Probab. Theory Relat. Fields (2012) 153:363–404
DOI 10.1007/s00440-011-0348-5

Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds

Gerard Kerkyacharian · Richard Nickl · Dominique Picard

Received: 8 November 2010 / Revised: 11 February 2011 / Published online: 18 March 2011

Abstract Let $X_1, \ldots, X_n$ be a random sample from some unknown probability density f defined on a compact homogeneous manifold $\mathbf{M}$ of dimension $d \geq 1$. Consider a 'needlet frame' $\{\phi_{j\eta}\}$ describing a localised projection onto the space of eigenfunctions of the Laplace operator on $\mathbf{M}$ with corresponding eigenvalues less than 2^{2j}, as constructed in Geller and Pesenson (J Geom Anal 2011). We prove non-asymptotic concentration inequalities for the uniform deviations of the linear needlet density estimator $f_n(j)$ obtained from an empirical estimate of the needlet projection $\sum_\eta \phi_{j\eta} \int f \phi_{j\eta}$ of f. We apply these results to construct risk-adaptive estimators and nonasymptotic confidence bands for the unknown density f. The confidence bands are adaptive over classes of differentiable and Hölder-continuous functions on $\mathbf{M}$ that attain their Hölder exponents.

Mathematics Subject Classification (2000) 62G07 · 60E15 · 42C40

G. Kerkyacharian · D. Picard
Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris-Diderot,
175, Rue de Chevaleret, 75013 Paris, France
e-mail: kerk@math.jussieu.fr

D. Picard
e-mail: picard@math.jussieu.fr

R. Nickl (✉)
Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge,
Cambridge CB30WB, UK
e-mail: r.nickl@statslab.cam.ac.uk

1 Introduction

We consider the problem of constructing confidence bands for an unknown probability density f based on a sample $X_1, \ldots, X_n$ from f observed on the d-dimensional compact homogeneous manifold $\mathbf{M}$.The classical statistical applications occur when $\mathbf{M}$ equals the d-dimensional unit sphere $\mathbb{S}^d$ of $\mathbb{R}^{d+1}$: If $d = 1$ this corresponds to estimating a periodic univariate density, and recent interest lies mostly in the case $d = 2$, strongly motivated by statistical problems in astrophysics, see Baldi et al. [2] for an account of typical problems and applications in astrophysics and directional statistics more generally. In Baldi et al. [2] a recent construction of wavelet type bases on $\mathbb{S}^d$—due to Narcowich et al. [24,25], who called these new basis functions *needlets*—was employed to construct risk-adaptive estimators for $f(x), x \in \mathbb{S}^d$, by a local needlet series with support concentrated in a neighborhood of x. See also Kerkyacharian et al. [16] for similar results in the spherical deconvolution problem. The main advantages of this approach are that they share none of the drawbacks of classical approaches: kernel methods do not take the manifold structure of the sphere well into account, orthogonal series methods associated with spherical harmonics have very poor pointwise (and even worse uniform) performance since spherical harmonics are not well localized but spread out all over the sphere, and methods based on stereographic projections of the sphere onto the plane use a distorted approximation-theoretic paradigm. In contrast needlets are a tight frame constructed on the spherical harmonics which are highly localized and allow for optimal approximation not only in L^2 but in general L^p-spaces, including in particular L^∞, which is particularly relevant in the problem of constructing confidence bands for f. Moreover the localization property is of crucial importance since in astrophysical data sets some parts of the sphere (sky) may not be covered by the observations, so that non-local procedures may suffer severely from missing data points.

The main contributions of the present article are three-fold. First, building on recent results on wavelets and approximation of functions on manifolds in [7,8], we show how needlet estimators $f_n(j, y), y \in \mathbf{M}$, with resolution level $j \geq 0$, can be defined also on the more general class of compact homogeneous differentiable manifolds $\mathbf{M}$, which includes, next to d-dimensional unit spheres, also other relevant examples such as real and complex projective spaces, or Grassmann and Stiefel manifolds. The main idea behind this construction is to use tools from harmonic analysis on compact Lie groups that allow to build a localized frame on the eigenfunctions of a second order elliptic differential Laplace operator on $\mathbf{M}$, which in the case of the sphere coincides with the construction of [24,25], where these eigenfunctions are precisely the spherical harmonics.

The second goal of this article is to prove non-asymptotic concentration inequalities for the uniform fluctuations

$$\sup_{y \in \mathbf{M}} |f_n(j, y) - Ef_n(j, y)|$$

of needlet estimators $f_n(j)$ around the needlet projections $Ef_n(j) = A_j(f)$ of the unknown density f. The constants in these concentration inequalities depend in a natural way on the manifold and we derive reasonably tight constants for the case $\mathbf{M} = \mathbb{S}^d, d \geq 1$. We present both Bernstein-type bounds and inequalities based on Rademacher-symmetrization in a similar vein as in recent work in [12,18,21].

The third goal is to use the above concentration inequalities to construct estimators and confidence bands for the unknown density $f : \mathbf{M} \to \mathbb{R}$. Even the problem of spherical confidence bands seems not to have been addressed in the literature so far—one reason may arise from the fact that the classical approach in the univariate case [3] via extreme value theory does not straightforwardly generalise to sample spaces with a different geometric structure. Our concentration inequalities hold on arbitrary compact homogeneous manifolds and can be used directly to construct estimators and nonasymptotic confidence bands for the unknown density f if one has a priori control of the approximation error of f by its needlet projection $A_j(f)$ (the 'bias' of estimation), which by results in Geller and Pesenson [8] is equivalent to classical Hölderian smoothness conditions for f on $\mathbf{M}$.

Since knowledge of the bias is usually not available, the question of how to choose j comes into sight, and to which extent *adaptive* estimators and confidence bands can be constructed. It is known on the one hand [22] that adaptive and honest confidence bands in nonparametric function estimation problems cannot exist over the entirety of the usual smoothness classes (in our case, Hölder-balls on $\mathbf{M}$). Recent work in this field, however, can be interpreted as a new way of looking at this problem: One can devise statistically relevant subsets of the usual smoothness function classes for which adaptive confidence bands *do exist*. One example comes from shape constrained nonparametric regression, see, e.g., [4]. Other examples are 'self-similar functions' that attain their Hölder exponent—see Picard and Tribouley [26] in the case of the Gaussian white noise model and regression framework and Giné and Nickl [11] in density estimation on the real line. Moreover, building on Jaffard's [15] work on the Frisch–Parisi conjecture [6], Giné and Nickl [11] proved that 'generic' subsets (in the Baire-sense) of the class of Hölder balls can be constructed for which asymptotically honest adaptive confidence bands exist.

In the present paper we follow the line of Picard and Tribouley [26] and Giné and Nickl [11], but take a nonasymptotic approach. We propose an adaptive procedure $\hat{j}_n$ based on Lepski's method [20] to choose the resolution level j for the needlet estimator $f_n(j)$ in a data-driven way. The resulting estimator $f_n(\hat{j}_n)$ adapts to the unknown smoothness of f in sup-norm risk. In our main result we devise an analytic condition on the approximation errors of f by its needlet projections $A_j(f)$ under which we can establish both an asymptotic and a nonasymptotic coverage result for confidence bands for f over arbitrary subsets Ω of $\mathbf{M}$ that are centered at $f_n(\hat{j}_n)$, and we show that this band adapts to the unknown smoothness of f in the minimax sense. Intuitively the results in Giné and Nickl [11] suggest that adaptation is possible for functions $f : \mathbf{M} \to \mathbb{R}$ that attain their Hölder exponent, and indeed we prove that our analytic condition can be interpreted in terms of classical Hölder regularity properties of f. The proof of this result is somewhat delicate and we detail it only in the case $\mathbb{S}^d$, where the representation of the projector onto spherical harmonics in terms of Gegenbauer polynomials allows for explicit derivations.

Let us finally remark that even in the univariate case $\mathbb{S}^1$ our nonasymptotic approach to confidence bands gives an alternative to the more classical asymptotic techniques based on extreme value theory, as initiated in the classical paper [3], and as also used in the adaptive context in Giné and Nickl [11]. Not surprisingly the nonasymptotic approach has limitations, but in contrast to the classical asymptotic theory referred to above, the present results give precise conditions for what is necessary to obtain coverage in finite samples.

2 Compact homogeneous manifolds and needlets

We summarize here some facts on compact homogeneous manifolds and Lie groups (see [5,13,14,30] for general references), and the construction and essential properties of the associated needlet frame due to [7,8], generalising the spherical case considered in Narcowich et al. [24].

2.1 Compact Lie groups and the Laplace operator

Let $\mathbf{M}$ be a compact connected differentiable (C^∞-) manifold of dimension $dim(\mathbf{M}) = d$. A Lie group G of dimension τ is said to act on $\mathbf{M}$ via

$$(g, x) \in G \times \mathbf{M} \mapsto g.x \in \mathbf{M}$$

if (a) this action is, for every $g \in G$, a diffeomorphism of $\mathbf{M}$, if (b) $g_1g_2.x = g_1.(g_2.x)$ holds for every $g_1, g_2 \in G$, $x \in \mathbf{M}$, if (c) the identity $e \in G$ satisfies $e.x = x$ and if (d) for every $g \in G$, $g \neq e$, there exists a point $x \in \mathbf{M}$ such that $g.x \neq x$. A group G acts *transitively* on $\mathbf{M}$ if in addition

$$\text{for every } x, y \in \mathbf{M} \;\; \text{there exists } g \in G \text{ s.t. } g.x = y.$$

A compact manifold $\mathbf{M}$ is said to be *homogeneous* if it is a compact connected differentiable manifold on which a compact Lie group acts transitively. Examples include the d-dimensional unit sphere $\mathbb{S}^d$ of $\mathbb{R}^{d+1}$, projective spaces, Stiefel and Grassmann manifolds, see page 125 in Warner [30] and also Wang [29] for the two-point homogeneous case.

Any compact homogeneous manifold $\mathbf{M}$ can be realised as a quotient G/K where K is a closed subgroup of G. More precisely, if we fix once and for all a point $x_0 \in \mathbf{M}$, and let $K = \{h \in G, h.x_0 = x_0\}$ be the closed isotropy subgroup at x_0, then $\mathbf{M}$ is diffeomorphic to G/K and the canonical projection $\pi : g \in G \mapsto \overline{g} = \{gh, h \in K\} \in G/K$ is continuous, onto and verifies $\pi(g_1g_2) = g_1\pi(g_2)$, see Warner [30], p. 123 onwards. Moreover the image of the Haar measure on G under π,

$$\int_G f(\pi(g))dg = \int_{G/K} f(x)dx = \int_{\mathbf{M}} f(x)dx,$$

is a natural "Haar" measure dx on $\mathbf{M}$, invariant under the action of G. (It is the unique G-invariant measure on $\mathbf{M}$ up to a scaling factor.) The usual Lebesgue spaces on $\mathbf{M}$

are denoted by $L^p(\mathbf{M}) := L^p(\mathbf{M}, dx)$, $1 \le p \le \infty$. Since G is compact, dx is bi-invariant: for $f \in L^1(\mathbf{M})$ and $g \in G$ let us define $L_g(f)(x) = f(g^{-1}x)$, $R_g(f)(x) = f(xg)$, then

$$\int_{\mathbf{M}} L_g(f)(x)dx = \int_{\mathbf{M}} f(x)dx = \int_{\mathbf{M}} R_g(f)(x)dx.$$

The Lie algebra $Lie(G)$ of G is characterized by the fact that

$$X \in Lie(G) \mapsto e^X \in G,$$

and since G is compact, this mapping is onto. Let us recall that we have the *Ad* representation of G in $Lie(G)$:

$$g \in G \mapsto Ad(g)X \equiv gXg^{-1} \in Lie(G), \quad \text{and} \quad ge^Xg^{-1} = e^{Ad(g)X},$$

and there exists an Euclidean structure $\langle \cdot, \cdot \rangle$ on $Lie(G)$ for which Ad is unitary, that is, such that

$$\forall g \in G, \quad \forall X \in Lie(G), \quad \langle Ad(g)X, Ad(g)Y\rangle = \langle X, Y\rangle, \quad |X|^2 = \langle X, X\rangle, \tag{1}$$

see Proposition 6.1.1 in [5].

Every $X \in Lie(G)$ generates a vector field on G so that we can define a one parameter group

$$t \mapsto e^{tX} \in G, t \in \mathbb{R},$$

and we can define a metric on G by the 'length' $|X|$ of the 'shortest geodesic' joining two points $g_1, g_2 \in G$,

$$d_G(g_1, g_2) = \inf\{|X|,\ e^X g_1 = g_2\} = \inf\{|X|,\ g_1 e^X = g_2\}. \tag{2}$$

The two previous definitions are equivalent, as:

$$e^X g_1 = g_2 \iff g_1 g_1^{-1} e^X g_1 = g_2 \iff g_1 e^{Ad(g_1^{-1})X} = g_2,\ |Ad(g_1^{-1})X| = |X|$$

and it is not difficult to verify that this metric is bi-invariant:

$$\forall g_1, g_2, g \in G,\ d_G(g_1, g_2) = d_G(gg_1, gg_2) = d_G(g_1 g, g_2 g).$$

Every $X \in Lie(G)$ also naturally generates a one parameter group on $\mathbf{M}$:

$$t \in \mathbb{R} \mapsto e^{tX}.x \in \mathbf{M}$$

which describes geodesics of the Riemannian structure on $\mathbf{M}$ associated to the Euclidean structure $\langle \cdot, \cdot \rangle$ on $Lie(G)$. The metric on $\mathbf{M}$ is given by

$$\begin{aligned} d_{\mathbf{M}}(x, y) &= \inf\{|X|,\ e^X.x = y\} = d_{G/K}(x, y) \\ &= \inf\{d_G(g_1, g_2),\ \pi(g_1) = x, \pi(g_2) = y\} \end{aligned}$$

So $d_{\mathbf{M}}(\pi(g), \pi(g')) \leq d_G(g, g')$. Moreover

$$\forall g \in G,\ x, y \in \mathbf{M},\ d_{\mathbf{M}}(g.x, g.y) = d_{\mathbf{M}}(x, y).$$

This is again due to (1) as

$$e^X.x = y \Longleftrightarrow g.e^X g^{-1}.g.x = g.y \Longleftrightarrow e^{Ad(g)X} g.x = g.y, \quad \text{and} \quad |X| = |Ad(g)X|.$$

Now similarly every $X \in Lie(G)$ gives rise to a one-parameter group on $L^p(\mathbf{M})$, $1 \leq p < \infty$, given by

$$f \mapsto T_t(f)(x) = f(e^{tX}.x);\ t \in \mathbb{R}, x \in \mathbf{M}, f \in L^p(\mathbf{M})$$

and we denote the infinitesimal generator of this one-parameter group by D_X, so

$$D_X f(x) = \frac{d}{dt} f(e^{tX}.x)|_{t=0}, \quad x \in \mathbf{M},$$

the derivative of f at x in the direction of the X-geodesic.

If $X_i, i = 1, \dots, \tau$, is an orthonormal basis of $Lie(G)$ with respect to the scalar product induced by the adjoint representation, the sum

$$\mathcal{L} = \sum_{i=1}^{\tau} X_i^2$$

defines the Casimir operator, which is independent of the choice of the basis, and which is a central element of the enveloping algebra of $Lie(G)$. Associated to the Casimir operator is the following operator on $L^2(\mathbf{M})$ (we keep the same notation $\mathcal{L}$)

$$\mathcal{L} = D_{X_1}^2 + D_{X_2}^2 + \cdots + D_{X_\tau}^2.$$

The operator $-\mathcal{L}$, which is often called the Laplace operator, is a second order, positive, elliptic differential operator defined on the space $C^\infty(\mathbf{M})$ of infinitely differentiable functions on $\mathbf{M}$. Moreover $-\mathcal{L}$ can be closed to give a positive, self-adjoint second order elliptic differential operator on $L^2(\mathbf{M})$ with a discrete spectrum of eigenvalues $\lambda_k, k \in \mathbb{N}$, arranged in increasing and divergent order. By the spectral theorem the corresponding eigenfunctions $\{e_k\}_{k \in \mathbb{N}}$ constitute an orthonormal basis of $L^2(\mathbf{M})$, and we define, for $n \in \mathbb{N}$, the closed finite-dimensional subspaces $E_n = E_n(\mathbf{M})$ of $L^2(\mathbf{M})$

spanned by eigenfunctions e_k whose corresponding eigenvalues λ_k do not exceed n. Formally

$$E_n(\mathbf{M}) := \left\{ x \mapsto \sum_{k:\lambda_k \leq n} c_k e_k(x) : c_k \in \mathbb{R}, \lambda_k \text{ an eigenvalue of } e_k \right\}.$$

2.2 Connection to the Laplace–Beltrami operator

The operator $\mathcal{L}$ need not necessarily coincide with the Laplace–Beltrami operator on $\mathbf{M}$, but it does in several important cases. If M is a two-point homogeneous space then $\mathcal{L}$ equals, up to a scaling constant, the Laplace–Beltrami operator, see Proposition 4.11 in Chapter II of Helgason [14]. By Wang's [29] classification of such spaces this includes, among others, the d-dimensional unit sphere, real and certain complex projective spaces. Further examples for manifolds where the Laplace–Beltrami operator coincides with $-\mathcal{L}$ are given in Geller and Pesenson [8]. Since this connection is of some interest in applications, we discuss this point here in some more detail.

The Laplace operator $\mathcal{L}$ is left- and right invariant and symmetric with respect to the inner product $\langle \cdot, \cdot \rangle$ induced by the adjoint representation, see [5, p. 162]. By the general theory of irreducible unitary representation of compact Lie groups (e.g., Theorem 6.4.1 and Proposition 8.2.1 in [5]):

$$L^2(\mathbf{M}) = \bigoplus_j V_j, \;\; V_j = ker(\mathcal{L} - c_j I)$$

for constants c_j, and $\forall g \in G, \;\; L_g(V_j) \subset V_j$,

$$g \in G \mapsto L_g \in Lin(V_j)$$

is a finite dimensional unitary representation of G, where $Lin(V_j)$ denotes the space of bounded linear operators on V_j.

Moreover, as a Riemannian manifold, $\mathbf{M}$ is equipped with a Laplace–Beltrami operator Δ which commutes with the G-action: $\forall g \in G, \;\; \Delta L_g = L_g \Delta$. If M is compact:

$$L^2(\mathbf{M}) = \bigoplus_k \mathcal{H}_k, \quad \mathcal{H}_k = ker(\Delta - \lambda_k I).$$

Moreover $\mathcal{H}_k$ is G-invariant ($\forall g \in G, \;\; L_g(\mathcal{H}_k) \subset \mathcal{H}_k$), so

$$g \in G \mapsto L_g \in Lin(\mathcal{H}_k)$$

is a finite dimensional unitary representation of G.

Clearly, if $\Phi_k(x, y)$ is the kernel of the projection operator onto $\mathcal{H}_k$, then $\phi_k(y) = \Phi_k(x_0, y)$ verifies $\|\phi_k\|_2^2 = \phi_k(x_0) = dim(\mathcal{H}_k)$ and is moreover a *zonal* function (recall that f is zonal if $\forall h \in K,\ L_h(f) = f$, see, e.g., [9,14]). If the space of zonal functions in $\mathcal{H}_k$ is of dimension 1 then $g \in G \mapsto L_g \in Lin(\mathcal{H}_k)$ is an irreducible representation. If this is the case for all $\mathcal{H}_k$ then $\mathcal{L}$ and the Laplace–Beltrami will coincide, if we can check that the eigenvalues are the same.

Let us illustrate this in the case of $\mathbf{M} = \mathbb{S}^d$, where

$$G = SO(d+1) = \{A \in M(d+1 \times d+1),\ A^{-1} = A^t\},$$
$$Lie(G) = so(d+1) = \{X \in M(d+1 \times d+1),\ -X = X^t\}$$

and we can take

$$\langle X, Y\rangle = \frac{1}{2} Tr(XY^t).$$

An orthonormal basis is then given by

$$X_{i,j} = E_{i,j} - E_{j,i},\ 1 \le i < j \le d+1, \quad E_{j,i} = (\alpha^{i,j}_{k,l})_{k,l},\ \alpha^{i,j}_{k,l} = \delta_{i,k}\delta_{j,l}.$$

We take $x_0 = (1, 0, \dots, 0)$ so $K \approx SO(d)$ and

$$\forall x, y \in M = \mathbb{S}^d,\ d_{\mathbb{S}^d}(x, y) = \arccos(\langle x, y\rangle_{\mathbb{R}^{d=1}})$$

The eigenvalues of Δ are $\lambda_k = -k(k+d-1)$, see Proposition 9.3.5 in [5], the space $\mathcal{H}_k$ equals the space of spherical harmonic functions of degree k, and there is only one zonal function in each $\mathcal{H}_k$ (which is given through Gegenbauer polynomials) so the induced representation are irreducible (and not equivalent). To see that $\Delta = -\mathcal{L}$ it is enough to compute the eigenvalue of $\mathcal{L}$ on $\mathcal{H}_k$ and this can be carried on in the case of the sphere using the explicit expression of $\mathcal{L} = \sum_{i<j} D^2_{X_{i,j}}$.

2.3 A smoothed projection onto the span of the eigenfunctions of $-\mathcal{L}$

We shall write $\langle g, h\rangle$ from now on for the standard $L^2(\mathbf{M})$-inner product of two functions $g, h \in L^2(\mathbf{M}) := L^2(\mathbf{M}, dx)$. We also denote by $\|g\|_\Omega = \sup_{y\in\Omega} |g(y)|$ the supremum norm of $g : \mathbf{M} \to \mathbb{R}$ over $\Omega \subseteq \mathbf{M}$, and we shall write $\|g\|_\infty$ when $\Omega = \mathbf{M}$.

Let $0 \le a \le 1$ be an infinitely differentiable nonnegative function defined on $[0, \infty)$. We require a to be identically 1 on $[0, 1/2]$ and compactly supported on $[0, 1]$. Define the sequence of linear operators $A_j,\ j \ge 0$, with

$$A_0 f = \int_{\mathbf{M}} f(x)dx, \quad A_j f(x) := A_j(f)(x) = \int_{\mathbf{M}} A_j(x, y) f(y) dy, \quad j > 0,$$

where, for $L_k(x, y) = e_k(x)\overline{e_k(y)}$,

$$A_j(x, y) := \sum_k a\left(\frac{\lambda_k}{2^{2j}}\right) L_k(x, y) = \sum_{k:\lambda_k < 2^{2j}} a\left(\frac{\lambda_k}{2^{2j}}\right) e_k(x)\overline{e_k(y)}.$$

Clearly

$$\langle A_j f, f\rangle = \sum_k a\left(\frac{\lambda_k}{2^{2j}}\right)\langle L_k f, f\rangle \le \|f\|_2^2, \quad \|A_j f\|_2 \le \|f\|_2$$

from Parseval's identity and since $|a| \le 1$. Since a is identically one on $[0, 1/2]$

$$h \in E_{2^{2j-1}}(\mathbf{M}) \text{ implies } A_j(h) = h \tag{3}$$

and since $E_n(\mathbf{M})$, $n \ge 1$, is dense in $L^2(\mathbf{M})$ we conclude

$$\lim_{j\to\infty} \|A_j f - f\|_2 = 0$$

for every $f \in L^2(\mathbf{M})$. Thus A_j furnishes us with an approximation of the identity operator on $L^2(\mathbf{M})$.

The kernel A can be 'split' as follows: If we define

$$C_j(x, y) = \sum_{k:\lambda_k<2^{2j}} \sqrt{a\left(\frac{\lambda_k}{2^{2j}}\right)} L_k(x, y)$$

then due to the orthogonality properties of the L_k's we see

$$A_j(x, y) = \int_{\mathbf{M}} C_j(x, u) C_j(u, y) du. \tag{4}$$

2.4 Gauss cubature formula and needlets on a manifold

The following quadrature formula holds on $E_k(\mathbf{M})$, see Theorem 5.3 in Geller and Pesenson [8]. For every $k \in \mathbb{N}$ there exists a finite subset χ_k of $\mathbf{M}$ of cardinality $|\chi_k| \le C k^{d/2}$ and positive real numbers $b_\eta := b_{\eta k} > 0$, indexed by the elements η of χ_k, such that

$$\forall f \in E_k(\mathbf{M}), \qquad \int_{\mathbf{M}} f(x)\, dx = \sum_{\eta\in\chi_k} b_\eta f(\eta). \tag{5}$$

The kernel C_j defined above clearly satisfies $z \mapsto C_j(x, z) \in E_{2^{2j}}(\mathbf{M})$ for every $x \in \mathbf{M}$, and Theorem 6.1 in Geller and Pesenson [8] states that

$$f, g \in E_n(\mathbf{M}) \Rightarrow fg \in E_{4\tau n}(\mathbf{M}), \tag{6}$$

so we deduce $z \mapsto C_j(x, z)C_j(z, y) \in E_{\tau 2^{2j+2}}(\mathbf{M})$. Note that it is property (6) where homogeneity of the manifold is used crucially. It is in the same spirit as (but not equivalent to) the addition formula for eigenfunctions of the Laplace–Beltrami operator on a Riemannian manifold (see [9]). Combining (4) with (5) thus implies

$$A_j(x, y) = \int_{\mathbf{M}} C_j(x, z)C_j(z, y)dz = \sum_{\eta \in \chi_{\tau 2^{2j+2}}} b_\eta C_j(x, \eta)C_j(\eta, y)$$

and the action of A_j on $L^2(\mathbf{M})$ can hence be represented as

$$A_j f(x) = \int_{\mathbf{M}} A_j(x, y) f(y)dy = \int_{\mathbf{M}} \sum_{\eta \in \chi_{\tau 2^{2j+2}}} b_\eta C_j(x, \eta)C_j(\eta, y) f(y)dy$$
$$= \sum_{\eta \in \chi_{\tau 2^{2j+2}}} \sqrt{b_\eta} C_j(x, \eta) \int_{\mathbf{M}} \sqrt{b_\eta} C_j(\eta, y) f(y)dy.$$

This motivates the definition of the *needlet scaling function* $\phi_{j\eta}$ indexed by the cubature points $\eta \in \mathcal{Z}_j$,

$$\phi_{j\eta}(x) := \sqrt{b_\eta}\, C_j(x, \eta); \qquad \eta \in \mathcal{Z}_j \equiv \chi_{\tau 2^{2j+2}}.$$

With this notation we can write

$$A_j f(x) = \sum_{\eta \in \mathcal{Z}_j} \langle \phi_{j\eta}, f\rangle \phi_{j\eta}(x), \tag{7}$$

and call this approximation the *needlet projection* of f onto $E_{\tau 2^{2j+2}}(\mathbf{M})$ at resolution level j.

We shall need below the following estimates on the cubature set, see Geller and Pesenson [8]

$$\tfrac{1}{k_1}\frac{1}{2^{dj}} \le b_{\eta j} \le k_1 \frac{1}{2^{dj}} \quad \forall \eta \in \mathcal{Z}_j, \quad \tfrac{1}{k_2}\, 2^{dj} \le |\mathcal{Z}_j| \le k_2\, 2^{dj} \tag{8}$$

for some explicit constants $k_1, k_2 > 0$.

Although we shall not explicitly use it in what follows, we can telescope the needlet projections in the usual way to obtain a wavelet-type multiresolution approximation

$$A_j f = A_0 f + \sum_{0 \le l \le j-1} \sum_{\eta} \langle f, \psi_{l\eta}\rangle \psi_{l\eta}$$

of a function f on a compact homogeneous manifold by needlets

$$\psi_{l\eta}(x) = \sqrt{b_{\eta l}} \sum_m c(\lambda_m/2^{2l}) L_k(x, \eta), \quad \eta \in \mathcal{Z}_l,$$

with $c(y) = \sqrt{a(y/2) - a(y)}$. See Section 8 of Geller and Pesenson [8] for details. In particular

$$f \in L^2(\mathbf{M}) \Rightarrow \left\| f - \sum_{l \le j} \sum_{\eta \in \mathcal{Z}_l} \langle f, \psi_{l\eta} \rangle \psi_{l\eta}, \right\|_2 \to 0 \quad \text{as } j \to \infty,$$

and the $(\psi_{j\eta})$'s form a tight frame of $L^2(\mathbf{M})$:

$$\forall f \in L^2(\mathbf{M}), \quad \|f\|_2^2 = \sum_j \sum_\eta |\langle f, \psi_{j\eta} \rangle|^2. \tag{9}$$

2.5 Properties of the needlet frame

We establish some key properties of needlets, including their near-exponential localization property.

Proposition 1 *We have, for some constant* $0 < D_1(\mathbf{M}) < \infty$ *and every* $j \ge 0, \eta \in \mathcal{Z}_j$,

$$\|\phi_{j\eta}\|_2 \le 1, \quad \|\phi_{j\eta}\|_\infty \le D_1(\mathbf{M}) 2^{jd/2}. \tag{10}$$

Moreover, for every $x \in \mathbf{M}, \eta \in \mathcal{Z}_j$ *and every* $N \in \mathbb{N}$ *there exists a constant* c_N *such that*

$$|\phi_{j\eta}(x)| \le \frac{c_N 2^{jd/2}}{(1 + 2^{jd} d_{\mathbf{M}}(\eta, x))^N}. \tag{11}$$

Proof For the first inequality in (10), let $\eta \in \mathbf{M}$, $n \in \mathbb{N}$ and note

$$\int_{\mathbf{M}} \left(\sum_{k:\lambda_k \le n} L_k(x, \eta) \right)^2 dx = \sum_{k:\lambda_k \le n} L_k(\eta, \eta).$$

On the other hand, by (6),

$$x \mapsto \left(\sum_{k:\lambda_k \le n} L_k(x, \eta) \right)^2 \in E_{4\tau n}(\mathbf{M}),$$

so if $\chi_{4\tau n}$ is the set of cubature points of $E_{4\tau n}(\mathbf{M})$ and $\eta \in \chi_{4\tau n}$

$$\int_{\mathbf{M}} \left(\sum_{k:\lambda_k \le n} L_k(x, \eta) \right)^2 dx = \sum_{\xi \in \chi_{4\tau n}} b_\xi \left(\sum_{k:\lambda_k \le n} L_k(\xi, \eta) \right)^2 \ge b_\eta \left(\sum_{k:\lambda_k \le n} L_k(\eta, \eta) \right)^2.$$

so, combining these estimates,

$$b_\eta \le \frac{1}{\sum_{k:\lambda_k<n} L_k(\eta,\eta)}$$

for every $\eta \in \chi_{4dn}$. This implies, for every $\eta \in \mathcal{Z}_j$,

$$\int_{\mathbf{M}} \phi_{j\eta}^2(x)dx = b_\eta \sum_{k:\lambda_k<2^{2j}} a(\lambda_k/2^{2j})L_k(\eta,\eta) \le 1.$$

To prove the remaining claims, recall that by definition

$$\phi_{j\eta}(x) = \sqrt{b_\eta} \sum_{k:\lambda_k<2^{2j}} \sqrt{a(\lambda_k/2^{2j})}L_k(x,y).$$

For f a function from the Schwartz-class on $\mathbb{R}^+$, Lemma 4.1 (and the remark after it) in Geller and Mayeli [7], applied to the elliptic operator $f(\mathcal{L}/2^{2j})$ (notation of functional calculus, $t = 2^{-2j}$ in their lemma), proves that for every integer $N \ge 0$ there exists a constant $c_N(f)$ such that

$$\sum_{k:\lambda_k<2^{2j}} f(\lambda_k/2^{2j})L_k(x,\eta) \le \frac{c_N(f)2^{jd}}{(1+2^{jd}d(\eta,x))^N}. \tag{12}$$

Applying this to $f = \sqrt{a}$ and using (8), we infer the second bound in (10) as well as (11) follows from (8) and (12). □

Proposition 2 *We have*

$$\sup_{x\in\mathbf{M}} \int_{\mathbf{M}} A_j^2(x,y)dy \le D_2(\mathbf{M})2^{jd}, \quad \sup_{x,y\in\mathbf{M}} |A_j(x,y)| \le D_2(\mathbf{M})2^{jd} \tag{13}$$

for some finite positive constant $D_2(\mathbf{M})$ that depends only on the manifold.

Proof As $A_j(x,y) := \sum_k a(\lambda_k/2^{2j})L_k(x,y)$, the second claim follows from (12) with $f = a$. For the first

$$\begin{aligned}\int_{\mathbf{M}} A_j^2(x,y)dy &= \int_{\mathbf{M}} \sum_{k,l} a\left(\frac{\lambda_k}{2^{2j}}\right) L_k(x,y)\overline{a\left(\frac{\lambda_l}{2^{2j}}\right) L_l(x,y)}dy \\ &= \sum_k a^2\left(\frac{\lambda_k}{2^{2j}}\right) L_k(x,x)\end{aligned}$$

and again using (12) with $f = a^2$ gives the result. □

2.6 The case of $\mathbb{S}^d$

In the case of the d-dimensional unit sphere $\mathbb{S}^d$ of $\mathbb{R}^{d+1}$ the above construction is effectively the one in Narcowich et al. [24]. On $\mathbb{S}^d$ the differential operator $\mathcal{L}$ coincides with the usual Laplace–Beltrami operator, and we have

$$L^2(\mathbb{S}^d) = \bigoplus_k \mathcal{H}_k, \quad \mathcal{H}_k \equiv \mathcal{H}_k(\mathbb{S}^d) = ker(\Delta - \lambda_k I), \ \lambda_k = -k(k+d-1).$$

The eigenfunctions e_k in this case are the spherical harmonics with eigenvalues $k(k+d-1)$ (e.g., Proposition 9.3.5 in [5]). Thus if we take the subsequence $N \equiv N_k$ of $\mathbb{N}$ for which $k(k+d-1) = N_k$ as k runs through the nonnegative integers, then the spaces $E_N(\mathbb{S}^d)$ correspond to the spaces $\mathcal{P}_N(\mathbb{S}^d)$ of spherical polynomials of degree less than or equal to N, which are spanned by the mutually orthogonal spaces $\mathcal{H}_k(\mathbb{S}^d)$, $0 \le k \le n$, of spherical harmonics, see [5,27].

If $\{e_i^k\}$ is any orthonormal basis of $\mathcal{H}_k$, then we write, in slight abuse of notation,

$$L_k(x, y) = \sum_i e_i^k(x)\overline{e_i^k(y)} = L_k(\langle x, y\rangle_{d+1}), \quad \langle x, y\rangle_{d+1} = \sum_{i=1}^{d+1} x_i y_i$$
$$|\mathbb{S}^d| L_k(u) = \left(1 + \frac{k}{\nu}\right) C_k^\nu(u), \quad \nu = \frac{d-1}{2}, \quad u \in [-1, 1]$$

where C_k^ν is the corresponding Gegenbauer polynomial, and $|\mathbb{S}^d|$ is the Lebesgue measure of $\mathbb{S}^d$, i.e., $|\mathbb{S}^d| = \int_{\mathbb{S}^d} dx = (2\pi^{(d+1)/2})/\,\Gamma((d+1)/2)$. We have furthermore ([27, p. 144]) for every $x \in \mathbb{S}^d$,

$$\sum_i |e_i^k(x)|^2 dx = \frac{dim(\mathcal{H}_k(\mathbb{S}^d))}{|\mathbb{S}^d|}$$

and thus

$$|\mathbb{S}^d| L_k(1) = \dim(\mathcal{H}_k(\mathbb{S}^d)). \tag{14}$$

Moreover, for $d \ge 2$ and any $n \in \mathbb{N}$, $\mathcal{P}_n(\mathbb{S}^d) = \bigoplus_{k=0}^n \mathcal{H}_k(\mathbb{S}^d)$ and as a consequence, by Stein and Weiss [27],

$$dim(\mathcal{H}_k(\mathbb{S}^d)) = C_{k+d}^d - C_{k-2+d}^d = \frac{(d+k-2)!(d+2k-1)}{k!(d-1)!}$$
$$dim(\mathcal{P}_n(\mathbb{S}^d)) = C_{n+d}^d + C_{n+d-1}^d = \frac{2}{d!}(n+1)(n+2)..(n+d-1)\left(n + \frac{d}{2}\right)$$
$$= \frac{2}{d!} n^d \left(1 + \frac{1}{n}\right)\left(1 + \frac{2}{n}\right) \ldots . \left(1 + \frac{d-1}{n}\right)\left(1 + \frac{d}{2n}\right)$$

$$= n^d \left(\frac{2}{d} + \frac{1}{n}\right) \prod_{j=1}^{d-1} \left(\frac{1}{j} + \frac{1}{n}\right).$$

So, for $d \geq 2, n \geq 2$,

$$\frac{2}{d!}(n+1)^d \leq dim(\mathcal{P}_n(\mathbb{S}^d)) \leq n^d \left(\frac{n+1}{n}\right)^2$$
$$\frac{2}{d!} n^d \leq dim(\mathcal{P}_{n-1}(\mathbb{S}^d)) \leq n^d \quad \text{and} \quad dim(\mathcal{P}_{n-1}(\mathbb{S}^1)) = n.$$

By virtue of these bounds the constants in Proposition 2 can be explicitly calculated. To obtain a unified notation define, for $j \in \mathbb{N}$, the integers $k(j) = \max\{k \in \mathbb{N} : \lambda_k = k(k+d-1) < 2^{2j}\}$ so that $k(j) < 2^j$ always holds. Then

$$\int_{\mathbb{S}^d} A_j^2(x,y)dx = \sum_{k:\lambda_k < 2^{2j}} [a(\lambda_k/2^{2j})]^2 L_k(1)$$
$$= \frac{1}{|\mathbb{S}^d|} \sum_{k:\lambda_k < 2^{2j}} [a(\lambda_k/2^{2j})]^2 dim(\mathcal{H}_k(\mathbb{S}^d))$$
$$\leq \frac{dim(\mathcal{P}_{k(j)}(\mathbb{S}^d))}{|\mathbb{S}^d|} \leq \frac{2^{jd}}{|\mathbb{S}^d|},$$

and these inequalities imply that the same bound holds for $|A_j(x,y)|$. We can also deduce, as in the proof of Proposition 1

$$\|\phi_{j\eta}\|_\infty = \sqrt{b_\eta} \sum_{k:\lambda_k < 2^{2j}} \sqrt{a(\lambda_k/2^{2j})} L_k(1) \leq \sqrt{\sum_{k:\lambda_k < 2^{2j}} L_k(1)} \leq \sqrt{\frac{2^{jd}}{|\mathbb{S}^d|}}.$$

Conclude that the key constants $D_1(\mathbf{M})$, $D_2(\mathbf{M})$ in the last subsection can be taken to be

$$D_1(\mathbb{S}^d) = \sqrt{\frac{1}{|\mathbb{S}^d|}}, \quad D_2(\mathbb{S}^d) = \frac{1}{|\mathbb{S}^d|} \tag{15}$$

in the case of the unit sphere. Finally we should remark that in the case of the unit sphere the addition formula (6) holds with $4\tau n$ replaced by $2n$ as one is multiplying spherical polynomials. (Indeed whenever the Laplace–Beltrami operator coincides with $\mathcal{L}$ one can use the addition formula for eigenfunctions of the Laplacian in [9].) Moreover, if $d = 2$, for each resolution level j, the HEALPix pixelisation (commonly used for astrophysical data) gives $12 \cdot 2^{2j}$ cubature points, so $k_2 = 12$ in (8).

3 Linear needlet density estimators and concentration properties of their uniform fluctuations

Let $X, X_1, \ldots, X_n$ be i.i.d. random variables taking values in a compact homogeneous manifold $\mathbf{M}$ of dimension d. Denote their common law by P and assume that P possesses a density $f : \mathbf{M} \to [0, \infty)$ w.r.t. dx on $\mathbf{M}$. Denote further by $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical measure of the sample. Let $A_j(x, y)$ be the needlet projection kernel. For $j \in \mathbb{N}$, the linear needlet density estimator of f is defined as

$$f_n(j, y) = \frac{1}{n}\sum_{i=1}^{n} A_j(X_i, y) = \int_{\mathbf{M}} A_j(x, y)dP_n(x), \quad y \in \mathbf{M}. \tag{16}$$

We shall often write, in slight abuse of notation, $f_n(j)$ for $f_n(\cdot, j)$.

3.1 A Bernstein-type concentration inequality for needlet estimators

We define now some quantities that measure the 'Gaussian' and 'Poissonian' fluctuations of the uniform deviations of the centered estimator $f_n(j)$. Recall the explicit constants $D_1(\mathbf{M})$, $|\mathcal{Z}_j| \le k_2 2^{jd}$ from (8), (10) in the previous section. Note moreover that the second estimate in Proposition 1 implies

$$2^{jd/2}c_0(\mathbf{M}, j) \equiv \sup_{x\in\mathbf{M}} \sum_{\eta\in\mathcal{Z}_j} |\phi_{j\eta}(x)| \le 2^{jd/2}C(\mathbf{M}). \tag{17}$$

The constant $c_0(\mathbf{M}, j) \equiv c_0(\mathbf{M}, j, a, \mathcal{Z}_j)$ (or an upper bound for it) can be computed explicitly after the regularizing function a and the quadrature set $\mathcal{Z}_j$ have been chosen, and a sharp numerical evaluation of it is important in application of Proposition 3 below.

Define then

$$\bar{\sigma}(n, l, x) := \bar{\alpha}(x, l)\sqrt{\frac{2^{ld}}{n}} + \bar{\alpha}'(x, l)\frac{2^{ld}}{n}$$

where

$$\bar{\alpha}(x, l) := \bar{\alpha}(\mathbf{M}, f, x, l) := c_0(\mathbf{M}, l)\sqrt{2(\log(2|\mathcal{Z}_l|) + x)\|f\|_\infty}$$

and

$$\bar{\alpha}'(x, l) := \bar{\alpha}'(\mathbf{M}, x, l) := c_0(\mathbf{M}, l)\frac{2}{3}D_1(\mathbf{M})(\log(2|\mathcal{Z}_l|) + x).$$

We now prove the following concentration inequality for the needlet density estimator.

Proposition 3 *Let* $\mathbf{M}$ *be a compact homogeneous manifold and suppose* $f : \mathbf{M} \to [0, \infty)$ *is bounded. We have, for every* $n \in \mathbb{N}$, *every* $j \in \mathbb{N}$ *and every* $x \geq 0$

$$\Pr\left\{\sup_{y\in\mathbf{M}} |f_n(j, y) - Ef_n(j, y)| \geq \bar{\sigma}(n, j, x)\right\} \leq e^{-x}.$$

Proof The explicit cubature formula for eigenfunctions of $\mathcal{L}$ allows to reduce the infinite supremum $\sup_{y\in\mathbf{M}} |f_n(j, y) - Ef_n(j, y)|$ to one over a finite set, so that finite-dimensional probabilistic methods can be applied. Indeed, the estimate (17) implies that the supremum of any $h \in E_{2^{2j-1}}(\mathbf{M})$ over $\mathbf{M}$ can be bounded by the (finite) maximum of the needlet coefficients of h: Clearly from (3)

$$\forall h \in E_{2^{2j-1}}(\mathbf{M}), \quad h(x) = A_j h(x) = \sum_{\eta\in\mathcal{Z}_j} \langle \phi_{j\eta}, h\rangle \phi_{j\eta}(x)$$

so that for $\mathcal{Z}_j$ a cubature set of $E_{\tau 2^{2j+2}}(\mathbf{M})$ one has

$$\sup_{x\in\mathbf{M}} |h(x)| \leq \max_{\eta\in\mathcal{Z}_j} \left|\langle \phi_{j\eta}, h\rangle\right| \sup_{x\in\mathbf{M}} \sum_{\eta\in\mathcal{Z}_j} |\phi_{j\eta}(x)| = 2^{jd/2} c_0(\mathbf{M}, j) \max_{\eta\in\mathcal{Z}_j} \left|\langle \phi_{j\eta}, h\rangle\right|. \tag{18}$$

Now using $\langle \cdot, \cdot \rangle$ notation also acting on finite signed measures,

$$\begin{aligned} \|f_n(j) - Ef_n(j)\|_\infty &= \sup_{y\in\mathbf{M}} \left| \sum_{\eta\in\mathcal{Z}_j} \phi_{j\eta}(y) \langle \phi_{j\eta}, P_n - P\rangle \right| \\ &\leq 2^{jd/2} c_0(\mathbf{M}, j) \max_{\eta\in\mathcal{Z}_j} \left|\langle \phi_{j\eta}, P_n - P\rangle\right| \end{aligned}$$

by (17) above. Consider the finite empirical process indexed by the class of functions $\{\phi_{j\eta_k}\}_{k=1}^{|\mathcal{Z}_j|}$ which has envelope $U = 2^{jd/2} D_1(\mathbf{M})$ in view of (10). The class of functions

$$\mathcal{G} := \left\{\phi_{j\eta_1}/2U, \ldots, \phi_{j\eta_{|\mathcal{Z}_j|}}/2U\right\},$$

is thus uniformly bounded by $1/2$ and its weak variances σ^2 satisfy

$$\sup_{g\in\mathcal{G}} Eg^2(X) \leq \sigma^2 = \frac{\|f\|_\infty}{2^{jd+2} D_1^2(\mathbf{M})}$$

since $\|\phi_{j\eta}\|_2 \leq 1$ (again (10)). Recall Bernstein's inequality (e.g., p. 26 in [23]): If $Z_1, \ldots, Z_n$ are i.i.d.centered random variables bounded in absolute value by 1 then

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| \geq \sqrt{\frac{2tv}{n}} + \frac{t}{3n}\right\} \leq 2e^{-t} \tag{19}$$

where $v \geq EZ_i^2$. Therefore, using the notation $\|\mu\|_{\mathcal{G}} \equiv \sup_{g\in\mathcal{G}} |\int g d\mu|$ for signed measures μ,

$$\Pr\left\{\|f_n(j) - Ef_n(j)\|_\infty \geq c_0(\mathbf{M}, j)\left(\sqrt{\frac{2(\log(2|\mathcal{Z}_j|) + x)2^{jd}\|f\|_\infty}{n}} + \frac{2U2^{jd/2}(\log(2|\mathcal{Z}_j|) + x)}{3n}\right)\right\}$$

$$\leq \Pr\left\{\max_{\eta\in\mathcal{Z}_j} \left|\langle\phi_{j,\eta}, P_n - P\rangle\right| \geq \sqrt{\frac{2(\log(2|\mathcal{Z}_j|)+x)\|f\|_\infty}{n}} + \frac{2U(\log(2|\mathcal{Z}_j|)+x)}{3n}\right\}$$

$$\leq \Pr\left\{\|P_n - P\|_{\mathcal{G}} \geq \sqrt{\frac{2(\log(2|\mathcal{Z}_j|) + x)\|f\|_\infty}{D_1^2(\mathbf{M})2^{jd+2}n}} + \frac{\log(2|\mathcal{Z}_j|) + x}{3n}\right\}$$

$$= \Pr\left\{\max_{m=1,\dots,|\mathcal{Z}_j|}\left|\frac{1}{n}\sum_{i=1}^n(g_m(X_i) - Eg_m(X))\right| \geq \sqrt{\frac{2(\log(2|\mathcal{Z}_j|) + x)\sigma^2}{n}} + \frac{\log(2|\mathcal{Z}_j|) + x}{3n}\right\}$$

$$\leq \sum_{m=1}^{|\mathcal{Z}_j|}\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^n(g_m(X_i) - Eg_m(X))\right| \geq \sqrt{\frac{2(\log(2|\mathcal{Z}_j|) + x)\sigma^2}{n}} + \frac{\log(2|\mathcal{Z}_j|) + x}{3n}\right\}$$

$$\leq 2|\mathcal{Z}_j|\exp\left\{-\log(|\mathcal{Z}_j|) - \log 2 - x)\right\} = e^{-x},$$

which completes the proof of Proposition 3. □

We should mention that a minor modification of the proof of Proposition 3 combined with the usual blocking arguments (as, e.g., in Theorem 1 in [10]) implies under standard conditions on j_n (including $2^{j_n} \approx n^\eta$ for some $0 < \eta < 1$) that

$$\limsup_n \sqrt{\frac{2^{j_n d} j_n}{n}} \sup_{y\in\mathbf{M}} |f_n(j, y) - Ef_n(y, d)| \leq D \quad \textit{almost surely} \tag{20}$$

where the constant D depends only on $\mathbf{M}$, k_2 and $\|f\|_\infty$.

In some proofs below we shall need that $\bar{\sigma}(n, l, x)$ is monotone increasing in $l \in \mathbb{N}$. In general whether this holds true or not depends on the cubature $\mathcal{Z}_l$ as well as on the function a. Monotonicity of $\bar{\sigma}(n, l, x)$ can be easily ensured if we replace $\bar{\alpha}(x, l)$ and $\bar{\alpha}'(x, l)$ by their upper bounds $\alpha(x, l), \alpha'(x, l)$ obtained from the inequalities

$|\mathcal{Z}_l| \le k_2 2^{ld}$, $c_0(\mathbf{M}, l) \le C(\mathbf{M})$. While we do not advocate this in practice, for the theoretical development we define

$$\sigma(n,l,x)=\alpha(x,l)\sqrt{\frac{2^{ld}}{n}}+\alpha'(x,l)\frac{2^{ld}}{n}, \quad A(n,l,x):=\left[\alpha(x,l)+\alpha'(x,l)\sqrt{2^{ld}/n}\right]. \tag{21}$$

The constant $A(n, l, x)$ allows for $\sigma(n, l, x)$ to be written as a constant multiple of the 'Gaussian component' $\sqrt{2^{ld}/n}$, that is, $\sigma(n, l, x) = A(n, l, x)\sqrt{2^{ld}/n}$.

3.2 Concentration inequalities via Rademacher processes on manifolds

Despite its conceptual simplicity the approach from the previous section has one drawback: the uniform deviations of $f_n - Ef_n$ are controlled globally on $\mathbf{M}$ by the function $\sigma(n, l, x)$ – constant on $\mathbf{M}$. For functions f that exhibit spatially inhomogeneous regularity properties it is of interest to have a 'localised' version of $\sigma(n, l, x)$. This could be achieved in Proposition 3 by means of proving a 'local' analogue of (18), which, however, is a rather intricate matter that we do not pursue here. Instead we show how a simple symmetrization technique can be used to deal with this problem. This is inspired by Koltchinskii [18] and also Giné and Nickl [12]. For Ω any subset of $\mathbf{M}$, define a Rademacher process $\{(1/n)\sum_i \varepsilon_i A_j(X_i, y)\}_{y\in\Omega}$ and set

$$R_n(\Omega, j) = \sup_{y\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i A_j(X_i, y)\right|$$

with $(\varepsilon_i)_{i=1}^n$ an i.i.d. Rademacher sequence, independent of the X_i's (and defined on a large product probability space). $R_n(\Omega, j)$ can be computed in practice by first simulating n i.i.d. random signs, applying these signs to the summands $A_j(X_i)$ of the needlet density estimator, and maximizing the resulting function. The idea is that the supremum $R_n(\Omega, j)$ of the symmetrized process serves as a random surrogate for the unknown supremum $\sup_{y\in\Omega}|f_n(j, y) - Ef_n(j, y)|$ of the centered process. Indeed Proposition 4 shows that $\sup_{y\in\Omega}|f_n(y) - Ef_n(y)|$ concentrates around (a constant multiple of) $R_n(\Omega, j)$. Define the deviation term

$$\sigma^R(\Omega, n, j, x) = 6R_n(\Omega, j) + 10\sqrt{\frac{2^{jd}D_2(\mathbf{M})\|f\|_\infty(x+\log 2)}{n}} + 22\frac{2^{jd}D_2(\mathbf{M})(2x+2\log 2)}{n}.$$

Proposition 4 *Let* $\mathbf{M}$ *be a compact homogeneous manifold and suppose* $f : \mathbf{M} \to [0, \infty)$ *is bounded. We have for every* $n \in \mathbb{N}$, *every* $j \in \mathbb{N}$, *every* $\Omega \subseteq \mathbf{M}$ *and every*

$x > 0$ *that*

$$\Pr\left\{\sup_{y\in\Omega}|f_n(y,j)-Ef_n(y,j)|\geq\sigma^R(\Omega,n,j,x)\right\}\leq e^{-x}.$$

Proof We use the following general result for empirical processes.

Proposition 5 *Let $\mathcal{F}$ be a countable class of real-valued measurable functions defined on* $\mathbf{M}$*, uniformly bounded by* $1/2$*. We have for every* $n\in\mathbb{N}$ *and* $x>0$

$$\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^{n}(f(X_i)-Pf)\right\|_{\mathcal{F}}\geq 6\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathcal{F}}+10\sqrt{\frac{(x+\log 2)\sigma^2}{n}}\right.$$
$$\left.+22\frac{x+\log 2}{n}\right\}\leq e^{-x}$$

The proof, which is based on Talagrand's [28] inequality with constants (e.g., [23]), is inspired by ideas in [12,18], and can be found in Proposition 5 in [21]. Now to prove Proposition 4 note that

$$\|f_n(j)-Ef_n(j)\|_\Omega=\sup_{y\in\Omega}\left|\frac{1}{n}\sum_{i=1}^{n}\left(A_j(X_i,y)-EA_j(X,y)\right)\right|$$

for $\Omega\subseteq\mathbf{M}$. This amounts to studying the empirical process indexed by the class of functions $\left\{A_j(\cdot,y):y\in\Omega\right\}$ for $\Omega\subseteq\mathbf{M}$. This class has envelope $2^{jd}D_2(\mathbf{M})$ in view of Proposition 2. Define thus

$$\mathcal{G}:=\mathcal{G}_j=\left\{A_j(\cdot,y)/(2^{jd+1}D_2(\mathbf{M})):y\in\Omega\right\}\tag{22}$$

which is uniformly bounded by $1/2$. (In fact, by continuity of the mapping $y\mapsto A_j(x,y)$ for every $x\in\mathbf{M}$ we can restrict ourselves to a countable subset of Ω, which we still denote by Ω.) Furthermore the upper bound for the weak variances can be taken to be

$$\sup_{g\in\mathcal{G}}Eg^2(X)\leq\frac{\|f\|_\infty}{D_2^2(\mathbf{M})2^{2jd+2}}\sup_{y\in\mathbf{M}}\int_{\mathbf{M}}A_j^2(x,y)dx\leq\frac{\|f\|_\infty}{D_2(\mathbf{M})2^{jd+2}}=:\sigma^2\tag{23}$$

in view of Proposition 2. Then, recalling the notation $\|\cdot\|_{\mathcal{G}}$ from the proof of Proposition 3

$$\Pr\left\{\|f_n(j,\cdot)-Ef_n(j,\cdot)\|_\Omega \geq 6R_n(\Omega,j)+10\sqrt{\frac{2^{jd}D_2(\mathbf{M})\|f\|_\infty(x+\log 2)}{n}}\right.$$
$$\left.+\,22\frac{2^{jd}D_2(\mathbf{M})(2x+2\log 2)}{n}\right\}$$
$$=\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^n(g(X_i)-Pg)\right\|_{\mathcal{G}} \geq \frac{6R_n(\Omega,j)}{2^{jd+1}D_2(\mathbf{M})}+10\sqrt{\frac{\|f\|_\infty(x+\log 2)}{D_2(\mathbf{M})2^{jd+2}n}}\right.$$
$$\left.+\,22\frac{x+\log 2}{n}\right\}$$
$$=\Pr\left\{\left\|\frac{1}{n}\sum_{i=1}^n(g(X_i)-Pg)\right\|_{\mathcal{G}} \geq 6\left\|\frac{1}{n}\sum_{i=1}^n\varepsilon_i g(X_i)\right\|_{\mathcal{G}}+10\sqrt{\frac{(x+\log 2)\sigma^2}{n}}\right.$$
$$\left.+\,22\frac{x+\log 2}{n}\right\}$$

and the last expression is less than or equal to e^{-x} using Proposition 5 with $\mathcal{G}$ as in (22) and σ specified by (23). □

It is interesting to compare σ^R to σ from Proposition 3. On the one hand the second and third terms defining $\sigma^R(\Omega,n,j,x)$ are of a smaller asymptotic order than $\sigma(n,j,x)$ for $j\to\infty$ due to the absence of $|\mathcal{Z}_j|$ in σ^R. On the other hand the term $R_n(\Omega,j)$ is random, and one is led to ask whether in average σ^R will be larger or smaller than σ. Our proofs imply, for some constant C independent of j,n, that

$$ER_n(\Omega,j)\leq C\left(\sqrt{\frac{2^{jd}j}{n}}+\frac{2^{jd}j}{n}\right)$$

so that σ^R has the same size as σ as a function of j,n, up to constants.

Inspection of the proofs and arguments similar to those in the proof of Proposition 2 in Giné and Nickl [12] show that $R_n(\Omega,j)$ in Proposition 4 can be replaced by its (conditional) expectation $E_\varepsilon R_n(\Omega,j)$—a quantity that may be more stable in applications. Moreover, the constants appearing in the definition of σ^R may still be fairly conservative: the proof is based on an application of Talagrand's [28] inequality with explicit constants (see [23]), and in the lower deviation version thereof the optimal constants are not known yet.

4 Confidence bands

If the size of the bias $\|Ef_n(j) - f\|_\infty$ were known, one could directly use Propositions 3 or 4 and a suitable choice of j to obtain confidence bands with prescribed finite sample coverage. For instance, if f is the uniform distribution (volume element) on $\mathbf{M}$, the bias $A_0(f) - f$ of the estimate $f_n(0)$ is exactly zero. In analogy, if $f \in E_N(\mathbf{M})$ is a finite linear combination of eigenfunctions of $\mathcal{L}$ (so in the spherical case a polynomial) then the estimator $f_n(J)$ for sufficiently large but finite J also has bias zero (cf. (3)). As usual, going beyond finite-dimensional smoothness classes is possible by considering spaces of differentiable functions on $\mathbf{M}$. For instance one defines $C^k(\mathbf{M})$ as the set of continuous functions $f \in C(\mathbf{M})$ such that for all $X_1, X_2, \dots, X_k$ in $Lie(G)$, $D_{X_1} D_{X_2} \dots D_{X_k} f \in C(\mathbf{M})$. It is a Banach space when equipped with the following norm:

$$\|f\|_{C^k} = \sup_{|X_1| \le 1, \dots, |X_k| \le 1} \|D_{X_1} D_{X_2} \dots D_{X_k} f\|_\infty + \|f\|_\infty,$$

and $C^\infty(\mathbf{M})$ is the intersection of all the spaces $C^k(\mathbf{M})$, $k \in \mathbb{N}$. One can define such spaces also for noninteger k by introducing a modulus of continuity along vectorial directions X, and the resulting scale of Hölder–Zygmund function spaces $\mathcal{C}^k(\mathbf{M})$ can be characterized by the decay of their needlet coefficients in very much the same way as in the case of Hölder–Zygmund spaces on Euclidean spaces: A k-regular function in $\mathcal{C}^k(\mathbf{M})$, $k > 0$ then satisfies the estimate

$$\|A_j(f) - f\|_\infty \le C 2^{-jk}. \tag{24}$$

See Geller and Pesenson [8] for these results. If the smoothness degree t of f is known such bounds can be used, together with Propositions 3, 4, in the construction of asymptotic confidence sets, proceeding in the same way as in the classical paper [3] via choosing a resolution level j_n that leads to 'undersmoothing', i.e., a bias of smaller order as a function of n than the random fluctuations of the centered estimators.

However, in the typical nonparametric function estimation problem the size of the bias is not known, and the above assumptions are far from realistic. So we have the more ambitious goal to obtain confidence sets for the needlet estimator with an automatic choice of the resolution level j.

4.1 Estimate of the resolution level

Split the sample into two parts $\mathcal{S}_1$ and $\mathcal{S}_2$, each of (integer) size $n_1 > 0$ and $n_2 > 0$ respectively. For asymptotic considerations we shall require that n_1/n_2 is bounded away from zero and infinity as $n \to \infty$. Denote by

$$P_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} \delta_{X_i}, \quad \text{and} \quad P_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} \delta_{X_{n_1+i}}$$

the empirical measures associated with the first and the second subsample, respectively, and define the associated needlet density estimators

$$f_{n_v}(j, y) = \int_{\mathbf{M}} A_j(x, y) dP_{n_v}(x), \quad y \in \mathbf{M}, \quad v = 1, 2.$$

We use the sample $\mathcal{S}_2$ to choose the resolution level j. For $n_2 > 1$, choose an integer $j_{\max} := j_{\max,n}$ and define the grid of candidate bandwidths as

$$\mathcal{J} := \mathcal{J}_n = \{[0, j_{\max}] \cap \mathbb{N}\}.$$

For asymptotic considerations we shall only require

$$2^{j_{\max}} \simeq \left(\frac{n_2}{(\log n_2)^2}\right)^{1/d}, \tag{25}$$

but a practical choice is to first choose l^* such that $\alpha(x, l^*)\sqrt{2^{l^*}/n_2} = \alpha'(x, l^*)$ $(2^{l^*}/n_2)$ and to define $j_{\max}$ such that $2^{j_{\max}} = 2^{l^*}/(\log n_2)^{1/d}$. Such a choice of $j_{\max}$ is just slightly below the boundary where the Poissonian term starts to dominate the Gaussian term in $\sigma(n_2, l, x)$ in Proposition 3, and choosing $j > j_{\max}$ would then result in inconsistent estimators, so that $j_{\max}$ is a natural upper bound for $\mathcal{J}$.

The goal is to select a data-driven bandwidth $\hat{j}_n$ from $\mathcal{J}_n$. Heuristically, for $l > j$,

$$f_{n_2}(j) - f_{n_2}(l) = [f_{n_2}(j) - Ef_{n_2}(j)] - [f_{n_2}(l) - Ef_{n_2}(l)] + [A_j(f) - f] - [A_l(f) - f]$$

and with large probability the first two terms should not exceed $2\sigma(n_2, l, x)$, a quantity that increases in l, and we would like to choose $\hat{j}_n$ to be the smallest j such that the approximation error $2(A_j(f) - f)$ (which decreases in j) does not exceed the size $2\sigma(n_2, l, x)$ of the random fluctuations.

We shall use the subsample $\mathcal{S}_2$ to select $\hat{j}_n$ following this idea, which is due to Lepskiĭ [20], formalised as follows:

$$\hat{j}_n = \min\left\{j \in \mathcal{J} : \|f_{n_2}(j) - f_{n_2}(l)\|_\Omega \leq 4\sigma(n, l) \ \ \forall l > j, l \in \mathcal{J}\right\}. \tag{26}$$

where $\sigma(n, l) = \sigma(n_2, l, \kappa \log n_2)$, cf. (21), where $\kappa > 0$ is any numerical constant. By definition $\hat{j}_n = j_{\max}$ if $\forall j, \ \exists\, l > j, \ l, j \ \in \mathcal{J}, \|f_{n_2}(j) - f_{n_2}(l)\|_\Omega > 4\sigma(n, l)$.

A few remarks about the constants involved in the definition of $\sigma(n, l)$ are in order: All these constants are explicit once the function a and the cubature $\mathcal{Z}_j$ have been chosen, except for the quantity $\|f\|_\infty$. If no upper bound for $\|f\|_\infty$ is known we advocate that $\|f\|_\infty$ be replaced by $\|f_n(j_{\max})\|_\infty$. Standard arguments imply that this random quantity exponentially concentrates around $\|f\|_\infty$, see for instance Giné and Nickl [12]. Consequently we neglect the case of $\|f\|_\infty$ unknown in what follows in order to reduce technicalities. Moreover we shall see below how the choice of the numerical constant κ influences the finite-sample performance, but our results hold for any choice $\kappa > 0$, in particular it does not have to be 'large enough' (as is often assumed in the adaptive estimation literature). See Remark 3 for discussion.

4.2 Confidence bands with random sizes

To construct the center of the corridor of the confidence band over $\Omega \subseteq \mathbf{M}$ we evaluate the linear estimator $f_{n_1}(\cdot, y)$ from (16) at the random bandwidth $\hat{j}_n$. It turns out that some undersmoothing is useful—in fact crucial—so let u_n be a sequence of natural numbers and define

$$\hat{f}_n(y) = f_{n_1}(\hat{j}_n + u_n, y), \quad y \in \Omega.$$

We shall see below how the sequence u_n influences our results but heuristically, and for asymptotic considerations, one may think of u_n of the order $\log \log n$.

The confidence band we propose is centered at $\hat{f}_n(y)$, $y \in \Omega$, and has random size

$$s_n(x) = 1.01\sigma(n_1, \hat{j}_n + u_n, x),$$

cf. (21), more precisely

$$C_n := C_n(x, y) = \left[\hat{f}_n(y) - s_n(x), \hat{f}_n(y) + s_n(x)\right], \quad x > 0, \ y \in \Omega \subseteq \mathbf{M}. \tag{27}$$

Alternatively one can use the band size $s_n^R(\Omega, x) = 1.01\sigma^R(\Omega, n_1, \hat{j}_n + u_n, x)$, and all results proved below go through by virtue of Proposition 4 and using techniques from Rademacher processes (as in [12]), but we abstain from this to reduce technicalities.

4.3 Coverage and adaptation properties of C_n

4.3.1 Coverage over eigenspaces of $\mathcal{L}$—the finite dimensional case

We first consider here the important case where f is a very smooth function, that is, a fixed linear combination of eigenfunctions of $\mathcal{L}$, so $f \in E_{2^J - 1}(\mathbf{M})$ for some fixed J. For simplicity of exposition let us consider the case of global confidence bands $\Omega = \mathbf{M}$ only in this subsection. We start with the case where f equals the volume element of $\mathbf{M}$.

Theorem 1 *If f is the volume element of* $\mathbf{M}$, $\int_{\mathbf{M}} f dx = 1$, *then we have, for every* $n \in \mathbb{N}$

$$\Pr(\hat{j}_n = 0) \geq 1 - 2 j_{\max} n_2^{-2\kappa}.$$

Furthermore, for every $n \in \mathbb{N}$ and every $x > 0$ we have

$$\Pr\{f(y) \in C_n(x, y) \ \ \textit{for every} \ \ y \in \mathbf{M}\} \geq 1 - e^{-x} \tag{28}$$

and, if $2^{u_n d}/n \to 0$ *as* $n \to \infty$ *then* $s_n(x) = O_{\Pr}\left(2^{u_n d/2}/\sqrt{n}\right)$.

In other words our automatic band C_n attains *exact* finite sample coverage if f is uniformly distributed, and in the usual situation where $u_n = \log\log n$ the size of the band shrinks almost at the parametric rate $1/\sqrt{n}$.

It is instructive to consider next the case where $f \in E_{2^{2J-1}}(\mathbf{M}) \setminus E_{2^{2J-2}}(\mathbf{M})$ for some fixed $J \in \mathbb{N}$. We would then hope that $\hat{j}_n = J$ with large probability, as then $A_J(f) - f = 0$ (see (3) above). In the following theorem we restrict ourselves to asymptotic considerations to highlight the main ideas.

Theorem 2 *Suppose $f \in E_{2^{2J-1}}(\mathbf{M}) \setminus E_{2^{2J-2}}(\mathbf{M})$ for some fixed $J \in \mathbb{N}$. We then have that*

$$\Pr(\hat{j}_n \notin [J-1, J]) = O(n^{-2\kappa} + e^{-cn})$$

as $n \to \infty$ for some constant c that depends on f only through $\|f\|_\infty$ and through

$$b_1(f) \equiv \inf_{p \in E_{2^{2J-2}}(\mathbf{M})} \|p - f\|_\infty > 0.$$

Moreover if $u_n > 1\ \forall n \in \mathbb{N}$ then

$$\Pr\{f(y) \in C_n(x, y) \quad \textit{for every} \quad y \in \mathbf{M}\} = 1 - e^{-x} - O(e^{-cn}) \tag{29}$$

and if $2^{u_n d}/n \to 0$ as $n \to \infty$ then $s_n(x) = O_{\Pr}\left(2^{u_n d/2}/\sqrt{n}\right)$.

Thus the confidence band C_n has asymptotic coverage for any fixed spherical polynomial, and the asymptotic size of the band C_n is of order $1/\sqrt{n}$ up to the undersmoothing factor.

Clearly we have neglected the question of honesty of C_n, that is we have not addressed the question whether (29) holds uniformly in $f \in \cup_{0 \le j \le J-1} E_{2^j}(\mathbf{M})$. Inspection of the proof implies that C_n is honest over linear combinations of eigenfunctions of $\mathcal{L}$ for which the separation constants $b_1(f)$ are bounded below by a constant multiple of $1/\sqrt{n}$. That uniformity over all densities between $E_{2^{2J-1}}$ and $E_{2^{2J-2}}$ cannot be attained for our 'adaptive' procedure is related to impossibility results for post-model selection estimators in finite-dimensional models, see Leeb and Pötscher [19].

4.3.2 *Asymptotic coverage over Hölder balls*

Theorem 2 just resembles the finite dimensional situation, and if it were indeed known a priori that $f \in E_{2^{J-1}}(\mathbf{M})$ for a fixed J one could simply use $f_n(J)$ as an estimator, circumventing the uniformity problems raised in the previous subsection. However if no finite-dimensional model seems realistic for f we may accept these uniformity problems for which $b_1(f)$ is not well-behaved if in return our procedure performs well in the infinite-dimensional setting. Note that in the usual infinite-dimensional nonparametric models the default estimator $f_n(j_{\max})$ has only a logarithmic rate of convergence to zero in supremum norm risk, and will lead to unnecessarily large confidence bands. In contrast our confidence band C_n adapts over an infinite-dimensional class of Hölder continuous densities f as we show in this section.

Our first main result is that the size of the band C_n equals, with large probability, the optimal band size that one would obtain from balancing approximation error $A_j(f) - f$ and random fluctuations $f_n(j) - A_j(f)$. For asymptotic considerations this will imply that our band shrinks at the optimal rate of convergence depending on the regularity of f. To formalize this statement we shall impose a regularity condition on the density f, namely that its approximation errors $\|A_j(f) - f\|_\Omega$ are bounded by a constant multiple of 2^{-jt} for some unknown $t > 0$. As mentioned in (24) above this is tantamount to assuming a classical t-Hölder condition on f. The theoretical bandwidth that balances bias and variance is then, up to additive constants (see (38) below for an exact definition)

$$j_n^*(t) = \frac{1}{2t+d}(\log_2 n - \log_2 \log n).$$

Theorem 3 (Size of the band) *Let Ω be any subset of $\mathbf{M}$. Suppose $f : \mathbf{M} \to [0, \infty)$ is bounded and that $\|A_j(f) - f\|_\Omega \le b_2 2^{-jt}$ for some $b_2 > 0$ and some $t > 0$. Let $2s_n(x)$ be the diameter of the band $C_n(x, y)$. Then, for every $n \in \mathbb{N}, x > 0$,*

$$\Pr\left\{s_n(x) > 1.01\sigma(n_1, j_n^*(t) + u_n + 1, x)\right\} \le 2(j_{\max} - j_n^*(t))n_2^{-\kappa}.$$

In particular, if the undersmoothing constants u_n are such that

$$r_n(t) := \left(\frac{\log n}{n}\right)^{\frac{t}{2t+d}} 2^{\frac{u_n d}{2}} = o(1)$$

as $n \to \infty$ then $s_n(x) = O_{\Pr}(r_n(t))$.

Note that the proof of the theorem, combined with standard arguments from adaptive estimation (e.g., [12]), implies as well that $\hat{f}_n$ is rate-adaptive in sup-norm loss, that is, for every $t > 0$,

$$\sup_{f : \|A_j(f) - f\|_{\mathbf{M}} \le b_2 2^{-jt}} E \sup_{x \in \mathbf{M}} |\hat{f}_n(x) - f(x)| = O(r_n(t)). \tag{30}$$

The rate of convergence $r_n(t)$ cannot be improved over classes of functions that are t-Hölder, see for instance [17] in the case $\mathbf{M} = \mathbb{S}^d$, and since these Hölder classes are, up to constants, sets of the form $\{f : \|A_j f - f\|_\infty \le b_2 2^{-jt}\}$ for suitable b_2 (see the results in [8]), this implies that (30) is optimal, and that the band C_n in Theorem 3 shrinks at the optimal rate in a minimax sense (up to the undersmoothing factor, which will typically be of size $\sqrt{\log n}$).

Clearly without a sharp evaluation of the probability of the event $\{f \in C_n\}$ Theorem 3 is useless for statistical inference. It is known (see [22]) that adaptive confidence bands for densities on $\mathbb{R}$ cannot have coverage over a continuous scale $\bigcup_{t>0} \Sigma(t, b_2)$

of Hölder balls $\Sigma(t, b_2)$. In a way Low's results can be seen as an infinite-dimensional analogue of the pathologies in finite dimensions mentioned above. On the other hand recent results in Giné and Nickl [11] show that adaptation is possible over 'generic' subsets of $\bigcup_{t>0} \Sigma(t, b_2)$ when densities are estimated on the real line. The idea is that even if some pathologies cannot be avoided there are still exhaustive classes of densities for which adaptation is possible, and we show how this applies to density estimation on $\mathbf{M}$.

To this end we assume the following crucial approximation condition. While the upper bound is standard, the quantity occurring in the lower bound can be viewed as an infinite-dimensional analogue to the constant b_1 that appeared in Theorem 2. Note that whereas b_1 is always positive the lower bound in the following condition may fail to hold for any t for a given continuous function f, at least for large enough j. We discuss this in Sect. 4.4.

Condition 1 *Assume that* $f : \mathbf{M} \to [0, \infty)$ *is bounded and let* $t, b_2 > 0$ *be real numbers. Suppose that there exists a sequence* $b(n)$ *such that* $0 < b(n) \leq b_2$ *for every* $n \in \mathbb{N}$ *and such that* f *satisfies, for every* $j \in \mathcal{J}_n$, *the inequalities*

$$b(n)2^{-jt} \leq \|A_j(f) - f\|_\Omega \leq b_2 2^{-jt}. \tag{31}$$

Under this condition we can prove asymptotic coverage of our nonparametric confidence band. We should note that inspection of the proof reveals that this coverage result is 'honest': it holds uniformly over classes of densities satisfying Condition 1.

Theorem 4 (Asymptotic Coverage) *Let* Ω *be any subset of* $\mathbf{M}$. *Suppose* f *satisfies Condition 1 and that the undersmoothing sequence* $u_n \in \mathbb{N}$ *is such that* $u_n + \frac{1}{t}\log_2(b(n)) \to \infty$ *as* $n \to \infty$. *Then we have, for every* $x > 0$,

$$\liminf_n \Pr\{f(y) \in C_n(x, y) \quad \textit{for every} \quad y \in \Omega\} \geq 1 - e^{-x}. \tag{32}$$

For instance if one knows that $\liminf_n b(n) > 0$ (we shall see generic examples for this below) then *any* undersmoothing sequence $u_n \to \infty$ gives asymptotic coverage of the band. On the other hand if $u_n \to \infty$ then $b_n \to 0$ is admissible and the lower bound requirement in Condition 1 becomes more and more lenient as sample size increases. This result and the discussion in Sect. 4.4 below shows that our nonparametric procedure does well asymptotically for 'typical' Hölder-continuous functions on the unit sphere.

4.3.3 A nonasymptotic coverage result

The asymptotic Theorem 4 is in fact a consequence of the following finite-sample result. While the stochastic terms are similarly well-behaved as in Theorems 1 and 2, the presence of nonnegligible approximation error is the reason why the following theorem is more intricate.

Theorem 5 (Finite Sample Coverage) *Let* Ω *be any subset of* $\mathbf{M}$. *Suppose* f *satisfies Condition 1 and let* $m^* := m_n^*(f)$ *be the smallest integer such that* $b(n)2^{tm^*} \geq 7b_2$.

Set $m := m_n(f) = \min(j_n^*(t), m^*)$. *Then we have, for every* $n \in \mathbb{N}$ *and every* $x > 0$

$$\Pr\{f(y) \in C_n(x, y) \quad \textit{for every} \quad y \in \Omega\} \geq 1 - e^{-x} - \upsilon_n \tag{33}$$

where

$$\upsilon_n = 2(j_{\max} - m)n_2^{-\kappa} + \mathcal{I}_n$$

with

$$\mathcal{I}_n = I\left\{100\sqrt{\frac{n_1}{n_2}}\frac{A(n_2, j_n^*(t) + 1, \kappa \log n_2)}{A(n_1, j_n^*(t) + u_n - m, x)} > 2^{(u_n - m - 1)(\frac{d}{2} + t)}\right\},$$

with $\kappa > 0$ *equal to the constant from after* (26) *and where* $A(n, l, x)$ *was defined in* (21).

Remark 1 (Undersmoothing in finite samples) Note first that if $u_n \geq m$, then the fraction on the l.h.s.of the inequality in the definition of $\mathcal{I}_n$ is bounded away from zero and infinity. Consequently the tradeoff between the constants u_n and $b(n)$ is such that if $u_n + t^{-1}\log_2(b(n)) \to \infty$ then $\mathcal{I}_n = 0$ for all n from some n_0 onwards, which in particular implies Theorem 4. Not surprisingly obtaining coverage in finite samples is more delicate, as n_0 depends on f: The undersmoothing constant u_n should be chosen so large that $\mathcal{I}_n = 0$ for every n. Closer inspection of $\mathcal{I}_n$ shows that this is possible if an upper bound for m is available, which can be obtained by requiring an a priori lower bound for the sequence $b(n)$ as well as for t. The discussion in Sect. 4.4 will show that such a priori bounds can indeed be obtained in relevant cases.

Remark 2 (Admissible lower bounds in Condition 1) Another point of view is to start with an undersmoothing sequence u_n and to ask which sequences of $b(n)$'s are admissible to obtain coverage. Assume for simplicity that the sample size is $2n$ and that $n_1 = n_2 = n$. Let $C_n(\kappa \log n, y)$ be the confidence band from (27) with undersmoothing sequence $u_n \in \mathbb{N}$ and $x = \kappa \log n$. If f satisfies Condition 1 and if

$$b(n) \geq 7b_2 \cdot (100)^{t/(t+d/2)} 2^{(-u_n+2)t},$$

then

$$\Pr\{f(y) \in C_n(\kappa \log n, y) \quad \text{for every} \quad y \in \Omega\} \geq 1 - (2j_{\max} + 3)n^{-\kappa}. \tag{34}$$

For instance if $d = 2$ and f is at least once differentiable, then finite sample coverage holds for the set of densities that satisfy Condition 1 for $1 \leq t < \infty$ and $b(n) \geq b_2 \cdot 2^{8.2-u_n}$.

Remark 3 (The role of the thresholding constant κ) The thresholding constant κ plays an important role in the construction of $\hat{j}_n$. Our results are presented for fixed κ without any restriction on this constant. This is an advantage since this constant has to be carefully chosen in applications. Our bounds typically contain a term of the form $n^{-\kappa}$, and one could be tempted to choose κ as large as possible, however it is important to notice that choosing κ very large will increase the difficulty of cancelling $\mathcal{I}_n$ in Theorem 5. An adaptive choice of this tuning constant is possible but beyond the scope of this paper.

4.4 Regularity of functions on the sphere and Condition 1

Condition 1 can be characterized in terms of classical Hölder regularity properties of the unknown density $f : \mathbf{M} \to \mathbb{R}$. We shall only discuss the case $\mathbf{M} = \mathbb{S}^d$, which is the case of primary statistical interest, but all findings below generalize to $\mathbf{M}$ with suitable modifications.

There are several ways to approximate unknown functions defined on $\mathbb{S}^d$, but it is a fortiori not clear whether a given method retrieves the natural intuition that the degree of smoothness of a function f is the driving quantity of the approximation properties of f. For instance, while $L^2(\mathbf{M})$-projections onto spherical harmonics constitute a way of approximating a continuous function $f : \mathbb{S}^d \to \mathbb{R}$, it is well known already from the special case $d = 1$ that this approximation may diverge at any given point x, which is particularly worrying when one is interested in the local or even uniform behavior of the approximation errors. Furthermore the important question arises whether the approximation method allows for very smooth (for instance infinitely differentiable) functions to be approximated in an optimal way.

The fact that needlets form a tight frame of $L^2(\mathbb{S}^d)$ implies good approximation properties in that space, similar to those of the spherical harmonics. Moreover, these approximations are also optimal approximands for differentiable and Hölder-continuous functions in the uniform norm on $\mathbb{S}^d$ (as follows from the results in [8]), so the upper bound in Condition 1 has a natural interpretation in terms of Hölder–Zygmund-norms on $\mathbb{S}^d$.

The lower bound in Condition 1 is more intricate. The results in Jaffard [15] and Giné and Nickl [11] for functions on $\mathbb{R}$ suggest that this condition should be satisfied if f 'attains t as its Hölder exponent' viewed as a function on the unit sphere (in fact a slightly stronger requirement is necessary). In the simplest case, if a real-valued function f defined on $\mathbb{R}$ scales like $|x - x_0|^t$ at some point x_0 (if $t > 1$ a similar property has to hold for the highest existing derivative), then f attains the Hölder exponent t, and the results in Jaffard [15] imply that 'quasi every' function (in a Baire sense) in $\mathcal{C}^t(\mathbb{R})$ does this. Indeed Proposition 4 in Giné and Nickl [11] implies that quasi-every function in $\mathcal{C}^t(\mathbb{R})$ satisfies the lower bound in the $\mathbb{R}$-analogue of Condition 1 (where $A_j(f)$ has to be replaced by a corresponding wavelet projection). Proving such general results in the case where f is defined on the sphere is technical, mostly since needlets only form a tight frame but not an orthonormal basis. We therefore return to the intuition of Hölder exponents and show that 'typical' α-Hölder functions on $\mathbb{S}^d$ satisfy Condition 1: let us consider spherical analogues of functions on $\mathbb{R}$ that scale like $|x - x_0|$: If x_0 is any point in $\mathbb{S}^d$, then the zonal functions $d_{\mathbb{S}^d}(x, x_0)$ or $(1 - \langle x, x_0\rangle_{d+1})^{1/2}$ are natural candidates for the class $\mathcal{C}^1(\mathbb{S}^d)$. More generally

$$f_\alpha(x) = (1 - \langle x, x_0\rangle_{d+1})^{\alpha/2}$$

for $0 < \alpha < \infty$, $\alpha/2 \notin \mathbb{N}$, is a natural candidate for $\mathcal{C}^{\alpha}(\mathbb{S}^d)$. We prove in Proposition 6 below

$$b_1 2^{-j\alpha} \leq \|A_j(f_\alpha) - f_\alpha\|_\infty \leq b_2 2^{-j\alpha}$$

for some fixed constants $0 < b_1 < b_2 < \infty$. Note that obviously, for $\alpha = 2k,\ k \in \mathbb{N}$, $f_\alpha(x) = (1 - \langle x, x_0\rangle_{d+1})^k = 1 - \cos(d_{\mathbb{S}^d}(x, x_0))^k$ is actually a polynomial on $\mathbb{S}^d$.

5 Proofs for Section 3

5.1 Proof of Theorem 1

If f is the volume element of $\mathbf{M}$, then

$$\|A_j(f) - f\|_\infty = 0 \tag{35}$$

for every $j \geq 0$. Clearly by definition of $\hat{j}_n$

$$\Pr\left\{\hat{j}_n \neq 0\right\} \leq \sum_{l \in \mathcal{J}: l > 0} \Pr\left\{\left\| f_{n_2}(0) - f_{n_2}(l)\right\|_\infty > 4\sigma(n, l)\right\}.$$

Now since $Ef_n(l) = A_l(f) = f$ for every $l \geq 0$, the l-th probability is bounded by

$$\begin{aligned}
&\Pr\left\{\left\| f_{n_2}(0) - f_{n_2}(l) - Ef_{n_2}(0) + Ef_{n_2}(l)\right\|_\infty > 4\sigma(n, l)\right\} \\
&\quad \leq \Pr\left\{\left\| f_{n_2}(0) - Ef_{n_2}(0)\right\|_\infty > 2\sigma(n, l)\right\} \\
&\qquad + \Pr\left\{\left\| f_{n_2}(l) - Ef_{n_2}(l)\right\|_\infty > 2\sigma(n, l)\right\} \leq 2n_2^{-2\kappa}
\end{aligned}$$

in view of Proposition 3, so that $\Pr\{\hat{j}_n \neq 0\} \leq 2 j_{\max} n_2^{-2\kappa}$ follows. To prove the second claim of the theorem, we have from independence of $\hat{j}_n$ and f_{n_1}, from (35) and from Proposition 3

$$\begin{aligned}
&\Pr\{f(y) \in C_n(x, y) \quad \text{for every} \quad y \in \mathbf{M}\} \\
&\quad = \Pr\left\{\sup_{y \in \mathbf{M}} \left|\hat{f}_n(y) - f(y)\right| \leq s_n(x)\right\} \\
&\quad \geq 1 - \Pr\left\{\sup_{y \in \mathbf{M}} \left| f_{n_1}(\hat{j}_n + u_n, y) - E_1 f_n(\hat{j}_n + u_n, y)\right| > \sigma(n_1, \hat{j}_n + u_n, x)\right\} \\
&\quad = 1 - \sum_{0 \leq l \leq j_{\max}} \Pr\left\{\| f_{n_1}(l + u_n, \cdot) - E_1 f_{n_1}(l + u_n, \cdot)\|_\infty > \sigma(n_1, l + u_n, x)\right\} \Pr\{\hat{j}_n = l\} \\
&\quad \geq 1 - e^{-x} \sum_{0 \leq l \leq j_{\max}} \Pr\{\hat{j}_n = l\} = 1 - e^{-x}.
\end{aligned}$$

The last claim of Theorem 1 follows from the first and definition of $\sigma(n, l, x)$.

5.2 Proof of Theorem 2

Since $j_{\max} \to \infty$ as $n \to \infty$ and since this theorem is of an asymptotic nature we assume $J \leq j_{\max}$ in what follows. We recall from (3) that $f \in E_{2^{2J-1}}$ implies

$$\|A_l(f) - f\|_\infty = 0 \tag{36}$$

for every $l \geq J$. Then

$$\Pr\left\{\hat{j}_n > J\right\} \leq \sum_{l \in \mathcal{J}: l > J} \Pr\left\{\left\| f_{n_2}(J) - f_{n_2}(l)\right\|_\infty > 4\sigma(n,l)\right\},$$

and the l'th summand is bounded by

$$\begin{aligned}&\Pr\left\{\left\| f_{n_2}(J) - f_{n_2}(l) - Ef_{n_2}(J) + Ef_{n_2}(l)\right\|_\infty > 4\sigma(n,l)\right\}\\&\quad\leq \Pr\left\{\left\| f_{n_2}(J) - Ef_{n_2}(J)\right\|_\infty > 2\sigma(n,l)\right\}\\&\qquad + \Pr\left\{\left\| f_{n_2}(l) - Ef_{n_2}(l)\right\|_\infty > 2\sigma(n,l)\right\} \leq 2n_2^{-2\kappa}\end{aligned}$$

in view of (36) and Proposition 3.

For integer $l < J-1$ (so that $2^l < 2^{J-1}$) we have

$$\|A_l(f) - f\|_\infty \geq \inf_{p \in E_{2^{J-2}}} \|p - f\|_\infty \equiv b_1 > 0$$

since $A_l(f) \in E_{2^{2J-2}}$ and since $E_{2^{2J-2}}$ is a closed proper subspace of $E_{2^{2J-1}}$. By definition we have

$$\Pr(\hat{j}_n = l) \leq \Pr\left(\left\| f_{n_2}(l) - f_{n_2}(J)\right\|_\infty \leq 4\sigma(n,J)\right). \tag{37}$$

The triangle inequality and (36) now give

$$\left\| f_{n_2}(l) - f_{n_2}(J)\right\|_\infty \geq \|A_l(f) - f\|_\infty - \left\| f_{n_2}(l) - Ef_{n_2}(l) - f_{n_2}(J) + Ef_{n_2}(J)\right\|_\infty$$

so that the probability in (37) is bounded by

$$\begin{aligned}&\Pr\left(\left\| f_{n_2}(l) - Ef_{n_2}(l) - f_{n_2}(J) + Ef_{n_2}(J)\right\|_\infty \geq b_1 - 4\sigma(n,J)\right)\\&\quad\leq \Pr\left(\left\| f_{n_2}(l) - Ef_{n_2}(l)\right\|_\infty \geq \frac{b_1}{2} - 2\sigma(n,J)\right)\\&\qquad + \Pr\left(\left\| f_{n_2}(J) - Ef_{n_2}(J)\right\|_\infty \geq \frac{b_1}{2} - 2\sigma(n,J)\right).\end{aligned}$$

For n large enough depending on b_1 we have $2\sigma(n,J) \leq b_1/4$ so that Proposition 3 implies, for J fixed, $\Pr\{\hat{j}_n < J-1\} \leq \sum_{0 \leq l < J-1} \Pr\{\hat{j}_n = l\} \leq 2Je^{-cn}$ for some constant $c > 0$ depending on b_1, J and those constants appearing in the definition of $\sigma(n,l,x)$ that do not depend on n, l. Summarizing we deduce $\Pr\{\hat{j}_n \notin$

$[J-1, J]\} \leq 2 j_{\max} n_2^{-2\kappa} + 2Je^{-cn}$ for n large enough. To prove coverage we proceed as in Theorem 1, noting $u_n > 1$,

$$\begin{aligned}
&\Pr\{f(y) \in C_n(x, y) \quad \text{for every} \quad y \in \mathbf{M}\} \\
&\geq 1 - \Pr\left\{\sup_{y \in \mathbf{M}} \left| f_{n_1}(\hat{j}_n + u_n, y) - f(y) \right| > \sigma(n_1, \hat{j}_n + u_n, x)\right\} \\
&\geq 1 - 2Je^{-cn} \\
&\quad - \sum_{J-1 \leq l \leq j_{\max}} \Pr\left\{\| f_{n_1}(l+u_n, \cdot) - E_1 f_{n_1}(l+u_n, \cdot)\|_\infty > \sigma(n_1, l+u_n, x)\right\} \Pr\{\hat{j}_n = l\} \\
&\geq 1 - 2Je^{-cn} - e^{-x} \sum_{J-1 \leq l \leq j_{\max}} \Pr\{\hat{j}_n = l\} \geq 1 - e^{-x} - 2Je^{-cn}
\end{aligned}$$

where we used (36) and Proposition 3. The last claim of the theorem is proved as in Theorem 1.

5.3 Proof of Theorems 4 and 5

We first prove Theorem 5. For f satisfying Condition 1 there exists a unique $t := t(f)$ such that f satisfies Condition 1 for this t. Define

$$B(j, t) = b_2 2^{-jt}, \; j_n^*(t) = \min\{j \in \mathcal{J} \setminus \{0\} : B(j, t) \leq \sigma(n_2, j)\} - 1. \qquad (38)$$

If no $j \in \mathcal{J}$ exists such that $B(j, t) \leq \sigma(n_2, j)$ we set $j_n^*(t) = j_{\max} - 1$. We shall assume without loss of generality that b_2 is large enough such that $b_2 \geq \sigma(1, 0)$. In this way $B(j_n^*(t)) \geq \sigma(n_2, j_n^*(t))$ also holds when $j_n^*(t) = 0$.

It is easy to see that $j_n^*(t)$ satisfies

$$2^{j_n^*(t)} \simeq \left(\frac{n_2}{\log n_2}\right)^{\frac{1}{2t+d}}, \qquad (39)$$

so is a 'rate optimal' resolution level for estimating f satisfying Condition 1 for the given t. The constants in the definition of $j_n^*(t)$ depend only on b_2, t, a, d, k_2 and $\|f\|_\infty$.

Lemma 1 (a) *For every $n \in \mathbb{N}$,*

$$\Pr(\hat{j}_n > j_n^*(t) + 1) \leq 2(j_{\max} - j_n^*(t)) n_2^{-\kappa}. \qquad (40)$$

(b) *Let $m := \min(j_n^*(t), m^*)$ where m^* is the smallest integer such that $(b(n)/b_2)2^{tm^*} \geq 7$. Then, for every $j \in \mathcal{J}$ satisfying $0 \leq j < j_n^*(t) - m$ and every $n \in \mathbb{N}$ we have $\Pr(\hat{j}_n = j) \leq 2n_2^{-\kappa}$. As a consequence, for every $n \in \mathbb{N}$,*

$$\Pr\left(\hat{j}_n < j_n^*(t) - m\right) \leq 2(j_n^*(t) - m) n_2^{-\kappa} \qquad (41)$$

Proof Since this lemma only involves the sample $\mathcal{S}_2$, we set $n = n_2$ for notational simplicity. We also put $j_n^{*+} = j_n^*(t) + 1$. If $j_n^{*+} = j_{\max}$ Part (a) is proved. Otherwise one has

$$\Pr(\hat{j}_n > j_n^{*+}) \leq \sum_{l \in \mathcal{J}: l > j_n^{*+}} \Pr\left(\left\| f_n(j_n^{*+}) - f_n(l)\right\|_\Omega > 4\sigma(n, l)\right).$$

We first observe that by Condition 1 (noting also $Ef_n(j) = A_j(f)$)

$$\left\| f_n(j_n^{*+}) - f_n(l)\right\|_\Omega \leq \left\| f_n(j_n^{*+}) - f_n(l) - Ef_n(j_n^{*+}) + Ef_n(l)\right\|_\Omega + B(j_n^{*+}, t) + B(l, t),$$

and that

$$B(j_n^{*+}, t) + B(l, t) \leq 2B(j_n^{*+}, t) \leq 2\sigma(n, j_n^{*+}) \leq 2\sigma(n, l)$$

by definition of $j_n^*(t)$ and since $l > j_n^{*+}$. Consequently, the l-th probability in the last sum is bounded by

$$\begin{aligned}
&\Pr\left(\left\| f_n(j_n^{*+}) - f_n(l) - Ef_n(j_n^{*+}) + Ef_n(l)\right\|_\Omega > 2\sigma(n, l)\right) \\
&\quad \leq \Pr\left(\left\| f_n(j_n^{*+}) - Ef_n(j_n^{*+})\right\|_\Omega > \sigma(n, l)\right) + \Pr\left(\| f_n(l) - Ef_n(l)\|_\Omega > \sigma(n, l)\right) \leq 2n^{-\kappa}
\end{aligned}$$

where we have used Proposition 3.

To prove the second claim, fix $j < j_n^*(t) - m$. Clearly we only have to consider the case $m = m^*$. Observe that

$$\Pr(\hat{j}_n = j) \leq \Pr\left(\left\| f_n(j) - f_n(j_n^*(t))\right\|_\Omega \leq 4\sigma(n, j_n^*(t))\right). \tag{42}$$

Now using Condition 1 and the triangle inequality we deduce

$$\begin{aligned}
\left\| f_n(j) - f_n(j_n^*(t))\right\|_\Omega \geq\ & \frac{b(n)}{b_2} B(j, t) - B(j_n^*(t), t) \\
& - \left\| f_n(j) - Ef_n(j) - f_n(j_n^*(t)) + Ef_n(j_n^*(t))\right\|_\Omega
\end{aligned}$$

so that the probability in (42) is bounded by

$$\begin{aligned}
\Pr\Bigg(&\left\| f_n(j) - Ef_n(j) - f_n(j_n^*(t)) + Ef_n(j_n^*(t))\right\|_\Omega \geq \frac{b(n)}{b_2} B(j, t) - B(j_n^*(t), t) \\
&- 4\sigma(n, j_n^*(t))\Bigg).
\end{aligned}$$

By definition of $j_n^*(t)$ and $B(j,t)$, we have

$$\frac{b(n)}{b_2}B(j,t)-B(j_n^*(t),t)=\frac{b(n)}{b_2}2^{t(j_n^*(t)-j)}B(j_n^*(t),t)-B(j_n^*(t),t)$$
$$>\left(\frac{b(n)}{b_2}2^{tm}-1\right)B(j_n^*(t),t)$$

as well as $B(j_n^*(t),t)\geq\sigma(n,j_n^*(t))\geq\sigma(n,j)$ so that the last probability is bounded by

$$\Pr\Bigg(\left\|f_n(j)-Ef_n(j)-f_n(j_n^*(t))+Ef_n(j_n^*(t))\right\|_\Omega$$
$$\geq\left[\left(\frac{b(n)}{b_2}2^{tm}-1\right)-4\right]\sigma(n,j_n^*(t))\Bigg)$$
$$\leq\Pr\left(\|f_n(j)-Ef_n(j)\|_\Omega\geq 2^{-1}\left(\frac{b(n)}{b_2}2^{tm}-5\right)\sigma(n,j)\right)$$
$$+\Pr\left(\left\|f_n(j_n^*(t))-Ef_n(j_n^*(t))\right\|_\Omega\geq 2^{-1}\left(\frac{b(n)}{b_2}2^{tm}-5\right)\sigma(n,j_n^*(t))\right)$$

By definition of m, the term in brackets is greater than or equal to two, and then—using Proposition 3—the last two probabilities do not exceed $2n^{-\kappa}$. Moreover,

$$\Pr\left(\hat{j}_n<j_n^*(t)-m\right)=\sum_{0\leq j<j_n^*(t)-m}\Pr(\hat{j}_n=j)\leq 2\sum_{0\leq j<j_n^*(t)-m}n^{-\kappa}$$
$$\leq 2(j_n^*(t)-m)n^{-\kappa},$$

which completes the proof. □

Combining (40) with (41) we have, for every $n\in\mathbb{N}$ and for m as in the lemma

$$\Pr\{\hat{j}_n\notin[j_n^*(t)-m,j_n^*(t)+1]\}\leq 2[(j_n^*(t)-m)+(j_{\max}-j_n^*(t))]n_2^{-\kappa}$$
$$=2(j_{\max}-m)n_2^{-\kappa}:=Z_n,\tag{43}$$

a fact we shall use below.

We now prove Theorem 5. Denoting by E_1 expectation w.r.t. $\mathcal{S}_1$, one has by definition of $s_n(x)$ that

$$\Pr\{f(y)\in C_n(x,y)\ \text{ for every }\ y\in\Omega\}$$
$$=\Pr\left\{\sup_{y\in\Omega}\left|\hat{f}_n(y)-f(y)\right|\leq s_n(x)\right\}$$
$$=1-\Pr\left\{\sup_{y\in\Omega}\left|\hat{f}_n(y)-E_1\hat{f}_n(y)+E_1\hat{f}_n(y)-f(y)\right|>1.01\sigma(n_1,\hat{j}_n+u_n,x)\right\}$$

$$\geq 1 - \Pr\left\{\|\hat{f}_n - E_1\hat{f}_n\|_\Omega > \sigma(n_1, \hat{j}_n + u_n, x)\right\}$$
$$- \Pr\left\{\|E_1\hat{f}_n - f\|_\Omega > 0.01\sigma(n_1, \hat{j}_n + u_n, x)\right\}$$
$$= 1 - I - II$$

About term I: This probability equals, by independence of $f_{n_1}(j, y)$ and $\hat{j}_n$,

$$\Pr\left\{\|f_{n_1}(\hat{j}_n + u_n, \cdot) - E_1 f_{n_1}(\hat{j}_n + u_n, \cdot)\|_\Omega > \sigma(n_1, \hat{j}_n + u_n, x)\right\}$$
$$= \sum_{0\leq l\leq j_{\max}} \Pr\left\{\|f_{n_1}(l+u_n, \cdot) - E_1 f_{n_1}(l+u_n, \cdot)\|_\Omega > \sigma(n_1, l+u_n, x)\right\} \Pr\{\hat{j}_n = l\}$$
$$\leq e^{-x} \sum_{0\leq l\leq j_{\max}} \Pr\{\hat{j}_n = l\} = e^{-x}$$

in view of Proposition 3.

About term II: Using Condition 1 as well as (43), and recalling (21), this quantity equals

$$\Pr\left\{\left\|Ef_{n_1}(\hat{j}_n + u_n) - f\right\|_\Omega > 0.01\sigma(n, \hat{j}_n + u_n, x)\right\}$$
$$\leq \Pr\left\{100 b_2 2^{-t(\hat{j}_n+u_n)} > \sigma(n_1, \hat{j}_n + u_n, x)\right\}$$
$$= \Pr\left\{100\sqrt{n_1} b_2 > 2^{(\hat{j}_n+u_n)\left(\frac{d}{2}+t\right)} A(n_1, \hat{j}_n + u_n, x)\right\}$$
$$\leq \sum_{j_n^*(t)-m\leq l\leq j_n^*(t)+1} I\left\{100\sqrt{n_1} b_2 > 2^{(l+u_n)\left(\frac{d}{2}+t\right)} A(n_1, l+u_n, x)\right\} \Pr\{\hat{j}_n = l\} + Z_n$$
$$\leq I\left\{\frac{100 b_2\sqrt{n_1}}{A(n_1, j_n^*(t) + u_n - m, x)} > 2^{(j_n^*(t)+1)\left(\frac{d}{2}+t\right)} 2^{(u_n-m-1)\left(\frac{d}{2}+t\right)}\right\} + Z_n$$
$$\leq I\left\{100\sqrt{\frac{n_1}{n_2}} \frac{A(n_2, , j_n^*(t) + 1, \kappa \log n_2)}{A(n_1, j_n^*(t) + u_n - m, x)} > 2^{(u_n-m-1)\left(\frac{d}{2}+t\right)}\right\} + Z_n$$

where we have used that (38) implies

$$2^{(j_n^*(t)+1)\left(\frac{d}{2}+t\right)} \geq \frac{\sqrt{n}_2 b_2}{A(n_2, j_n^*(t) + 1, \kappa \log n_2)}$$

in the last inequality. This proves Theorem 5. Theorem 4 follows from Theorem 5 using that tradeoff between $b(n)$ and u_n through the constant m (cf. also Remark 1).

5.4 Proof of Theorem 3

The size of the band is $2.02\sigma(n_1, \hat{j}_n + u_n, x)$. In view of (40)—whose proof only requires the hypotheses of Theorem 3—we have $\hat{j}_n \leq j_n^*(t) + 1$ with probability larger than $1 - 2(j_{\max} - j_n^*(t))n_2^{-\kappa}$, so the size of this band is less than or equal to $2.02\sigma(n_1, j_n^*(t) + u_n + 1, x)$ with the same probability bound. The second claim of Theorem 3 then follows from definition of $\sigma(n, l, x)$ (cf. (21)) and of $j_n^*(t)$ (cf. (39)).

6 Precise validity of condition (31)

In this section we investigate examples of functions verifying condition (31) if $\mathbf{M} = \mathbb{S}^d$. Let us recall that the projection kernel on $\mathcal{H}_k(\mathbb{S}^d)$ is given by

$$L_k(\langle x, y\rangle_{d+1}) = \frac{1}{|\mathbb{S}^d|}\left(1 + \frac{k}{\nu}\right) C_k^\nu(\langle x, y\rangle_{d+1}), \quad \nu = \frac{d-1}{2}$$

where $C_k^\nu(x)$ is the corresponding Gegenbauer polynomial. For ease of notation we shall redefine $A_j(x, y) = \sum_{k<2^j} a(k/2^j)L_k(x, y)$, to be in line with the notation in [2,24,25]. (For $j \to \infty$ this modification is immaterial.) We shall use the classical symbol

$$\forall k \in \mathbb{N},\, (a)_k = a(a+1)\cdot(a+k-1)\left(= \frac{\Gamma(a+k)}{\Gamma(a)} \text{if } -a \notin \mathbb{N}\right),\, (a)_0 = 1.$$

The following Olindes Rodrigues formula defines the Gegenbauer polynomials and is useful for integration by parts: for $t \in I = [-1, 1]$

$$C_k^\nu(t) = (-1)^k \frac{1}{k!2^k} \frac{(2\nu)_k}{\left(\nu + \frac{1}{2}\right)_k} \frac{D^k\{(1-t^2)^k)\omega^\nu(t)\}}{\omega^\nu(t)}, \quad \omega^\nu(t) = (1-t^2)^{\nu-1/2}. \tag{44}$$

Proposition 6 *For* $0 < \alpha < \infty,\ \frac{\alpha}{2} \notin \mathbb{N}$, *we define the following functions:*

$$f_\alpha(x) = \left(\sqrt{1 - \langle x, x_0\rangle_{d+1}}\right)^\alpha = \left(\sqrt{1 - \cos(d_{\mathbb{S}^d}(x, x_0))}\right)^\alpha$$

where $d_{\mathbb{S}^d}$ *is the geodesic distance on* $\mathbb{S}^d$. *Then there exist constants* $c_1 > 0,\ c_2 > 0$ *independent of* j *such that*

$$c_1 2^{-j\alpha} \le \|A_j(f_\alpha) - f_\alpha\|_\infty \le c_2 2^{-j\alpha}.$$

Proof of the upper bound: Let us consider first the case $0 < \alpha \leq 1$. We have

$$
\begin{aligned}
|A_j(f_\alpha)(x) - f_\alpha(x)| &= \left| \int_{\mathbb{S}^{d-1}} A_j(x, y) f_\alpha(y) dy - f_\alpha(x) \right| \\
&= \left| \int_{\mathbb{S}^d} A_j(x, y)(f_\alpha(y) - f_\alpha(x)) dy \right| \\
&\leq \int_{\mathbb{S}^d} |A_j(x, y)||f_\alpha(y) - f_\alpha(x)| dy
\end{aligned}
$$

But

$$
\forall \theta, \theta' \in [0, \pi], \quad |\sqrt{1 - \cos\theta} - \sqrt{1 - \cos\theta'}| = \sqrt{2} \left| \sin\frac{\theta}{2} - \sin\frac{\theta'}{2} \right| \leq \frac{1}{\sqrt{2}} |\theta - \theta'|,
$$

so

$$
\begin{aligned}
|f_1(x) - f_1(y)| &= |\sqrt{1 - \cos(d_{\mathbb{S}^d}(x, x_0))} - \sqrt{1 - \cos(d_{\mathbb{S}^d}(y, x_0))}| \\
&\leq \frac{1}{\sqrt{2}} |d_{\mathbb{S}^d}(x, x_0) - d_{\mathbb{S}^d}(y, x_0)| \leq \frac{1}{\sqrt{2}} d_{\mathbb{S}^d}(x, y)
\end{aligned}
$$

And, by the subadditivity of $x \mapsto x^\alpha$ for $0 < \alpha \leq 1$

$$
|f_\alpha(x) - f_\alpha(y)| = |f_1^\alpha(x) - f_1^\alpha(y)| \leq |f_1(x) - f_1(y)|^\alpha \leq \frac{1}{2^{\alpha/2}} (d_{\mathbb{S}^d}(x, y))^\alpha
$$

So, by the integration formula for zonal functions on the sphere (Section 9.1 in [5]):

$$
\begin{aligned}
\forall x \in \mathbb{S}^d, \quad & \int_{\mathbb{S}^d} |A_j(x, y)||f_\alpha(y) - f_\alpha(x)| dy \\
&\leq 2^{-\alpha/2} \int_{\mathbb{S}^d} |A_j(\langle x, y \rangle_{d+1})| (d_{\mathbb{S}^d}(x, y))^\alpha dy \\
&= 2^{-\alpha/2} |\mathbb{S}^{d-1}| \int_0^\pi A_j(\cos\theta) \theta^\alpha (\sin\theta)^{d-1} d\theta \\
&\leq 2^{-\alpha/2} |\mathbb{S}^{d-1}| \int_0^\pi A_j(\cos\theta) \theta^{d-1+\alpha} d\theta
\end{aligned}
$$

But using the following concentration result from [25]

$$
\forall K > 0, \exists\, C_K < \infty, \quad A_j(\cos\theta) \leq C_K 2^{jd} [1 \wedge 1/(2^j \theta)^K]
$$

Taking $K > d + \alpha$, we obtain

$$\|A_j(f) - f\|_\infty \leq 2^{-\alpha/2}|\mathbb{S}^{d-1}|C_K 2^{jd}\left(\int_0^{2^{-j}} \theta^{d-1+\alpha}d\theta + \int_{2^{-j}}^1 \theta^{d-1+\alpha}\frac{1}{(2^j\theta)^K}d\theta\right)$$
$$\leq 2^{-\alpha/2}|\mathbb{S}^{d-1}|C_K 2^{-j\alpha}\frac{K}{(d+\alpha)(K-d-\alpha)}$$

Let us now consider the case $\alpha > 1$. Taking $d = 2$ the previous proof shows that, on the classical torus $\mathbb{T}$, for $0 < \alpha \leq 1$, the 2π-periodical function $\phi_\alpha(\theta) = (\sqrt{1-\cos\theta})^\alpha = 2^\alpha|\sin\frac{\theta}{2}|^\alpha$ belongs to $\mathcal{C}^\alpha(\mathbb{T})$. But, if for k in $\mathbb{N}, \alpha$ equals $\alpha = k+\beta \leq k+1$, it is clear that $\phi_\alpha(\theta)$ is k-times differentiable, and $D^k\phi_\alpha(\theta)$ as a linear combination of C_∞ periodical functions times $|\sin\frac{\theta}{2}|^{\beta+j}, \ j = 0, 1, \ldots, k$ belongs to $\mathcal{C}^\beta(\mathbb{T})$. So, $\phi_\alpha \in \mathcal{C}^\alpha(\mathbb{T})$, and, as moreover $\phi_\alpha(\theta)$ is even, there exists $P_j(\cos\theta)$, a sequence of trigonometrical polynomials of degree less than 2^j such that:

$$\|(\sqrt{1-\cos\theta})^\alpha - P_j(\cos\theta)\|_\infty \leq C2^{-j\alpha}$$

But $P_j(\cos\langle x, x_0\rangle_{d+1})$ is a polynomial on the sphere of degree less than 2^j and

$$\|(\sqrt{1-\cos\langle x, x_0\rangle_{d+1}})^\alpha - P_j(\langle x, x_0\rangle_{d+1})\|_\infty \leq C2^{-j\alpha}.$$

Proof of the lower bound We only have to consider the case j large enough since f_α is not a spherical polynomial and thus not in $E_N(\mathbb{S}^d)$ for any finite N. Using again the integration fomulae for zonal functions

$$\|A_j(f_\alpha) - f_\alpha\|_\infty \geq |A_j(f_\alpha)(x_0) - f_\alpha(x_0)|$$
$$= \left|\int_{\mathbb{S}^d} A_j(x_0, y)(f_\alpha(y) - f_\alpha(x_0))dy\right|$$
$$= \left|\int_{\mathbb{S}^d} A_j(x_0, y)(\sqrt{1-\langle y, x_0\rangle_{d+1}})^\alpha dy\right|$$
$$= |\mathbb{S}^{d-1}|\int_0^\pi A_j(\cos\theta)(\sqrt{1-\cos\theta})^\alpha(\sin\theta)^{d-1}d\theta|$$
$$= |\mathbb{S}^{d-1}|\int_I A_j(t)(1-t)^{\alpha/2}(1-t^2)^{\nu-1/2}dt|$$
$$= \frac{|\mathbb{S}^{d-1}|}{|\mathbb{S}^d|}\left|\sum_{0\leq k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)\int_I C_k^\nu(t)(1-t)^{\alpha/2}(1-t^2)^{\nu-1/2}dt\right|$$

But, using (44)

$$\int_I C_k^\nu(t)(1-t)^{\alpha/2}(1-t^2)^{\nu-1/2}dt = \int_I C_k^\nu(t)(1-t)^{\alpha/2}\omega^\nu(t)dt$$
$$= (-1)^k \frac{1}{k!2^k}\frac{(2\nu)_k}{\left(\nu+\frac{1}{2}\right)_k}\int_I (1-t)^{\alpha/2} D^k\{(1-t^2)^k\omega^\nu(t)\}dt$$
$$= \frac{1}{k!2^k}\frac{(2\nu)_k}{\left(\nu+\frac{1}{2}\right)_k}\int_I D^k\{(1-t)^{\alpha/2}\}(1-t^2)^k\omega^\nu(t)dt$$
$$= \frac{1}{k!2^k}\frac{(2\nu)_k}{\left(\nu+\frac{1}{2}\right)_k}\left(-\frac{\alpha}{2}\right)_k\int_I (1-t)^{\alpha/2-k}(1-t^2)^k\omega^\nu(t)dt$$
$$= \frac{1}{k!2^k}\frac{(2\nu)_k}{\left(\nu+\frac{1}{2}\right)_k}\left(-\frac{\alpha}{2}\right)_k\int_I (1-t)^{\alpha/2}(1+t)^k\omega^\nu(t)dt = u_k$$

Clearly $\forall k \geq 0, u_k \neq 0$ (because $\frac{\alpha}{2} \notin \mathbb{N}$), $u_k = (-1)^k|u_k|$ for $0 \leq k < \frac{\alpha}{2}+1$ and $u_k = -(-1)^{[\alpha/2]}|u_k|$ for $k > \frac{\alpha}{2}+1$. By the upper bound, and for j large enough:

$$C2^{-j} \geq \|A_j f_\alpha - f_\alpha\|_\infty \geq \frac{|\mathbb{S}^{d-1}|}{|\mathbb{S}^d|}\left|\sum_{0\leq k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)u_k\right|$$
$$= \frac{|\mathbb{S}^{d-1}|}{|\mathbb{S}^d|}\left|\sum_{0\leq k<\alpha/2+1} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)(-1)^k|u_k| - (-1)^{[\alpha/2]}\right.$$
$$\left.\times\sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)|u_k|\right|$$
$$= \frac{|\mathbb{S}^{d-1}|}{|\mathbb{S}^d|}\left|\sum_{0\leq k\leq[\alpha/2]}\left(1+\frac{k}{\nu}\right)u_k - (-1)^{[\alpha/2]}\sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)|u_k|\right|$$

So if $[\alpha/2]$ is even, and j large enough

$$\left|\sum_{0\leq k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)u_k\right| = \sum_{0\leq k\leq[\alpha/2]}\left(1+\frac{k}{\nu}\right)u_k$$
$$- \sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right)|u_k|$$

$$= \sum_{0\le k\le[\alpha/2]} \left(1+\frac{k}{\nu}\right) u_k + \sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k.$$

So

$$\sum_{0\le k<2^j} \left(1+\frac{k}{\nu}\right) u_k \le \left|\sum_{0\le k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k\right| \le \sum_{0\le k<2^{j-1}} \left(1+\frac{k}{\nu}\right) u_k.$$

Now, if $[\alpha/2]$ is odd, and j large enough

$$\begin{aligned}\left|\sum_{0\le k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k\right| &= -\left(\sum_{0\le k\le[\alpha/2]} \left(1+\frac{k}{\nu}\right) u_k\right. \\ &\quad \left. + \sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) |u_k|\right) \\ &= -\left(\sum_{0\le k\le[\alpha/2]} \left(1+\frac{k}{\nu}\right) u_k\right. \\ &\quad \left. + \sum_{\alpha/2+1<k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k\right)\end{aligned}$$

So

$$-\sum_{0\le k<2^j} \left(1+\frac{k}{\nu}\right) u_k \le \left|\sum_{0\le k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k\right| \le -\sum_{0\le k<2^{j-1}} \left(1+\frac{k}{\nu}\right) u_k,$$

and in any case

$$\left|\sum_{0\le k<2^j} a\left(\frac{k}{2^j}\right)\left(1+\frac{k}{\nu}\right) u_k\right| \sim \left|\sum_{0\le k<2^j} \left(1+\frac{k}{\nu}\right) \int_I C_k^\nu(t)(1-t)^{\alpha/2}(1-t^2)^{\nu-1/2} dt\right|$$

Denote now by $\langle\cdot,\cdot\rangle_\nu$ the $L^2([-1,1])$-inner product w.r.t. ω^ν and recall (see [1, p. 343])

$$\sum_{0\le k\le n} \left(1+\frac{k}{\nu}\right) C_k^\nu(x) = \frac{(n+2\nu)C_n^\nu(x) - (n+1)C_{n+1}^\nu(x)}{2\nu(1-x)}$$

so that

$$2\nu\left\langle \sum_{0\le k\le n}\left(1+\frac{k}{\nu}\right)C_k^\nu(x),(1-x)^{\alpha/2}\right\rangle_\nu = (n+2\nu)\langle C_n^\nu(x),(1-x)^{\alpha/2-1}\rangle_\nu$$
$$-(n+1)\langle C_{n+1}^\nu(x),(1-x)^{\alpha/2-1}\rangle_\nu$$

$$\langle C_k^\nu(x),(1-x)^{\alpha/2}\rangle_\nu = (-1)^k\frac{1}{k!2^k}\frac{(2\nu)_k}{(\nu+\frac12)_k}\int_I (1-t)^{\alpha/2-1}D^k((1-t^2)^k\omega_\nu(t))dt$$
$$=\frac{1}{k!2^k}\frac{(2\nu)_k}{\left(\nu+\frac12\right)_k}\left(1-\frac{\alpha}{2}\right)_k\int_I(1-t)^{\alpha/2-1-k}(1-t^2)^k(1-t^2)^{\nu-1/2}dt$$
$$=\frac{1}{k!2^k}\frac{\Gamma(2\nu+k)}{\Gamma(2\nu)}\frac{\Gamma\left(\nu+\frac12\right)}{\Gamma\left(\nu+k+\frac12\right)}\frac{\Gamma\left(-\frac{\alpha}{2}+k+1\right)}{\Gamma\left(1-\frac{\alpha}{2}\right)}\int_I(1-t)^{\nu+\alpha/2-3/2}(1+t)^{\nu-1/2+k}dt$$
$$=\frac{\sin\frac{\pi\alpha}{2}}{\pi}\Gamma(\alpha/2)\frac{1}{k!2^k}\frac{\Gamma(2\nu+k)}{\Gamma(2\nu)}\frac{\Gamma(\nu+\frac12)}{\Gamma\left(\nu+k+\frac12\right)}\Gamma\left(-\frac{\alpha}{2}+k+1\right)2^{2\nu+k-1+\frac{\alpha}{2}}$$
$$\times\frac{\Gamma\left(\nu+\frac{\alpha}{2}-\frac12\right)\Gamma\left(\nu+k+\frac12\right)}{\Gamma\left(2\nu+k+\frac{\alpha}{2}\right)}$$
$$=\frac{2^{\alpha/2}\sin\left(\frac{\pi\alpha}{2}\right)\Gamma\left(\nu+\frac{\alpha}{2}-\frac12\right)\Gamma(\alpha/2)}{\Gamma(\nu)\sqrt{\pi}}\frac{\Gamma\left(k+1-\frac{\alpha}{2}\right)\Gamma(2\nu+k)}{k!\Gamma\left(2\nu+k+\frac{\alpha}{2}\right)}.$$

Using the following standard formulaes

$$\Gamma(1-\alpha/2)\Gamma(\alpha/2)=\frac{\pi}{\sin\pi\alpha/2};\quad \Gamma(2\nu)\sqrt{\pi}=2^{2\nu-1}\Gamma(\nu)\Gamma(\nu+1/2).$$

We deduce

$$\left\langle\sum_{0\le k\le n}\left(1+\frac{k}{\nu}\right)C_k^\nu(x),(1-x)^{\alpha/2}\right\rangle_\nu$$
$$=\frac{2^{\alpha/2}\sin\left(\frac{\pi\alpha}{2}\right)\Gamma\left(\nu+\frac{\alpha}{2}-\frac12\right)\Gamma(\alpha/2)}{2\nu\Gamma(\nu)\sqrt{\pi}}\frac{1}{n!}.\frac{(n+2\nu)\Gamma\left(n+1-\frac{\alpha}{2}\right)\Gamma(2\nu+n)}{\Gamma\left(2\nu+n+\frac{\alpha}{2}\right)}$$
$$\times\left\{1-\frac{\left(n+1-\frac{\alpha}{2}\right)}{\left(2\nu+n+\frac{\alpha}{2}\right)}\right\}$$
$$=\frac{(2\nu-1+\alpha)2^{\alpha/2}\sin\left(\frac{\pi\alpha}{2}\right)\Gamma\left(\nu+\frac{\alpha}{2}-\frac12\right)\Gamma(\alpha/2)}{2\nu\Gamma(\nu)\sqrt{\pi}}\frac{(n+2\nu)\Gamma\left(n+1-\frac{\alpha}{2}\right)\Gamma(2\nu+n)}{(n+2\nu+\alpha/2)n!\Gamma\left(2\nu+n+\frac{\alpha}{2}\right)}$$
$$\times\sin\left(\frac{\pi\alpha}{2}\right)C(\alpha,\nu)\frac{(n+2\nu)}{(n+2\nu+\alpha/2)}\frac{\Gamma\left(n+1-\frac{\alpha}{2}\right)\Gamma(2\nu+n)}{n!\Gamma\left(2\nu+n+\frac{\alpha}{2}\right)}$$

Clearly $\sin\left(\frac{\pi\alpha}{2}\right)$ determines the sign, and by Stirling's formula:

$$\frac{\Gamma\left(n+1-\frac{\alpha}{2}\right)\Gamma(2\nu+n)}{n!\Gamma\left(2\nu+n+\frac{\alpha}{2}\right)} \sim n^{-\alpha}$$

So the lower bound of $\|A_j(f_\alpha) - f_\alpha\|_\infty$ is of order $2^{-j\alpha}$.

Acknowledgments The second author gratefully acknowledges the hospitality of the Café Bräunerhof in Vienna.

References

1. Andrews, G.E., Askey, R., Roy, R.: Special functions. In: Encyclopedia of Mathematics and its Applications, vol. 71. Cambridge University Press, Cambridge (1999)
2. Baldi, P., Kerkyacharian, G., Marinucci, D., Picard, D.: Adaptive density estimation for directional data using needlets. Ann. Stat. **37**(6A), 3362–3395 (2009)
3. Bickel, P.J., Rosenblatt, M.: On some global measures of the deviations of density function estimates. Ann. Stat. **1**, 1071–1095 (1973). ISSN:0090-5364
4. Dümbgen, L.: Optimal confidence bands for shape-restricted curves. Bernoulli **9**(3), 423–449 (2003). ISSN:1350-7265
5. Faraut, J.: Analysis on Lie groups. In: Cambridge Studies in Advanced Mathematics, vol. 110. Cambridge University Press, Cambridge (2008). ISBN:978-0-521-71930-8
6. Frisch, U., Parisi, G.: On the singularity structure of fully developed turbulence; appendix to fully developed turbulence and intermittency. In: Proc. Int. Summer School Phys. Enrico Fermi (1985)
7. Geller, D., Mayeli, A.: Continuous wavelets on compact manifolds. Math. Z. **262**(4), 895–927 (2009). ISSN:0025-5874
8. Geller, D., Pesenson, I.Z.: Band-limited localized parseval frames and besov spaces on compact homogeneous manifolds. J. Geom. Anal. **21**:334–371 (2011)
9. Giné, E.: The addition formula for the eigenfunctions of the Laplacian. Adv. Math. **18**(1), 102–107 (1975). ISSN:0001-8708
10. Giné, E., Nickl, R.: Uniform limit theorems for wavelet density estimators. Ann. Probab. **37**, 1605–1646 (2009)
11. Giné, E., Nickl, R.: Confidence bands in density estimation. Ann. Stat. **38**, 1122–1170 (2010)
12. Giné, E., Nickl, R.: Adaptive estimation of a distribution function and its density in sup-norm loss by wavelet and spline projections. Bernoulli **16**, 1137–1163 (2010)
13. Helgason, S.: Differential geometry, Lie groups, and symmetric spaces. In: Pure and Applied Mathematics, vol. 80 (1978)
14. Helgason, S.: Groups and geometric analysis. In: Mathematical Surveys and Monographs, vol. 83. American Mathematical Society, Providence (2000). ISBN:0-8218-2673-5 [Integral geometry, invariant differential operators, and spherical functions, Corrected reprint of the 1984 original]
15. Jaffard, S.: On the Frisch-Parisi conjecture. J. Math. Pures Appl. (9) **79**(6), 525–552 (2000). ISSN:0021-7824
16. Kerkyacharian, G., Pham Ngoc, T.M., Picard, D.: Localized deconvolution on the sphere. Ann. Stat. (2011, to appear)
17. Klemelä, J.: Asymptotic minimax risk for the white noise model on the sphere. Scand. J. Stat. **26**(3), 465–473 (1999). ISSN:0303-6898
18. Koltchinskii, V.: Local Rademacher complexities and oracle inequalities in risk minimization. Ann. Stat. **34**(6), 2593–2656 (2006)
19. Leeb, H., Pötscher, B.M.: Can one estimate the conditional distribution of post-model-selection estimators?. Ann. Stat. **34**(5), 2554–2591 (2006)
20. Lepskiĭ, O.V.: Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. Teor. Veroyatnost. i Primenen. **36**(4), 645–659 (1991). ISSN:0040-361X
21. Lounici, K., Nickl, R.: Global uniform risk bounds for wavelet deconvolution estimators. Ann. Stat. **39**, 201–231 (2011)

22. Low, M.: On nonparametric confidence intervals. Ann. Stat. **25**(6), 2547–2554 (1997). ISSN:0090-5364
23. Massart, P.: Concentration inequalities and model selection. Lecture Notes in Mathematics, vol. 1896. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. Springer (2007). ISBN:978-3-540-48497-4; 3-540-48497-3
24. Narcowich, F., Petrushev, P., Ward, J.: Local tight frames on spheres. SIAM J. Math. Anal. **38**, 574–594 (2006)
25. Narcowich, F.J., Petrushev, P., Ward, J.: Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. J. Funct. Anal. **238**, 530–564 (2006)
26. Picard, D., Tribouley, K.: Adaptive confidence interval for pointwise curve estimation. Ann. Stat. **28**(1), 298–335 (2000). ISSN:0090-5364
27. Stein, E., Weiss, G.: Introduction to Fourier Analysis on Euclidean spaces. Princeton University Press, Princeton (1971)
28. Talagrand, M.: New concentration inequalities in product spaces. Invent. Math. **126**(3), 505–563 (1996)
29. Wang, H.-C.: Two-point homogeneous spaces. Ann. Math. (2) **55**, 177–191 (1952). ISSN:0003-486X
30. Warner, F.W.: Foundations of Differentiable Manifolds and Lie Groups, vol. 94. Springer-Verlag, New York (1983). ISBN:0-387-90894-3. Corrected reprint of the 1971 edition