Post-print of: Ćmiel B., Dziedziul K., Jarzębkowska N.: Multiresolution analysis and adaptive estimation on a sphere using stereographic wavelets NONLINEAR ANALYSIS-THEORY METHODS & APPLICATIONS Vol. 179, (2019), pp.41-71 DOI: 10.1016/j.na.2018.08.003

MULTIRESOLUTION ANALYSIS AND ADAPTIVE ESTIMATION ON A SPHERE USING STEREOGRAPHIC WAVELETS

BOGDAN ĆMIEL, KAROL DZIEDZIUL, AND NATALIA JARZĘBKOWSKA

Abstract. We construct an adaptive estimator of a density function on d dimensional unit sphere $\mathbb{S}^d$ ($d \geq 2$), using a new type of spherical frames. The frames, or as we call them, stereografic wavelets are obtained by transforming a wavelet system, namely Daubechies, using some stereographic operators. We prove that our estimator achieves an optimal rate of convergence on some Besov type class of functions by adapting to unknown smoothness. Our new construction of stereografic wavelet system gives us a multiresolution approximation of $L^2(\mathbb{S}^d)$ which can be used in many approximation and estimation problems. In this paper we also demonstrate how to implement the density estimator in $\mathbb{S}^2$ and we present a finite sample behavior of that estimator in a numerical experiment.

1. Introduction

In this paper, we consider an adaptive estimator of a density function on the d-dimensional unit sphere $\mathbb{S}^d$, $d \geq 2$ using a new type of Parseval frame. To construct the estimator we create a new stereografic wavelet system which gives us a multiresolution approximation of $L^2(\mathbb{S}^d)$. Since our construction uses a standard wavelet system (namely Daubechies) and some stereografic operators one can only make some modifications of existing algorithms in $\mathbb{R}^d$, which is relatively easy, to enjoy the benefits of mutiresolution analysis on a sphere and solve many approximation and estimation problems.

Let us start from the definition

Definition 1.1. Let $\{K_j : j \geq j_0\}$ be a family of measurable functions (called kernels) $K_j : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$. Let $X_1, \ldots X_n$ be i.i.d. with density function f on $\mathbb{S}^d$ with respect to Lebesgue'a measure. For $j \geq j_0$ we define an estimator of f

$$f_n(j)(x) = \frac{1}{n} \sum_{i=1}^{n} K_j(x, X_i).$$

We denote the balls of density functions in Besov spaces on sphere

$$\Sigma(s, \tilde{B}) = \{f \in B^s_{2,\infty}(\mathbb{S}^d) : \int_{\mathbb{S}^d} f(x) d\sigma_d(x) = 1,\ f \geq 0,\ \|f\|_{s,2} \leq \tilde{B}\}.$$

Since we want to obtain an adaptive estimator, we want to construct kernels K_j on a sphere for which we have an optimal rate of estimation. Namely,

Theorem 1.1. *Let $d/2 < r < R$ and let $X_1, \ldots X_n$ be i.i.d. with density function $f \in B^s_{2,\infty}(\mathbb{S}^d)$, where s is unknown and $r \leq s \leq R$. Then there is a family of kernels $\{K_j :$*

Date: October 15, 2018.

2000 *Mathematics Subject Classification.* 42C40.

Key words and phrases. smooth orthogonal projection, sphere, Parseval frame, adaptive estimator, Talagrand's inequality, Besov spaces, parameter of smoothness.

$j \geq j_0\}$ such that for any $U > 0$ there are constants $c = c(r, R, U)$ and $C = C(U)$ such that for all s, n and $\tilde{B} > 1$ we have

$$\sup_{f \in \Sigma(s,\tilde{B}), \|f\|_\infty \leq U} \mathbb{E}\|f_n(j_n) - f\|_2^2 \leq c\tilde{B}^{2d/(2s+d)} n^{-2s/(2s+d)},$$

where

$$j_n = \min\left\{ j \in [j_{min}, j_{max}] : \forall_{l,\ j<l\leq j_{max}}\ \|f_n(j) - f_n(l)\|_2^2 \leq C\frac{2^{ld}}{n} \right\}$$

and

$$j_{min} = \left\lfloor \frac{\log_2 n}{2R+d} \right\rfloor, \qquad j_{max} = \left\lceil \frac{\log_2 n}{2r+d} \right\rceil.$$

In the above theorem the smoothness parameter s is unknown but for choosing the resolution level we use a lower bound r and an upper bound R. Let us discuss some consequences of choosing different values for r and R. It seems that it is good idea to take r as small as possible and R as big as possible to consider a very wide range for the unknown smoothness. The first part of that is true since there are no serious consequences of taking small r. Unfortunately if we take a big value for R, then we need to use in our construction some very smooth wavelets (smoother than R). The smoother the wavelets are, the bigger support they have and if one scales them to a smaller area, then they change values very rapidly. In the asymptotic point of view this is not a problem but for fixed n the estimator loses its efficiency if the value R is too big. The same problem we can observe in case of a wavelet estimation on $\mathbb{R}$.

It is well-known (see Hall, Kerkyacharian and Picard (1998) [22] Theorem 4.1) that on the real line if one considers wavelets estimators with a block thresholding procedure, one attains minimax rate of convergence without extraneous logarithmic factors for $B^s_{2,\infty}$ Besov spaces and L^2-loss, i.e., $n^{-2s/(1+2s)}$. Similar result was given in [6]. We follow the arguments presented there.

The problem of estimating nonparametrically a density on the d-dimensional unit sphere $\mathbb{S}^d$ over Besov classes is not new (see Baldi, Kerkyacharian, Marinucci and Picard (2009) [2] for a direct setting and for an indirect setting see Kerkyacharian, Pham Ngoc and Picard (2011), [24]). In particular, in Baldi, Kerkyacharian, Marinucci and Picard (2009), the authors had already dealt with the considered problem in a more general framework, namely, by considering $B^s_{q,r}$ Besov spaces. They constructed an adaptive estimator based on a set of spherical wavelets, named needlets, with a hard thresholding procedure. They obtained minimax rates of convergence for $B^s_{q,r}$ Besov spaces, L^p-loss and sup-norm loss up to a logarithmic factor. This deep approach was continued in [12] but in regression case. Moreover the rates are without the logarithmic factor. We obtain the minimax rate of convergence for the L^2-loss, i.e., $n^{-2s/(2s+d)}$ without the logarithmic factor that one usually gets with adaptive methods of estimating density function on the sphere. We want to emphasize that section 2 can be rewritten for a compact smooth manifold. So if one can construct a family of kernels such that they satisfy conditions of Theorem 2.5, one obtains a method of minimax rate of convergence for the L^2-loss, i.e., $n^{-2s/(2s+d)}$ without the logarithmic factor on the manifold M. The first step is done in [5] i.e., a smooth orthogonal decomposition of identity in $L^2(M)$ is constructed.

In section 2 we formulate general conditions on K_j which guarantee Talagrand's inequality in Bousquet's version. This inequality is a key argument to prove Theorem 1.1.

In section 3 we construct kernels K_j using a new type of frames on sphere which gives an optimal rate of the estimation. The new frame, called stereographic wavelets, inherits all properties of the classical multivariate Daubechies wavelets and we know that such basis is an excellent tool in the process of approximation and estimation mentioned in [6, Theorem 2] or [22, Theorem 4.1].

In great amount of literature, tight frames (needlets) are used as a tool in approximation as well as in estimation of densities (see [9], [16], [23], [29], [30]). Unfortunately this approach does not give an optimal rate of estimation (see [25]). The new frames on sphere, introduced by Bownik M., Dziedziul K. in [4] give a construction of K_j such that we achieve the optimal rate of convergence on Besov spaces via adaptive estimation. The method of constructing Parseval frame on sphere consists of two steps. In first step we obtain a localized wavelet system on sphere by transforming Daubechies wavelet system on $[1-\varepsilon, 1+\varepsilon]^d$ using two stereographic operators. Next we create a Parseval frame by applying P. Auscher, G. Weiss, M. V. Wickerhauser (AWW) operator ([1]) on sphere (see [4]). Consequently we create a multivariate approximation on $L^2(\mathbb{S}^d)$, i.e.,

$$V_{j_0}(\mathbb{S}^d) \subset V_{j_0+1}(\mathbb{S}^d) \subset \cdots \subset L^2(\mathbb{S}^d),$$

for all $j \geq j_0$

$$\dim V_j(\mathbb{S}^d) < \infty,$$

and $\bigcup_{j\geq j_0} V_j(\mathbb{S}^d)$ is dense in $L^2(\mathbb{S}^d)$. The functions $K_j(x,y)$, $x, y \in \mathbb{S}^d$, $j \geq j_0$ in the main theorem are kernels of the orthogonal projection

$$K_j : L^2(\mathbb{S}^d) \to V_j(\mathbb{S}^d),$$

$$K_j(f)(x) = \int_{\mathbb{S}^d} f(y) K_j(x,y) d\sigma_d(y),$$

where σ_d is Lebesgue measure on $\mathbb{S}^d$.

In section 4 we present a technical version of Theorem 1.1. In section 5 we show a numerical example of such estimation for $\mathbb{S}^2$ (classical sphere). All the proofs are given in the appendix A.

Since the coefficients of the frame give us characterization of Besov spaces $B^s_{2,\infty}(\mathbb{S}^d)$, it is possible to use our approach from earlier papers to estimate the smoothness of density function or to construct a smoothness test (see [7], [8] and [15]), but this is not the aim of this paper.

2. Talagrand's inequality

In this section we present Talagrand's inequality (see [3], [17] and Theorem 3.3.9 (Upper tail of Talagrand's inequality, Bousquet's version [18]) and its consequences. Let us cite from [17]: in the special case "Talagrand's inequality becomes exactly the Bernstein and Prohorov inequalities. Clearly then, Talagrand's inequality is essentially a best possible exponential bound for the empirical process." We start our consideration with general type of kernels $K_j(x,y)$, $x, y \in \mathbb{S}^d$. In the next section we focus our attention on kernels which arise from the Parseval frame.

Theorem 2.1. *Let $X, X_1, \ldots, X_n$ be i.i.d. random variables with law μ on a measurable space $(M, \mathcal{M})$. Let $\mathcal{K}$ be a countable class of real measurable functions on M, uniformly* *bounded by a constant U and $\mu-$ centered, i.e.,*

$$\mu(k) := \int k d\mu = 0, \quad k \in \mathcal{K}. \tag{2.1}$$

For $H : \mathcal{K} \to \mathbb{R}$ define

$$\|H\|_{\mathcal{K}} \;=\; \sup_{k\in\mathcal{K}} |H(k)|. \tag{2.2}$$

Let ω be a positive number such that

$$\omega^2 \;\geq\; \sup_{k\in\mathcal{K}} \mathbb{E}k^2(X) \tag{2.3}$$

and

$$V \;:=\; n\omega^2 + 2U\mathbb{E}\left\|\sum_{i=1}^{n} k(X_i)\right\|_{\mathcal{K}}. \tag{2.4}$$

Then for every $x \geq 0$ and $n \in \mathbb{N}$

$$P\left\{\left\|\sum_{i=1}^{n} k(X_i)\right\|_{\mathcal{K}} \geq \mathbb{E}\left\|\sum_{i=1}^{n} k(X_i)\right\|_{\mathcal{K}} + \sqrt{2Vx} + Ux/3\right\} \;\leq\; 2e^{-x}. \tag{2.5}$$

Following [17] we adapt this theorem to our situation. Let $X_1, \ldots X_n$ be i.i.d. random variables with density f on $\mathbb{S}^d$ with respect to Lebesgue measure σ_d on $\mathbb{S}^d$ (the surface measure),

$$\mathbb{S}^d = \left\{x \in \mathbb{R}^{d+1} : ||x|| = \sqrt{x_1^2 + \ldots + x_{d+1}^2} = 1\right\} \subset \mathbb{R}^{d+1}.$$

Let us define

$$d\mu = f d\sigma_d.$$

We assume that

$$\|f\|_\infty < \infty.$$

Since $L^2(\mathbb{S}^d)$ is a separable set, then there is a countable subset B_0 of the unit ball $B \subset L^2(\mathbb{S}^d)$ such that for all $\beta \in L^2(\mathbb{S}^d)$

$$\|\beta\|_2 = \left(\int_{\mathbb{S}^d} |\beta(t)|^2 d\sigma_d(t)\right)^{1/2} = \sup_{g\in B_0}\left|\int_{\mathbb{S}^d} \beta(t)g(t)d\sigma_d(t)\right| = \sup_{g\in B_0} |\langle\beta, g\rangle| \tag{2.6}$$

We assume that a family of symmetric kernels $K_j(\cdot,\cdot)$, $j \geq j_0$ i.e. $K_j(x, y) = K_j(y, x)$ for all x, y, j, satisfies the following three conditions:

$$\forall_{j\geq j_0} \quad \sup_{x,y\in\mathbb{S}^d} |K_j(x, y)| < \infty, \tag{2.7}$$

$$\exists_{D>0}\forall_{y\in\mathbb{S}^d}\forall_{j\geq j_0} \quad \int_{\mathbb{S}^d} K_j^2(x, y)d\sigma_d(x) \leq D2^{jd}, \tag{2.8}$$

$$\exists_{C_{\mathcal{K}}>0}\forall_{j\geq j_0}\forall_{g\in L^2(\mathbb{S}^d)} \quad \|K_j(g)\|_2 \leq C_{\mathcal{K}} \|g\|_2, \tag{2.9}$$

where

$$K_j(g)(t) = \int_{\mathbb{S}^d} g(x)K_j(x, t)d\sigma_d(x),$$

for $g \in L^1(\mathbb{S}^d)$. Note that

$$K_j(f)(t) = \mathbb{E}[K_j(t, X)]. \tag{2.10}$$

For simplicity we will assume further that $C_{\mathcal{K}} = 1$. A classical estimator of density f is given by

$$f_n(j)(x) = \frac{1}{n}\sum_{i=1}^{n} K_j(x, X_i).$$

From (2.8) we obtain the following lemma

Lemma 2.2. *Let $X_1, \dots X_n$ be i.i.d. with common density f on $\mathbb{S}^d$ with respect to Lebesgue measure. Let symmetric kernels $K_j(\cdot,\cdot)$ satisfy (2.7), (2.8), (2.9). Then, there exists $D > 0$ such that*

$$\forall_{j \geq j_0} \quad \mathbb{E}\, \|f_n(j) - \mathbb{E} f_n(j)\|_2^2 \leq D\frac{2^{jd}}{n}, \tag{2.11}$$

and

$$\forall_{j\geq j_0} \quad \mathbb{E}\left\|\sum_{i=1}^{n}\Big(K_j(\cdot, X_i) - \mathbb{E}K_j(\cdot, X_i)\Big)\right\|_2 \leq \sqrt{Dn2^{jd}}. \tag{2.12}$$

[proof in the appendix A.1]

For $j_0 \in \mathbb{Z}$ and $j \geq j_0$ we define the following family of kernels

$$\mathcal{K} = \mathcal{K}_j := \left\{k_g = \int_{\mathbb{S}^d} g(t)K_j(t,\cdot)d\sigma_d(t) - \int_{\mathbb{S}^d} g(t)K_j(f)(t)d\sigma_d(t) : g \in B_0\right\}. \tag{2.13}$$

Lemma 2.3. *Let $X_1, \dots X_n$ be i.i.d. with common density f on $\mathbb{S}^d$ with respect to Lebesgue measure. Let symmetric kernels $K_j(\cdot,\cdot)$ satisfy (2.7), (2.8), (2.9). Then $\mathcal{K}$ satisfies the assumptions of Theorem 2.1 i.e., for all $k_g \in \mathcal{K}$*

$$\|k_g\|_\infty \leq \sqrt{D}2^{jd/2} + \|f\|_\infty^{1/2} =: U_{\mathcal{K}_j} < \infty. \tag{2.14}$$

$$\mu(k_g) = \int_{\mathbb{S}^d} k_g(x)f(x)d\sigma_d(x) = 0. \tag{2.15}$$

$$\mathbb{E}\big[k_g(X)\big]^2 \leq \|f\|_\infty =: \omega_{\mathcal{K}}^2. \tag{2.16}$$

[proof in the appendix A.2]

To transform the thesis of Theorem 2.1 in a case of density function estimation note that if we define $\mu_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$, then for every $k_g \in \mathcal{K}$

$$\begin{aligned}
\mu_n(k_g) &= \frac{1}{n}\sum_{i=1}^{n} k_g(X_i) \\
&= \int_{\mathbb{S}^d} g(t)\left(\frac{1}{n}\sum_{i=1}^{n}K_j(t,X_i) - K_j(f)(t)\right)d\sigma_d(t) && (2.17)\\
&= \int_{\mathbb{S}^d} g(t)\Big(f_n(j)(t) - \mathbb{E}f_n(j)(t)\Big)d\sigma_d(t). && (2.18)
\end{aligned}$$

Hence taking $H = \mu_n - \mu$ in Theorem 2.1 we have

$$
\begin{aligned}
\langle\ (2.2)\ \rangle \qquad \|\mu_n - \mu\|_{\mathcal{K}} &= \sup_{k_g \in \mathcal{K}} \Big|(\mu_n - \mu)k_g\Big| \\
\langle\ (2.15) \text{ and } (2.18)\ \rangle \qquad &= \sup_{g \in B_0} \left| \int_{\mathbb{S}^d} g(t)\Big(f_n(j)(t) - \mathbb{E}f_n(j)(t)\Big) d\sigma_d(t) \right| \\
\langle\ (2.6)\ \rangle \qquad &= \|f_n(j) - \mathbb{E}f_n(j)\|_2 = \frac{1}{n} \left\| \sum_{i=1}^{n} \Big(K_j(\cdot, X_i) - \mathbb{E}K_j(\cdot, X_i)\Big) \right\|_2 .
\end{aligned}
$$

Note that by (2.17)

$$
\left\| \sum_{i=1}^{n} k_g(X_i) \right\|_{\mathcal{K}} = n \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2 . \tag{2.19}
$$

Now taking into account the above result we can formulate the following one.

Corollary 2.4. *Let $X_1, \dots X_n$ be i.i.d. with common density f on $\mathbb{S}^d$ with respect to Lebesgue measure. Let symmetric kernels $K_j(\cdot,\cdot)$ satisfy (2.7), (2.8), (2.9). For the family $\mathcal{K} = \mathcal{K}_j$* (2.13) *we have the following inequality (Talagrand's inequality* (2.5) *from Theorem 2.1)*

$$
\forall_{x \geq 0} \forall_{n \in \mathbb{N}} \forall_{j \geq j_0}
$$

$$
P\left\{ n \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2 \geq n\mathbb{E} \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2 + \sqrt{2Vx} + U_{\mathcal{K}_j} x/3 \right\} \leq 2e^{-x}, \tag{2.20}
$$

where

$$
V = n\omega_{\mathcal{K}}^2 + 2nU_{\mathcal{K}_j} \mathbb{E} \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2 \tag{2.21}
$$

and

$$
\omega_{\mathcal{K}}^2 = \|f\|_\infty, \tag{2.22}
$$

$$
U_{\mathcal{K}_j} = \sqrt{D} 2^{jd/2} + \|f\|_\infty^{1/2}. \tag{2.23}
$$

We want to transform (2.20) into a formula which will be convenient in our later calculation. Note that

$$
\begin{aligned}
(2.24) \qquad \sqrt{2Vx} &= \sqrt{2nx\omega_{\mathcal{K}}^2 + 4xU_{\mathcal{K}_j} n\mathbb{E} \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2} \\
\langle\ \sqrt{a+b} \leq \sqrt{a} + \sqrt{b}\ \rangle \qquad &\leq \sqrt{2nx\omega_{\mathcal{K}}^2} + 2\sqrt{xU_{\mathcal{K}_j} n\mathbb{E} \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2} \\
\langle\ \sqrt{ab} \leq \tfrac{a+b}{2}\ \rangle \qquad &\leq \sqrt{2nx\omega_{\mathcal{K}}^2} + xU_{\mathcal{K}_j} + n\mathbb{E} \left\| f_n(j) - \mathbb{E}f_n(j) \right\|_2 .
\end{aligned}
$$

By (2.24), (2.22), (2.23)

$$RHS := n\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2 + \sqrt{2Vx} + U_{\mathcal{K}_j}x/3$$
$$\leq 2n\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2 + \frac{4}{3}(\sqrt{D2^{jd}} + \|f\|_\infty^{1/2})x + \sqrt{2xn\|f\|_\infty}$$

Note that by Lemma 2.2

$$n\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2 = \mathbb{E}\left\|\sum_{i=1}^n (K_j(\cdot, X_i) - \mathbb{E}K_j(\cdot, X_i))\right\|_2 \leq \sqrt{Dn2^{jd}}.$$

Consequently, for $x = 2^{jd}$ we obtain

$$\text{(2.25)} \qquad RHS \leq n\frac{2^{jd/2}}{\sqrt{n}}\left(2\sqrt{D} + \frac{4}{3}\sqrt{D}\frac{2^{jd}}{\sqrt{n}} + \frac{4}{3}\|f\|_\infty^{1/2}\frac{2^{jd/2}}{\sqrt{n}} + \sqrt{2\|f\|_\infty}\right).$$

It follows from Lemma 2.2, that we need to assume that the relation between level j and the size of sample n is such that

$$\frac{2^{jd}}{n} \to 0$$

as $j, n \to \infty$. This condition is justified to guarantee balance between stochastic and deterministic error. So through this paper we assume that there is $C_E \geq 0$ such that we work in the range of parameter j and n such that

$$\text{(2.26)} \qquad \frac{2^{jd}}{n} \leq C_E.$$

Consequently

$$\text{(2.27)} \qquad RHS \leq n\frac{2^{jd/2}}{\sqrt{n}}M\sqrt{1 \vee \|f\|_\infty},$$

where $M = 2(\sqrt{D}\left(2 + \frac{4}{3}C_E\right) \vee (\frac{4}{3}C_E + \sqrt{2}))$.

Finally from (2.27) and Corollary 2.4 we get the main estimation. If we denote by $\mathcal{L}(\mathbb{S}^d)$ a $\sigma-$ algebra of Lebesgue sets contained in $\mathbb{S}^d$, then the following theorem is true.

Theorem 2.5. *Let $K_j(\cdot, \cdot), j \geq j_0$ be a family of real symmetric measurable function with respect to $\mathcal{L}(\mathbb{S}^d) \times \mathcal{L}(\mathbb{S}^d)$, satisfying:*

$$\text{(2.28)} \qquad \forall_{j \geq j_0} \sup_{x,y \in \mathbb{S}^d} |K_j(x, y)| < \infty,$$

$$\text{(2.29)} \qquad \exists_{D>0}\forall_{y\in\mathbb{S}^d}\forall_{j\geq j_0} \int_{\mathbb{S}^d} K_j^2(x, y)d\sigma_d(x) \leq D2^{jd}.$$

$$\text{(2.30)} \qquad \forall_{j\geq j_0}\forall_{g\in L^2(\mathbb{S}^d)} \quad \|K_j(g)\|_2 \leq \|g\|_2.$$

Let $X_1, \ldots X_n$ be i.i.d. with common density $f \in L^\infty(\mathbb{S}^d)$. Let

$$f_n(j)(x) = \frac{1}{n}\sum_{i=1}^n K_j(x, X_i).$$

Then for $j \geq j_0$ and $n \in N$ such that $\frac{2^{jd}}{n} \leq C_E$ we have

$$P\left\{\|f_n(j) - \mathbb{E}f_n(j)\|_2 \geq \frac{2^{jd/2}}{\sqrt{n}} M\sqrt{1 \vee \|f\|_\infty}\right\} \leq 2e^{-2^{jd}},$$

where

$$M = 2\max\left\{\sqrt{D}\left(2 + \frac{4}{3}C_E\right), \frac{4}{3}C_E + \sqrt{2}\right\}.$$

3. Besov spaces and stereographic wavelets on sphere

In this section we construct a Parseval frame on $\mathbb{S}^d, d \geq 2$ using Bownik-Dziedziul construction (see [4]). For any fixed angle $0 < \delta < \pi/2$ let us decompose the sphere onto two Patches A_- and A_+, depending on the angle (see Figure 1), where

$$A_- = \{x \in \mathbb{S}^d : x_{d+1} \leq \cos(\pi/2 - \delta)\},$$
$$A_+ = \{x \in \mathbb{S}^d : x_{d+1} \geq \cos(\pi/2 + \delta)\}.$$

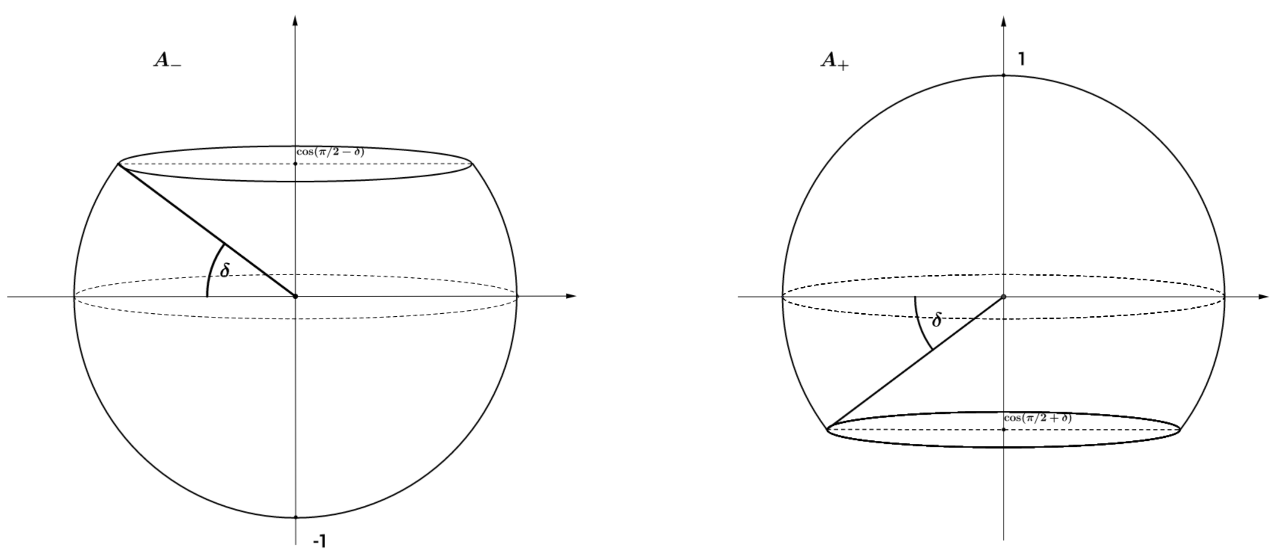

Figure 1. Patches A_- i A_+ on $\mathbb{S}^d$

Consider a natural parametrization of the sphere

$$\Phi_d : [0, \pi] \times \mathbb{S}^{d-1} \to \mathbb{S}^d, \qquad \Phi_d(\theta, \xi) = (\xi \sin\theta, \cos\theta), \quad (\theta, \xi) \in [0, \pi] \times \mathbb{S}^{d-1}.$$

The function $\Phi_d : (0, \pi) \times \mathbb{S}^{d-1} \to \mathbb{S}^d \setminus \{\mathbf{1}^d, -\mathbf{1}^d\}$ is a diffeomorphism, where

$$\mathbf{1}^d = (0, \ldots, 0, 1) \in \mathbb{S}^d \quad \text{is the "North Pole", see Figure 2.}$$

Then $f \in L^1(\mathbb{S}^d, d\sigma_d)$ and we have (see [10, (1.5.4)])

$$\int_{\mathbb{S}^d} f(u) d\sigma_d(u) = \int_{\mathbb{S}^{d-1}} \int_0^\pi f \circ \Phi_d(\theta, \xi)(\sin\theta)^{d-1} d\theta d\sigma_{d-1}(\xi).$$

It makes sense to introduce the notation $g(\theta,\xi) = g(\Phi_d(\theta,\xi))$. Let us take some real-valued, smooth function $s \in C^\infty(\mathbb{R})$ such that

$$\operatorname{supp} s \subset [-\delta,\infty), \qquad s^2(t)+s^2(-t)=1, \quad t\in\mathbb{R}.$$

Now we can define Auscher-Weiss-Wickerhouser (AWW) operator $E = E_{\delta,s}$, pointwise for every $g:\mathbb{S}^d\to\mathbb{R}$

$$E(g)(\theta,\xi) = \begin{cases} g(\theta,\xi), & \theta > \pi/2+\delta \\ s^2(\theta-\pi/2)g(\theta,\xi)+s(\theta-\pi/2)s(\pi/2-\theta)g(\pi-\theta,\xi) \\ 0, & \theta < \pi/2-\delta, \end{cases} \tag{3.1}$$

where $\xi\in\mathbb{S}^{d-1}$ (see (3.5) and (3.6) in [4]).

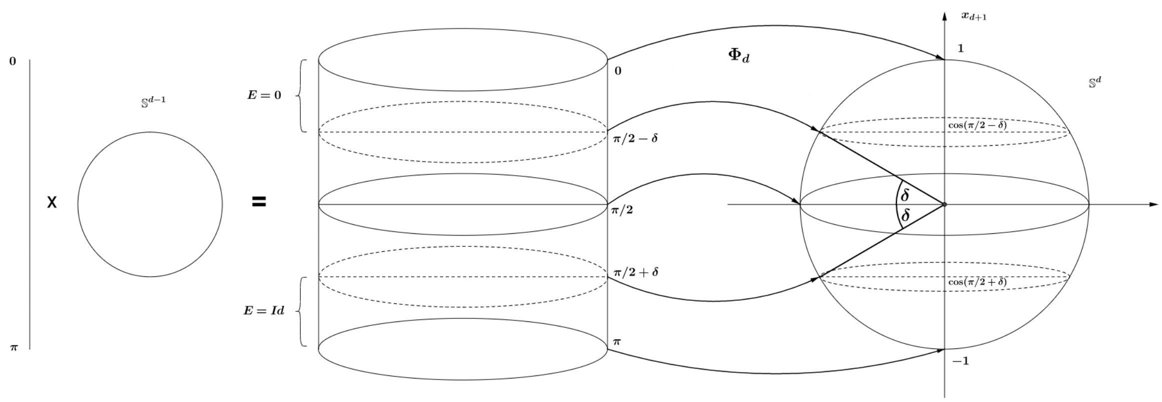

FIGURE 2. Function Φ_d and operator E

Lemma 3.1. [4, Lemma 3.3] $E : L^2(\mathbb{S}^d, d\sigma_d) \to L^2(\mathbb{S}^d, d\sigma_d)$ *is an orthogonal projection.*

Since we consider two patches A_- and A_+ we have very simple decomposition of identity operator I (orthogonal partition of unity), see [4] Theorem 1.1

$$E^+ = Id - E \quad\text{ and }\quad E^- = E.$$

Definition 3.1. We say that an operator P is localized on an open set U, i.e., for any $f:\mathbb{S}^d\to\mathbb{R}$ we have

$$Pf(x) = 0 \qquad \text{for } x\in\mathbb{S}^d\setminus U.$$

Now we are ready to reformulate [4, Theorem 5.1]. We recall that Sobolev space [14].

Definition 3.2. For $k\in\mathbb{N}$ and $f:\mathbb{S}^d\to\mathbb{R}$ we denote by $\nabla^k f(x)$, $x\in\mathbb{S}^d$, the covariant derivative of f of order k in some local chart. We let $|\nabla^k f|$ to be its norm (which is independent of a choice of chart). Given $1\le p<\infty$ we define the norm

$$||f||_{W_p^r} = \sum_{k=0}^{r}\left(\int_{\mathbb{S}^d} |\nabla^k f(x)|^p d\sigma_d(x)\right)^{1/p} < \infty.$$

The Sobolev space $W_p^r(\mathbb{S}^d)$ is the completion of $C^r(\mathbb{S}^d)$ with respect to the norm $||\cdot||_{W_p^r}$.

To see more direct definition see [10] or [4]. See also [5, Definition 2,3 and Lemma 2.7]. Despite our characterization also holds for fractional Sobolev spaces, Theorem 3.5, we omit these considerations.

Theorem 3.2.

(1) The operator E^+ is localized on A_+ and the operator E^- on A_-,

(2) The both operators $E^\pm$ (i.e., E^+, E^-) $E^\pm : L^2(\mathbb{S}^d) \to L^2(\mathbb{S}^d)$ are orthogonal projections

(3) For all $r = 0, 1, \dots$ and $1 \le p < \infty$, each $E^\pm$ is a continuous operator

$$E^\pm : W^r_p(\mathbb{S}^d) \to W^r_p(\mathbb{S}^d).$$

Consider Daubechies multivariate wavelets. For $N \ge 2$, let ${}_N\phi$ be a univariate, compactly supported scaling function with support $\operatorname{supp} {}_N\phi = [0, 2N-1]$ associated with the compactly supported, orthogonal univariate Daubechies wavelet ${}_N\psi$, see [11, Section 6.4]. Moreover it is known that a smoothness of Daubechies wavelets $\varrho = \varrho(N) \approx 0.2N$, ${}_N\phi \in C^{\varrho(N)}(\mathbb{R})$, $\varrho(N) \in \mathbb{N}$, see [11, Section 7.1.2]. Note that we take ${}_N\psi$ such that $\operatorname{supp} {}_N\psi = [0, 2N-1]$. For convenience, let $\psi^0 = {}_N\phi$ and $\psi^1 = {}_N\psi$. Let $\mathcal{E}' = \{0,1\}^d$ be the vertices of the unit cube and let $\mathcal{E} = \mathcal{E}' \setminus \{0\}$ be the set of nonzero vertices. For each $\mathbf{e} = (e_1, \dots, e_d) \in \mathcal{E}'$, define

$$\psi^{\mathbf{e}}(x) = \psi^{e_1}(x_1) \cdots \psi^{e_d}(x_d), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Observe that $\operatorname{supp} \psi^{\mathbf{e}} = [0, 2N-1]^d$.

Let $\mathcal{D}$ be the set of dyadic cubes in $\mathbb{R}^d$ of the form $I = 2^{-j}(k + [0,1]^d)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^d$. Denote the side length of I by $\ell(I) = 2^{-j}$. For any $\mathbf{e} \in \mathcal{E}'$ define scaled wavelet, related to I by

$$\psi^{\mathbf{e}}_I(x) = 2^{jd/2}\psi^{\mathbf{e}}(2^j x - k), \qquad x \in \mathbb{R}^d.$$

It is well-known that $\{\psi^{\mathbf{e}}_I(x) : I \in \mathcal{D}, \mathbf{e} \in \mathcal{E}\}$ is an orthonormal basis of $L^2(\mathbb{R}^d)$, [28, chapter 3]. Then to characterize classical Sobolev spaces (as well as fractional Sobolev spaces called also Bessel potential spaces) $W^s_p(\mathbb{R}^d)$ for $1 < p < \infty$, $s \ge 0$ or Besov spaces $B^s_{pq}(\mathbb{R}^d)$ for $1 \le p \le \infty$, $1 \le q \le \infty$, $s > 0$ by the magnitude of coefficients of d-dimensional Daubechies wavelets ψ^e_I we need to assume that

$$\varrho(N) > s$$

[28, chapter 6, section 2 and section 10], compare also [13, Theorem 9.4 and Theorem 9.5].

For our purposes it is convenient to consider a localized wavelet systems on a cube.

Definition 3.3. Suppose that $J = [-1,1]^d$ is a cube in $\mathbb{R}^d$ and $\varepsilon > 0$. Define its ε enlargement by $J_\varepsilon = [-1-\varepsilon, 1+\varepsilon]^d$. Let $j_0 \in \mathbb{Z}$ be the smallest integer such that

$$(2N-1)2^{-j_0} \le \varepsilon/2. \tag{3.2}$$

For any $j \ge j_0$, consider families of dyadic cubes

$$\mathcal{D}_j = \mathcal{D}_j(e) = \{I \in \mathcal{D} : \ell(I) = 2^{-j} \text{ and } \operatorname{supp} \psi^{\mathbf{e}}_I \subset J_\varepsilon\}$$

and

$$\mathcal{D}^+_{j_0} = \mathcal{D}^+_{j_0}(e) = \bigcup_{j=j_0}^{\infty} \mathcal{D}_j.$$

Define a localized wavelet system, related to the cube J and $\varepsilon > 0$ by

$$S(J,\varepsilon) := \{\psi_I^{\mathbf{e}} : \mathbf{e} \in \mathcal{E}, I \in \mathcal{D}_{j_0}^{+}(e)\} \cup \{\psi_I^0 : I \in \mathcal{D}_{j_0}(0)\}. \tag{3.3}$$

By appropriate choice of ε, ($\varepsilon = k$ or $\varepsilon = 2^{-k}$, $k \in \mathbb{N} \setminus \{0\}$) we get a sequence of finite dimensional spaces

$$V_{j_0} \subset \cdots \subset V_j \subset \cdots \subset L^2(J_\varepsilon), \tag{3.4}$$

where

$$V_j = \operatorname{span}_{L^2(J_\varepsilon)}\{\psi_I^0 : I \in \mathcal{D}_j(0)\}. \tag{3.5}$$

We have usual a dilation and translation properties. We consider only the dilation by two, for $j \geq j_0$ if

$$f \in V_j \Rightarrow f(2\cdot) \in V_{j+1}. \tag{3.6}$$

We will transport that sequence by two stereographic projections on sphere. After using AWW operators we obtain MRA on $L^2(\mathbb{S}^d)$, (see below).

Lemma 3.3. *The localized wavelet system $S(J,\varepsilon)$ has following properties:*

- *$S(J,\varepsilon)$ is an orthonormal sequence in $L^2 = L^2(J_\varepsilon)$,*
- *for every $f \in L^2(J_\varepsilon)$ with $\operatorname{supp} f \subset J_{\varepsilon/2}$ we have*

$$\|f\|_{L^2}^2 = \sum_{\mathbf{e}\in\mathcal{E}} \sum_{I\in\mathcal{D}_{j_0}^+} |\langle f, \psi_I^{\mathbf{e}}\rangle_{L^2}|^2 + \sum_{I\in\mathcal{D}_{j_0}} |\langle f, \psi_I^0\rangle_{L^2}|^2. \tag{3.7}$$

- *magnitudes of coefficients $\{|\langle f, g\rangle|\}_{g\in S(J,\varepsilon)}$ characterize functions $f \in \mathcal{F}(\mathbb{R}^d)$ satisfying $\operatorname{supp} f \subset J_{\varepsilon/2}$, where $\mathcal{F}$ is either the Sobolev space $W_p^s(\mathbb{R}^d)$, $0 \leq s < \varrho(N)$, $1 < p < \infty$ or the Besov space $B_{p,q}^s(\mathbb{R}^d)$, $0 < s < \varrho(N)$, $1 \leq p, q \leq \infty$.*

Proof is similar to the proof of Lemma 6.1 from [4] since there is no restriction for $p, q = \infty$.

Localized wavelet system $S(J,\epsilon)$ is transformed to $\mathbb{S}^d$ by stereographic projections (see Figure 3)

$$S_- : \mathbb{S}^d \setminus \{\mathbf{1}^d\} \to \mathbb{R}^d, \qquad S_-(x_1,\ldots,x_{d+1}) = \left(\frac{x_1}{1-x_{d+1}},\ldots,\frac{x_d}{1-x_{d+1}}\right),$$

$$S_+ : \mathbb{S}^d \setminus \{-\mathbf{1}^d\} \to \mathbb{R}^d, \qquad S_+(x_1,\ldots,x_{d+1}) = \left(\frac{x_1}{1+x_{d+1}},\ldots,\frac{x_d}{1+x_{d+1}}\right).$$

For $\epsilon > 0$ we define variable change operators (for $+$ and $-$)

$$T_d^{\pm} : L^2([-1-\varepsilon, 1+\varepsilon]^d) \to L^2(\mathbb{S}^d)$$

given by

$$T_d^{\pm}(\psi)(u) = \frac{\psi(S_\pm(u))}{\sqrt{J_d(S_\pm(u))}}, \qquad u \in \mathbb{S}^d,$$

where J_d the Jacobian of $S_\pm^{-1}$

$$J_d(x_1,\ldots,x_d) = \left(\frac{2}{1+x_1^2+\ldots+x_d^2}\right)^d.$$

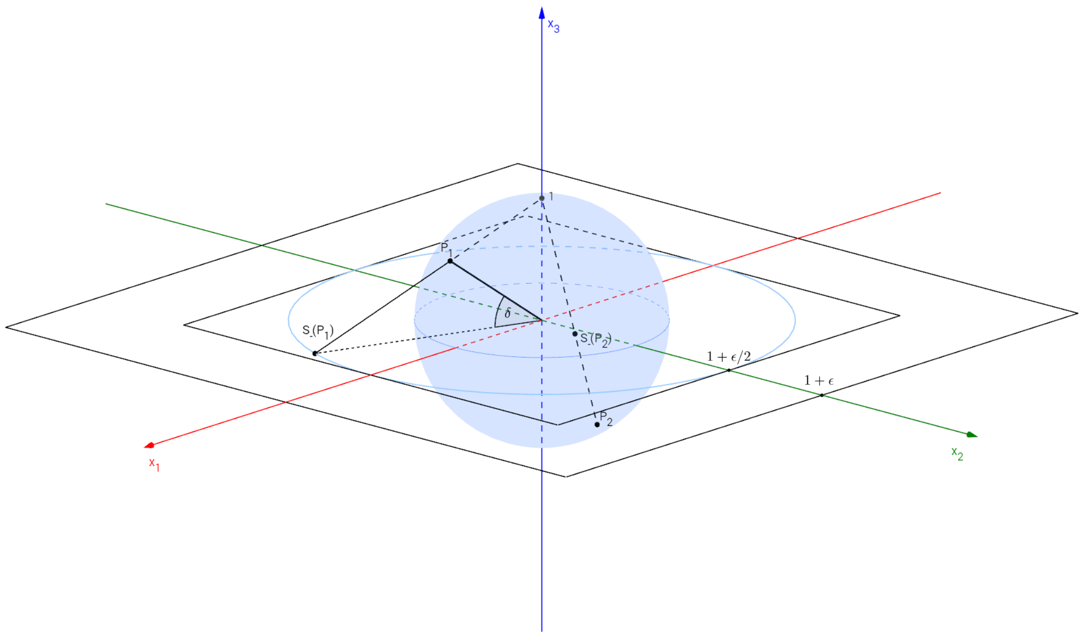

FIGURE 3. Stereographic projection S_- for $d = 2$

Both operators $T_d^{\pm}$ are isometric isomorphisms. This leads us to two a local wavelet system on $\mathbb{S}^d$. Namely,

$$\begin{aligned}\mathcal{S}_{\pm} = \mathcal{S}_{\pm}(A_{\pm}) = T_d^{\pm}(S(J,\epsilon)) = \\ = \{T_d^{\pm}(\psi_I^{\mathbf{e}}) : \mathbf{e} \in \mathcal{E}, I \in \mathcal{D}_{j_0}^{+}\} \cup \{T_d^{\pm}(\psi_I^{\mathbf{0}}) : I \in \mathcal{D}_{j_0}\},\end{aligned}$$

Various equivalent norms of Besov spaces $B_{p,q}^s = B_{p,q}^s(\mathbb{S}^d)$ are given in [27]. Let us recall the definition of Besov space (called Nikolskij-Besov space) form [27]. For $r \in \mathbb{N}$ let $\omega_r(f,\tau)_p$ be a modulus of smoothness on sphere, i.e.,

$$\omega_r(f,\tau)_p = \sup_{0<t\leq\tau} \|\Delta_t^r f\|_p$$

and

$$\Delta_t^r = (Id - T_t)^r, \quad 0 < t < \pi.$$

Here Id is an identity operator and T_t is a translation operator, compare Definition 2.1.4 [10],

$$T_t(f)(\theta) = \frac{\Gamma(d/2)}{2\pi^{d/2}(\sin t)^{d-1}} \int_{\theta\circ y=\cos t} f(y) dl_{\theta,t}(y)$$

and $dl_{\theta,t}$ denotes Lebesgue measure on the set $\{y \in \mathbb{S}^d : \theta \circ y = \cos t\}$. Let $s > 0$ and $1 \leq p, q \leq \infty$. Then

$$B_{p,q}^s(\mathbb{S}^d) = \{f \in L^p(\mathbb{S}^d) : \left(\int_0^{\pi} \left(\frac{\omega_r(f,t)_p}{t^s}\right)^q \frac{dt}{t}\right)^{1/q} < \infty\},$$

where $r > s/2$. We have an analogue of [4, Lemma 6.2]

Lemma 3.4. *The system $\mathcal{S}_{\pm}$ has following properties:*

(1) $\mathcal{S}_{\pm}$ is an orthogonal system in $L^2(\mathbb{S}^d)$,

(2) *If* $\epsilon \geq 2\left(\frac{\cos\delta}{1-\sin\delta} - 1\right)$, *then* $E^{\pm}(\mathcal{S}_{\pm})$ *is Parseval frame for* $E^{\pm}(L^2(\mathbb{S}^d))$, *a.e. for all* $f \in E^{\pm}(L^2(\mathbb{S}^d))$

$$\|f\|_2^2 = \sum_{g\in\mathcal{S}_{\pm}} |\langle f, E^{\pm}(g)\rangle|^2.$$

(3) *the coefficients* $\{|\langle f, g\rangle|\}_{g\in E^{\pm}(\mathcal{S}_{\pm})}$ *characterize functions* $f \in E^{\pm}(\mathcal{F}(\mathbb{S}^d))$, *where* $\mathcal{F}$ *is a Sobolev space* $W_p^s(\mathbb{S}^d)$, *where* $0 \leq s < \varrho(N)$, $1 < p < \infty$ *or Besov space* $B^s_{p,q}(\mathbb{S}^d)$, $0 < s < \varrho(N)$, $1 \leq p, q \leq \infty$.

Proof is a consequence of Lemma 3.3 and Lemma 6.2 from [4].

We define a wavelet system called stereographic wavelets corresponding to A_+ and A_-

$$\mathcal{S} := E^+(\mathcal{S}_+) \cup E^-(\mathcal{S}_-), \tag{3.8}$$

where

$$E^{\pm}(\mathcal{S}_{\pm}) = \{E^{\pm} \circ T_d^{\pm}(\psi_I^{\mathbf{e}}) : \mathbf{e} \in \mathcal{E}, I \in \mathcal{D}^+_{j_0}\} \cup \{E^{\pm} \circ T_d^{\pm}(\psi_I^{\mathbf{0}}) : I \in \mathcal{D}_{j_0}\}.$$

We can define a sequence of finite dimensional spaces $j \geq j_0$

(3.9)

$$V_j(\mathbb{S}^d) = \mathrm{span}_{L^2(\mathbb{S}^d)}\{E^+ \circ T_d^+(\psi_I^{\mathbf{0}}) : I \in \mathcal{D}_j(0)\} \oplus \mathrm{span}_{L^2(\mathbb{S}^d)}\{E^- \circ T_d^-(\psi_I^{\mathbf{0}}) : I \in \mathcal{D}_j(0)\},$$

$$W_j(\mathbb{S}^d) = \mathrm{span}_{L^2(\mathbb{S}^d)}\left\{E^{\pm} \circ T_d^{\pm}(\psi_I^e) : e \in \mathcal{E}, I \in \mathcal{D}_j\right\}.$$

Note that

$$V_j(\mathbb{S}^d) = E^+ \circ T_d^+(V_j) \oplus E^- \circ T_d^-(V_j).$$

and

$$W_j(\mathbb{S}^d) = E^+ \circ T_d^+(W_j) \oplus E^- \circ T_d^-(W_j).$$

For a function $f \in V_j(\mathbb{S}^d)$ we will use a notation $f = E^+(f^+) \oplus E^-(f^-)$ where $f^+ \in T_d^+(V_j)$ and $f^- \in T_d^-(V_j)$.

We have an analogue of Theorem 6.2 from [4] with multiresolution structure on $L^2(\mathbb{S}^d)$. To formulate a dilation property we need two spherical dilations $\vartheta^{\pm}$. We take natural parametrization. Let $\Phi_d(\theta, \xi) = x \in \mathbb{S}^d \setminus \{-\mathbf{1}^d\}$. Then

$$\vartheta^+ : \mathbb{S}^d \setminus \{-\mathbf{1}^d\} \to \mathbb{S}^d \setminus \{-\mathbf{1}^d\}$$

is given by

$$\vartheta^+(x) = \vartheta^+(\Phi_d(\theta, \xi)) = \Phi_d(\phi(\theta), \xi) = (\xi \sin(\phi(\theta)), \cos(\phi(\theta))),$$

where $\phi : (0, \pi) \to (0, \pi)$ is a diffeomorphism. To find a formula for $\phi = \phi(\theta)$ we need to solve

$$\frac{2\sin\theta}{1 + \cos\theta} = \frac{\sin\phi}{1 + \cos\phi}.$$

Since in the natural parametrization

$$S_+(\xi \sin\theta, \cos\theta) = \frac{\xi \sin\theta}{1 + \cos\theta}, \tag{3.10}$$

then for the dilation operator $dial(y) = 2y$, where $y \in \mathbb{R}^d$ we have

$$S_+ \circ \vartheta^+ = dial \circ S_+. \tag{3.11}$$

We have a diagram

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{\quad dial \quad} & \mathbb{R}^d \\ \uparrow S_+ & & \uparrow S_+ \\ \mathbb{S}^d \setminus \{-\mathbf{1}^d\} & \xrightarrow[\quad \vartheta^+ \quad]{} & \mathbb{S}^d \setminus \{-\mathbf{1}^d\} \end{array}$$

Similar for ϑ^-.

Theorem 3.5. *Let $\epsilon = k$ or $\epsilon = 2^{-k}$, $k \in \mathbb{N}\backslash\{0\}$. If $\epsilon \geq 2\left(\frac{\cos\delta}{1-\sin\delta} - 1\right)$, the wavelet system $\mathcal{S}$ is a Parseval frame in $L^2(\mathbb{S}^d)$. The sequence of $\{V_j(\mathbb{S}^d)\}_{j\geq j_0}$ has following properties*

$$V_{j_0}(\mathbb{S}^d) \subset V_{j_0+1}(\mathbb{S}^d) \subset \cdots \subset L^2(\mathbb{S}^d), \tag{3.12}$$

$$V_{j+1}(\mathbb{S}^d) = V_j(\mathbb{S}^d) \oplus W_j(\mathbb{S}^d), \tag{3.13}$$

and the set $\bigcup_{j\geq j_0} V_j(\mathbb{S}^d)$ is dense in $L^2(\mathbb{S}^d)$. We have also spherical dilations property, if $j \geq j_0$ and $f = E^+(f^+) \oplus E^-(f^-) \in V_j(\mathbb{S}^d)$, then there are functions $H^\pm$ independent of j such that

$$E^\pm\left(H^\pm(\cdot)f^\pm(\vartheta^\pm(\cdot))\right) \in V_{j+1}(\mathbb{S}^d). \tag{3.14}$$

Moreover, the magnitudes of the coefficients $\{|\langle f, g\rangle|\}_{g\in\mathcal{S}}$ characterize $f \in \mathcal{F}(\mathbb{S}^d)$, where $\mathcal{F}$ is Sobolev space $W^s_p(\mathbb{S}^d)$, where $0 \leq s < \varrho(N)$, $1 < p < \infty$ or Besov space $B^s_{p,q}(\mathbb{S}^d)$, $0 < s < \varrho(N)$, $1 \leq p, q \leq \infty$.
[proof in the appendix A.3]

In the next section we will consider $B^s_{2,\infty}(\mathbb{S}^d)$ for $0 < s < \varrho(N)$. We will use a characterization of functions from $B^s_{2,\infty}(\mathbb{S}^d)$ by frame coefficients. Namely, from Theorem 3.5 we get that a function $f \in L^2(\mathbb{S}^d)$ belongs to $B^s_{2,\infty}(\mathbb{S}^d)$, $0 < s < \varrho(N)$, if $\|f\|_{s,2} < \infty$, where

$$\|f\|_{s,2} := \max\Bigg(\sup_{j\geq j_0} 2^{js}\sqrt{\sum_{e\in\mathcal{E}, I\in\mathcal{D}_j}\langle f, E^+\circ T^+_d(\psi^{\mathbf{e}}_I)\rangle^2},\ 2^{j_0 s}\sqrt{\sum_{I\in\mathcal{D}_{j_0}}\left\langle f, E^+\circ T^+_d(\psi^{\mathbf{0}}_I)\right\rangle^2},$$
$$\sup_{j\geq j_0} 2^{js}\sqrt{\sum_{e\in\mathcal{E}, I\in\mathcal{D}_j}\langle f, E^-\circ T^-_d(\psi^{\mathbf{e}}_I)\rangle^2},\ 2^{j_0 s}\sqrt{\sum_{I\in\mathcal{D}_{j_0}}\left\langle f, E^-\circ T^-_d(\psi^{\mathbf{0}}_I)\right\rangle^2}\Bigg). \tag{3.15}$$

In fact we will use an equivalent with the same notation

$$\|f\|_{s,2} := \max\Bigg(2^{j_0 s}\sqrt{\sum_{I\in\mathcal{D}_{j_0}}\left\langle f, E^+\circ T^+_d(\psi^{\mathbf{0}}_I)\right\rangle^2 + \sum_{I\in\mathcal{D}_{j_0}}\left\langle f, E^-\circ T^-_d(\psi^{\mathbf{0}}_I)\right\rangle^2},$$
$$\sup_{j\geq j_0} 2^{js}\sqrt{\sum_{e\in\mathcal{E}, I\in\mathcal{D}_j}\langle f, E^+\circ T^+_d(\psi^{\mathbf{e}}_I)\rangle^2 + \sum_{e\in\mathcal{E}, I\in\mathcal{D}_j}\langle f, E^-\circ T^-_d(\psi^{\mathbf{e}}_I)\rangle^2}\Bigg), \tag{3.16}$$

We define a family of operators

$$K_j : L^2(\mathbb{S}^d) \to V_j, \quad K_j(f)(x) = \int_{\mathbb{S}^d} K_j(x,y)f(y)dy,$$

where

$$K_j(x,y) = \sum_{I\in\mathcal{D}_j} E^{\pm}\circ T_d^{\pm}(\psi_I^{\mathbf{0}})(x)\cdot E^{\pm}\circ T_d^{\pm}(\psi_I^{\mathbf{0}})(y). \tag{3.17}$$

If $\{x_n\}, n\in\mathbb{N}$ is the Parseval frame in separable Hilbert space H, then for all $x\in H$ and $I\subset\mathbb{N}$

$$\|x - \sum_{n\in\mathbb{N}\setminus I}\langle x,x_n\rangle x_n\|^2 = \|\sum_{n\in I}\langle x,x_n\rangle x_n\|^2 \le \sum_{n\in I}\langle x,x_n\rangle^2. \tag{3.18}$$

For completeness of arguments let us prove this inequality. We use the theorem that for Parseval frame there is orthogonal basis $\{f_n\}$ in H_1 such that $H\subset H_1$ and $x_n = P(f_n)$, where P is an orthogonal projection

$$P : H_1 \to H.$$

Hence for all $f\in H_1$ and all $I\subset\mathbb{N}$

$$\|f - \sum_{n\in\mathbb{N}\setminus I}\langle f,f_n\rangle f_n\|^2 = \|\sum_{n\in I}\langle f,f_n\rangle f_n\|^2 = \sum_{n\in I}\langle f,f_n\rangle^2.$$

If we use orthogonal projection, we obtain

$$\|P(f - \sum_{n\in\mathbb{N}\setminus I}\langle f,f_n\rangle f_n)\|^2 = \|P(f) - \sum_{n\in\mathbb{N}\setminus I}\langle f,f_n\rangle x_n\|^2 \le \|f - \sum_{n\in\mathbb{N}\setminus I}\langle f,f_n\rangle f_n\|^2.$$

If we use this for $f = Px = x$, we get

$$\|x - \sum_{n\in\mathbb{N}\setminus I}\langle x,Pf_n\rangle x_n)\|^2 = \|x - \sum_{n\in\mathbb{N}\setminus I}\langle x,x_n\rangle x_n\|^2 \le \sum_{n\in I}\langle x,x_n\rangle^2,$$

which proves (3.18).

Using (3.18) we obtain corollary

Corollary 3.6. *For all functions from Besov space $B^s_{2,\infty}$, $0<s<\varrho(N)$ and all $j\ge j_0$*

$$\|f - K_jf\|_2 \le 2^{-js}\|f\|_{s,2}.$$

[proof in the appendix A.4]

4. Adaptive estimator of density function

In this section we present a technical version of Theorem 1.1. Note that

Lemma 4.1. *For all $N\ge 2$ there is D_N such that*

$$\forall_{y\in\mathbb{S}^d}\forall_{j\ge j_0} \quad \int_{\mathbb{S}^d} K_j^2(x,y)d\sigma_d(x) \le D_N 2^{jd}. \tag{4.1}$$

[proof in the appendix A.5]

One can see that the kernels $K_j(\cdot,\cdot)$ fulfill the conditions from Theorem 2.5. For $f_n(j)(x) = \frac{1}{n}\sum_{i=1}^n K_j(x,X_i)$, where $K_j(\cdot,\cdot)$ is given by (3.17), we formulate the analogue of [6, Theorem 2] (our proof is more precise and gives all needed arguments). The idea of choosing the resolution level is taken from Lepski [26].

Theorem 4.2. *Let $d/2 < r < R$ and let $X_1, \dots X_n$ be i.i.d. with density function $f \in B^s_{2,\infty}(\mathbb{S}^d)$, where $r \le s \le R$. We assume that Daubechies wavelet is smooth enough, i.e., $R < \varrho(N)$.*

Let j_{min} and j_{max} be such that $j_0 \le j_{min} \le j_{max}$ and

$$j_{min} = \left\lfloor \frac{\log_2 n}{2R+d} \right\rfloor, \qquad j_{max} = \left\lceil \frac{\log_2 n}{2r+d} \right\rceil. \tag{4.2}$$

Define $\mathcal{J} := \mathcal{J}_n = [j_{min}, j_{max}] \cap \mathbb{N}$ and

$$j_n = \min \left\{ j \in \mathcal{J} : \forall_{l \in \mathcal{J}, l > j} \ \|f_n(j) - f_n(l)\|_2^2 \le C(\mathcal{S})(U \vee 1)\frac{2^{ld}}{n} \right\}, \tag{4.3}$$

where $C(\mathcal{S})$ is a constant such that

$$\sqrt{C(\mathcal{S})} \ge 2 + M_{N} 2 \sqrt{D_{N}}, \tag{4.4}$$

where the constant D_N is from Lemma 4.1 and M_N is the constant from theorem 2.5 depending on $C_E > 0$ and D_N for kernels (3.17).

Then for any $r, R, U > 0$ there is $c = c(r, R, U)$ such that for all s, n and $\tilde{B} > 1$ if j_n is defined by (4.3) we have

$$\sup_{f \in \Sigma(s, \tilde{B}), \|f\|_\infty \le U} \mathbb{E}\|f_n(j_n) - f\|_2^2 \le c\tilde{B}^{2d/(2s+d)} n^{-2s/(2s+d)}. \tag{4.5}$$

[proof in the appendix A.6]

In practice constant $C(\mathcal{S})$, from the above Theorem, can be chosen as

$$C(\mathcal{S}) = \left(2 + 4\max\{10\sqrt{D_N}/3\ ,\ 4/3 + \sqrt{2}\}\sqrt{D_N}\right)^2,$$

where

$$D_N = \sup_{j \in \{j_{min}, \dots, j_{max}\}} \sup_{y \in \mathbb{S}^d} \int_{\mathbb{S}^d} K_j^2(x, y) d\sigma_d(x)/2^{jd}.$$

It can be calculated numerically for y from some grid on $\mathbb{S}^d$. It is also known that the Daubechies wavelets smoothness increases with N approximately like $0.2N$ (see [11] chapter 7). It means that for the estimation one should take $N \approx 5R$.

In the proof of the above theorem the following lemma was used.

Lemma 4.3. *Under the above construction we have*

$$\mathbb{E}\,\|f_n(j) - \mathbb{E}f_n(j)\|_2^4 \le 4\left(32 D_{N}^2 \sigma^4(j, n) + (\sqrt{2} \cdot 2^{-js}\,\|f\|_{s,2})^4\right).$$

[proof in the appendix A.7]

5. Numerical results

The first purpose of this section is to explain how one can implement our estimator in $\mathbb{S}^2$, by presenting the exact formula of the estimator with a special choice of functions and parameters. The second purpose is to show, in a numerical experiment, that the estimator works and can be used in practice.

For an estimation the Daubechies wavelets "DB8" with the support $[0, 15]$ are used. Two values of the experiment size are used: $n = 100$ and $n = 10000$. The maximum resolution levels j_{max} from our main Theorem are $j_{max} = 2$ for $n = 100$ and $j_{max} = 3$

for $n = 10000$, where $r = 3/2$ (see Theorem 4.2). Since our wavelet support length is 15 on the resolution level 0, then we decide to use minimal resolution level $j_{min} \geq 2$ which allows us to take a lower value for ϵ. Because of that, we set $j_{min} = \max\{2, \left\lfloor \frac{\log_2 n}{2R+d} \right\rfloor\}$ which gives us $j_{min} = 2$ for $n = 100$ and $j_{min} = 2$ for $n = 10000$, where $R = 2$. The resolution levels of the estimator for generated data was $j_n = 2$ for $n = 100$ and $j_n = 3$ for $n = 10000$ (the method form the Theorem 4.2 for choosing j_n can be more useful in practice for bigger sample sizes n, when there is a bigger set of possible resolution levels). In the estimator formula (see definition 1.1) we choose $\delta = \pi/6$ (see figure 1), $\epsilon = 4$ (see definition 3.3) and the distribution function $s \in C^\infty(\mathbb{R})$ (see AWW operator (3.1))

$$s(t) = \left[\exp\left(\frac{t-\delta}{t+\delta}\right) \Big/ \sqrt{\exp\left(2\,\frac{t-\delta}{t+\delta}\right) + \exp\left(2\,\frac{-t-\delta}{-t+\delta}\right)} \;\right] \mathbb{1}_{(-\delta,\delta)}(t) + \mathbb{1}_{[\delta,\infty)}(t).$$

Notice that the choice of ϵ is justified since the length of the effective support of DB8 scaling function on the resolution level 2 is smaller than $\epsilon/2 = 2$. For the calculation of the wavelets values, a dyadic discretization is used. The distance between discretization points on the resolution level j is $2^{-(j+10)}$. Data samples $(X_1, Y_1, Z_1), ..., (X_n, Y_n, Z_n)$ are generated from the following density functions on the sphere $x^2 + y^2 + z^2 = 1$:

- $f_1(x, y, z) = 0,3785\ (\arcsin z - \pi/8)^2\ (\arcsin z - 7\pi/8)^2\ \mathbb{1}_{[\sin(\pi/8),1]}(z)$,
- $f_2(x, y, z) = 42.2126\ (\arcsin z - \pi/4)^2\ (\arcsin z - \pi/2)^2\ \mathbb{1}_{[\sin(\pi/4),1]}(z)$,

using the elimination method, which is the following. First we generate an observation $(U_x, U_y, U_z) = (N_x/\sqrt{N_x^2 + N_y^2 + N_z^2}, N_y/\sqrt{N_x^2 + N_y^2 + N_z^2}, N_z/\sqrt{N_x^2 + N_y^2 + N_z^2})$ from the uniform distribution on the unit sphere, where N_x, N_y, N_z are independent and have standard normal distribution. Then we generate M which is independent of (U_x, U_y, U_z) and has uniform distribution on $[0; 1 + \sup f]$, where f is a density on sphere. If $M < f(U_x, U_y, U_z)$ then we keep the observation (U_x, U_y, U_z). If not, we repeat the procedure. We repeat this until we have the whole sample of size n. The sample is i.i.d. with the density f. The estimator can be calculated in any point (x, y, z) of the unit sphere by the following formula:

$$\hat{f}_n(j_n)(x, y, z) = \frac{1}{n} \sum_{i=1}^{n} K_j((x, y, z), (X_i, Y_i, Z_i)),$$

where

$$K_j((x, y, z), (X_i, Y_i, Z_i)) = \sum_{I \in D_j} \Big[E^+ \circ T_d^+(\psi_I^{\mathbf{0}})(x, y, z) \cdot E^+ \circ T_d^+(\psi_I^{\mathbf{0}})(X_i, Y_i, Z_i) + E^- \circ T_d^-(\psi_I^{\mathbf{0}})(x, y, z) \cdot E^- \circ T_d^-(\psi_I^{\mathbf{0}})(X_i, Y_i, Z_i) \Big],$$

$$T_d^+(\psi_I^{\mathbf{0}})(x, y, z) = \frac{1}{2}\psi_I^{\mathbf{0}}\left(\frac{x}{1+z}, \frac{y}{1+z}\right)\left(1 + \left(\frac{x}{1+z}\right)^2 + \left(\frac{y}{1+z}\right)^2\right),$$

$$T_d^-(\psi_I^{\mathbf{0}})(x, y, z) = \frac{1}{2}\psi_I^{\mathbf{0}}\left(\frac{x}{1-z}, \frac{y}{1-z}\right)\left(1 + \left(\frac{x}{1-z}\right)^2 + \left(\frac{y}{1-z}\right)^2\right),$$

$$E^- \circ T_d^-(\psi_I^{\mathbf{0}})(x,y,z) = \begin{cases} T_d^-(\psi_I^{\mathbf{0}})(x,y,z), \quad \text{if} \quad z < \sin\delta \\ s^2(-\arcsin z)T_d^-(\psi_I^{\mathbf{0}})(x,y,z) + s(-\arcsin z)s(\arcsin z) \\ \cdot T_d^-(\psi_I^{\mathbf{0}})(x,y,-z), \quad \text{if} \quad -\sin\delta \le z \le \sin\delta \\ 0, \quad \text{if } z > \sin\delta \end{cases},$$

$$E^+ \circ T_d^+(\psi_I^{\mathbf{0}})(x,y,z) = \begin{cases} 0, \quad \text{if} \quad z < \sin\delta \\ T_d^+(\psi_I^{\mathbf{0}})(x,y,z) - s^2(-\arcsin z)T_d^+(\psi_I^{\mathbf{0}})(x,y,z) - s(-\arcsin z) \\ \cdot s(\arcsin z)T_d^+(\psi_I^{\mathbf{0}})(x,y,-z), \quad \text{if} \quad -\sin\delta \le z \le \sin\delta \\ T_d^+(\psi_I^{\mathbf{0}})(x,y,z), \quad \text{if } z > \sin\delta \end{cases}.$$

In our simulations the estimator values are calculated on the following discrete set of points:

$$\begin{aligned} \{(x,y,z) : &\big[(x,y) \in \{-0.98, -0.96, ..., 0.98\}^2 \wedge x^2 + y^2 \le 1 \wedge z^2 = 1 - x^2 - y^2\big] \\ &\vee \big[(x,z) \in \{-0.98, -0.96, ..., 0.98\}^2 \wedge x^2 + z^2 \le 1 \wedge y^2 = 1 - x^2 - z^2\big] \\ &\vee \big[(y,z) \in \{-0.98, -0.96, ..., 0.98\}^2 \wedge y^2 + z^2 \le 1 \wedge x^2 = 1 - y^2 - z^2\big]\}, \end{aligned}$$

which is approximately uniformly distributed on the sphere and quite comfortable in implementation. The results of our estimation are presented in figures 4 and 5.

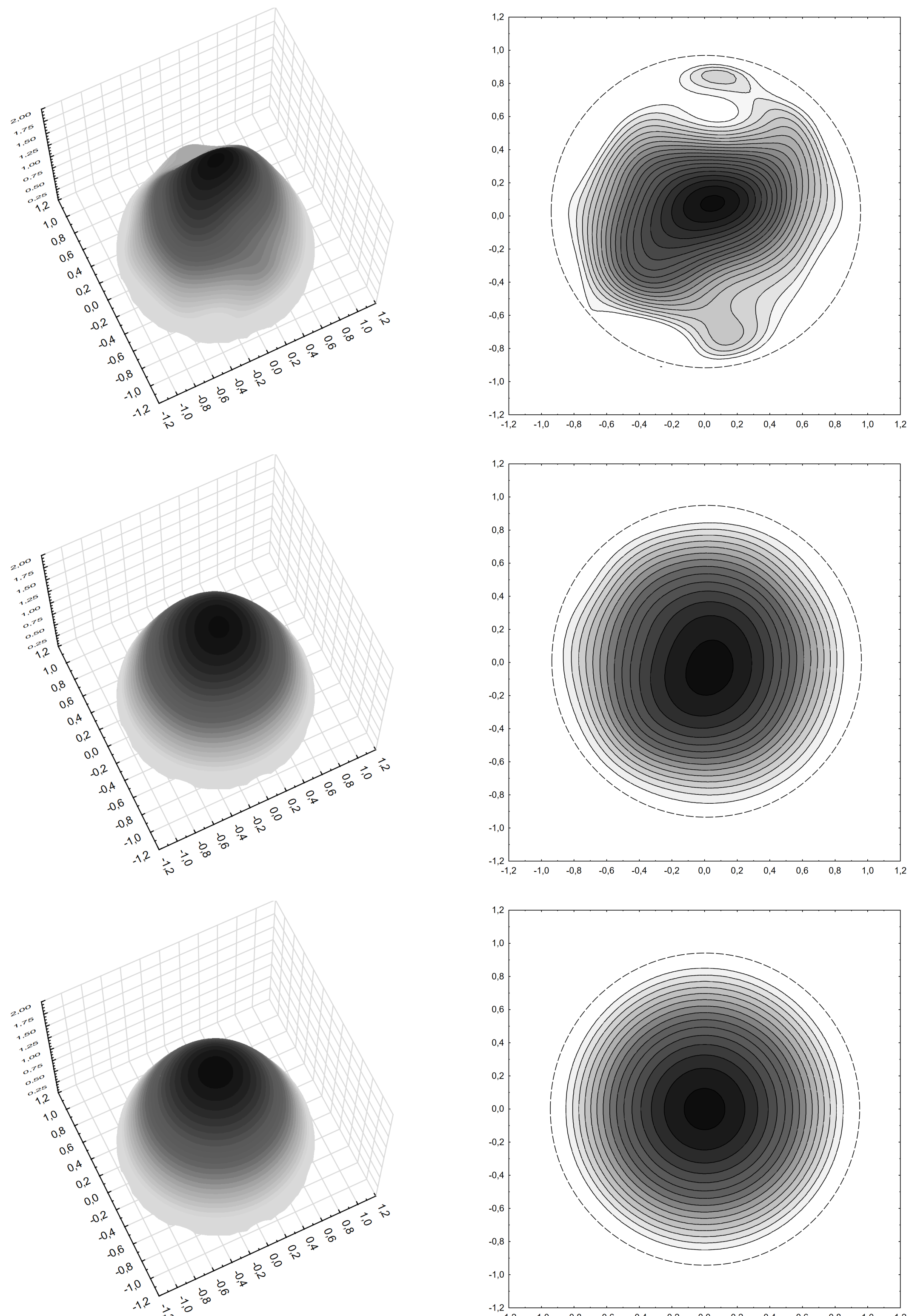

Figure 4. From top: estimator of function f_1 for $n = 100$, $n = 10000$ and true function f_1 on the bottom.

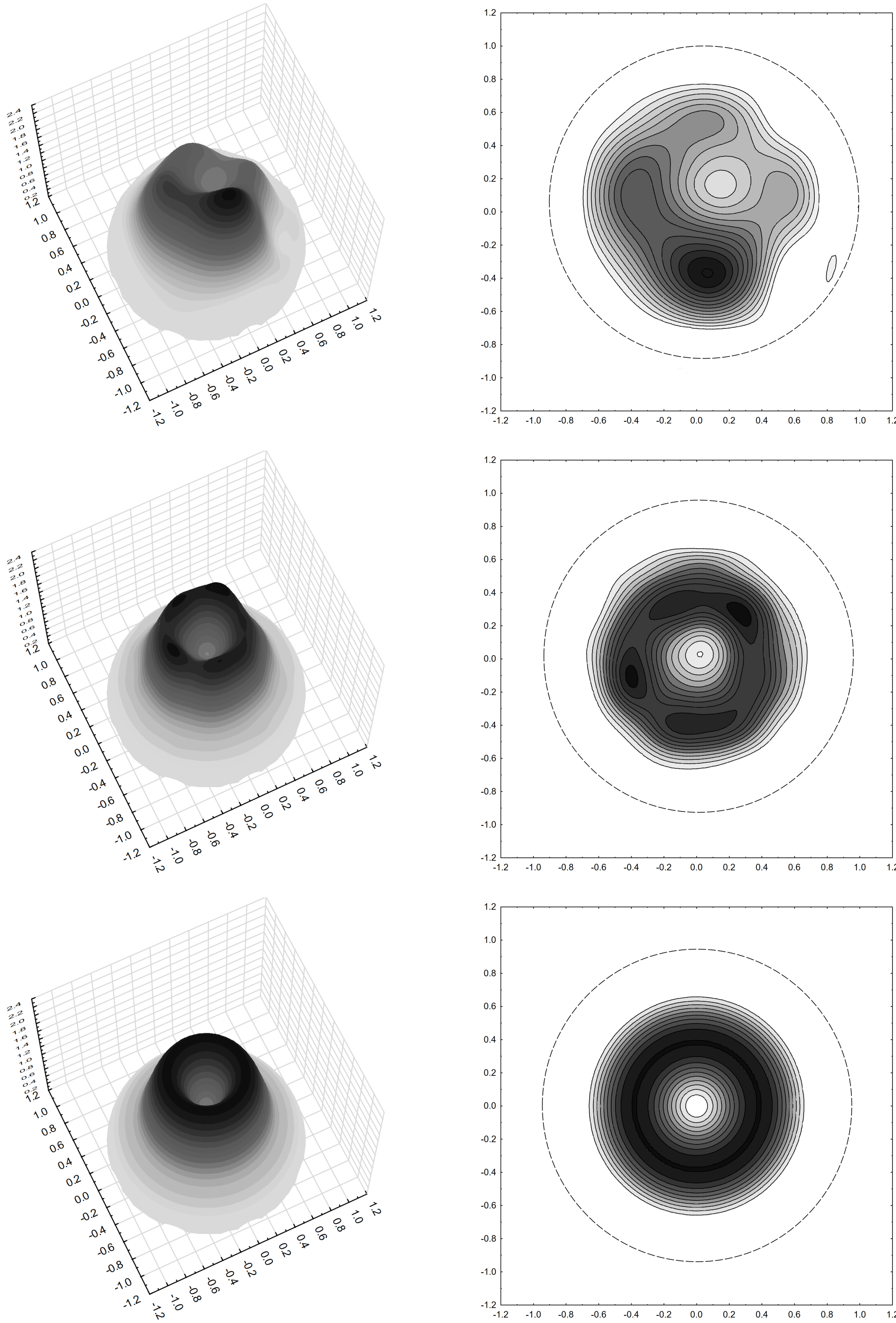

FIGURE 5. From top: estimator of function f_2 for $n = 100$, $n = 10000$ and true function f_2 on the bottom.

Appendix A. Mathematical Proofs

A.1. **Proof of Lemma 2.2.** First we prove (2.11). Let $j \geq j_0$ be fixed. For $x \in \mathbb{S}^d$ define

$$Y_i(x) = K_j(x, X_i) - \mathbb{E}K_j(x, X_i) = K_j(x, X_i) - K_j f(x).$$

Since $Y_i(x)$ are i.i.d. and $\mathbb{E}Y_i(x) = 0$ by (2.8) we get

$$\begin{aligned}\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2^2 &= \frac{1}{n}\int_{\mathbb{S}^d} \mathbb{E}\left(Y_i(x)\right)^2 d\sigma_d(x) \\ &\leq \frac{1}{n}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_j^2(x,y)f(y)d\sigma_d(y)d\sigma_d(x) \leq D\frac{2^{jd}}{n}.\end{aligned}$$

Now we prove (2.12). From Jensen's inequality and (2.11) we have

$$\begin{aligned}\left(\mathbb{E}\left\|\sum_{i=1}^n \Big(K_j(\cdot, X_i) - \mathbb{E}K_j(\cdot, X_i)\Big)\right\|_2\right)^2 &\leq n^2\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n \Big(K_j(\cdot, X_i) - \mathbb{E}K_j(\cdot, X_i)\Big)\right\|_2^2 \\ &= n^2\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2^2 \leq Dn2^{jd}.\end{aligned}$$

A.2. **Proof of Lemma 2.3.** Let us check that $\mathcal{K}$ satisfies the assumptions of Theorem 2.1. We start with assumption that $k_g \in \mathcal{K}$ are uniformly bounded. If we take $k_g \in \mathcal{K}$, then

$$\begin{aligned}
&\|k_g\|_\infty = \sup_{x\in\mathbb{S}^d} |k_g(x)| \\
\langle \text{ Schwarz's inequality } \rangle \quad &\leq \sup_{x\in\mathbb{S}^d}\left(\|g\|_2 \|K_j(\cdot,x)\|_2 + \|g\|_2 \|K_j(f)\|_2\right) \\
\langle \text{ since } g \in B_0 \rangle \quad &\leq \sup_{x\in\mathbb{S}^d}\left(\|K_j(\cdot,x)\|_2 + \|K_j(f)\|_2\right) \\
\langle \ (2.9.) \ \rangle \quad &\leq \sup_{x\in\mathbb{S}^d}\left(\|K_j(\cdot,x)\|_2 + \|f\|_2\right).
\end{aligned}$$

Note that by (2.8) we get

$$\forall_{j\geq j_0}\forall_{x\in\mathbb{S}^d} \quad \|K_j(\cdot,x)\|_2 = \left(\int_{\mathbb{S}^d} K_j^2(t,x)d\sigma_d(t)\right)^{1/2} \leq \sqrt{D}2^{jd/2}. \tag{A.1}$$

Moreover

$$\|f\|_2 = \left(\int_{\mathbb{S}^d} f^2(t)d\sigma_d(t)\right)^{1/2} \leq \left(\|f\|_\infty \int_{\mathbb{S}^d} f(t)d\sigma_d(t)\right)^{1/2} = \|f\|_\infty^{1/2} < \infty. \tag{A.2}$$

By (A.1) and (A.2)

$$\|k_g\|_\infty \leq \sqrt{D}2^{jd/2} + \|f\|_\infty^{1/2} =: U_{\mathcal{K}_j} < \infty.$$

We have proved that a function from $\mathcal{K} = \mathcal{K}_j$ is bounded by $U_{\mathcal{K}_j}$. By (2.14) we get

$$\forall_{k_g\in\mathcal{K}} \quad \int_{\mathbb{S}^d} |k_g(x)|f(x)d\sigma_d(x) < \infty.$$

Consequently,

$$\begin{aligned}\mu(k_g) &= \int_{\mathbb{S}^d} k_g(x) f(x) d\sigma_d(x) \\ &= \int_{\mathbb{S}^d} g(t)\left(K_j(f)(t) - K_j(f)(t)\int_{\mathbb{S}^d} f(x)d\sigma_d(x)\right) d\sigma_d(t) = 0.\end{aligned}$$

Let us show (3.6) i.e., for $\omega_{\mathcal{K}}^2 = \|f\|_\infty$

$$\forall_{g\in B_0} \quad \mathbb{E}\big[k_g(X)\big]^2 \le \omega_{\mathcal{K}}^2.$$

If $g \in B_0$, then by (2.9)

$$\begin{aligned}\mathbb{E}\big[k_g(X)\big]^2 &= Var\left(\int_{\mathbb{S}^d} g(t)K_j(t,X)d\sigma_d(t)\right) \\ &\le \mathbb{E}\left[\int_{\mathbb{S}^d} g(t)K_j(t,X)d\sigma_d(t)\right]^2 \\ &\le \|f\|_\infty \int_{\mathbb{S}^d}\left(\int_{\mathbb{S}^d} g(t)K_j(t,x)d\sigma_d(t)\right)^2 d\sigma_d(x) \\ &\le \|f\|_\infty \|g\|_2^2. \qquad \text{(A.3)}\end{aligned}$$

Consequently,

$$\sup_{g\in B_0\subset B} \mathbb{E}\big[k_g(X)\big]^2 = \|f\|_\infty \sup_{g\in B_0\subset B} \|g\|_2^2 \le \omega_{\mathcal{K}}^2.$$

A.3. **Proof of Theorem 3.5.** From Lemma 3.4 we get that $\mathcal{S}$ is a Parseval frame in $L^2(\mathbb{S}^d)$. From (3.4) we obtain (3.12). Since $\mathcal{S}$ is a Parseval frame in $L^2(\mathbb{S}^d)$ we get that the sum $\bigcup_{j\ge j_0} V_j(\mathbb{S}^d)$ is dense in $L^2(\mathbb{S}^d)$. Let

$$W_j = \mathrm{span}_{L^2(J_\varepsilon)}\{\psi_I^e : e \in \mathcal{E}, I \in \mathcal{D}_j(e)\}.$$

Then by definition of W_j and (3.5)

$$V_j \oplus W_j \subset V_{j+1}.$$

Hence by (3.9)

$$V_j(\mathbb{S}^d) + W_j(\mathbb{S}^d) \subset V_{j+1}(\mathbb{S}^d).$$

Note that both orthogonal projection E^+ and E^- are localized. Then if $f \in L^2(J_\varepsilon)$ is such that

$$\mathrm{supp}\, f \cap (-1-\varepsilon/2, 1+\varepsilon/2)^d = \emptyset$$

then

$$E^\pm \circ T_d^\pm(f) = 0.$$

Hence the space $V_{j+1}(\mathbb{S}^d)$ is spanned by functions $E^\pm \circ T_d^\pm \psi_L^0$ such that $l(L) = 2^{-j-1}$ and

$$\mathrm{supp}\, \psi_L^0 \cap (-1-\varepsilon/2, 1+\varepsilon/2)^d \ne \emptyset.$$

From definition of j_0 we get that such ψ_L^0 are

$$\psi_L^0 = \sum_{I\in\mathcal{D}_j} \langle \psi_L^0, \psi_I^0\rangle \psi_I^0 + \sum_{e\in\mathcal{E}, I\in\mathcal{D}_j} \langle \psi_L^0, \psi_I^e\rangle \psi_I^e.$$

Consequently

$$V_{j+1}(\mathbb{S}^d) \subset V_j(\mathbb{S}^d) + W_j(\mathbb{S}^d)$$

and we prove (3.12). Now we prove (3.14). Let us take ψ_I^0 and $I \in \mathcal{D}_j(0)$. Then

$$T_d^+(\psi_I^0)(\vartheta^+(u)) = \frac{\psi_I^0(S_+(\vartheta^+(u))}{\sqrt{J_d(S_+(\vartheta^+(u)))}}, \qquad u \in \mathbb{S}^d,$$

where

$$J_d(x_1, \dots, x_d) = \left(\frac{2}{1 + x_1^2 + \dots + x_d^2}\right)^d.$$

By (3.11) we have

$$T_d^+(\psi_I^0)(\vartheta^+(u)) = \frac{\psi_I^0(2S_+(u))}{\sqrt{J_d(2S_+(u))}}, \qquad u \in \mathbb{S}^d.$$

Consequently there is $\tilde{I} \in \mathcal{D}_{j+1}(0)$ such that for $y \in \mathbb{R}^d$

$$\psi_I^0(2y) = \psi_{\tilde{I}}^0(y).$$

The function H^+ we obtain comparing $T_d^+(\psi_I^0)(\vartheta^+(u))$ with

$$T_d^+(\psi_{\tilde{I}}^0)(u) = \frac{\psi_{\tilde{I}}^0(S_+(u))}{\sqrt{J_d(S_+(u))}}, \qquad u \in \mathbb{S}^d.$$

Hence

$$H^+(u) = \sqrt{\frac{J_d(2S_+(u))}{J_d(S_+(u))}}.$$

By (3.10) in the natural parametrization $u = (\xi \sin\theta, \cos\theta)$ the function H^+ depends on $a = a(\theta) = \sin\theta/(1 + \cos\theta)$ since

$$J_d(S_+(u)) = J_d\left(\frac{\xi \sin\theta}{1 + \cos\theta}\right) = J_d(a\xi) = \left(\frac{1}{1 + a^2}\right)^d.$$

From Theorem 6.1 [4] we get that there is $C > 0$ such that for $f \in \mathcal{F}(\mathbb{S}^d)$.

$$\|E^\pm f\|_{\mathcal{F}(\mathbb{S}^d)} \le C\|f\|_{\mathcal{F}(\mathbb{S}^d)}.$$

Since both operators E^+ i E^- are projection in Banach spaces, then

$$\mathcal{F}(\mathbb{S}^d) = E^+(\mathcal{F}(\mathbb{S}^d)) \oplus E^-(\mathcal{F}(\mathbb{S}^d))$$

with equivalence of norm

$$\|f\|_{\mathcal{F}(\mathbb{S}^d)} \simeq \|E^+ f\|_{\mathcal{F}(\mathbb{S}^d)} + \|E^- f\|_{\mathcal{F}(\mathbb{S}^d)}.$$

Since $E^+ f \in E^+(\mathcal{F}(\mathbb{S}^d))$ a $E^- f \in E^-(\mathcal{F}(\mathbb{S}^d))$, consequently from Lemma 3.4 we get the Theorem.

A.4. **Proof of Corollary 3.6.** Let

$$K_{j+1}(f) = K_{j+1}^+(f) + K_{j+1}^-(f),$$

where

$$K_{j+1}^+(f) = \sum_{I \in \mathcal{D}_{j+1}} \langle f, E^+ T_d^+(\psi_I^0)\rangle E^+ T_d^+(\psi_I^0).$$

Then

$$K_{j+1}^+(f) = E^+\left(\sum_{I \in \mathcal{D}_{j+1}} \langle E^+ f, T_d^+(\psi_I^0)\rangle T_d^+(\psi_I^0)\right)$$

But for $j \geq j_0$ we have

$$V_j \oplus W_j \subset V_{j+1}.$$

So since

$$supp\left((T_d^+)^{-1}E^+ f\right) \subset [-1-\varepsilon/2, 1+\varepsilon/2]^d$$

we get

$$K_{j+1}^+(f) = E^+\left(\sum_{I\in\mathcal{D}_j}\langle E^+ f, T_d^+(\psi_I^0)\rangle T_d^+(\psi_I^0)\right) + E^+\left(\sum_{e\in\mathcal{E}}\sum_{I\in\mathcal{D}_j(e)}\langle E^+ f, T_d^+(\psi_I^e)\rangle T_d^+(\psi_I^e)\right)$$

If we introduce

$$Q_j(f) = \int_{\mathbb{S}^d} G_j(x,y) f(y)\sigma_d(dy)$$

where

$$G_j(x,y) = \sum_{e\in\mathcal{E}}\sum_{I\in\mathcal{D}_j} E^{\pm}\circ T_d^{\pm}(\psi_I^e)(x)\cdot E^{\pm}\circ T_d^{\pm}(\psi_I^e)(y). \tag{A.4}$$

then for $j \geq j_0$

$$K_{j+1} = K_j + Q_j. \tag{A.5}$$

Hence

$$K_{j+1} = K_{j_0} + \sum_{k=j_0}^{j} Q_k. \tag{A.6}$$

We get the corollary by (3.18), (3.15) and Theorem 3.5.

A.5. **Proof of Lemma 4.1.** Since $E^{\pm}$ are orthogonal projections and $T_d^{\pm}$ are isometric isomorphisms we have

$$\left\|E^{\pm}\circ T_d^{\pm}(\psi_I^{\mathbf{0}})\right\|_2 \leq \left\|T_d^{\pm}(\psi_I^{\mathbf{0}})\right\|_2 = \left\|\psi_I^{\mathbf{0}}\right\|_2 = 1. \tag{A.7}$$

Let

$$\tilde{E}_I^+(x,y) := E^+\circ T_d^+(\psi_I^{\mathbf{0}})(x)\cdot E^+\circ T_d^+(\psi_I^{\mathbf{0}})(y)$$

and

$$\tilde{E}_I^-(x,y) := E^-\circ T_d^-(\psi_I^{\mathbf{0}})(x)\cdot E^-\circ T_d^-(\psi_I^{\mathbf{0}})(y).$$

Then for $y \in \mathbb{S}^d$ and $j \geq j_0$

$$\int_{\mathbb{S}^d} K_j^2(x,y)d\sigma_d(x) = \int_{\mathbb{S}^d}\left[\sum_{I\in\mathcal{D}_j}\left(\tilde{E}_I^+(x,y)+\tilde{E}_I^-(x,y)\right)\right]^2 d\sigma_d(x). \tag{A.8}$$

Note that there is a constant $\tilde{D}$ such that for all $x, y \in \mathbb{S}^d$

$$D(x,y) := \#\{I\in\mathcal{D}_j : \tilde{E}_I^+(x,y)+\tilde{E}_I^-(x,y) \neq 0\} \leq \tilde{D} < \infty.$$

By Jensen's inequality (A.8)

$$\int_{\mathbb{S}^d} K_j^2(x,y)d\sigma_d(x) \le \int_{\mathbb{S}^d} \tilde{D} \sum_{I\in\mathcal{D}_j} \left[\tilde{E}_I^+(x,y) + \tilde{E}_I^-(x,y)\right]^2 d\sigma_d(x)$$

$$\le 2\tilde{D} \int_{\mathbb{S}^d} \sum_{I\in\mathcal{D}_j} \left[\left(\tilde{E}_I^+(x,y)\right)^2 + \left(\tilde{E}_I^-(x,y)\right)^2\right] d\sigma_d(x). \tag{A.9}$$

From (A.7) we get

$$\int_{\mathbb{S}^d} \sum_{I\in\mathcal{D}_j} \left(\tilde{E}_I^\pm(x,y)\right)^2 d\sigma_d(x) = \sum_{I\in\mathcal{D}_j} \left(E^\pm \circ T_d^\pm(\psi_I^{\mathbf{0}})(y)\right)^2 \int_{\mathbb{S}^d} \left(E^\pm \circ T_d^\pm(\psi_I^{\mathbf{0}})(x)\right)^2 d\sigma_d(x)$$

$$\le \sum_{I\in\mathcal{D}_j} \left(E^\pm \circ T_d^\pm(\psi_I^{\mathbf{0}})(y)\right)^2 \le C_1 \left(2^{jd/2}\right)^2 = C_1 2^{jd}. \tag{A.10}$$

Consequently from (A.9) i (A.10)

$$\int_{\mathbb{S}^d} K_j^2(x,y)d\sigma_d(x) \le 4\tilde{D}C_1 \cdot 2^{jd} = D_{_N} 2^{jd}.$$

A.6. **Proof of Theorem 4.2.** From Lemma 2.2 we get that for all $j \ge j_0$

$$\mathbb{E}\left\|f_n(j) - \mathbb{E}f_n(j)\right\|_2^2 \le D_{_N}\frac{2^{jd}}{n} =: D_{_N}\sigma^2(j,n). \tag{A.11}$$

If $f \in B^s_{2,\infty}(\mathbb{S}^d)$ by (3.15) and Corollary 3.6 we get

$$\left\|\mathbb{E}f_n(j) - f\right\|_2^2 \le (\sqrt{2}\cdot 2^{-js}\left\|f\right\|_{s,2})^2 =: B^2(j,f). \tag{A.12}$$

We define

$$j^* = \min\left\{j \in \mathcal{J} : B(j,f) \le \sqrt{D_{_N}}\sigma(j,n)\right\}. \tag{A.13}$$

Note that

$$\mathbb{E}\left\|f_n(j_n) - f\right\|_2^2 = \mathbb{E}\left(\left\|f_n(j_n) - f\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} + \left\|f_n(j_n) - f\right\|_2^2 \mathbf{1}_{\{j_n> j^*\}}\right). \tag{A.14}$$

To show (4.5) we will estimate both components.

(I) For first component we use $(a+b)^2 \le 2(a^2+b^2)$, thus

$$\mathbb{E}\left\|f_n(j_n) - f\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} \le \mathbb{E}\left(\left\|f_n(j_n) - f_n(j^*)\right\|_2 + \left\|f_n(j^*) - f\right\|_2\right)^2 \mathbf{1}_{\{j_n\le j^*\}}$$

$$\le 2\mathbb{E}\left\|f_n(j_n) - f_n(j^*)\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} + 2\mathbb{E}\left\|f_n(j^*) - f\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} \tag{A.15}$$

If $j_n \le j^*$, then by (4.3) we get

$$\mathbb{E}\left\|f_n(j_n) - f_n(j^*)\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} \le \mathbb{E}\left(C(\mathcal{S})(||f||_\infty \vee 1)\frac{2^{j^*d}}{n}\mathbf{1}_{\{j_n\le j^*\}}\right)$$

$$= C(\mathcal{S})(||f||_\infty \vee 1)\sigma^2(j^*,n)$$

Applying (A.11), (A.12) and (A.13) we get

$$\mathbb{E}\left\|f_n(j^*) - f\right\|_2^2 \mathbf{1}_{\{j_n\le j^*\}} = \mathbb{E}\left\|f_n(j^*) - \mathbb{E}f_n(j^*)\right\|_2^2 + \left\|\mathbb{E}f_n(j^*) - f\right\|_2^2$$

$$\le D_{_N}\sigma^2(j^*,n) + B^2(j^*,f) \le 2D_{_N}\sigma^2(j^*,n). \tag{A.16}$$

Note that for $f \in \Sigma(s, \tilde{B})$,

$$\sigma^2(j^*, n) \le D' \tilde{B}^{\frac{2d}{2s+d}} n^{\frac{-2s}{2s+d}}. \tag{A.17}$$

Indeed for $j^{*-} = j^* - 1$ from (A.13) we have

$$B(j^{*-}, f) \ge \sqrt{D_N}\sigma(j^{*-}, n).$$

Consequently using (A.11) and (A.12) we get

$$\|f\|_{s,2}^2 \, 2^{-2sj^{*-}+1} \ge D_N \frac{2^{dj^{*-}}}{n}$$

$$\|f\|_{s,2}^2 \, 2^{-2s(j^*-1)+1} \ge D_N \frac{2^{d(j^*-1)}}{n}$$

$$\frac{2}{D_N} \|f\|_{s,2}^2 \, 2^{2s+d} \ge \frac{1}{n} 2^{j^*(2s+d)}.$$

Hence by standard calculation we get

$$\left(\frac{2}{D_N} \|f\|_{s,2}^2\right)^d 2^{d(2s+d)} \ge \left(\frac{2^{j^* d}}{n}\right)^{2s+d} n^{2s}.$$

From (A.11)

$$\left(\frac{2}{D_N} \|f\|_{s,2}^2\right)^{\frac{d}{2s+d}} 2^d n^{\frac{-2s}{2s+d}} \ge \sigma^2(j^*, n).$$

Finally for $f \in \Sigma(s, \tilde{B})$ and $s \ge d/2$

$$\sigma^2(j^*, n) \le \left(\frac{2}{D_N}\right)^{\frac{d}{2s+d}} 2^d \cdot \tilde{B}^{\frac{2d}{2s+d}} n^{\frac{-2s}{2s+d}} \le \left(\frac{2}{D_N}\right)^{1/2} 2^d \cdot \tilde{B}^{2\frac{d}{2s+d}} n^{\frac{-2s}{2s+d}}.$$

Note that the constant in the inequality (A.17) is the following

$$D' = \sqrt{\frac{2}{D_N}} 2^d. \tag{A.18}$$

By (A.15), (A.16) and (A.17) for $C'' = C(\mathcal{S})(||f||_\infty \vee 1)$.

$$\mathbb{E} \|f_n(j_n) - f\|_2^2 \mathbf{1}_{\{j_n \le j^*\}} \le 2(C'' + D_N)\sigma^2(j^*, n) \le 2(C'' + D_N) D' \tilde{B}^{\frac{2d}{2s+d}} n^{\frac{-2s}{2s+d}}, \tag{A.19}$$

(II) For second component of (A.14) using Schwartz inequality we get

$$\begin{aligned} \mathbb{E}\left(\|f_n(j_n) - f\|_2^2 \mathbf{1}_{\{j_n > j^*\}}\right) &= \sum_{j \in \mathcal{J}, j > j^*} \mathbb{E}\left(\|f_n(j) - f\|_2^2 \mathbf{1}_{\{j_n = j\}}\right) \\ &\le \sum_{j \in \mathcal{J}, j > j^*} \left(\mathbb{E}\|f_n(j) - f\|_2^4\right)^{1/2} \sqrt{P(j_n = j)}. \end{aligned} \tag{A.20}$$

Note that

$$\|f_n(j) - f\|_2 \le \|f_n(j) - \mathbb{E}f_n(j)\|_2 + \|\mathbb{E}f_n(j) - f\|_2 \le \|f_n(j) - \mathbb{E}f_n(j)\|_2 + B(j, f).$$

Hence

$$\mathbb{E}\left(\|f_n(j)-f\|_2\right)^4 \le \mathbb{E}\left(\|f_n(j)-\mathbb{E}f_n(j)\|_2 + B(j,f)\right)^4$$

$$\le 8\left(\mathbb{E}\,\|f_n(j)-\mathbb{E}f_n(j)\|_2^4 + B^4(j,f)\right). \tag{A.21}$$

By Lemma 4.3 and (A.13) for $j > j^*$ and $j \in \mathcal{J}$ there is C such that

$$\mathbb{E}\,\|f_n(j)-\mathbb{E}f_n(j)\|_2^4 \le 4\left(32D_N^2\sigma^4(j,n) + (\sqrt{2}\cdot 2^{-js}\,\|f\|_{s,2})^4\right) \le C\sigma^4(j_{\max},n). \tag{A.22}$$

Consequently by (A.20), (A.22) and the definition of $j_{\max}$ there is C'' such that

$$\mathbb{E}\left(\|f_n(j_n)-f\|_2^2\,\mathbf{1}_{\{j_n>j^*\}}\right) = \sqrt{C''}\sum_{j\in\mathcal{J}:j>j*}\sqrt{P(j_n=j)}. \tag{A.23}$$

Let us fix $j \in \mathcal{J}$ such that $j > j*$. We will estimate $P(j_n = j)$. Since $j_n = j$, then from the definition of j_n and for $j^- = j-1$ we get that there is $l > j-1 \ge j*$ such that

$$\left\|f_n(j^-)-f_n(l)\right\|_2^2 > C(\mathcal{S})(||f||_\infty \vee 1)\frac{2^{ld}}{n}. \tag{A.24}$$

Note that

$$\begin{aligned}
&\left\|f_n(j^-)-f_n(l)\right\|_2 \le \\
&\le \left\|f_n(j^-)-\mathbb{E}f_n(j^-)-f_n(l)+\mathbb{E}f_n(l)\right\|_2 + \left\|f-\mathbb{E}f_n(j^-)\right\|_2 + \|f-\mathbb{E}f_n(l)\|_2 \\
&\le \left\|f_n(j^-)-\mathbb{E}f_n(j^-)-f_n(l)+\mathbb{E}f_n(l)\right\|_2 + B(j^-,f)+B(l,f).
\end{aligned} \tag{A.25}$$

From the definitions of $B(l,f)$ and $\sigma(l,n)$ we get

$$B(j^-,f)+B(l,f) \le 2B(j^*,f) \le 2\sqrt{D_N}\sigma(j^*,f) \le 2\sqrt{D_N}\sigma(l,f).$$

By (A.25)

$$\left\|f_n(j^-)-f_n(l)\right\|_2 \le \left\|f_n(j^-)-\mathbb{E}f_n(j^-)-f_n(l)+\mathbb{E}f_n(l)\right\|_2 + 2\sqrt{D_N}\sigma(l,f). \tag{A.26}$$

Hence by (A.24) and (A.26) we have

$$\begin{aligned}
&P(j_n=j)\le \\
&\le \sum_{l\in\mathcal{J}:l\ge j} P\left(\left\|f_n(j^-)-f_n(l)\right\|_2 > \sqrt{C(\mathcal{S})(||f||_\infty\vee 1)}\sigma(l,n)\right) \\
&\le \sum_{l\in\mathcal{J}:l\ge j} P\Big(\left\|f_n(j^-)-\mathbb{E}f_n(j^-)-f_n(l)+\mathbb{E}f_n(l)\right\|_2 > \\
&\qquad\qquad > \left(\sqrt{C(\mathcal{S})(||f||_\infty\vee 1)}-2\sqrt{D_N}\right)\sigma(l,n)\Big) \\
&\le \sum_{l\in\mathcal{J}:l\ge j}\Big[P\Big(\left\|f_n(j^-)-\mathbb{E}f_n(j^-)\right\|_2 > G\sigma(l,n)\Big) + P\Big(\|f_n(l)-\mathbb{E}f_n(l)\|_2 > G\sigma(l,n)\Big)\Big],
\end{aligned} \tag{A.27}$$

where the constant G is given by

$$G = \frac{\sqrt{C(\mathcal{S})(||f||_\infty\vee 1)}-2\sqrt{D_N}}{2}.$$

Since we assume that $\sqrt{C(\mathcal{S})} \geq 2M_N + 2\sqrt{D_N}$, see (4.4) we get

$$\begin{aligned} G &= \frac{1}{2}\left(\sqrt{C(\mathcal{S})(||f||_\infty \vee 1)} - 2\sqrt{D_N}\right) \\ &\geq \frac{1}{2}\left(2M_N\sqrt{(||f||_\infty \vee 1)} + 2\sqrt{D_N(||f||_\infty \vee 1)} - 2\sqrt{D_N}\right) \\ \langle\ (||f||_\infty \vee 1) \geq 1\ \rangle \quad &\geq M_N\sqrt{(||f||_\infty \vee 1)} \end{aligned}$$

Thus, it holds

$$\begin{aligned} P(j_n = j) \leq \sum_{l \in \mathcal{J}: l \geq j} \Big[& P\left(\left\|f_n(j^-) - \mathbb{E}f_n(j^-)\right\|_2 > M_N\sqrt{(||f||_\infty \vee 1)}\sigma(l,n)\right) \\ & + P\left(\|f_n(l) - \mathbb{E}f_n(l)\|_2 > M_N\sqrt{(||f||_\infty \vee 1)}\sigma(l,n)\right)\Big]. \end{aligned}$$

Since for $l \geq j > j^-$ we have $\sigma(l,n) > \sigma(j^-,n)$ then

$$\begin{aligned} P(j_n = j) \leq \sum_{l \in \mathcal{J}: l \geq j} \Big[& P\left(\left\|f_n(j^-) - \mathbb{E}f_n(j^-)\right\|_2 > M_N\sqrt{(||f||_\infty \vee 1)}\sigma(j^-,n)\right) \\ & + P\left(\|f_n(l) - \mathbb{E}f_n(l)\|_2 > M_N\sqrt{(||f||_\infty \vee 1)}\sigma(l,n)\right)\Big]. \end{aligned}$$

By Theorem 2.5

$$P(j_n = j) \leq \sum_{l \in \mathcal{J}: l \geq j} \left(2e^{-2^{dj^-}} + 2e^{-2^{dl}}\right) \leq 4(j_{max} - j_{min})e^{-2^{dj_{min}}}.$$

Using this in (A.23) we obtain

$$\begin{aligned} \mathbb{E}\left(\|f_n(j_n) - f\|_2^2 \mathbf{1}_{\{j_n > j^*\}}\right) &= \sqrt{C''} \sum_{j \in \mathcal{J}: j > j*} \sqrt{4(j_{max} - j_{min})} e^{-2^{dj_{min}}/2} \\ &\leq 2\sqrt{C''}(j_{max} - j_{min})^{\frac{3}{2}} e^{-\frac{1}{2}2^{dj_{min}}}. \end{aligned} \tag{A.28}$$

But from (4.2)

$$j_{max} - j_{min} \leq \left(\frac{1}{2r+d} - \frac{1}{2R+d}\right)\log_2 n \leq \left(\frac{1}{2r+d} - \frac{1}{2R+d}\right) n$$

and

$$e^{-\frac{1}{2}2^{dj_{min}}} \simeq e^{-\frac{1}{2}n^{\frac{d}{2R+d}}}.$$

Now (A.28) we estimate by

$$\begin{aligned} \mathbb{E}\left(\|f_n(j_n) - f\|_2^2 \mathbf{1}_{\{j_n > j^*\}}\right) &\leq \tilde{C}\left(\frac{1}{2r+d} - \frac{1}{2R+d}\right)^{\frac{3}{2}} n^{\frac{3}{2}} e^{-\frac{1}{2}n^{\frac{d}{2R+d}}} \\ &= \tilde{C}_1 \frac{n^{\frac{3}{2}+\frac{2s}{2s+d}}}{\exp(\frac{1}{2}n^{\frac{d}{2R+d}})} \cdot n^{-\frac{2s}{2s+d}}. \end{aligned} \tag{A.29}$$

But for any $\alpha, \beta, \gamma > 0$ we have

$$\lim_{x \to \infty} \frac{x^\beta}{e^{\alpha x^\gamma}} = 0.$$

Consequently (A.29) we estimate by

$$\mathbb{E}\left(\|f_n(j_n) - f\|_2^2 \mathbf{1}_{\{j_n > j^*\}}\right) \leq C''' \tilde{B}^{2d/(2s+d)} n^{-2s/(2s+d)}, \tag{A.30}$$

since $\tilde{B} > 1$.
Finally by (A.19), (A.30) and (A.14) we obtain the Theorem.

A.7. **Proof of Lemma 4.3.** The idea of proof is from [21]. Note that

$$\mathbb{E}\,\|f_n(j) - \mathbb{E}f_n(j)\|_2^4 =$$
$$= \mathbb{E}\left(\int_{\mathbb{S}^d} |f_n(j)(x) - \mathbb{E}f_n(j)(x)|^2\, d\sigma_d(x)\right)^2$$
$$= \mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left|\frac{1}{n}\sum_{i=1}^n (K_j(x,X_i) - K_jf(x))\right|^2 \left|\frac{1}{n}\sum_{l=1}^n (K_j(y,X_l) - K_jf(y))\right|^2 d\sigma_d(x)d\sigma_d(y)\right)^2.$$

For $x \in \mathbb{S}^d$ we denote $Y_i(x) = K_j(x,X_i) - K_jf(x)$. Hence $\mathbb{E}Y_i(x) = 0$. Thus

$$\mathbb{E}\,\|f_n(j) - \mathbb{E}f_n(j)\|_2^4 = \frac{1}{n^4}\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left|\sum_{i=1}^n Y_i(x)\right|^2 \left|\sum_{l=1}^n Y_l(y)\right|^2 d\sigma_d(x)d\sigma_d(y)\right)$$
$$= \frac{1}{n^4}\mathbb{E}\Bigg(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\Bigg(\sum_{i,l=1}^n Y_i^2(x)Y_l^2(y) + 2\sum_{i=1}^n\sum_{l<m} Y_i^2(x)Y_l(y)Y_m(y) +$$
$$+ 2\sum_{l=1}^n\sum_{i<k} Y_l^2(y)Y_i(x)Y_k(x) + 4\sum_{i<k}\sum_{l<m} Y_i(x)Y_k(x)Y_l(y)Y_m(y)\Bigg)d\sigma_d(x)d\sigma_d(y)\Bigg).$$

Put

$$\mathbb{E}\,\|f_n(j) - \mathbb{E}f_n(j)\|_2^4 = \frac{1}{n^4}(I_1 + I_2 + I_3 + I_4). \tag{A.31}$$

Now we will estimate all components of (A.31). We will use the following inequality which is a consequence of Jensen's inequality for all $x \in \mathbb{S}^d$
(A.32)
$$(K_jf(x))^2 = \left(\int_{\mathbb{S}^d} K_j(x,u)f(u)d\sigma_d(u)\right)^2 \le \int_{\mathbb{S}^d} K_j^2(x,u)f(u)d\sigma_d(u) = \mathbb{E}K_j^2(x,X).$$

We start with I_1.

$$I_1 = \mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\sum_{i,l=1}^n Y_i^2(x)Y_l^2(y)d\sigma_d(x)d\sigma_d(y)\right)$$
$$= \sum_{i=1}^n \mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} Y_i^2(x)Y_i^2(y)d\sigma_d(x)d\sigma_d(y)\right) + \sum_{1\le i\ne l\le n}\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} Y_i^2(x)Y_l^2(y)d\sigma_d(x)d\sigma_d(y)\right)$$

Put

$$I_1 = \sum_{i=1}^n I_1^{i,i} + \sum_{1\le i\ne l\le n} I_1^{i,l}. \tag{A.33}$$

For $i = 1, 2, \ldots, n$ using the inequality $(a+b)^2 \le 2(a^2+b^2)$, and (A.32) we get for any i, l

$$\begin{aligned} I_1^{i,l} &= \mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} Y_i^2(x)Y_l^2(y)d\sigma_d(x)d\sigma_d(y)\right) \\ &= \mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\Big[K_j(x,X_i) - K_jf(x)\Big]^2\Big[K_j(y,X_l) - K_jf(y)\Big]^2 d\sigma_d(x)d\sigma_d(y)\right) \\ &\le 4\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\Big[K_j^2(x,X_i) + (K_jf(x))^2\Big]\Big[K_j^2(y,X_l) + (K_jf(y))^2\Big]d\sigma_d(x)d\sigma_d(y)\right) \\ &\le 4\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\Big\{\mathbb{E}\Big[K_j^2(x,X_i)K_j^2(y,X_l)\Big] + 3\mathbb{E}\Big[K_j^2(x,X_i)\Big]\mathbb{E}\Big[K_j^2(y,X_l)\Big]\Big\}d\sigma_d(x)d\sigma_d(y). \end{aligned}$$

Now

$$\begin{aligned} I_1^{i,i} \le 4\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_j^2(x,u)K_j^2(y,u)f(u)d\sigma_d(x)d\sigma_d(y)d\sigma_d(u) + \\ + 12\int_{\mathbb{S}^d}\mathbb{E}\Big[K_j^2(x,X_i)\Big]d\sigma_d(x)\int_{\mathbb{S}^d}\mathbb{E}\Big[K_j^2(y,X_i)\Big]d\sigma_d(y). \end{aligned} \tag{A.34}$$

For all $x \in \mathbb{S}^d$ from Lemma 4.1

$$\begin{aligned} \int_{\mathbb{S}^d}\mathbb{E}\Big[K_j^2(x,X_i)\Big]d\sigma_d(x) &= \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_j^2(x,u)f(u)d\sigma_d(x)d\sigma_d(u) \\ &\le \int_{\mathbb{S}^d} f(u)\cdot D_N 2^{jd}d\sigma_d(u) = D_N 2^{jd}. \end{aligned} \tag{A.35}$$

Hence using (A.34) and Lemma 4.1 we get

$$\begin{aligned} I_1^{i,i} &\le 4\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} K_j^2(x,u)K_j^2(y,u)f(u)d\sigma_d(x)d\sigma_d(y)d\sigma_d(u) + 12(D_N 2^{jd})^2 \\ &\le 4\int_{\mathbb{S}^d} f(u)\left(D_N 2^{jd}\right)^2 d\sigma_d(u) + 12D_N^2 2^{2jd} = 16D_N^2 2^{2jd}. \end{aligned} \tag{A.36}$$

For all $1 \le i \ne l \le n$ random variable $K_j^2(x,X_i)$ and $K_j^2(y,X_l)$ are independent using (A.35), we obtain

$$I_1^{i,l} = 16\int_{\mathbb{S}^d}\mathbb{E}\Big[K_j^2(x,X_i)\Big]d\sigma_d(x)\int_{\mathbb{S}^d}\mathbb{E}\Big[K_j^2(y,X_l)\Big]d\sigma_d(y) \le 16(D_N 2^{jd})^2. \tag{A.37}$$

Consequently (A.33) and (A.36) with (A.37) give

$$I_1 = \sum_{i=1}^n I_1^{i,i} + \sum_{1\le i\ne l\le n} I_1^{i,l} \le \sum_{i=1}^n 16D_N^2 2^{2jd} + \sum_{1\le i\ne l\le n} 16D_N^2 2^{2jd} = 16n^2D_N^2 2^{2jd}. \tag{A.38}$$

Let us estimate I_2 and I_3. Since $Y_1, \ldots, Y_n$ are independent, we get

$$I_2 = 2\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\sum_{i=1}^n\sum_{l<m} Y_i^2(x)Y_l(y)Y_m(y)d\sigma_d(x)d\sigma_d(y)\right) = 0. \tag{A.39}$$

Similar

$$I_3 = 2\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\sum_{l=1}^n\sum_{i<k} Y_l^2(y)Y_i(x)Y_k(x)d\sigma_d(x)d\sigma_d(y)\right) = 0. \tag{A.40}$$

To finish the proof of Lemma we need to estimate I_4.,

$$I_4 = 4\sum_{i<k}\sum_{l<m}\mathbb{E}\left(\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} Y_i(x)Y_k(x)Y_l(y)Y_m(y)d\sigma_d(x)d\sigma_d(y)\right) = 4\sum_{i<k}\sum_{l<m} I_4^{i,k,l,m}. \tag{A.41}$$

Note that for $i<k$, $l<m$ and assuming $i=l$ i $k=m$ we get

$$I_4^{i,k,l,m} = \int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left\{\mathbb{E}\big[Y_i(x)Y_i(y)\big]\right\}^2 d\sigma_d(x)d\sigma_d(y). \tag{A.42}$$

From Jensen's inequality (A.36) we get

$$I_4^{i,k,l,m} \le I_1^{i,i} \le 16D_{_N}^2 2^{2jd}. \tag{A.43}$$

If $i<k$ and $l<m$ assuming that $i\neq l$ or $k\neq m$ we get

$$I_4^{i,k,l,m} = \int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\mathbb{E}\big[Y_i(x)Y_k(x)Y_l(y)Y_m(y)\big]d\sigma_d(x)d\sigma_d(y) = 0. \tag{A.44}$$

Now by (A.43), (A.44) and (A.41) we get

$$I_4 = 4\sum_{i<k}\sum_{l<m} I_4^{i,k,l,m} \le n^2\cdot 16D_{_N}^2 2^{2jd} = 16n^2D_{_N}^2 2^{2jd}. \tag{A.45}$$

By (A.38), (A.39), (A.40) and (A.45) applying (A.31) we get

$$\begin{aligned}\mathbb{E}\,\|f_n(j)-\mathbb{E}f_n(j)\|_2^4 &= \frac{1}{n^4}(I_1+I_2+I_3+I_4) \le \frac{32}{n^4}n^2D_{_N}^2 2^{2jd}\\ &= 32D_{_N}^2\left(\frac{2^{jd}}{n}\right)^2 = 32D_{_N}^2\sigma^4(j,n).\end{aligned}$$

Finally to finish the proof we use above inequality and (A.12) with (A.21). Indeed

$$\mathbb{E}\,\|f_n(j)-f\|_2^4 \le 4\left(32D_{_N}^2\sigma^4(j,n) + (\sqrt{2}\cdot 2^{-js}\,\|f\|_{s,2})^4\right).$$

Acknowledgements. The work of B. Ćmiel was partially supported by the Faculty of Applied Mathematics AGH UST dean grant for PhD students and young researchers within subsidy of Ministry of Science and Higher Education.

References

1. P. Auscher, G. Weiss, M. V. Wickerhauser, *Local sine and cosine bases of Coifman and Meyer and the construction of smooth wavelets.* Wavelets, 237-256, Wavelet Anal. Appl., 2, Academic Press, Boston, MA, 1992.
2. P. Baldi, G. Kerkyacharian, D. Marinucci, and D. Picard, *Adaptive density estimation for directional data using needlets.* Ann. Statist. Volume 37, Number 6A (2009), 3362-3395.
3. S. Boucheron, G. Lugosi G. and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford, UK 2013
4. M. Bownik, K. Dziedziul, *Smooth Orthogonal Projections on Sphere*, Constructive Approximation vol. 41 (1), 23-48, 2015.
5. M. Bownik, K. Dziedziul and A. Kamont, *Smooth orthogonal projections on Riemannian manifold*, 2018, arXiv:1803.03634.
6. A. Bull, R. Nickl, *Adaptive confidence sets in L^2*, Probability Theory and Related Fields, vol. 156 (3-4), 889-919, 2013.
7. B. Ćmiel, K. Dziedziul, *Density smoothness estimation problem using a wavelet approach*, ESAIM: Probability and Statistics, 18, 130-144, 2014.

8. B. Ćmiel, K. Dziedziul, B. Wolnik, *The smoothness test for a density function*, Nonlinear Analysis: Theory, Methods & Applications, 104, 21-39, 2014.
9. F. Dai, *Characterizations of function spaces on the sphere using frames*, Transactions of the American Mathematical Society, vol. 359 (2), 567-589, 2007.
10. F. Dai, Y. Xu, Approximation theory and harmonic analysis on spheres and balls, Springer-Verlag, New York, 2013.
11. I. Daubechies, *Ten lectures on wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 61, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
12. C. Durastanti, *Adaptive Global Thresholding on the Sphere*, Journal of Multivariate Analysis, Vol. 151, 110-132, 2016.
13. W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, *Wavelets, Approximation, and Statistical Applications.* Lecture Notes in Statistics, 129. Springer-Verlag, New York, 1998.
14. E. Hebey, *Nonlinear analysis on manifolds: Sobolev spaces and inequalities.* Courant Lecture Notes in Mathematics, 5. American Mathematical Society, Providence, RI, 1999.
15. K. Dziedziul, M. Kucharska, B. Wolnik, *Estimation of the smoothness of density*, Journal of Nonparametric Statistics, vol. 23 (4), 991-1001, 2011.
16. D. Geller, I. Pesenson, *Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds*, Journal of Geometric Analysis, vol. 21 (2), 334-371, 2011.
17. E. Giné, R. Nickl, *Rates of contraction for posterior distributions in L^r-metrics,* $1 \leq r \leq \infty$, The Annals of Statistics, 2883-2911, 2011.
18. E. Giné and R. Nickl *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press 2015.
19. E. Hernández, G. Weiss, *A First Course on Wavelets*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1996.
20. M. R. Hestenes, *Extension of the range of a differentiable function*, Duke Math. J., 8, 183-192, 1941.
21. P. Hall, *Central limit theorem for integrated square error of multivariate nonparametric density estimators*, Journal of Multivariate Analysis, vol. 14 (1), 1-16, 1984.
22. P. Hall, G. Kerkyacharian, and D. Picard, *Block threshold rules for curve estimation using kernel and wavelet methods*, Ann. Statist. Volume 26, Number 3 (1998), 922-942.
23. G. Kerkyacharian, R. Nickl, D. Picard, *Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds*, Probability Theory and Related Fields, vol. 153 (1-2), 363-404, 2012.
24. G. Kerkyacharian, T. M. Pham Ngoc, and D. Picard, *Localized spherical deconvolution*, Ann. Statist. Volume 39, Number 2 (2011), 1042-1068.
25. A. Kueh, *Locally adaptive density estimation on the unit sphere using needlets.* Constr. Approx. 36 (3), 433-458, 2012.
26. O.V. Lepski, *Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates.*, Theory Probab. Appl. 36, (1991) 682-697.
27. P. I. Lizorkin, and Kh. P. Rustamov, *Nikol'skij-Besov spaces on the sphere in connection with approximation theory*, Trudy Matematicheskogo Instituta im. VA Steklova, vol. 204, 172-200, 1993
28. Y. Meyer, *Wavelets and Operators* (Cambridge Studies in Advanced Mathematics) (D. Salinger, Trans.). Cambridge: Cambridge University Press, 1993.
29. F. Narcowich, P. Petrushev, J. Ward, *Decomposition of Besov and Triebel-Lizorkin spaces on the sphere*, Journal of Functional Analysis, vol. 238 (2), 530-564, 2006.
30. F. Narcowich, P. Petrushev, J. Ward, *Localized tight frames on spheres.* SIAM J. Math. Anal. 38 (2), 574-594, 2006.

Faculty of Applied Mathematics, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracow, Poland.
E-mail address: cmielbog@gmail.com

Faculty of Applied Mathematics, Gdańsk University of Technology, ul. G. Narutowicza 11/12, 80-952 Gdańsk, Poland
E-mail address: karol.dziedziul@pg.edu.pl

Faculty of Applied Mathematics, Gdańsk University of Technology, ul. G. Narutowicza 11/12, 80-952 Gdańsk, Poland