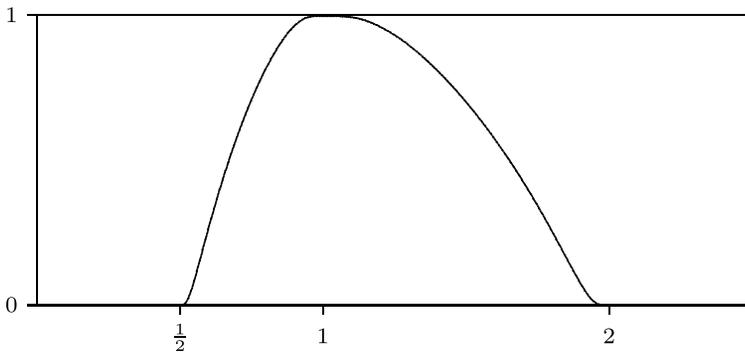


FIG. 4. *The function b .*

TABLE 1
Number of coefficients surviving thresholding for various values of k_0 , $n = 2000$

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$k_0 = 1$	8 (0.89)	29 (0.45)	96 (0.38)	471 (0.46)
$k_0 = 1.5$	7 (0.78)	16 (0.25)	45 (0.18)	264 (0.26)
$k_0 = 2$	4 (0.44)	4 (0.06)	29 (0.11)	126 (0.12)

Womersley’s set [recall that on the sphere the polynomials of degree d form a vector space of dimension $(2d + 1)^2$]. This gives $2^{10} = 1024$ cubature points for $j = 3$, $2^{12} = 4096$ for $j = 4$ and $2^{14} = 16384$ for $j = 5$. To avoid to have more coefficients than observations, we decided to set $J = 3$ for $n = 2000$ and $J = 4$ for $n = 8000$.

As for the value of κ , we shall give the result with $\kappa = k_0\sqrt{0.107}M$, where M is a bound for $\|f\|_\infty$, trying different values of k_0 . Recall that this means that the threshold kills all coefficients $\beta_{j\xi}$ such that $|\beta_{j\xi}| < \kappa\sqrt{\frac{\log n}{n}}$.

EXAMPLE 13. $f = \frac{1}{4\pi}$, the uniform density. In this case in the development (10) it holds $\beta_{j\xi} = \langle f, \psi_{j\xi} \rangle_{L^2} = 0$ for every j and ξ . Therefore, a first simple way of assessing the performance of the procedure is to count the number of coefficients that survive thresholding. Of course in this case a good estimate is such that the coefficients $\beta_{j,\xi}$ fall below the threshold. Taking into account Lemma 2 the square root of the sum of the squares of the coefficients surviving thresholding gives an estimate of $\|\hat{f} - f\|_2$. Therefore a measure of the goodness of the fit is obtained by taking the sum of their squares. Tables 1 and 2 give the number of surviving coefficients for different values of the constant k_0 . In order to kill all the coefficients one should choose $k_0 = 5.4$ for $n = 2000$ and $k_0 = 2.8$ for $n = 8000$. The estimate of the L^2 norm of the difference between \hat{f} and f by taking the square root of the sum of the squares of the coefficients is

k_0	1	1.5	2
$n = 2000$	0.146	0.131	0.107
$n = 8000$	0.108	0.0834	0.060.

EXAMPLE 14. Let us consider a mixture f of two densities of the form $f_i(x) = c_i e^{-k_i|x-x_i|^2}$, $i = 1, 2$, for $k_1 = 0.6$ and $k_2 = 4$ and with weights 0.7 and 0.3, respectively. Here the centers x_i of the two bell-shaped densities were taken to be $x_1 = (0, 1, 0)$ and $x_2 = (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0) = (-0.707, -0.707, 0)$. With these choices it turns out that $\|f\|_\infty = 0.4$. The graph of f in the coordinates (φ, θ)

TABLE 2
Number of coefficients surviving thresholding for various values of k_0 , $n = 8000$

	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$
$k_0 = 1$	4 (0.44)	28 (0.44)	96 (0.38)	413 (0.40)	1610 (0.39)
$k_0 = 1.5$	2 (0.22)	12 (0.19)	50 (0.20)	207 (0.20)	921 (0.20)
$k_0 = 2$	1 (0.11)	4 (0.06)	16 (0.06)	97 (0.09)	368 (0.09)

($\varphi =$ longitude, $\theta =$ colatitude) is given in Figure 5. The following Figures 6, 7 and 8 provide the graphs of the needlet estimators obtained with various choices of the value of k_0 . The best results appear to be given by choosing $k_0 = 0.45$, as at $k_0 = 0.25$ some ripples appear whereas for $k_0 = 0.7$ the shape looks good, but the graph is considerably flattened (see also Table 3 below).

Upon closer inspection, for $k_0 = 0.45$ only 6 coefficients at level $j = 3$ were not killed by thresholding (out of 1024) and only 1 survived for $j = 4$. All of these were related to cubature points near the location of the highest peak and actually the flat part of the graph is reconstructed essentially at the level $J = 1$ of resolution. For $k_0 = 0.7$ all coefficients for $j = 3$ and $j = 4$ were killed. The multiresolution approach with threshold shows nicely its flexibility: finer levels of resolution are used only at locations where they are needed.

As a comparison in Figures 9 and 10 we give the result of the reconstruction using a Gaussian kernel $f_2(x) = c_2 e^{-k|x-x_i|^2}$, which is similar to those already considered in the literature (see [13, 20] and [6]), and also similar to the form of the original density. We tried different values of the bandwidth k . One can remark that the value $k = 10$ produces an agreeable shape, but also a large difference in L^∞

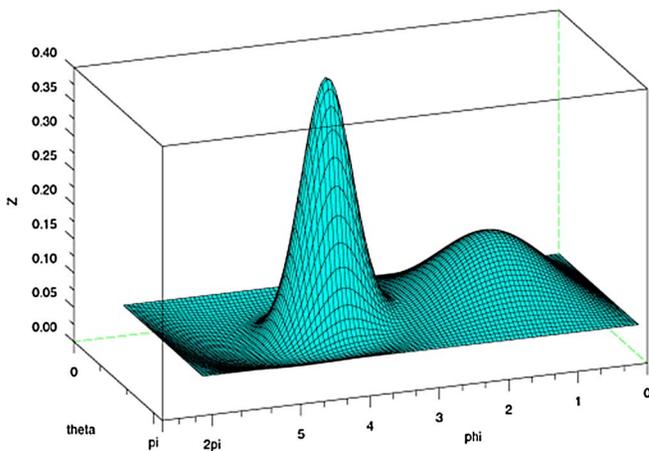


FIG. 5. *The target density.*

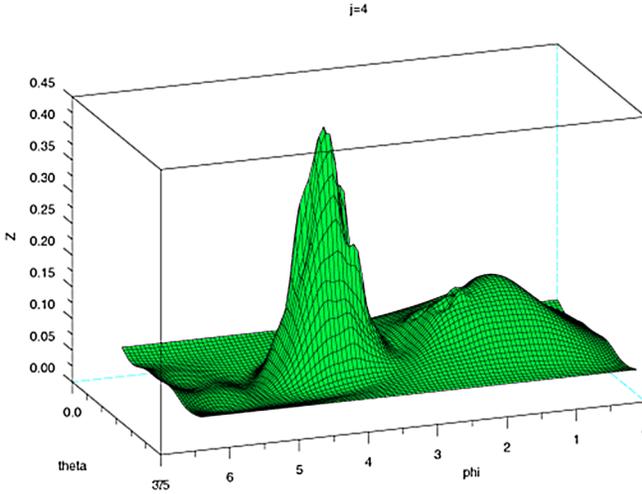


FIG. 6. The estimated density using needlets for $k_0 = 0.25$.

norm (the top of the highest peak is really too low). Larger values of the bandwidth appear to improve the L^∞ distance, but at the cost of the appearance of artifacts.

Table 3 gives the estimates of the L^∞ distance with the different methods.

It is therefore easy to point out here the advantages of the multiresolution approach. For the kernel estimation it can be difficult to adjust the value of the bandwidth parameter, even for a regular function. Indeed this value is fixed for all regions of the sphere and in the example above it appears that the value that is good at the points near the location of the highest peak are not the good ones in the flat regions.

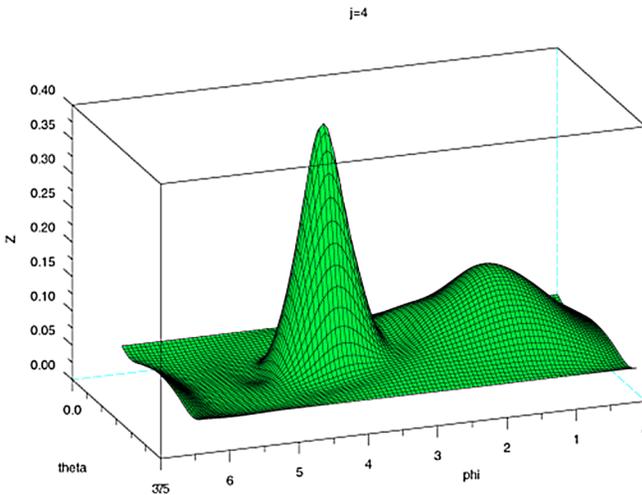


FIG. 7. The estimated density using needlets for $k_0 = 0.45$.

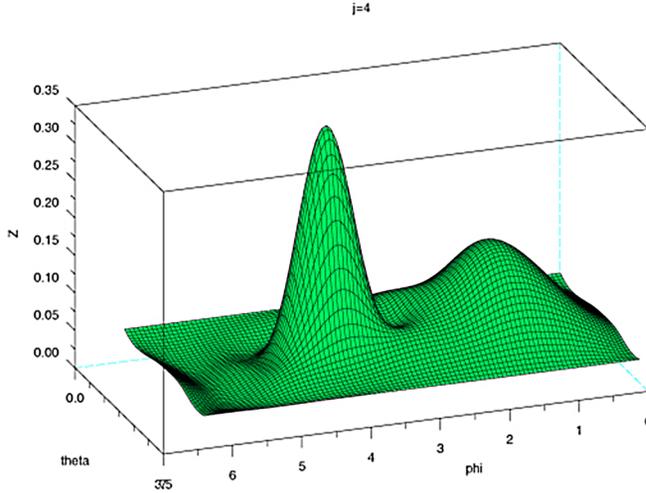


FIG. 8. The estimated density using needlets for $k_0 = 0.7$. The shape is good, but the height of the highest peak is notably lower than the target.

Conversely, thresholding and multiresolution provide a very flexible tool when dealing with this kind of situations.

To sum up, the multiresolution properties of the thresholding needlet estimator allow for local adaptation in the presence of multiple peaks and different slopes, whereas this possibility is ruled out for kernel estimators.

7. Proof of Theorem 8. In the sequel we note $t(\hat{\beta}_{j,\xi}) = \hat{\beta}_{j\eta} 1_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}}$ with $t_n = \sqrt{\frac{\log n}{n}}$, so that the needlet estimator (17) is

$$\hat{f} = \frac{1}{|\mathbb{S}^d|} + \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} t(\hat{\beta}_{j\eta}) \psi_{j\eta}.$$

TABLE 3
Estimates of the L^∞ distance with different methods

Method	L^∞ distance
Needlets $k_0 = 0.25$	0.073
Needlets $k_0 = 0.45$	0.045
Needlets $k_0 = 0.7$	0.066
Gaussian kernel $k = 10$	0.112
Gaussian kernel $k = 20$	0.064

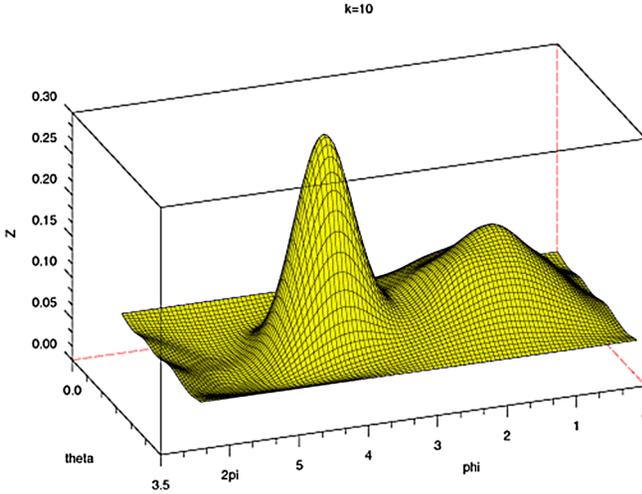


FIG. 9. The estimated density with a Gaussian kernel, bandwidth $k = 10$. The maximum is below 0.3.

In this section and in the next one the density f is fixed and we shall write \mathbb{E} instead of \mathbb{E}_f , as there is no danger of confusion.

The following proposition collects the main estimates needed in the proof.

PROPOSITION 15. Let $J_1 \leq J$ be such that, for all $J_1 \leq j \leq J$, $|\beta_{j\eta}| \leq \frac{\kappa}{2} t_n$ (possibly $J_1 = J$; obviously, when f belongs to a Besov class, J_1 depends on the “regularity” s). Then for any $\gamma > 0$, $s > 0$, $z \geq 1$, we have:

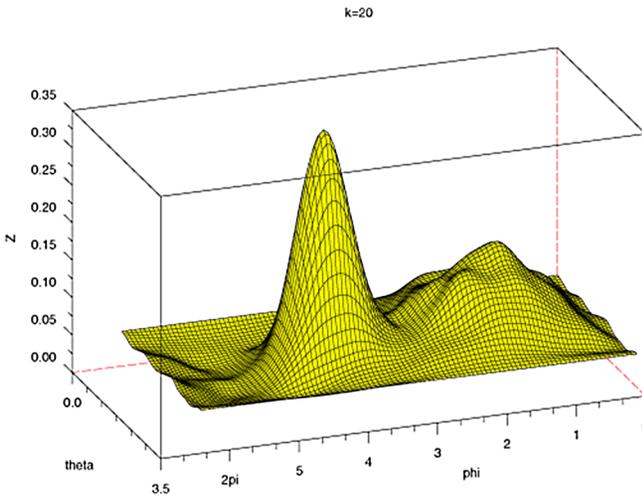


FIG. 10. The estimated density with a Gaussian kernel, bandwidth $k = 15$.

(1) If $\kappa > \frac{1}{3}(\frac{\gamma}{d} + \frac{1}{2})$

$$(22) \quad \sum_{j=0}^J 2^{\gamma j} \mathbb{E} \left[\sup_{\eta} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \right] \leq C \left[2^{J_1 \gamma} (J_1 + 1)^z n^{-z/2} + \sum_{j=J_1+1}^J 2^{\gamma j} \sup_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z + n^{-z/2} \right].$$

(2) If $\kappa > \frac{\gamma}{6d}$

$$(23) \quad \sum_{j=0}^J 2^{(\gamma-d)j} \mathbb{E} \sum_{\eta} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \leq C \sum_{j=0}^{J_1} 2^{(\gamma-d)j} n^{-z/2} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| > \kappa/2t_n\}} + C \sum_{j=0}^{J_1} 2^{(\gamma-d)j} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \leq 2\kappa t_n\}} + C n^{-z/2}$$

and

$$(24) \quad \sum_{j=0}^J 2^{(\gamma-d)j} \mathbb{E} \sum_{\eta} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \leq C \left[2^{J_1 \gamma} (J_1 + 1)^z n^{-z/2} + \sum_{j=J_1+1}^J 2^{(\gamma-d)j} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z + n^{-z/2} \right].$$

We delay the proof of Proposition 15 to Section 8 and derive from it the proof of Theorem 8. In this proof C will denote an absolute constant which may change from line to line. Let us now prove that Proposition 15 yields to the statements of Theorem 8.

Let us prove the L^∞ upper bound (19), first under the condition $q = r = \infty$

$$(25) \quad \begin{aligned} & \mathbb{E} \|\hat{f} - f\|_\infty^z \\ & \leq C \left[\mathbb{E} \left\| \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\xi}) - \beta_{j\eta}) \psi_{j\eta} \right\|_\infty^z + \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_\infty^z \right] \\ & := I + II. \end{aligned}$$

The term II is easy to analyze: as f belongs to $B_{\infty,\infty}^s(M)$, we have using (12) and (13),

$$\begin{aligned} \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_{\infty} &\leq \sum_{j>J} \left\| \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_{\infty} \leq C \sum_{j>J} \sup_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}| \|\psi_{j\eta}\|_{\infty} \\ &\leq C \sum_{j>J} 2^{-j[s+d/2]} 2^{jd/2} \leq \left(\frac{\log n}{n} \right)^{s/d}. \end{aligned}$$

Then we only need to remark that $\frac{s}{d} \geq \frac{s}{d+2s}$ for $s > 0$.

As for I , using the triangle inequality together with Hölder’s inequality, then (12), and (22) with $\gamma = \frac{dz}{2}$, $z > 1$, we get

$$\begin{aligned} I &\leq C(J+1)^{z-1} \sum_{j=0}^J 2^{jdz/2} \mathbb{E} \sup_{\eta \in \mathcal{Z}_j} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \\ &\leq C(J+1)^{z-1} \left[2^{J_1 dz/2} (J_1+1)^z n^{-z/2} + C \sum_{j=J_1+1}^J 2^{dz/2j} \sup_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z + n^{-z/2} \right]. \end{aligned}$$

As f belongs to $B_{\infty,\infty}^s(M)$, $|\beta_{j\eta}| \leq M2^{-j(s+d/2)}$ and we can see that

$$2^{J_1} = \frac{\kappa}{2M} \left[\frac{n}{\log n} \right]^{1/(2s+d)} \wedge 2^J$$

is adequate, it is easy to conclude.

For arbitrary q, r (19) is now easy to deduce from the previous computation by the Besov embedding (Theorem 5) $B_{r,q}^s(M) \subset B_{\infty,\infty}^{s-d/r}(M)$. Let us prove (21), that is the regular case. We observe first that since $B_{r,q}^s(M) \subset B_{p,q}^s(M)$ for $r \geq p$, this case will be assimilated to the case $p = r$, and from now on, we only consider $r \leq p$. We follow the same arguments as above. Equation (25) can be replaced by

$$\begin{aligned} &\mathbb{E} \|\hat{f} - f\|_p^p \\ (26) \quad &\leq C \left(\mathbb{E} \left\| \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p + \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_p^p \right) \\ &=: I + II. \end{aligned}$$

For II using the embedding $B_{r,q}^s(M) \subset B_{p,q}^{s-d/p+d/p}(M)$, for $r \leq p$, we have

$$II^{1/p} \leq C \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_p \leq C 2^{-J(s-d/r+d/p)}.$$

And it is easy to verify that $\frac{s}{d} - \frac{1}{r} + \frac{1}{p} \geq \frac{s}{2s+d}$ on the zone that we are considering in this part. In effect as $s \geq \frac{p}{2}(\frac{d}{r} - \frac{d}{p})$, $\frac{s}{2s+d} \leq \frac{sr}{dp}$ we have $\frac{s}{d} - \frac{1}{r} + \frac{1}{p} - \frac{sr}{dp} = (\frac{1}{r} - \frac{1}{p})(\frac{s}{d}r - 1) \geq 0$.

For I , we have using the triangle inequality together with Hölder’s inequality,

$$\begin{aligned} & \mathbb{E} \left\| \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p \\ & \leq C(J+1)^{p-1} \sum_{j=0}^J 2^{jd(p/2-1)} \times \sum_{\eta \in \mathcal{Z}_j} \mathbb{E} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^p. \end{aligned}$$

Then we need only to use (24), with $\gamma = \frac{dp}{2}$, $z = p$, to obtain

$$\begin{aligned} I \leq C(J+1)^{p-1} & \left[2^{J_1 dp/2} (J_1+1)^p n^{-p/2} \right. \\ & \left. + \sum_{j=J_1+1}^J 2^{d(p/2-1)j} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^p + n^{-p/2} \right]. \end{aligned}$$

It is easy to realize that again

$$2^{J_1} = \frac{\kappa}{2M} \left[\frac{n}{\log n} \right]^{1/(2s+d)} \wedge 2^J$$

is adequate, and to observe that the first term in the sum has the right order. For the second term, it can be bounded (as $p \geq r$) by

$$\begin{aligned} & C(J+1)^{p-1} \sum_{j=J_1+1}^J 2^{d(p/2-1)j} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^r [\kappa t_n]^{p-r} \\ & \leq C(J+1)^{p-1} \sum_{j=J_1+1}^J 2^{d(p/2-1)j} M^p 2^{-jr(s+d/2-d/r)} t_n^{p-r} \\ & \leq C t_n^{p-r} 2^{-J_1(sr-d/2(p-r))} \end{aligned}$$

as, $sr - \frac{d}{2}(p-r) \geq 0$. Now, this term obviously is of the right order.

Again we proceed as above and observe first that in order to have $s > 0$ as well as $s \leq \frac{pd}{2}(\frac{1}{r} - \frac{1}{p})$, it is necessary that $p \geq r$.

$$\begin{aligned} & \mathbb{E} \|\hat{f} - f\|_p^p \\ (27) \quad & \leq C \mathbb{E} \left\| \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p + \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_p^p \\ & =: I + II. \end{aligned}$$

For II using the embedding, $B_{r,q}^s(M) \subset B_{p,q}^{s-d/r+d/p}(M)$, for $r \leq p$, we have

$$II^{1/p} \leq C \left\| \sum_{j>J} \sum_{\eta \in \mathcal{Z}_j} \beta_{j\eta} \psi_{j\eta} \right\|_r \leq C 2^{-J(s-d/r+d/p)}.$$

And it is easy to verify that $\frac{s}{d} - \frac{1}{r} + \frac{1}{p} \geq \frac{(s-d(1/r-1/p))}{2(s-d(1/r-1/2))}$, since $2(s-d(\frac{1}{r}-\frac{1}{2})) \geq d$, when $s > \frac{d}{r}$.

For I , again, we have using the triangle inequality together with Hölder’s inequality,

$$\begin{aligned} & \mathbb{E} \left\| \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} (t(\hat{\beta}_{j\eta}) - \beta_{j\eta}) \psi_{j\eta} \right\|_p^p \\ & \leq C(J+1)^{p-1} \sum_{j=0}^J 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} \mathbb{E} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^p. \end{aligned}$$

Then we need to use (23), with $\gamma = d\frac{p}{2}, z = p$, to obtain

$$\begin{aligned} I & \leq C(J+1)^{p-1} \sum_{j \leq J_1} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| > \kappa/2t_n\}} |\beta_{j\eta}|^r \left[\frac{\kappa}{2} t_n \right]^{-r} n^{-p/2} \\ & \quad + C(J+1)^{p-1} \sum_{j \leq J} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \leq 2\kappa t_n\}} |\beta_{j\eta}|^p + n^{-z/2} \\ & \leq 2C(J+1)^{p-1} \sum_{j \leq J_1} 2^{jd(p/2-1)} M^r 2^{-jr\{s+d(1/2-1/r)\}} n^{(r-p)/(2)} \\ & \quad + C(J+1)^{p-1} \sum_{j \geq J_1} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \leq 2\kappa t_n\}} |\beta_{j\eta}|^p + n^{-z/2} \\ & \leq 2C(J+1)^{p-1} 2^{J_1\{d(p/2-r/2)-sr\}} n^{(r-p)/(2)} \\ & \quad + C(J+1)^{p-1} \sum_{j \geq J_1} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \leq 2\kappa t_n\}} |\beta_{j\eta}|^p + n^{-z/2} \end{aligned}$$

as we are in the sparse region. It is easy to realize that now, again because we are in the sparse region

$$2^{J_1} = \frac{\kappa}{2M} \left[\frac{n}{\log n} \right]^{1/(2s+2d(1/2-1/r))},$$

is adequate, and to observe then that the first term in the sum has the right order. For the second term, let us introduce

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

We easily observe that $p - m = \frac{p(s+d/p-d/r)}{s+d/2-d/r} > 0$, and that $m - r = \frac{-sr+(p-r)d/2}{s+d/2-d/r} \geq 0$. Then, as $B_{r,q}^s(M) \subset B_{m,q}^{s-d/r+d/m}(M)$

$$\begin{aligned} & \sum_{J \geq j \geq J_1} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \leq 2\kappa t_n\}} |\beta_{j\eta}|^p \\ & \leq \sum_{J \geq j \geq J_1} 2^{jd(p/2-1)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^m t_n^{p-m} \leq \sum_{J \geq j \geq J_1} M^m t_n^{p-m} \\ & \leq J t_n^{p(s+d/p-d/r)/(s+d/2-d/r)}, \end{aligned}$$

which gives the right order. Observe that the term J (which is of logarithmic order), can be avoided by choosing \tilde{m} instead of m in such a way that $\tilde{m} > m$, but $r < \tilde{m}$. This can be done except for the case where $r = \frac{dp}{2s+d}$ where this logarithmic term is unavoidable.

8. Proof of Proposition 15. The proof of Proposition 15 relies on the following lemma:

LEMMA 16. *There exist constants $\sigma^2 > 0, C, c$, such that, as soon as $2^j \leq \lfloor \frac{n}{\log n} \rfloor^{1/d}$,*

$$(28) \quad \mathbb{P}\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq v\} \leq 2 \exp\left\{-\frac{nv^2}{2(\sigma^2 + vc2^{jd/2}/3)}\right\} \quad \forall v > 0,$$

$$(29) \quad \mathbb{E}|\hat{\beta}_{j\eta} - \beta_{j\eta}|^q \leq s_q n^{-q/2} \quad \forall q \geq 1,$$

$$(30) \quad \mathbb{E} \sup_{\eta} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^q \leq s'_q (j+1)^q n^{-q/2} \quad \forall q \geq 1,$$

$$(31) \quad \mathbb{P}\left(|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \frac{\kappa}{2} t_n\right) \leq C 2n^{-6\kappa} \quad \forall \kappa \geq 6\sigma^2.$$

PROOF. Equation (28) is simply Bernstein’s inequality, noticing that

$$\mathbb{E}(\psi_{j\eta}(X_i))^2 \leq \|f\|_{\infty} \|\psi_{j\eta}\|_2^2 \leq MC =: \sigma^2$$

and $\|\psi_{j\eta}(X_i)\|_{\infty} \leq c2^{jd/2}$. The following inequality directly follows from (28), when $2^{jd} \leq n$:

$$(32) \quad \mathbb{P}\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq v\} \leq 2[e^{-nv^2/(4\sigma^2)} + e^{-3\sqrt{nv}/(4c)}].$$

Equation (29) follows from (32):

$$\begin{aligned} \mathbb{E}|\hat{\beta}_{j\eta} - \beta_{j\eta}|^q &= \int_{\mathbb{R}_+^*} v^{q-1} \mathbb{P}\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq v\} dv \\ &\leq \int_{\mathbb{R}_+^*} v^{q-1} 2[e^{-nv^2/(4\sigma^2)} + e^{-3\sqrt{nv}/(4c)}] dv \leq s_q n^{-q/2}, \end{aligned}$$

using the change of variables $u = \sqrt{nv}$.

Equation (30) also follows from (32): take $a = \max\{\frac{8cd}{3}, 2\sqrt{2}d\sigma\}$

$$\begin{aligned} & \mathbb{E} \sup_{\eta} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^q \\ &= \int_{\mathbb{R}^+} v^{q-1} \mathbb{P}(\sup_{\eta} |\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq v) dv \\ &\leq \int_{0 \leq v \leq aj/\sqrt{n}} v^{q-1} dv \\ &\quad + 2c \int_{v \geq aj/\sqrt{n}} v^{q-1} 2^{jd} [e^{-nv^2/(4\sigma^2)} + e^{-3\sqrt{nv}/(4c)}] dv. \end{aligned}$$

Now, if $v \geq \frac{aj}{\sqrt{n}}$, $2^{jd} e^{-nv^2/(4\sigma^2)} \leq e^{-nv^4/(8\sigma^2) - nv^4/(8\sigma^2) + jd} \leq e^{-nv^4/(8\sigma^2)}$. Similarly $2^{jd} e^{-3\sqrt{nv}/(4c)} \leq e^{-3\sqrt{nv}/(8c)}$, so that

$$\begin{aligned} & \mathbb{E} \sup_{\eta} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^q \\ &\leq \frac{1}{q} \left[\frac{aj}{\sqrt{n}} \right]^q + 2c \int_{v \geq aj/\sqrt{n}} v^{q-1} [e^{-nv^4/(8\sigma^2)} + e^{-3\sqrt{nv}/(8c)}] dv. \quad \square \end{aligned}$$

Let us now turn to the proof of the proposition. We partition our sum in four regions:

$$\begin{aligned} & \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \\ &= \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \{ \mathbf{1}_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}} + \mathbf{1}_{\{|\hat{\beta}_{j\eta}| < \kappa t_n\}} \} \\ &\leq \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| \geq \kappa/2 t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| < \kappa/2 t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| < \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| \geq 2\kappa t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| < \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| < 2\kappa t_n\}} \\ &=: Bb + Bs + Sb + Ss. \end{aligned}$$

We use extensively Lemma 16 in order to bound separately each of the four terms Bb, Ss, Sb, Bs .

Using (30)

$$\begin{aligned}
 Bb &\leq \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z 1_{\{|\beta_{j\eta}| \geq \kappa/2t_n\}} \\
 &\leq \sum_{j=0}^J 1_{\{\exists \eta \in \mathcal{X}_j, |\beta_{j\eta}| \geq \kappa/2t_n\}} 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \\
 &\leq \sum_{j=0}^J 1_{\{\exists \eta \in \mathcal{X}_j, |\beta_{j\eta}| \geq \kappa/2t_n\}} 2^{j\gamma} s'_1(j+1)^z n^{-z/2} \\
 &\leq C 2^{J_1\gamma} (J_1+1)^z n^{-z/2},
 \end{aligned}$$

where J_1 is chosen such that for $j \geq J_1, |\beta_{j\eta}| \leq \kappa/2t_n$. Also

$$\begin{aligned}
 Ss &\leq \sum_{j=0}^J 2^{j\gamma} \sup_{\eta \in \mathcal{X}_j} |\beta_{j\eta}|^z 1_{\{|\beta_{j\eta}| < 2\kappa t_n\}} \\
 &\leq \sum_{j=0}^{J_1} 2^{j\gamma} [2\kappa t_n]^z + \sum_{j=J_1+1}^J 2^{j\gamma} \sup_{\eta \in \mathcal{X}_j} |\beta_{j\eta}|^z
 \end{aligned}$$

which gives the proper rate of convergence. Moreover, using (30) and (31),

$$\begin{aligned}
 Bs &\leq \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z 1_{\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \kappa/2t_n\}} 1_{\{|\beta_{j\eta}| < \kappa/2t_n\}} \\
 &\leq \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z 1_{\{\exists \eta \in \mathcal{X}_j, |\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \kappa/2t_n\}} \\
 &\leq \sum_{j=0}^J 2^{j\gamma} \left[\mathbb{E} \sup_{\eta \in \mathcal{X}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^{2z} \right]^{1/2} \\
 &\quad \times \mathbb{P} \left\{ \exists \eta \in \mathcal{X}_j, |\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \frac{\kappa}{2} t_n \right\}^{1/2} \\
 &\leq \sum_{j=0}^J 2^{j\gamma} [s'_2(j+1)^{2z} n^{-z}]^{1/2} [c 2^{jd} n^{-6\kappa}]^{1/2} \leq n^{-z/2},
 \end{aligned}$$

where $\kappa > \frac{1}{3}(\frac{\gamma}{d} + \frac{1}{2})$. Finally, using (31), and the fact that for f bounded, $|\beta_{j\eta}| \leq C2^{-jd/2}$

$$\begin{aligned} Sb &\leq \sum_{j=0}^J 2^{j\gamma} \mathbb{E} \sup_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z \mathbf{1}_{\{|\beta_{j\eta} - \hat{\beta}_{j\eta}| \geq \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| \geq 2\kappa t_n\}} \\ &\leq \sum_{j=0}^J 2^{j\gamma} M 2^{-jzd/2} \mathbb{P}\{\exists \eta \in \mathcal{Z}_j, |\beta_{j\eta} - \hat{\beta}_{j\eta}| \geq \kappa t_n\} \\ &\leq \sum_{j=0}^J [c 2^{j[d(1-z/2)+\gamma]} n^{-6\kappa}] \leq C [2^{J[d(1-z/2)+\gamma]} n^{-6\kappa}] \leq n^{-z/2}, \end{aligned}$$

for $\kappa > \frac{1}{6}(\frac{\gamma}{d} + 1)$.

8.1. *Proof of (23) and (24).* This proof follows along the lines of the previous one. Equation (24) is a consequence of (23), and the two inequalities will be proved together. We again separate the four cases:

$$\begin{aligned} &\sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |t(\hat{\beta}_{j\eta}) - \beta_{j\eta}|^z \\ &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| \geq \kappa/2t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| \geq \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| < \kappa/2t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| < \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| \geq 2\kappa t_n\}} \\ &\quad + \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z \mathbf{1}_{\{|\hat{\beta}_{j\eta}| < \kappa t_n\}} \mathbf{1}_{\{|\beta_{j\eta}| < 2\kappa t_n\}} \\ &=: Bb + Bs + Sb + Ss. \end{aligned}$$

Let us now bound separately each of the four terms Bb, Ss, Sb, Bs . Using (29),

$$\begin{aligned} Bb &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \mathbf{1}_{\{|\beta_{j\eta}| \geq \kappa/2t_n\}} \\ &\leq \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} \mathbf{1}_{\{|\beta_{j\eta}| \geq \kappa/2t_n\}} 2^{j(\gamma-d)} \mathbb{E} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z \end{aligned}$$

$$\begin{aligned} &\leq \sum_{j=0}^J \sum_{\eta \in \mathcal{Z}_j} 1_{\{|\beta_{j\eta}| \geq \kappa/2t_n\}} 2^{j(\gamma-d)} s_z n^{-z/2} \quad (*) \\ &\leq C 2^{J_1 \gamma} n^{-z/2}, \end{aligned}$$

where J_1 again is chosen such that for $j \geq J_1$, $|\beta_{j\eta}| \leq \kappa/2t_n$. To prove (23), we stop in (*), the next bound yields (24):

$$\begin{aligned} Ss &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z 1_{\{|\beta_{j\eta}| < 2\kappa t_n\}} \quad (*) \\ &\leq \sum_{j=0}^{J_1} 2^{j(\gamma)} [2\kappa t_n]^z + \sum_{j=J_1+1}^J 2^{j(\gamma-d)} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z \end{aligned}$$

which gives the proper rate of convergence. Again, to prove (23), we stop in (*), the next bound yields (24). Moreover, using (30) and (31),

$$\begin{aligned} Bs &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z 1_{\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \kappa/2t_n\}} 1_{\{|\beta_{j\eta}| < \kappa/2t_n\}} \\ &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^z 1_{\{|\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \kappa/2t_n\}} \\ &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \sum_{\eta \in \mathcal{Z}_j} [\mathbb{E} |\hat{\beta}_{j\eta} - \beta_{j\eta}|^{2z}]^{1/2} \mathbb{P} \left\{ |\hat{\beta}_{j\eta} - \beta_{j\eta}| \geq \frac{\kappa}{2} t_n \right\}^{1/2} \\ &\leq \sum_{j=0}^J 2^{j(\gamma)} [s_{2z} n^{-z}]^{1/2} [c 2^{jd} n^{-6\kappa}]^{1/2} \leq C n^{-z/2} \end{aligned}$$

for $\kappa > \frac{\gamma}{6d}$. Finally, using (31), and the fact that for f bounded, $|\beta_{j\eta}| \leq C 2^{-jd/2}$,

$$\begin{aligned} Sb &\leq \sum_{j=0}^J 2^{j(\gamma-d)} \mathbb{E} \sum_{\eta \in \mathcal{Z}_j} |\beta_{j\eta}|^z 1_{\{|\beta_{j\eta} - \hat{\beta}_{j\eta}| \geq \kappa t_n\}} 1_{\{|\beta_{j\eta}| \geq 2\kappa t_n\}} \\ &\leq \sum_{j=0}^J 2^{j(\gamma-d)} M 2^{-jzd/2} \mathbb{P} \{ |\beta_{j\eta} - \hat{\beta}_{j\eta}| \geq \kappa t_n \} \\ &\leq \sum_{j=0}^J [c 2^{j[-dz/2+\gamma]} n^{-6\kappa}] \leq C [2^{J[-dz/2+\gamma]} n^{-6\kappa}] \leq C n^{1/2}, \end{aligned}$$

for $\kappa > 0$.

9. Proof of the lower bound. Let us recall that given two probabilities P, Q on some measure space their Kullback–Leibler distance is

$$K(P, Q) = \begin{cases} \int \log \frac{dP}{dQ} dP = \int \frac{dP}{dQ} \log \frac{dP}{dQ} dQ, & \text{if } P \ll Q, \\ +\infty, & \text{otherwise.} \end{cases}$$

If P, Q are probabilities on \mathbb{S}^d having densities f, g respectively with respect to Lebesgue measure, then if g is bounded below by some constant $c > 0$

$$\begin{aligned} (33) \quad K(P, Q) &= \int \log \frac{f}{g} f dx = \int \log \left(\frac{f-g}{g} + 1 \right) f dx \\ &\leq \int \frac{f-g}{g} f dx = \int \frac{(f-g)^2}{g} dx \leq \frac{1}{c} \|f-g\|_2^2. \end{aligned}$$

We make use of Fano’s lemma below; see [23] and the references therein. We use the point of view introduced in [3].

THEOREM 17 (Fano’s lemma). *Let \mathcal{F} be a sigma algebra on the space Ω . Let $F_i \in \mathcal{F}, i \in \{0, 1, \dots, m\}$ such that $\forall i \neq j, F_i \cap F_j = \emptyset$. Let $P_i, i = 0, \dots, m$ be probability measures on (Ω, \mathcal{A}) . If*

$$(34) \quad p \stackrel{\text{def}}{=} \sup_{i=0, \dots, m} P_i(F_i^c),$$

$$(35) \quad \kappa(P_0, \dots, P_m) \stackrel{\text{def}}{=} \inf_{j=0, \dots, m} \frac{1}{m} \sum_{i \neq j} K(P_i, P_j)$$

then

$$(36) \quad p \geq \frac{1}{2} \wedge C(\sqrt{m}e^{-\kappa(P_0, \dots, P_m)}), \quad C = e^{3/e}.$$

We prove first that the minimax L^p -loss is $\gtrsim n^{-\alpha p}$ with $\alpha = \frac{s}{2s+d}$. For every j let us consider the family \mathcal{A}_j of densities

$$(37) \quad f_\varepsilon = \frac{1}{|\mathbb{S}^d|} + \gamma \sum_{\xi \in A_j} \varepsilon_\xi \psi_{j,\xi},$$

where A_j is a subset of \mathcal{X}_j to be made precise later, $\varepsilon_\xi \in \{0, 1\}$ and γ is chosen so that all these functions are positive. We are going to show that for every estimator \hat{f} ,

$$\sup_{f_\varepsilon} \mathbb{E}_{f_\varepsilon} \|\hat{f} - f_\varepsilon\|_p^p \geq cn^{-sp/(2s+d)}.$$

Throughout this section we shall note $x \lesssim y, x \gtrsim y$ whenever it holds $x \leq cy$ or $x \geq cy$ respectively, c being a strictly positive constant independent of j, ξ .

We shall note $x \simeq y$ whenever both $x \lesssim y$ and $x \gtrsim y$ hold. Thanks to (12) for these functions to be positive it is enough that $|\gamma| \lesssim 2^{-jd/2}$. Such a γ can even be chosen in such a way that all the densities (37) are bounded from below by a strictly positive constant. If the functions $(\psi_{j\xi})_{\xi \in \mathcal{Z}_j}$ were orthonormal we would have immediately that

$$\left\| \sum_{\xi \in \mathcal{Z}_j} \lambda_\xi \psi_{j,\xi} \right\|_p \geq c \left(\sum_{\xi \in \mathcal{Z}_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}.$$

Needlets are not a basis, but their scalar product is close enough to 0 if the respective cubature points are far enough. Hence one can get the following lemma that states that a subset $A_j \subset \mathcal{Z}_j$ can be chosen so that it is quite large and inequalities (14) and (13), in a sense, can be reversed.

LEMMA 18. *There exists a subset $A_j \subset \mathcal{Z}_j$ such that $\text{card } A_j \gtrsim 2^{jd}$ and*

$$\left\| \sum_{\xi \in A_j} \lambda_\xi \psi_{j,\xi} \right\|_p \geq \begin{cases} c \sup_{\xi \in A_j} |\lambda_\xi| \|\psi_{j,\xi}\|_\infty, & \text{if } p = \infty, \\ c \left(\sum_{\xi \in A_j} |\lambda_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}, & \text{if } p < +\infty. \end{cases}$$

Let us now impose conditions that ensure that f_ε belongs to the ball $B_{r,q}^s(M)$. Now, recalling (15),

$$\|f_\varepsilon\|_{B_{r,q}^s} = |\gamma| 2^{j(s+d(1/2-1/r))} \left(\sum_{\xi \in \mathcal{Z}_j} |\varepsilon_\xi|^r \right)^{1/r} \lesssim |\gamma| 2^{j(s+d(1/2-1/r))} 2^{jd/r},$$

where we use the fact that $|\varepsilon_\xi| = 1$. Therefore the condition $\|f_\varepsilon\|_{B_{r,q}^s} \leq M$ follows from

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

In order to apply Fano’s lemma and get a lower bound of the left-hand side let us first get an upper bound for the Kullback–Leibler distances $K(f_\varepsilon; f_{\varepsilon'})$, which comes from (33) and (12) for $p = 2$,

$$(38) \quad \|f_\varepsilon - f_{\varepsilon'}\|_2^2 \leq \gamma^2 \sum_{\xi \in A_j} |\varepsilon_\xi - \varepsilon'_\xi|^2 < \gamma^2 2^{jd} \leq 2^{-2js}.$$

By Lemma 18

$$\|f_\varepsilon - f_{\varepsilon'}\|_p \geq \left(\sum_{\xi \in A_j} |\varepsilon_\xi - \varepsilon'_\xi|^p \|\psi_{j,\xi}\|_p^p \right)^{1/p}.$$

Thanks to Lemma 18, the set of functions \mathcal{A}_j has a cardinality that is $\geq 2^{c2^{jd}}$. By the Varshanov–Gilbert lemma (e.g., [23]) there exists a subset $\mathcal{A}'_j \subset \mathcal{A}_j$ such

that card $\mathcal{A}'_j \geq 2^{c'2^{jd}}$ and such that if $f_\varepsilon, f'_{\varepsilon'} \in \mathcal{A}'_j$, then $\sum_{\xi \in A'_j} |\varepsilon_\xi - \varepsilon'_{\xi}| > \frac{1}{4}2^{jd}$. Therefore, as $|\varepsilon_\xi - \varepsilon'_{\xi}|$ can be = 0 or = 1 only and by (12),

$$\|f_\varepsilon - f_{\varepsilon'}\|_p \gtrsim |\gamma|2^{jd(1/2-1/p)} \left(\frac{2^{jd}}{4}\right)^{1/p} \simeq 2^{-js}$$

which implies that the events $\{\|\widehat{f} - f_{\varepsilon'}\|_p \geq \frac{1}{2}2^{-js}\}$ are disjoint. The family of densities $f_\varepsilon \in \mathcal{A}'_j$ given by the Varsharov–Gilbert lemma has cardinality $m \simeq 2^{c'2^{jd}}$ and by (39) and (38)

$$K(f_\varepsilon, f_{\varepsilon'}) \lesssim \|f_\varepsilon - f_{\varepsilon'}\|_2^2 \lesssim 2^{-2js}.$$

We apply now Fano’s lemma to the probabilities P_ε that are the n times product of $f_\varepsilon dx$ and to the events $A_\varepsilon = \{\|\widehat{f} - f_{\varepsilon'}\|_p \geq \frac{1}{2}2^{-js}\}$. It is well known that

$$K(P_\varepsilon; P_{\varepsilon'}) = nK(f_\varepsilon; f_{\varepsilon'}).$$

By Markov inequality and Fano’s lemma,

$$\begin{aligned} \sup_{f_\varepsilon \in \mathcal{A}'_j} \mathbb{E}\|\widehat{f} - f_\varepsilon\|_p^p &\geq 2^{-p}2^{-jsp} \sup_{i \leq m} P_\varepsilon(\|\widehat{f} - f_\varepsilon\|_p > \delta) \\ &\gtrsim 2^{-jsp} \left(\frac{1}{2} \wedge e^{-n2^{-2js}} \underbrace{\sqrt{\#\mathcal{A}'_j}}_{\simeq 2^{c2^{jd}}}\right). \end{aligned}$$

Now let j be so that $n2^{-2js} \simeq 2^{jd}$, that is $2j \simeq n^{1/(2s+d)}$. With this choice one has

$$\frac{1}{2} \wedge (e^{-n2^{-2js}} e^{c2^{jd}}) \geq c > 0.$$

Therefore

$$\sup_{f_\varepsilon \in \mathcal{A}'_j} \mathbb{E}\|\widehat{f} - f_\varepsilon\|_p^p \gtrsim c2^{-jsp} \sim n^{-sp/(2s+d)}.$$

We prove now that

$$\text{the minimax } L^p\text{-loss is } \gtrsim n^{-p(s+d(1/p-1/r))/(2(s+d(1/2-1/r)))}.$$

Let us consider the two densities

$$f_0 = \frac{1}{|\mathbb{S}^d|} + \gamma\psi_{j,\xi}, \quad f_1 = \frac{1}{|\mathbb{S}^d|} + \gamma\psi_{j,\xi'}$$

with γ such that the above are positive ($|\gamma| \lesssim 2^{-jd/2}$ is enough). If $|\gamma| \leq 2^{-js}2^{-jd(1/2-1/r)}M$, then thanks to (15) both f_0 and f_1 belong to the ball $B_{r,q}^s(M)$. Remark that this condition implies $|\gamma| \lesssim 2^{-jd/2}$, as we assume $s \geq \frac{d}{r}$. Also

$$K(f_0 dx, f_1 dx) \leq \|f_0 - f_1\|_2^2 \approx \gamma^2$$

so that, if we denote by P_0, P_1 the n -times product of $f_0 dx$ and $f_1 dx$ by itself respectively, $K(P_0, P_1) \approx n\gamma^2$. By (13) and Lemma 18 we have

$$(39) \quad \begin{aligned} \|f_0 - f_1\|_p &= |\gamma| \|\psi_{j,\xi} - \psi_{j,\xi'}\|_p \geq |\gamma| 2^{jd(1/2-1/p)} \\ &\sim 2^{-j(s+d(1/2-1/r))} 2^{jd(1/2-1/p)} = r = 2^{-j(s+d(1/p-1/r))}. \end{aligned}$$

We choose $\gamma = \frac{1}{\sqrt{n}} = 2^{-j(s+d(1/2-1/r))}$, so that $K(P_0, P_1) \approx n$. Moreover with this choice of $n, j \approx \log n((2(s + d(\frac{1}{2} - \frac{1}{r})))^{-1})$, so that again by Fano’s lemma,

$$\sup_{i=1,2} \mathbb{E} \|\hat{f} - f_i\|_p^p \geq \delta^p \sup_{i=1,2} P_i(\|\hat{f} - f_i\|_p \geq \delta).$$

Thanks to (39) the events $\{\|\hat{f} - f_i\|_p \geq \delta\}$ are disjoint if $\delta \lesssim 2^{-j(s+d(1/p-1/r))}$. Therefore by Fano’s lemma,

$$\sup_{i=1,2} \mathbb{E} \|\hat{f} - f_i\|_p^p \gtrsim \delta^p \gtrsim 2^{-j(s+d(1/p-1/r))} = n^{-p(s+d(1/p-1/r))/(2(s+d(1/2-1/r)))}.$$

We have therefore proved that $\sup_{f \in B_{r,q}^s(M)} \min_{\hat{f}} \mathbb{E}_f(\|\hat{f} - f\|_p^p)$ is

$$\gtrsim n^{-sp/(2s+d)} \quad \text{and} \quad \gtrsim n^{-p(s+d(1/p-1/r))/(2(s+d(1/2-1/r)))}.$$

Putting things together and checking for which values of the parameters one rate is larger than the other one concludes the proof of Theorem 11. Note that, as $s > \frac{d}{r}$, if $1 \leq p \leq 2$

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

REFERENCES

[1] BALDI, P., KERKYACHARIAN, G., MARINUCCI, D. and PICARD, D. (2009). Asymptotics for spherical needlets. *Ann. Statist.* **37** 1150–1171.
 [2] BALDI, P., KERKYACHARIAN, G., MARINUCCI, D. and PICARD, D. (2009). Subsampling needlet coefficients on the sphere. *Bernoulli* **15** 438–463.
 [3] BIRGÉ, L. (2001). A new look at an old result: Fano’s lemma. *LPMA*. To appear. Available at www.proba.jussieu.fr/mathdoc/preprints/index.html#2001.
 [4] THE PIERRE AUGER COLLABORATION (2008). Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei. *Astroparticle Physics* **29** 188. doi:10.1016/j.astropartphys.2008.01.002.
 [5] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974
 [6] HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843
 [7] HEALY, D. M., JR and KIM, P. T. (1996). An empirical Bayes approach to directional data and efficient computation on the sphere. *Ann. Statist.* **24** 232–254. MR1389889
 [8] HENDRIKS, H. (2003). Application of fast spherical Fourier transform to density estimation. *J. Multivariate Anal.* **84** 209–221. MR1965218

- [9] HENDRIKS, H., JANSSEN, J. H. M. and RUYMGAART, F. H. (1993). Strong uniform convergence of density estimators on compact Euclidean manifolds. *Statist. Probab. Lett.* **16** 305–311. [MR1212054](#)
- [10] JUPP, P. E. (1995). Some applications of directional statistics to astronomy. In *New Trends in Probability and Statistics, Vol. 3 (Tartu/Pühajärve, 1994)* 123–133. VSP, Utrecht. [MR1655007](#)
- [11] JUPP, P. E. (2008). Data-driven Sobolev tests of uniformity on compact Riemannian manifolds. *Ann. Statist.* **36** 1246–1260. [MR2418656](#)
- [12] KERKYACHARIAN, G., PETRUSHEV, P., PICARD, D. and WILLER, T. (2007). Needlelet algorithms for estimation in inverse problems. *Electron. J. Stat.* **1** 30–76 (electronic). [MR2312145](#)
- [13] KLEMELÄ, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.* **73** 18–40. [MR1766119](#)
- [14] KLEMELÄ, J. (2003). Lower bounds for the asymptotic minimax risk with spherical data. *J. Statist. Plann. Inference* **113** 113–136. [MR1963037](#)
- [15] KLEMELÄ, J. (1999). Asymptotic minimax risk for the white noise model on the sphere. *Scand. J. Statist.* **26** 465–473. [MR1712031](#)
- [16] MARDIA, K. V. (1972). *Statistics of Directional Data. Probability and Mathematical Statistics* **13**. Academic Press, London. [MR0336854](#)
- [17] MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Chichester. [MR1828667](#)
- [18] NARCOWICH, F., PETRUSHEV, P. and WARD, J. (2006). Local tight frames on spheres. *SIAM J. Math. Anal.* **38** 574–594. [MR2237162](#)
- [19] NARCOWICH, F. J., PETRUSHEV, P. and WARD, J. M. (2006). Decomposition of Besov and Triebel–Lizorkin spaces on the sphere. *J. Funct. Anal.* **238** 530–564. [MR2253732](#)
- [20] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. [MR2179289](#)
- [21] RUYMGAART, F. H. (1989). Strong uniform convergence of density estimators on spheres. *J. Statist. Plann. Inference* **23** 45–52. [MR1029239](#)
- [22] STEIN, E. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidian Spaces*. Princeton Univ. Press, Princeton. [MR0304972](#)
- [23] TSYBAKOV, A. B. (2004). *Introduction à l'estimation non-paramétrique. Mathématiques & Applications (Berlin) [Mathematics & Applications]* **41**. Springer, Berlin. [MR2013911](#)
- [24] VARSHALOVICH, D. A., MOSKALEV, A. N. and KHERSONSKIĬ, V. K. (1988). *Quantum Theory of Angular Momentum*. World Scientific, Teaneck, NJ. [MR1022665](#)

P. BALDI
 D. MARINUCCI
 DIPARTIMENTO DI MATEMATICA
 UNIVERSITÀ DI ROMA TOR VERGATA
 VIA DELLA RICERCA SCIENTIFICA
 00161 ROMA
 ITALY
 E-MAIL: baldi@mat.uniroma2.it
 marinucci@mat.uniroma2.it

G. KERKYACHARIAN
 LABORATOIRE DE PROBABILITÉS ET
 MODÈLES ALÉATOIRES
 2, PL. JUSSIEU
 75251 PARIS CEDEX 05
 FRANCE
 E-MAIL: kerk@math.jussieu.fr

D. PICARD
 UNIVERSITÉ PARIS-DIDEROT
 CNRS-LPMA
 175 RUE DU CHEVALERET
 75013 PARIS
 FRANCE
 E-MAIL: picard@math.jussieu.fr