Convolution on the n-sphere with application to pdf modeling

Ivan Dokmanić, *Student Member, IEEE,* Davor Petrinović, *Member, IEEE*

Abstract—In this paper we derive an explicit form of the convolution theorem for functions on an n-sphere. Our motivation comes from the design of a probability density estimator for n-dimensional random vectors. We propose a pdf estimation method that uses the derived convolution result on S^n. Random samples are mapped onto the n-sphere and estimation is performed in the new domain by convolving the samples with the smoothing kernel density. The convolution is carried out in the spectral domain. Samples are mapped between the n-sphere and the n-dimensional Euclidean space by the generalized stereographic projection. We apply the proposed model to several synthetic and real world datasets and discuss the results.

Index Terms—n-sphere, hypersphere, spherical harmonics, hyperspherical harmonics, rotations, convolution, density estimation

I. INTRODUCTION

EXTENSION of signal processing methodologies to non-euclidean spaces is a natural way of dealing with directional data. A common example is a 2-sphere and many Euclidean paradigms that have found their spherical analogue. The most basic is the notion of Fourier transform, that on the sphere corresponds to the expansion of functions into series of familiar spherical harmonics. Vast amount of literature is available on such expansions, mostly from quantum mechanics and mathematical physics [1]. It is known from group representation theory that the Fourier transform can be defined on any compact Lie group and consequently on homogeneous spaces of these groups [2]. An example is the group of rotations of a three-dimensional space about the origin, $SO(3)$, and the 2-sphere, S^2, as a corresponding homogeneous space. Application of signal processing methodologies to different manifolds is facilitated by the discovery of efficient algorithms, analogous to the Fast Fourier Transform (FFT). Efficient algorithms have been designed and implemented for Fourier transform on the 2-sphere [3], [4] and on the $SO(3)$ rotation group [5]. Some FFT generalization ideas are surveyed in [6]. The generalized concept of convolution on groups is intimately related to the concept of filtering on homogeneous spaces. Some insight into spherical filtering can be found in [7] and with particular emphasis on wavelet transform in [8]–[12]. Computation of the Fourier transform and convolution on

groups is studied within the theory of noncommutative harmonic analysis. Examples of applications of noncommutative harmonic analysis in engineering are analysis of the motion of a rigid body, workspace generation in robotics, template matching in image processing, tomography, etc. A comprehensive list with accompanying theory and explanations is given in [13].

Statistics of random vectors whose realizations are observed along manifolds embedded in Euclidean spaces are commonly termed directional statistics. An excellent review may be found in [14]. It is of interest to develop tools for the directional statistics in analogy with the ordinary Euclidean. In particular, one can address a pdf estimation problem from the observations on a manifold. For the ubiquitous spherical case, a common domain example is the earth surface. Typical pdf estimation methods are Parzen windowing [15], also called kernel density estimation, and mixture models, with the most popular Gaussian mixture model (GMM), [16]–[18]. Application of the GMM to source coding is discussed in [19]. A discussion of Parzen windowing on manifolds is available in [20].

In this paper we discuss convolution on a sphere in an n-dimensional space, $S^{n-1} \subset \mathbb{R}^n$, regarded as a homogeneous space of the group of rotations of the n-dimensional space, $SO(n)$. A sphere in the n-dimensional space is a manifold of dimension $n-1$, hence the seemingly odd notation. We derive the explicit expression for convolution in the frequency domain. Our motivation comes from the design of a probability density estimator for data of arbitrary dimension, more generally from source coding.

We propose a pdf estimation model that maps $(n-1)$-dimensional samples onto the S^{n-1}. The involved stereographic mapping is beneficial for a class of bell-shaped densities as discussed in Section IV-D. For the discussion of benefits of transformations in density estimation see [21]. Pdf estimation is performed in the new domain using the kernel density estimation technique. This corresponds to putting the smoothing kernel function at the place of each sample and summing their contributions. Since the described technique can be regarded as a convolution between the symmetric kernel and the dataset, we can use the derived convolution result to realize the estimation in the frequency domain. The spatial estimate can be efficiently computed from the spectral model using the fast inverse spherical Fourier transform, [3], [4], [6].

Modern work on kernel density estimation deals with dependent random variables. It has been proposed by several authors [22]–[27] that the convolution formula can be used to improve the asymptotic properties of such estimators. The main idea in this setting is to improve the estimator by using

the structure of the model (i.e. by convolution). Note that we use the convolution only to compute the estimate, not to handle the dependency.

Pdf estimation is important in source coding where it is interesting to have compact representations of the data that enable adaptation of a quantizer to the changing statistics of the source through a narrow side-channel.

Another class of pdf estimators are those based on expansions of the pdf into series of orthogonal functions. If the smoothing kernel in the proposed approach degenerates to a Dirac delta function, an orthogonal series estimator is obtained that corresponds to Fourier expansions. An excellent general treatment of the density estimation theory along with L_1 treatment of orthogonal series estimators can be found in [28]. L_2 and L_∞ cases are treated in [29]. In this context the expansion coefficients, i.e. Fourier transform of the pdf estimate, may be regarded as model parameters. On the sphere, this corresponds to spherical harmonic expansions. Density estimation from samples on S^2 using Fourier transform is treated in [30]. Deconvolution density estimation is examined in [31] for the case of observations on S^2 corrupted by random rotations and in [32] for observations on $SO(3)$.

This paper is organized as follows. In Section II some background on hyperspherical harmonics and rotations in $\mathbb{R}^n$ is given. In Section III we discuss convolution on S^n and how it can be applied to density estimation. We derive the exact expression for convolution transform in Appendix B. In Section IV we lay out the details of the proposed estimation algorithm on S^n. Section V gives some results on the performance of the estimator. Experiments that verify the sanity of the model are described in Section VI.

II. DEFINITIONS AND PREREQUISITES

By a unit sphere in $\mathbb{R}^n$ we refer to a set of points whose Euclidean distance from the origin is equal to unity.

$$S^{n-1} = \left\{ \mathbf{x} \in \mathbb{R}^n | \mathbf{x}^T \mathbf{x} = 1 \right\}. \tag{1}$$

As already mentioned, the superscript $n-1$ is due to the fact that a sphere in $\mathbb{R}^n$ as defined by (1) is a manifold of dimension $n-1$. S^{n-1} may be parameterized by a set of hyperspherical polar coordinates. If $(x^{(1)}, \cdots, x^{(n)})$ are cartesian coordinates in $\mathbb{R}^n$, then we define the angles $\theta_1, \cdots, \theta_{n-1}$ and the radius r as

$$
\begin{aligned}
x^{(1)} &= r \sin \theta_{n-1} \sin \theta_{n-2} \cdots \sin \theta_2 \sin \theta_1 \\
x^{(2)} &= r \sin \theta_{n-1} \sin \theta_{n-2} \cdots \sin \theta_2 \cos \theta_1 \\
x^{(3)} &= r \sin \theta_{n-1} \sin \theta_{n-2} \cdots \cos \theta_2 \\
&\vdots \\
x^{(n-1)} &= r \sin \theta_{n-1} \cos \theta_{n-2} \\
x^{(n)} &= r \cos \theta_{n-1}.
\end{aligned} \tag{2}
$$

By setting $r = 1$ we get the coordinates of a point on a unit sphere S^{n-1}. This choice of coordinates is not unique, but it is a natural generalization of spherical polar coordinates in $\mathbb{R}^3$. In the familiar case of $S^2 \subset \mathbb{R}^3$, θ_2 corresponds to the elevation and θ_1 corresponds to the azimuth, often denoted $(\theta, \varphi) = (\theta_2, \theta_1)$.

In the space of square integrable functions on S^{n-1} we can choose a particular orthonormal basis $\{\Xi_\mathbf{K}^l\}$ called hyperspherical harmonics [2], [33]. Thus, for $f \in L^2(S^{n-1})$ we have

$$f(\xi) = \sum_l \sum_\mathbf{K} \widehat{f}(l, \mathbf{K}) \Xi_\mathbf{K}^l(\xi), \tag{3}$$

where $\xi = \xi(\theta_1, \cdots, \theta_{n-1})$, and $\widehat{f}(l, \mathbf{K})$ are projections of f onto the basis functions,

$$\widehat{f}(l, \mathbf{K}) = \int_{S^{n-1}} f(\xi) \overline{\Xi_\mathbf{K}^l(\xi)} d\xi. \tag{4}$$

In (3) and (4), l and $\mathbf{K}$ index the basis functions. A compound index $\mathbf{K}$ is a sequence $(k_1, \ldots, k_{n-2})$ of integers such that $l \equiv k_0 \geq k_1 \geq \cdots \geq k_{n-3} \geq |k_{n-2}| \geq 0$. For functions that are not band-limited, $0 \leq l < \infty$. In practice, we have to limit the range of l. That limit corresponds to the spectral bandwidth. Intuitively, one can think of l as of a master index that determines some global level of detail modeled by corresponding basis functions and $\mathbf{K}$ as a way of counting all the basis functions at the l-th detail level. A rotation invariant area element on S^{n-1} is given in the introduced coordinates as

$$d\xi = \frac{\Gamma(n/2)}{2\pi^{n/2}} \prod_{j=1}^{n-1} \sin^{n-j-1} \theta_{n-j} \ d\theta_j \tag{5}$$

where normalization is chosen such that the measure of the whole sphere is equal to 1. Note that in the remainder of the paper the hat notation $\widehat{f}$ will be used for the Fourier transform of the function f while tilde notation $\widetilde{f}$ for the estimate of the density f. Therefore, $\widehat{\widetilde{f}}$ implies the spectral representation of the estimate. Furthermore, we will use the Greek letter ξ to denote points on S^{n-1} when spherical polar parameterization is assumed.

Analysis in the sense of (3) is a Fourier series for functions from $L^2(S^{n-1})$. In the 2-sphere case, functions $\Xi_K^l(\xi)$ coincide with familiar spherical harmonics $Y_l^m(\theta, \varphi)$. An explicit expression for hyperspherical harmonics $\Xi_\mathbf{K}^l(\xi)$ is given in Appendix A.

Rotations of S^{n-1}, or equivalently, rotations in $\mathbb{R}^n$ form a group $SO(n)$ whose elements are $n \times n$ real orthogonal matrices of unit determinant. $SO(n)$ may be parameterized by $\frac{1}{2}n(n-1)$ real numbers. Natural choice are the Euler angles. It can be shown that any rotation $g \in SO(3)$ may be written as a product of three matrices, $g = u(\varphi)a(\theta)u(\psi)$ where $\varphi \in [0, 2\pi]$, $\theta \in [0, \pi]$, $\psi \in [0, 2\pi]$, and u represents the matrix of a rotation about the z-axis while a is the rotation about the y-axis. Rotation of the sphere parameterized by the Euler angles φ, θ and ψ takes the north pole η with cartesian coordinates $(0, 0, 1)$ to the point

$$w(\theta, \varphi) = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta), \tag{6}$$

with spherical polar coordinates (θ, φ). This can be visualized in the following fashion: first twist the function about the z-axis for the amount specified by ψ, then tilt the sphere so that the old north pole points in the (θ, φ) direction.

The analogy can be drawn to the $\mathbb{R}^n$ case, although not without some careful reflection: rotations should be regarded as rotations in a plane, not rotations about an axis. That is the reason that dimension of $SO(n)$ is $\frac{1}{2}n(n-1)$, the number of mutually orthogonal (coordinate) planes in $\mathbb{R}^n$. Rotation about an axis in $\mathbb{R}^n$ is not defined because there are several mutually nonparallel (in fact orthogonal) planes perpendicular to each axis. While all the $\frac{1}{2}n(n-1)$ angles are needed to specify the rotation of S^{n-1}, only $n-1$ determine the new location of the generalized north pole.

III. CONVOLUTION ON S^{n-1}

The notion of convolution is often associated with convolution on the line. Given two signals $f_1, f_2 : \mathbb{R} \to \mathbb{C}$, their convolution is defined as

$$f_1 * f_2(x) = \int_{-\infty}^{\infty} f_1(y) f_2(x-y)\, dy. \tag{7}$$

In the context of group representation theory, this is interpreted as a convolution of square integrable functions on a real additive group, $L^2(\mathbb{R})$ with the ordinary addition as a group composition law. Similarly we define the convolution of two functions $f_1(g)$ and $f_2(g)$ given on a general compact group G,

$$f_1 * f_2(g) = \int_G f_1(g_1) f_2(g_1^{-1} g)\, dg_1. \tag{8}$$

Group action in the argument of f_2 in (7) is called right, while the one in (8) is called left. Unlike the case of a real line, convolution in general is not commutative, i.e. $f_1 * f_2(g) \neq f_2 * f_1(g)$.

S^{n-1} is a homogeneous space of $SO(n)$ group. This means that for any two points ξ_1, $\xi_2 \in S^{n-1}$ there exists some rotation $g \in SO(n)$ such that $g\xi_1 = \xi_2$ i.e. $SO(n)$ acts transitively on S^{n-1}. Note that here we employ $SO(n)$ as a transformation group of $\mathbb{R}^n$ that acts on vectors from $\mathbb{R}^n$ by rotating them. Therefore, $g\xi$ denotes the rotation of the point ξ by g, regardless of the parameterization of ξ and g.

So far we have defined the convolution of functions whose arguments are group elements. What we really want is the convolution of functions on homogeneous spaces under the action of a group, in particular on S^{n-1}. First let's define the rotation operator for functions on S^{n-1}. Each rotation $g \in SO(n)$ induces a rotation operator on $L^2(S^{n-1})$ by

$$R^n(g) f(\xi) = f(g^{-1}\xi). \tag{9}$$

We define a (left) convolution operator for an arbitrary function $k \in L^2(S^{n-1})$ by using k as a weighing factor for operators $R^n(g)$ (definition adopted from [4]),

$$\begin{aligned} \mathcal{C}_k f(\xi) &= \left(\int_{SO(n)} dg\, k(g\eta) R^n(g) \right) f(\xi) \\ &= \int_{SO(n)} k(g\eta) f(g^{-1}\xi)\, dg \\ &= k * f(\xi). \end{aligned} \tag{10}$$

Group of motions of S^{n-1} is the group of rotations of $\mathbb{R}^n$, $SO(n)$. Therefore, we could allow that $k \in L^2(SO(n))$ and

have the result that lives on $SO(n)$. If we wanted the result on S^{n-1}, we would have to integrate out (sum the contributions of) the indistinguishable motions. This is implicit in (10).

The most interesting property of the convolution on S^{n-1} is that it corresponds to multiplication of Fourier coefficients as follows. Let $f_1, f_2 \in L^2(S^{n-1})$, and $f = f_1 * f_2$. Then

$$\widehat{f}(l, \mathbf{K}) = \sqrt{\frac{\Gamma(l+n-2)(2l+n-2)}{l!\Gamma(n-1)}}\, \widehat{f_1}(l, \mathbf{K}) \widehat{f_2}(l, \mathbf{0}). \tag{11}$$

This property has been proved for the special case of convolution on S^2 [4]. It is natural to assume that the similar property is valid in the general case, but we haven't been able to find the formula in the literature so we derived it ourselves. Details are given in Appendix B. The proof is straightforward from the definition, and uses the notion of matrix elements of the left quasi-regular representation of $SO(n)$ and their mutual orthogonality. Key part of the convolution formula is the normalization factor that depends on the dimension n and on the index l. Notice also the degeneration of $\mathbf{K}$ to $\mathbf{0}$ in the second index of $\widehat{f_2}$. This means that we only need information about the θ_{n-1} dependence of f_2. For the 2-sphere it means that after we apply all the motions given by the Euler angle ψ and sum the results, we always get a function symmetric about the z-axis, in other words a function of the elevation only. Regardless of the azimuthal dependence of f_2, from the viewpoint of the spherical convolution we always treat it like an axially symmetric function - just spin it around the z-axis. It is also important to notice that (11) depends on the choice of the measure, and needs to be adjusted by a scalar factor if different measure is chosen.

IV. ESTIMATION MODEL

We describe the model for the general n-dimensional case, and then give examples for two and three dimensional processes. Figures that attempt to illustrate the concept are given for 2-D case, that is for the 2-sphere embedded in three dimensional Euclidean space, since these are rather intuitive and suitable for visualization.

Probability densities of processes that are of our interest carry most of the information near their mean. Details that are farther away from the mean become less important in some sense defined by the application. For example, when modeling data source statistics with quantization and coding in mind it is important to have compact models that require comparatively narrow side-channels for forward adaptation. Intuitively, it makes sense to model the pdf regions that contribute less to the distortion with less care than those that significantly affect the distortion. Consider a mapping from a plane to a sphere that maps regions where process really exists to say, southern hemisphere and everything else to the northern hemisphere. That way the sphere area allocated for the relevant (interesting) part of the process domain is equal to the sphere area allocated for the whole remaining domain where pdf becomes relatively flat. Now we can model the pdf on the sphere as if it was the natural domain of a process. The estimation process is illustrated in Fig. 1. We propose a generalization of this
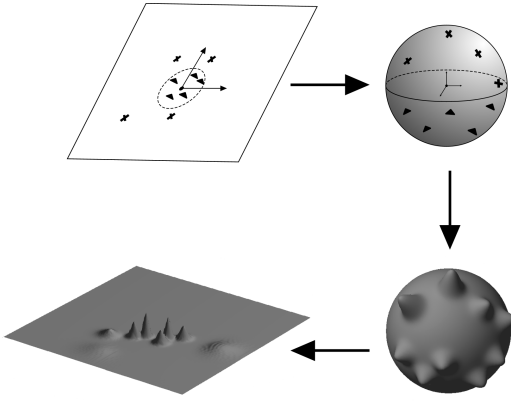
Fig. 1.   Illustration of the estimation process. A dataset that lives in $\mathbb{R}^n$ is mapped onto the n-sphere via the inverse stereographic projection. On a sphere, density estimation is performed by convolving the samples with a kernel density; convolution is carried out in the spectral domain. Density estimate is then projected back into the $\mathbb{R}^n$ with the proper measure adjustment (Jacobian).
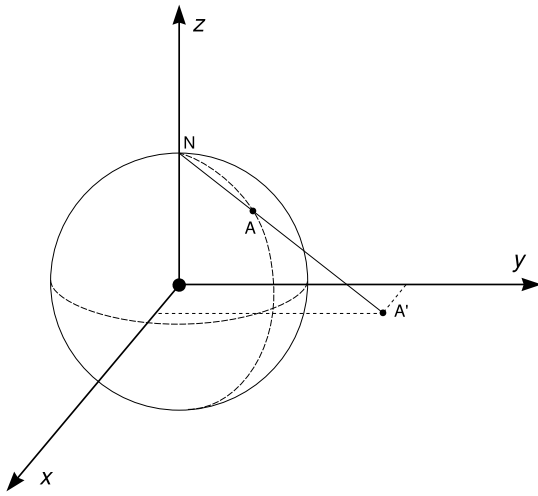

Fig. 2.   Stereographic projection from S^2 to equatorial plane in $\mathbb{R}^3$. $A^{'}$ is the projection of A from N, where N is the north pole. That is, a straight line through N and A intersects the $x\mathrm{O}y$ plane at $A^{'}$.

approach to higher dimensional spaces, e.g. mapping the 3-D space onto the hypersurface of a hypersphere S^3.

This intuitive reasoning is formalized for the univariate case in [21]. The authors discuss the benefits of transforming the original data and using a global width kernel density estimator in the new domain. The transformation approach is again in some sense equivalent to the locally adaptive kernel density estimation, as proposed in [34].

A. Mapping the process onto the sphere

The first step is to map the n-dimensional dataset $\{\mathbf{x}_k\}_{k=1}^N$, $\mathbf{x}_k = \left[x_k^{(1)}, \cdots, x_k^{(n)}\right]^T$ onto the n-sphere. We want the result of this mapping to be the set of spherical polar angle coordinates that represent points on the unit n-sphere, $\{\boldsymbol{\theta}_k\}_{k=1}^N$, $\boldsymbol{\theta}_k = \left[\theta_k^{(1)}, \cdots, \theta_k^{(n)}\right]^T$. This can be achieved by the inverse stereographic projection.

Stereographic projection is given by

$$(x^{(1)}, \cdots, x^{(n+1)}) \mapsto \left(\frac{x^{(1)}}{1 - x^{(n+1)}}, \cdots, \frac{x^{(n)}}{1 - x^{(n+1)}}, 0\right),$$
(12)

where $(x^{(1)}, \cdots, x^{(n+1)})$ are the coordinates of the point on the sphere $S^n \subset \mathbb{R}^{n+1}$ that project to a hyperplane perpendicular to $x^{(n+1)}$ axis (through the hyperequator). This is depicted in Fig. 2 for the S^2 case. We actually use the inverse stereographic projection to map the points from the plane onto the sphere. Before projection we employ a preparation step that involves debiasing and scaling the dataset,

$$\mathbf{x} \mapsto \mathrm{diag}(s_1, \ldots, s_n)(\mathbf{x} - \overline{\mathbf{x}}),$$
(13)

where $s_1, \ldots, s_n$ are scaling factors and $\overline{\mathbf{x}}$ is the mean of the samples. Scaling factors are a function of dataset variance, and are chosen so that the important region of the process domain fits approximately within a unit sphere, in accordance with the previous discussion.

It is also possible to design a custom mapping that is better suited to some special situation. Particular choice of the mapping affects the estimation performance, and can be tuned to the properties of the target process. This can be regarded as a plug-in component of the proposed approach.

B. Parzen windowing on a sphere

Now that we have the dataset on a sphere, we want to make a pdf estimation by centering a smoothing window function around each sample and summing the individual contributions. We propose an efficient method to obtain the transform coefficients of the estimated pdf.

When using Parzen windows on a line, it is clear that the estimation can be regarded as a convolution of the window, and the function $\Delta(x)$ constructed by putting a Dirac delta function $\delta(x)$ at the place of each sample,

$$\begin{aligned} \Delta(x) &= \frac{1}{N} \sum_{k=1}^N [T(x_k)\delta](x) \\ &= \frac{1}{N} \sum_{k=1}^N \delta(x - x_k), \end{aligned}$$
(14)

where $T(x_k)$ is the operator of translation by x_k.

If we know the Fourier transform of the applied window, $\widehat{w}$ then the convolution, i.e. the estimate,

$$\begin{aligned} \widetilde{f}(x) &= (\Delta * w)(x) \\ &= \int_{-\infty}^{\infty} \Delta(y)w(x - y)\,dy \\ &= \frac{1}{N} \sum_{k=1}^N w(x - x_k). \end{aligned}$$
(15)

has a Fourier transform that can be written as a product

$$
\begin{aligned}
\widehat{\overline{f}}(\omega) &= \widehat{w}(\omega)\widehat{\Delta}(\omega) \\
&= \widehat{w}(\omega)\left[\widehat{\sum_{k=1}^{N} T(x_k)\delta}\right](\omega) \\
&= \widehat{w}(\omega)\sum_{k=1}^{N} e^{-j\omega x_k}.
\end{aligned} \tag{16}
$$

In (15) we assumed that the window w is symmetric.

Thus, if we want to compute the value of the transform for a particular ω,

1) transform the window,
2) evaluate the conjugate of the basis function corresponding to ω for all x_k and sum the results,
3) multiply the obtained spectra.

We see that in order to obtain the transform of a Parzen window estimate, i.e. the model, we don't really need to find the estimate in the spatial domain.

By analogy, we regard the Parzen windowing on S^{n-1} as a convolution of a desired window with a function obtained by putting the spherical analogy of a Dirac delta function at the place of each sample.

To derive a spherical analogy of the $\Delta(x)$ function, we need to properly define a Dirac delta function on a sphere. Furthermore, we need to determine its spectrum. A spherical delta function can be defined by its property of extracting a function value under integration. Let $\mathring{\delta}$ denote the delta function on the sphere. Then,

$$
\int_{S^{n-1}} f(\xi)\mathring{\delta}(\xi)\, d\xi = f(\eta), \tag{17}
$$

In (17), η represents the north pole and $d\xi$ is the essentially unique, rotation invariant area-element on the S^{n-1} given by (5). On the line or generally in Euclidean spaces we use translation to move the Dirac delta function around, while on the sphere it is natural to use rotations as described in Section II.

So, to extract a function value from somewhere else than the north pole, we would use the rotated spherical delta function,

$$
\int_{S^{n-1}} f(\xi)\left[R^n(g)\mathring{\delta}(\xi)\right] d\xi = f(g^{-1}\eta) \tag{18}
$$

Now we can define the $(n-1)$-spherical analogy of the function $\Delta(x)$ as

$$
\mathring{\Delta}(\xi) = \frac{1}{N}\sum_{k=1}^{N} R^n(g_k)\mathring{\delta}(\xi) \tag{19}
$$

where g_k is any rotation that takes the north pole to the position of the k-th sample.

The Fourier transform of the $\mathring{\Delta}$ function is

$$
\begin{aligned}
\widehat{\mathring{\Delta}}(l, \mathbf{K}) &= \int_{S^{n-1}} \mathring{\Delta}(\xi)\overline{\Xi_{\mathbf{K}}^{l}(\xi)}\, d\xi \\
&= \int_{S^{n-1}} \left(\frac{1}{N}\sum_{k=1}^{N} R^n(g_k)\mathring{\delta}(\xi)\right)\overline{\Xi_{\mathbf{K}}^{l}(\xi)}\, d\xi \\
&= \frac{1}{N}\sum_{k=1}^{N}\int_{S^{n-1}} \overline{\Xi_{\mathbf{K}}^{l}(\xi)}R^n(g_k)\mathring{\delta}(\xi)\, d\xi \\
&= \frac{1}{N}\sum_{k=1}^{N} \overline{\Xi_{\mathbf{K}}^{l}(\xi_k)}
\end{aligned} \tag{20}
$$

where $\xi_k = g_k\eta$.

In the spherical setting, we can again find the transform of the dataset by simply evaluating the conjugate of the basis function at the points where we have the samples. Therefore, the spherical harmonic transform of the convolution of Parzen window $w(\xi)$ with $\mathring{\Delta}(\xi)$ becomes

$$
\begin{aligned}
\widehat{\mathring{f}}(l, \mathbf{K}) &= \left(\frac{\Gamma(l+n-2)(2l+n-2)}{l!\,\Gamma(n-1)}\right)^{1/2} \times \\
&\quad \times \left(\frac{1}{N}\sum_{k=1}^{N}\overline{\Xi_{\mathbf{K}}(\xi_k)}\right)\widehat{w}(l, \mathbf{0}).
\end{aligned} \tag{21}
$$

Spatial estimate is obtained by applying the inverse spherical harmonic transform to $\widehat{\mathring{f}}(l, \mathbf{K})$.

Expression (21) is definitely a more efficient way of obtaining the estimate spectrum than computing the spatial estimate and then transforming it. For large datasets, this expression combined with the fast inverse transform (e.g. [3]) is a more efficient way of computing the spatial estimate as well.

C. Computing the pdf in the original domain

In order to compute the pdf value for some point $\mathbf{x} \in \mathbb{R}^n$, it is not sufficient to project $\mathbf{x}$ onto the S^n and find the value of reconstruction. If we do that, we do not account for the area-element stretch introduced by the stereographic transform: this results in an invalid probability measure. To fix it, we multiply the yet-to-be pdf by the stereographic transform Jacobian and get the following result,

$$
\tilde{f}(x^{(1)}, \cdots, x^{(n)}) = \tilde{f}(\theta^{(1)}, ..., \theta^{(n)})(1 - \cos\theta^{(n)})^n, \tag{22}
$$

so that

$$
\int_{\mathbb{R}^n} \tilde{f}(\mathbf{x})d\mathbf{x} = \int_{S^n} \tilde{f}(\xi)d\xi = 1, \tag{23}
$$

where $\mathbf{x} = (x^{(1)}, \cdots, x^{(n)})$, $d\mathbf{x} = dx^{(1)}\cdots dx^{(n)}$, $\xi = (\theta^{(1)}, \cdots, \theta^{(n)})$ and $d\xi$ as in (5).

D. Relation between transformation and density

We use the stereographic projection to focus attention on a certain subset of the data. There are other ways of achieving the same effect. For example, the variable window width kernel density estimation where the window width varies with the local data density. Stereographic projection is appropriate for a certain class of densities, while for some densities it doesn't work well. To get the general idea of the density

for which the proposed transform is beneficial, we can try to use the result of Abramson, [34]. He demonstrated that if the window width is proportional to $f^{-1/2}$, the bias of the estimate will be of a smaller order than with the global window width estimator. Since we use a fixed kernel width on the sphere, we can approximate the width of a back-projected kernel by the n-th root of the Jacobian (22) where n is the dimension. It is straightforward to find the density that varies with the back-projected window width optimally in the Abramson sense. We get

$$f(\mathbf{x}) \propto \left(\frac{1}{1 + \|\mathbf{x}\|^2} \right)^2. \tag{24}$$

On the sphere this corresponds to the density

$$f(\xi) \propto (1 - \cos\theta_n)^{2-n}. \tag{25}$$

For the S^2, (24) gives the result that is intuitively expected. That is, if we use the same window width for the whole sphere, then the underlying density should be uniform. Observe that this reasoning is not valid in the topologically different Euclidean case. However, as the dimensionality grows, the tails of the density (24) get fatter, and in fact the north pole becomes a singular point of (25). This is not intuitive since it means that the majority of the samples would be outside of a desired region, i.e. on the wrong half of the sphere with regard to the stereographic projection. If we make the natural requirement that the density be uniform on the sphere, the back-projected density becomes

$$f(\mathbf{x}) \propto \left(\frac{1}{1 + \|\mathbf{x}\|^2} \right)^n. \tag{26}$$

Sphere is a compact manifold without boundary. An interesting consequence of this fact is that the convolution between uniform density on a sphere and arbitrary kernel function that integrates to unity is again a uniform density. Thus, the expected bias for the estimate of a uniform density on a sphere is exactly zero. The result (26) makes more sense since the number of the samples being mapped to either half of the sphere remains proportionally the same regardless of the dimension. For $n = 1$, (26) gives the univariate Cauchy distribution, although this is not true for the multivariate case.

E. Dealing with negative reconstructed pdf values

Problem inherent to orthogonal series estimates of a probability density function are negative reconstruction values. Given that the basis functions are oscillatory, summing only a finite number of basis functions due to spectral bandwidth limitation obviously gives a result that takes both positive and negative values. The same problem arises when using kernel estimates with kernels that assume negative values [35]. Such kernels may reduce the estimator bias, but again introduce negative estimate values. One possibility is to use the function

$$\widetilde{f}^*(x) = \frac{\widetilde{f}^+(x)}{\int \widetilde{f}^+(x)}. \tag{27}$$

where $\widetilde{f}^+(x) = \max(\widetilde{f}(x), 0)$. The approximation given by (27) is optimal in the L^1 sense i.e. for any probability density g,

$$\|\widetilde{f}^* - g\|_1 \leq \|\widetilde{f} - g\|_1, \tag{28}$$

where $\|f\|_1 = \int |f|$. This means that $\widetilde{f}^*$ is closer than $\widetilde{f}$ to the density that is being estimated, thus we can only improve our estimate by replacing $\widetilde{f}$ with $\widetilde{f}^*$ [28]. Density $\widetilde{f}^*$ is sometimes referred to as a *nonnegative projection*, although it is not a projection in a true sense, since L^1 is not a Hilbert space. There is also a possibility of using a proper L^2 projection, but we use the above schema due to its simplicity.

V. WINDOW WIDTH AND THE ESTIMATOR PERFORMANCE

A. Choice of the smoothing window width

In this paper we aim at the illustration of the applicability of the derived convolution formula and at the empirical assessment of the proposed density estimation model without fine-tuning the model. However, some discussion with regard to the choice of the smoothing window is needed.

The effect of the windowing should be regarded both in the spatial and in the frequency domain. We will use the term *bandwidth* by itself when referring to width in the spectral domain. When referring to the width of the kernel in the spatial domain we will use the term *window width*. If the window is very wide in the spatial domain, then its spectrum is narrow. Since the estimation corresponds to a convolution with the window, and spherical convolution is a pointwise multiplication in the spectral domain according to (11), it only makes sense to use relatively low model bandwidths. In contrast, if the window is very narrow in the spatial domain, we have to increase the model bandwidth or the estimate will suffer from excessive ringing. This is in complete analogy with the classical Fourier transform.

Large bandwidths (narrow smoothing windows) make sense only if the size of the dataset is proportionally large, otherwise the estimate will overfit the dataset. But if we use large model bandwidth we can indeed use spatially narrow windows with corresponding wide spectra.

Previous discussion is in accordance with intuition, since it makes sense to use a spatially narrower window if we have more samples in order not to oversmooth the pdf. The other way around, if we have fewer data we need to use a wider window so that we avoid having an estimate that resembles a collection of spikes.

Note that the choice of the window width is not limited to intuitive arguments only. Window width selection has a rich literature based on rigorous mathematical arguments, especially in the Euclidean case. For example [36] for the spatial domain perspective and [37] for the spectral domain (characteristic function approach). Comments on the window width choice for the S^n case can be found in [38], [39].

B. Notes on the estimator performance

The performance of the proposed estimator is largely governed by the performance of kernel density estimators on the n-dimensional sphere. Note that the effect of finite spectral bandwidth can largely be put aside if the bandwidth is adapted to the chosen window width. This means that we should

let the spectral magnitude smoothly approach zero, and not brick-wall it too early (see Fig. 3). Detailed discussion of spherical density estimators can be found in [38], [39]. Various results concerning the consistency of spherical estimators were established in [40]. We only point out the main results related to convergence. Expected squared-error under certain assumptions is given in [38] as

$$\mathcal{L}_2(\kappa) \sim N^{-1} b_1(\kappa) + \{1 - b_2(\kappa)\}^2 \rho(f), \qquad (29)$$

where b_1 and b_2 are some constants that depend on the chosen kernel, and ρ depends on the second derivative of the underlying density function. The first term on the right-hand side of (29) corresponds to the integrated variance and the second one corresponds to the integrated squared bias. Window width is controlled by the parameter κ. As κ tends to infinity, the kernel gets narrower. We could say that the kernel tends to a Dirac delta function. Furthermore, $b_2(\kappa) \to 1$ as $\kappa \to \infty$, in other words the bias term tends to zero. If the optimal choice for κ is made in the sense of minimizing (29), then the expected squared-error becomes

$$\inf_{\kappa > 0} \mathcal{L}_2(\kappa) \sim N^{-4/(n+4)} \{\rho(f)^{(n-2)/2}\}^{2/(n+2)} B_n, \qquad (30)$$

where B_n is a constant that depends on the dimensionality n. We see that the rate achieved in the spherical case equals the rate in the Euclidean case (see for example [41], [42]). That is, the rates for S^n and $\mathbb{R}^n$ are equal. For example, if we put $n = 1$, we get the familiar convergence rate of $N^{-4/5}$.

To understand the overall performance of the estimator it is important to note that the squared bias and the variance of the back-transformed estimate are proportional to the squared bias and the variance on the sphere. For instance, if $n = 2$,

$$\{E\{\widehat{f}(\mathbf{x})\} - f(\mathbf{x})\}^2 = (1 - \cos\theta_2)^4 \{E\{\widehat{f}(\xi)\} - f(\xi)\}^2, \quad (31)$$

and

$$\mathrm{var}\{\widehat{f}(\mathbf{x})\} = (1 - \cos\theta_2)^4 \, \mathrm{var}\{\widehat{f}(\xi)\} \qquad (32)$$

where $\theta_2 = \theta_2(\mathbf{x})$ and $\xi = (\theta_1, \theta_2)$ is the inverse stereographic projection of $\mathbf{x}$. Transformation of the dataset introduces a multiplicative factor into the asymptotic MISE (see [21]), but the rate of $N^{-4/(n+4)}$ remains. If the transform is matched to the underlying pdf, the introduced factor reduces the MISE accordingly. This implies that if the pointwise bias on the sphere is zero, the back-transformed bias will be zero as well. For the proposed estimator this occurs when the density is given by (26).

Since the expression for the optimal smoothing parameter requires the knowledge of the density that is being estimated, different empirical methods for the choice of this parameter have been proposed. Asymptotic optimality of these methods has also been proved for the spherical case [38], [39]. Application of such data-driven kernel width selectors should be straightforward for the proposed estimator.

C. Window samples

For computation of the bandlimited spherical harmonic transform in the 2-sphere case we used the discretization scheme proposed in [3], [4] and the corresponding fast algorithm implemented in *S2Kit* [43]. The sampling scheme used in *S2Kit* is equiangular with the distance between sampling angles inversely proportional to the bandwidth. In light of the previous discussion, we decided to use a fixed set of window samples for all bandwidths. For higher bandwidths, this resulted in a smoothing window of the same shape, but narrower in the spatial domain since the samples lie closer to the north pole. In fact, spatial and spectral width are reciprocal. The assumption that this would generate the same window spectra, only scaled in frequency was experimentally confirmed as is shown in Fig. 3. The assertion that the straightforward n-dimensional generalization is possible was also experimentally validated.

VI. SIMULATION/EXPERIMENTS

A. 2-D case

In the two-dimensional case, random samples are mapped onto the surface of a regular 2-sphere. Fourier analysis on S^2 corresponds to expansions into familiar spherical harmonics that are special cases of hyperspherical harmonics. The exact formulation of these functions is provided in Appendix A.

The smoothing kernel density is given by

$$w(\mathbf{x}) \propto \exp(\kappa \mu^T \mathbf{x}), \qquad (33)$$

where μ is the mean direction unit vector and $\mathbf{x}^T \mathbf{x} = 1$. $\kappa > 0$ is a smoothing parameter that controls the window width. It corresponds to κ in (29) and (30). This is actually a von Mises-Fisher distribution on S^2. If μ is the north pole, (33) reduces to the following form in spherical coordinates,

$$w(\theta, \varphi) \propto \exp(\kappa \cos\theta). \qquad (34)$$

As described in Section V, we actually sampled the smoothing window and used a fixed set of samples. We experimented with three different windows

$$w_{s1}[j] \propto \exp\left(10 \cos\left(\pi(2j+1)/20\right)\right), \quad j = 0, \ldots, 2$$
$$w_{s2}[j] \propto \exp\left(10 \cos\left(\pi(2j+1)/28\right)\right), \quad j = 0, \ldots, 4$$
$$w_{s3}[j] \propto \exp\left(10 \cos\left(\pi(2j+1)/40\right)\right), \quad j = 0, \ldots, 6. \quad (35)$$

For indices j larger than the indicated maximum, we assume that the window value is zero.

Model estimate in the spectral domain is now given by

$$\hat{\tilde{f}}(l, m) = \sqrt{\frac{1}{2l+1}} \left(\frac{1}{N} \sum_{k=1}^{N} Y_l^m(\theta_k, \varphi_k)\right) \hat{w}(l, 0) \qquad (36)$$

We verified the model on two random processes, one artificial, and one obtained from the Line Spectrum Frequencies (LSF) data [44], [45]. Furthermore, we compared it with GMM pdf estimates of the same process realizations, using GMM with the same total number of scalar parameters.
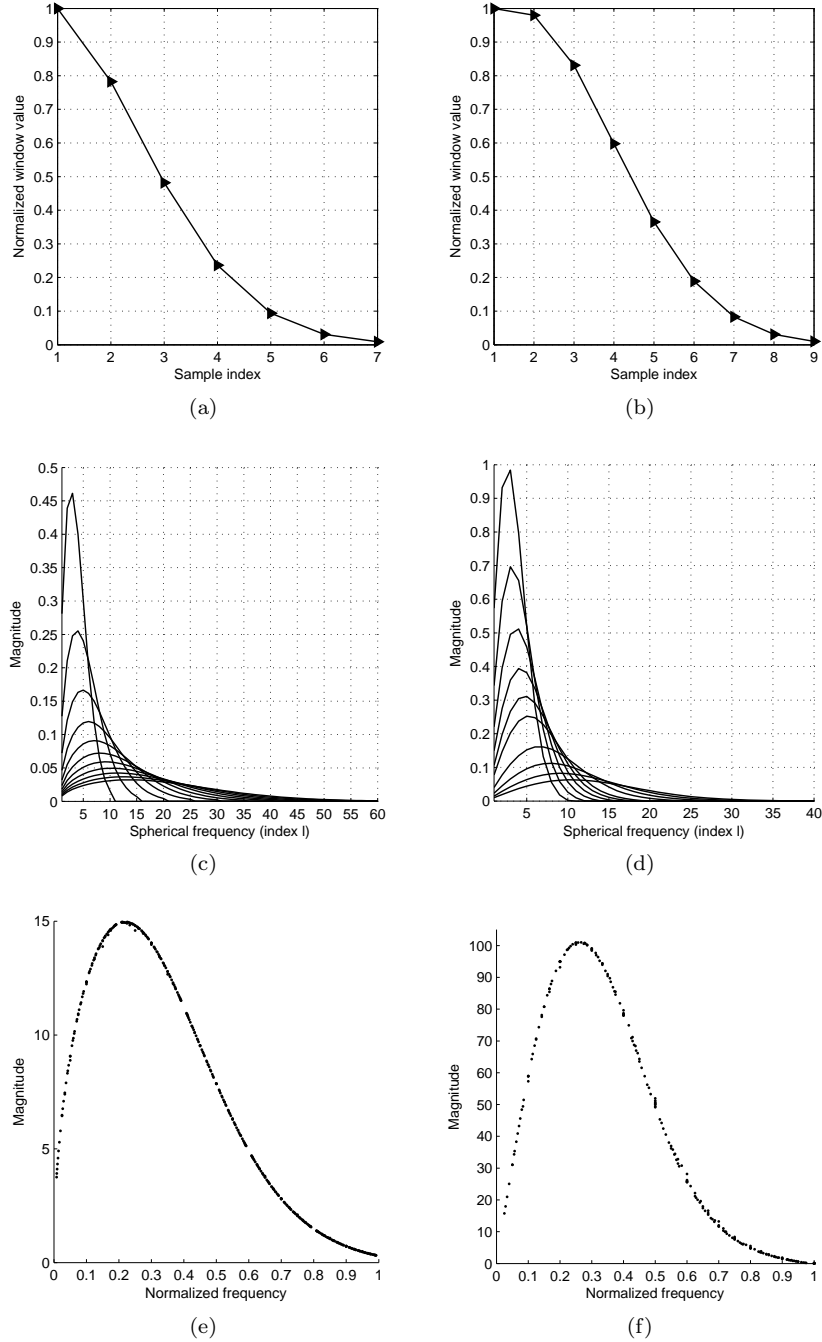
Fig. 3.   Spatial and spectral plots of the used smoothing windows: a) window samples used in 2-D case, b) windows samples used in 3-D case, c) and d) window spectra shown for different bandwidths for 2-D and 3-D cases, e) and f) suitably normalized spectrum magnitude - it can be observed that all spectral magnitude responses are of the same shape, only differently normalized. Spectral magnitudes are given as functions of the *master* index l, with $\mathbf{K}$ set to $\mathbf{0}$.

B. Ring-shaped process

In this experiment, the proposed model was used to estimate the density of a synthetic ring-shaped process depicted in Fig. 4. The process pdf can be analytically written down as

$$f_{\text{RING}}(r) = \frac{1}{2\pi} \left(\frac{\chi_{[\tau_2, \sigma_2]}(r)}{\sigma_2^2 - \tau_2^2} + \frac{\chi_{[\tau_1, \sigma_1]}(r)}{\sigma_1^2 - \tau_1^2} \right), \quad (37)$$

where $r = \|\mathbf{x}\|$, $\chi_A(r)$ is an indicator function of a set A, $\tau_{1,2}$ are the inner radii and $\sigma_{1,2}$ outer the radii of the rings. The

scaling factor (13) is computed as

$$s_j = \frac{1}{2 \sqrt{\text{var}\{x_k^{(j)}\}_{k=1}^N}}.$$

Training dataset comprised 20000 random samples. Our model successfully describes the original density from samples. Notice that spherical estimation is not an iterative process, and it always gives a usable model, while GMM makes sense only for a moderate number of components and the model has problems with overfitting if the number of components
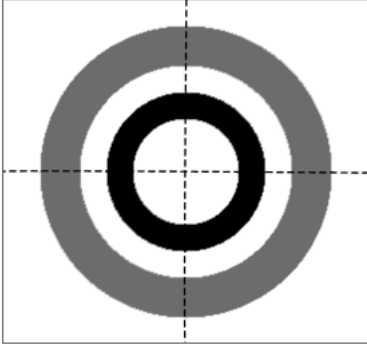
Fig. 4. Synthetic pdf used in the first 2-D experiment. Probability of generating a sample in either ring is 0.5.
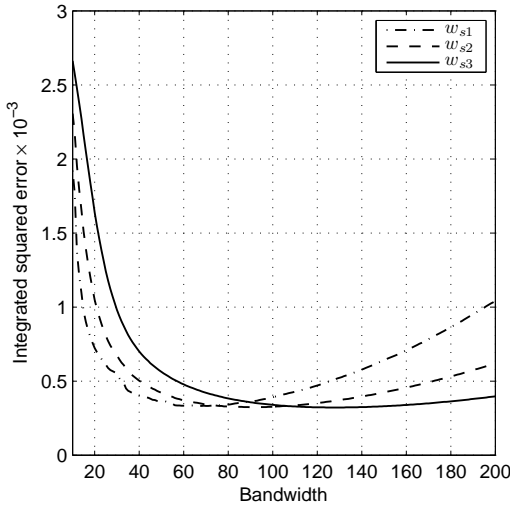

Fig. 5. Integrated squared L^2 distance between estimates and original density of a ring process, versus spherical harmonic bandwidth. Distance is shown for different kernels as specified in (35). Note that the spectral bandwidth axis is actually inversely proportional to the spatial kernel width. For each of the kernels, optimal choice of the smoothing parameter and the spectral bandwidth corresponds to the minimum of the error curve. Error for smaller bandwidths increases due to the oversmoothing while for higher bandwidths the kernels get too narrow.

is large relative to the sample size. L^2 error as a function of bandwidth is shown in Fig. 5.

C. Line spectrum frequencies (LSF)

In the second experiment we modeled the probability density of the first two LSF coefficients. Our intention was to verify visually whether it is possible to model more complicated densities from samples. Fig. 7 shows the scatter graph of the original process, density obtained with our model and density obtained with the GMM with same number of scalar parameters. GMM is shown for comparison with a different space-filling strategy. As in the previous experiment, we used a training set of 20000 samples.

D. 3-D case

Since little concrete and practical work is available on the usage of hyperspherical harmonics in signal processing, we demonstrate here the application of our generalized model to three dimensions. In three-dimensional case samples are mapped onto the surface of a hypersphere (3-sphere), therefore we use the hyperspherical harmonics to estimate the pdf. Hyperspherical harmonics can be neatly written down with the help of Gegenbauer polynomials (see Appendix A).

We chose the $(\theta_1, \theta_2, \theta_3)$ coordinates on S^3 according to (2). In these coordinates, 3-sphere harmonics are given by (39) by putting $n = 4$.

Same as in the 2-D case, after mapping the samples onto a hypersphere, we smooth them out with a window. Here, we experimented with the window of the form

$$w \propto \exp(-\theta_3^2 \kappa) \qquad (38)$$

where $\kappa > 0$ is a smoothing parameter.

Window samples were determined as follows,

$$w_s[j] \propto \exp(-\left(\pi(2j+1)\right)^2 / 480), \qquad j = 0, \ldots, 8.$$

We simulated the model for two datasets. The first one was drawn from an artificial, spiral shaped distribution, while the second one was obtained from speech Line Spectrum Frequencies data and corresponds to the first three LSF coefficients. In both experiments the sample size was 50000.

E. Spiral-shaped process

In this experiment samples are generated according to the pdf illustrated in Fig. 8. First a random point P is chosen uniformly along the length of the spiral, then a sample is generated in the plane that contains the z-axis and P according to a bivariate normal distribution with the mean P.

In Fig. 9 it can be observed from pdf isosurfaces that the proposed model successfully estimates the pdf from samples of the spiral process. Again, we include the estimate obtained by GMM for comparison. Convergence of the estimate towards the original pdf is shown in Fig. 10. We experimented with different prescaling factors (13) when mapping the process onto the 3-sphere. All three factors yield virtually identical L^2 distances for given bandwidths. In particular, prescales (13) were computed as

$$s_j = \frac{k^{-1}}{2\sqrt{\text{var}\{x_k^{(j)}\}_{k=1}^N}},$$

where k is a parameter that controls the magnitude of scaling.

F. LSF premodeled by GMM

In order to numerically assess the performance of the proposed model for a real-world dataset, we used the realizations of the lower three LSFs. Since the actual pdf of these coefficients is unknown, we first modeled the dataset with GMM comprising 25 Gaussian components. We used our model to estimate the pdf of a new process generated by the GMM, and computed the L^2 distance from the original GMM density for different bandwidths. Isosurfaces of the GMM density and densities obtained from the proposed model are shown in Fig. 11. One may observe that it is possible to describe a relatively complicated density shape using our model. Bear in mind that
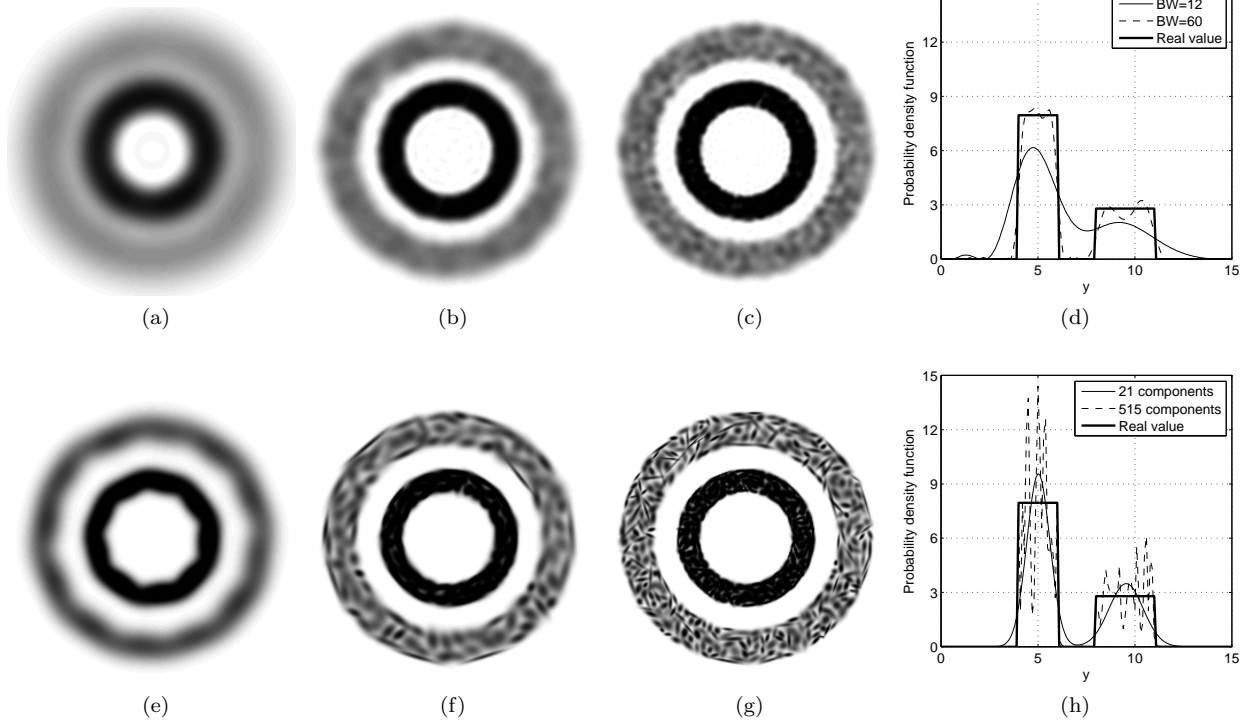
Fig. 6. Density estimates of a synthetic ring process, compared with GMM estimates with the same number of scalar coefficients. (a) BW=12 (144 coefficients), (b) BW=34 (1156 coefficients), (c) BW=60 (3600 coefficients). (e) 21 mixture component, (f) 166 mixture components, (g) 515 mixture components. (d) and (h) show the crossections along the y-axis (vertical).
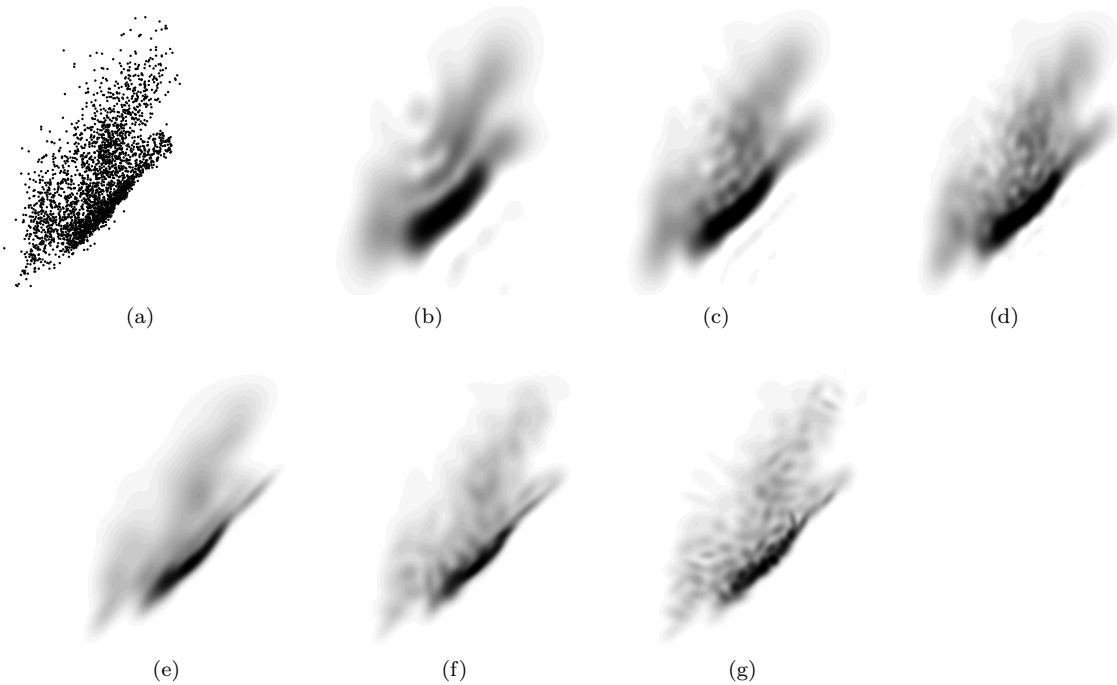

Fig. 7. Estimates of distribution density of lowest two LSF coefficients, shown for different bandwidths and compared with GMM estimates for the same number of scalar coefficients. (a) scatter graph for 3000 samples (b) BW=12 (144 coefficients), (c) BW=34 (1156 coefficients), (d) BW=60 (3600 coefficients). (e) 21 mixture component, (f) 166 mixture components, (g) 515 mixture components.

while Fig. 11a shows the isosurface of the analytical GMM density, densities in the remaining figures are estimated from samples. L^2 distance from the GMM is shown in Fig. 12. Note that differently from Fig. 5 only the decreasing part of error curves in shown in Fig. 10 and Fig. 12 (this would correspond to the bias term due to oversmoothing). Although the optimum is achieved for higher bandwidths, this part of the curve is interesting in the context of compact data modeling.
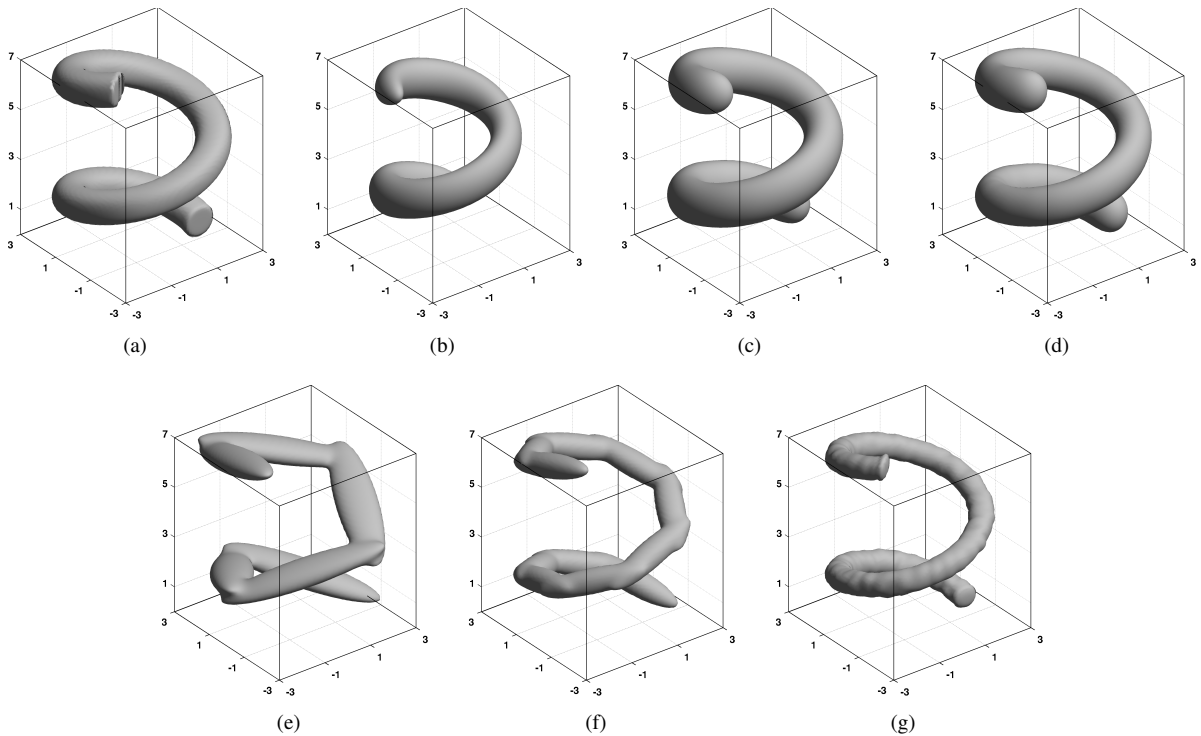
Fig. 9.   Isosurfaces of spiral density estimates. Isovalue of pdf used for figure generation is 0.008. (a) original density, (b) BW=14 (1015) coefficients), (c) BW=25 (5525 coefficients), (d) BW=40 (22140 coefficients). GMM estimates are shown in (e) for 6 mixture components, (f) for 11 mixture components and (g) for 100 mixture components. Here, GMM is shown as an example of a different space-filling strategy, and corresponding number of scalar components doesn't match.
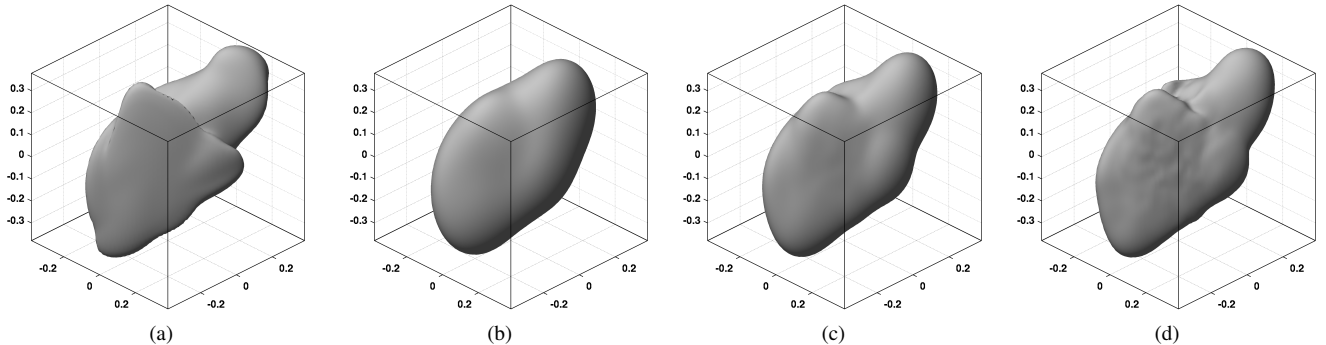

Fig. 11.   Isosurfaces of density estimate for a process generated by a GMM of LSF coefficients. Isovalue of pdf used for figure generation is 2. (a) *original* density comprising 25 Gaussian mixture components, (b) BW=14 (1015 coefficients), (c) BW=25 (5525 coefficients), (d) BW=40 (22140 coefficients).

VII. CONCLUSION AND FUTURE WORK

We have derived the formula for the Fourier transform of a convolution of functions on an n-sphere, S^n. Furthermore, we proposed a probability density modeling technique that uses the obtained result for the convolution. The samples are mapped onto the sphere, where they are convolved with the smoothing kernel density.

The model was applied to several synthetic and real-world processes in two and three dimensions, i.e. on the sphere and on the hypersphere. The obtained results suggest that the model successfully describes processes of different structures. Three dimensional experiments demonstrate how the familiar spherical harmonic techniques may be extended to higher dimensions (hyperspherical harmonics). The model is readily extendable to dimensions higher than 3 using the provided n-dimensional framework.

Future work includes fine tuning of the prescaling step and the mapping parameters, as well as determining the optimal window shape for specific datasets. Also, the information content of the quantized expansion coefficients and the effect of the quantization on the reconstruction will be investigated. Another functionality that should be easy to implement is adding the information from new samples into the existing estimate. Since the model is of a one-shot kind, the timings required to obtain the estimate may be accurately predicted. This property may be useful in source coding applications.

The biggest current drawback is the lack of readily available implementations of the involved higher dimensional fast transforms, even though the theory exists.
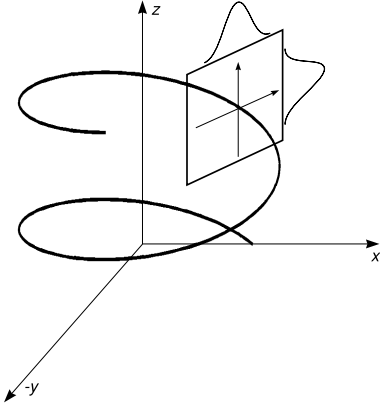
Fig. 8. Illustration of the density from which data were drawn in the first 3-D experiment. It can be viewed as a bivariate normal distribution superposed onto the spiral.

APPENDIX A
SPHERICAL AND HYPERSPHERICAL HARMONICS

In literature there are several definitions of spherical and hyperspherical harmonics that differ in normalization, phase convention, choice of coordinates, etc. We generally follow the conventions of Vilenkin [2]. Let l and $\mathbf{K} = (k_1, \ldots, k_{n-2})$ be defined as in Section II. In the coordinates (2), hyperspherical harmonics are given by

$$\Xi_{\mathbf{K}}^l(\xi) = A_{\mathbf{K}}^l \times$$
$$\times \prod_{i=0}^{n-3} C_{k_i-k_{i+1}}^{\frac{n-i-2}{2}+k_{i+1}}(\cos \theta_{n-i-1}) \sin^{k_{i+1}} \theta_{n-i-1} e^{\pm j k_{n-2}\theta_1}.$$
$$(39)$$

The normalization factor is

$$(A_{\mathbf{K}}^l)^2 = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \prod_{i=0}^{n-3} 2^{2k_{i+1}+n-i-4} \times$$
$$\times \frac{(k_j - k_{i+1})!(n-i+2k_i-2)\Gamma^2\left(\frac{n-i-2}{2}+k_{i+1}\right)}{\sqrt{\pi}\ \Gamma(k_i+k_{i+1}+n-i-2)}.$$
$$(40)$$

In [33] Avery uses a convention where the area of a unit $(n-1)$-sphere is $\frac{2\pi^{n/2}}{\Gamma(n/2)}$. In contrast, we follow the convention where that area is equal to unity, hence the normalization in (5). In (39), $C_n^\lambda(t)$ are the Gegenbauer polynomials of degree n. They can be defined as the coefficients of α^n in the power-series expansion of the following function [46],

$$(1 - 2t\alpha + \alpha^2)^{-\lambda} = \sum_{n=0}^{\infty} C_n^\lambda(t)\alpha^n. \qquad (41)$$

A reliable method for computation of Gegenbauer polynomials is the three-term recurrence relation,

$$(n+2)C_{n+2}^\lambda(t) = 2(\lambda+n+1)tC_{n+1}^\lambda(t)-(2\lambda+n)C_n^\lambda(t), \quad (42)$$

with $C_0^\lambda(t) = 1$ and $C_1^\lambda(t) = 2\lambda t$.

A special case of (39) are the familiar spherical harmonics on S^2. In fact, Gegenbauer polynomials are the generalization of Legendre polynomials. Spherical harmonics on S^2 are given by

$$Y_l^m(\theta, \varphi) = N_{lm} P_l^m(\cos \theta) e^{jm\varphi} \qquad (43)$$

with the normalization factor

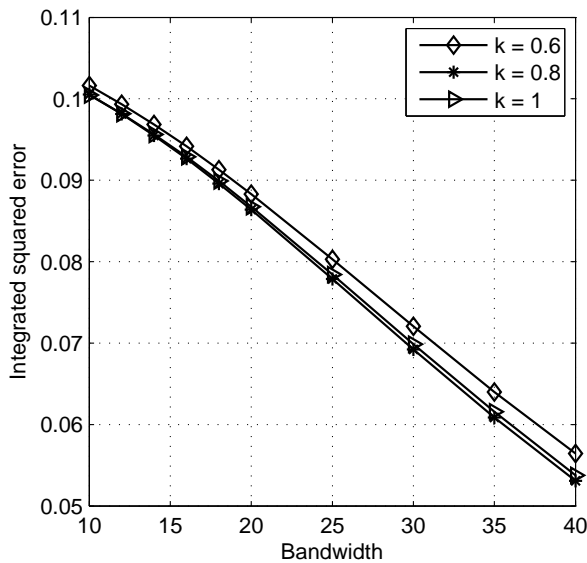$$N_l^m = \sqrt{(2l+1)\frac{(l-m)!}{(l+m)!}} \qquad (44)$$

Fig. 10. Integrated squared L^2 distance between estimates and original density of a spiral process, versus spherical harmonic bandwidth. Error is shown for three prescaling factors.
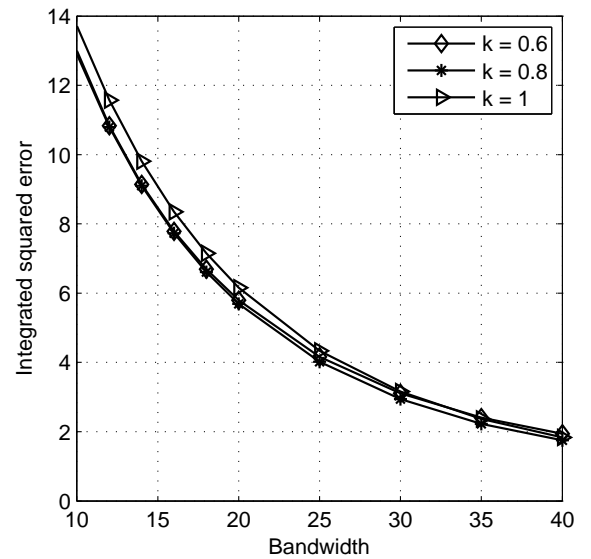

Fig. 12. Integrated squared L^2 distance between estimates and original density of the lower three LSF coefficients premodeled by GMM, versus spherical harmonic bandwidth. GMM premodeling is used to obtain an analytical model of a real-world process. Error is shown for three prescaling factors.

chosen so that the norm of each harmonic would be unity with respect to the measure (5). P_l^m are the associated Legendre polynomials that can also be computed via the three-term recurrence relation

$$(l-m+1)P_{l+1}^m(x) = (2l+1)xP_l^m(x) - (l+m)P_{l-1}^m(x), \quad (45)$$

with m and l integers such that $|m| \le l$. The initial values are given by[1] $P_l^l = (-1)^l(2l-1)!!(1-x^2)^{l/2}$ and $P_{l+1}^l(x) = x(2l+1)P_l^l(x)$.

APPENDIX B
DERIVATION OF THE CONVOLUTION FORMULA

Let $f_1, f_2 \in L^2(S^2)$ and $f = f_1 * f_2$. By definition,

$$\widehat{f}(l, \mathbf{K}) = \int_{S^{n-1}} f(\xi) \, \overline{\Xi_K^l(\xi)} \, d\xi$$

$$= \int_{S^{n-1}} \left(\int_{SO(n)} f_1(g\eta) f_2(g^{-1}\xi) \, dg \right) \overline{\Xi_K^l(\xi)} \, d\xi. \quad (46)$$

After rearranging the integrals we get

$$\widehat{f}(l, \mathbf{K}) = \int_{SO(n)} \left(\int_{S^{n-1}} f_2(g^{-1}\xi)\overline{\Xi_K^l(\xi)} \, d\xi \right) f_1(g\eta) \, dg \quad (47)$$

First, let's evaluate the inner integral. We introduce the matrix elements of the rotation operator $t_{\mathbf{MK}}^l(g)$ via their action on the canonical basis elements $\Xi_{\mathbf{K}}^l(\xi)$ (see (1) in [2], page 469),

$$\int_{S^{n-1}} f_2(g^{-1}\xi)\overline{\Xi_{\mathbf{K}}^l(\xi)} \, d\xi$$

$$= \int_{S^{n-1}} f_2(\xi)\overline{\Xi_{\mathbf{K}}^l(g\xi)} \, d\xi$$

$$= \int_{S^{n-1}} f_2(\xi)\overline{\sum_{\mathbf{M}} t_{\mathbf{MK}}^l(g^{-1})\Xi_{\mathbf{M}}^l(\xi)} \, d\xi$$

$$= \sum_{\mathbf{M}} \overline{t_{\mathbf{MK}}^l(g^{-1})} \int_{S^{n-1}} f_2(\xi)\overline{\Xi_{\mathbf{M}}^l(\xi)} \, d\xi$$

$$= \sum_{\mathbf{M}} \overline{t_{\mathbf{MK}}^l(g^{-1})}\widehat{f}_2(l, \mathbf{M}). \quad (48)$$

Now we substitute (48) into (47),

$$\widehat{f}(l, \mathbf{K}) = \int_{SO(n)} \sum_{\mathbf{M}} \overline{t_{\mathbf{MK}}^l(g^{-1})}\widehat{f}_2(l, \mathbf{M})f_1(g\eta) \, dg$$

$$= \sum_{\mathbf{M}} \widehat{f}_2(l, \mathbf{M}) \int_{SO(n)} f_1(g\eta)\overline{t_{\mathbf{MK}}^l(g^{-1})} \, dg. \quad (49)$$

Since

$$T^l(g)f_1(\eta) = \sum_{\mathbf{K}} \widehat{f}_1(l, \mathbf{K})t_{\mathbf{MK}}^l(g)\Xi_{\mathbf{K}}^l(\eta)$$

$$= \sum_{\mathbf{M}} \sum_{\mathbf{K}} t_{\mathbf{MK}}^l(g)\widehat{f}_1(l, \mathbf{K})\Xi_{\mathbf{M}}(\eta). \quad (50)$$

and $t_{\mathbf{MK}}^l(g) = \overline{t_{\mathbf{KM}}^l(g^{-1})}$ (unitarity of the representation),

$$\widehat{f}(l, \mathbf{K}) = \sum_{\mathbf{M}} \widehat{f}_2(l, \mathbf{M}) \times$$

$$\times \int_{SO(n)} \sum_{\mathbf{R}} \sum_{\mathbf{S}} \overline{t_{\mathbf{SR}}^l(g)}\widehat{f}_1(l, \mathbf{S})\Xi_R^l(\eta)t_{\mathbf{KM}}^l(g) \, dg$$

$$= \sum_{\mathbf{M}} \sum_{\mathbf{R}} \sum_{\mathbf{S}} \widehat{f}_2(l, \mathbf{M})\widehat{f}_1(l, \mathbf{S})\Xi_{\mathbf{R}}^l(\eta) \times$$

$$\times \int_{SO(n)} \overline{t_{\mathbf{SR}}^l(g)}t_{\mathbf{KM}}^l(g) \, dg$$

$$= \sum_{\mathbf{M}} \widehat{f}_2(l, \mathbf{M})\widehat{f}_1(l, \mathbf{K})\Xi_{\mathbf{M}}^l(\eta) \quad (51)$$

but $\Xi_{\mathbf{M}}^l(\eta) = 0$ for $\mathbf{M} \ne (0, ..., 0)$, and

$$\Xi_{\mathbf{O}}^l(\eta) = A_{\mathbf{O}}^l C_l^{\frac{n-2}{2}}(1), \quad (52)$$

where $\mathbf{O} = (0, ..., 0)$ and $A_{\mathbf{O}}^l$ is defined in (40), so

$$\widehat{f}(l, \mathbf{K}) = A_{\mathbf{O}}^l C_l^{\frac{n-2}{2}}(1)\widehat{f}_1(l, \mathbf{K})\widehat{f}_2(l, \mathbf{0})$$

$$= \sqrt{\frac{\Gamma(l+n-2)(2l+n-2)}{l!\Gamma(n-1)}} \widehat{f}_1(l, \mathbf{K})\widehat{f}_2(l, \mathbf{0}) \quad (53)$$

REFERENCES

[1] P. Szekeres, *A Course in Modern Mathematical Physics : Groups, Hilbert Space and Differential Geometry*. Cambridge University Press, December 2004.

[2] N. Y. Vilenkin, *Special functions and the theory of group representations*, 1st ed. American Mathematical Society, 1968.

[3] D. Healy, D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-sphere-improvements and variations," *Journal of Fourier Analysis and Applications*, vol. 9, no. 4, pp. 341–385, Jul. 2003. [Online]. Available: http://dx.doi.org/10.1007/s00041-003-0018-9

[4] J. R. Driscoll and J. Dennis M. Healy, "Computing Fourier transforms and convolutions on the 2-sphere," *Adv. Appl. Math.*, vol. 15, no. 2, pp. 202–250, 1994.

[5] P. Kostelec and D. Rockmore, "FFTs on the rotation group," *Journal of Fourier Analysis and Applications*, vol. 14, no. 2, pp. 145–179, Apr. 2008. [Online]. Available: http://dx.doi.org/10.1007/s00041-008-9013-5

[6] D. K. Maslen, Daniel, and D. N. Rockmore, "Generalized ffts - a survey of some recent results," 1995.

[7] B. Yeo, W. Ou, and P. Golland, "On the construction of invertible filter banks on the 2-sphere," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 283–300, March 2008.

[8] J.-P. Antoine, "Coherent states, wavelets and beyond," pp. 3–46, 2000.

[9] J. Antoine, L. Demanet, L. Jacques, and P. Vandergheynst, "Wavelets on the sphere : Implementation and approximations," *Applied and Computational Harmonic Analysis*, vol. 13, no. 3, pp. 177–200, 2002.

[10] J. p. Antoine and C. B. Torresani, "Wavelets on the 2-sphere: A group-theoretical approach," *Appl. Comput. Harmon. Anal*, vol. 7, pp. 1–30, 1999.

[11] J. Antoine and P. Vandergheynst, "Wavelets on the 2-sphere and related manifolds," *Reports on Mathematical Physics*, vol. 43, no. 1-2, pp. 13–24, 1999.

[12] I. Bogdanova, P. Vandergheynst, J. Antoine, L. Jacques, and M. Morvidone, "Stereographic Wavelet Frames on the Sphere," *Applied and Computational Harmonic Analysis*, vol. 19, no. 2, pp. 223–252, 2005.

[13] G. S. Chirikjian and A. B. Kyatkin, *Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups*, 1st ed. CRC Press, September 2000.

---

[1]The double exclamation mark symbol denotes the double factorial, $n!! = n(n-2)(n-4)\cdots, (-1)!! = 0!! = 1$.

[14] P. E. Jupp and K. V. Mardia, "A unified view of the theory of directional statistics, 1975-1988," *International Statistical Review*, vol. 57, no. 3, pp. 261–294, Dec. 1989.

[15] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962. [Online]. Available: http://www.jstor.org/stable/2237880

[16] X. Zhuang, Y. Huang, K. Palaniappan, and Y. Zhao, "Gaussian mixture density modeling, decomposition, and applications," *Image Processing, IEEE Transactions on*, vol. 5, no. 9, pp. 1293–1302, Sep 1996.

[17] R. Gray, "Gauss mixture vector quantization," *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 3, pp. 1769–1772 vol.3, 2001.

[18] C. Archambeau, M. Valle, A. Assenza, and M. Verleysen, "Assessment of probability density estimation methods: Parzen window and finite gaussian mixtures," *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pp. 4 pp.–, 21-24 May 2006.

[19] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 4, pp. 385–401, 2000.

[20] B. Pelletier, "Kernel density estimation on Riemannian manifolds," *Statistics & Probability Letters*, vol. 73, no. 3, pp. 297 – 304, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/B6V1D-4G1R0JK-1/2/e224139e5b322c9023139924310f1212

[21] M. P. Wand, J. S. Marron, and D. Ruppert, "Transformations in density-estimation," *Journal Of The American Statistical Association*, vol. 86, no. 414, pp. 343–353, Jun. 1991.

[22] A. Saavedra and R. Cao, "Rate of convergence of a convolution-type estimator of the marginal density of a ma(1) process," *Stochastic Processes And Their Applications*, vol. 80, no. 2, pp. 129–155, Apr. 1999.

[23] ——, "On the estimation of the marginal density of a moving average process," *Canadian Journal Of Statistics-Revue Canadienne De Statistique*, vol. 28, no. 4, pp. 799–815, Dec. 2000.

[24] A. Schick and W. Wefelmeyer, "Uniformly root-n consistent density estimators for weakly dependent invertible linear processes," *Annals Of Statistics*, vol. 35, no. 2, pp. 815–843, Apr. 2007.

[25] ——, "Root n consistent density estimators for sums of independent random variables," *Journal Of Nonparametric Statistics*, vol. 16, no. 6, pp. 925–935, Dec. 2004.

[26] B. Støve, "Some new approaches to smoothing: convolution estimators in regression models and backfitting in panels of time series," Ph.D. dissertation, The University of Bergen, Nov. 2005.

[27] J. C. Escanciano and D. T. Jacho-Chavez, "Root-n Uniformly Consistent Density Estimation in Nonparametric Regression Models," *SSRN eLibrary*, 2008.

[28] L. Devroye and L. Györfi, *Nonparametric Density Estimation: The L1 View*.   John Wiley & Sons, November 1984.

[29] H. Hendriks, "Nonparametric estimation of a probability density on a riemannian manifold using fourier expansions," *Annals Of Statistics*, vol. 18, no. 2, pp. 832–849, Jun. 1990.

[30] ——, "Application of fast spherical Fourier transform to density estimation," *Journal of Multivariate Analysis*, vol. 84, no. 2, pp. 209–221, Feb. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/B6WK9-4806GJ3-4/1/624c8d7780b02a23a804b263734c26ae

[31] D. M. Healy, H. Hendriks, and P. T. Kim, "Spherical deconvolution," *Journal Of Multivariate Analysis*, vol. 67, no. 1, pp. 1–22, Oct. 1998.

[32] P. T. Kim, "Deconvolution density estimation on SO(N)," *Annals Of Statistics*, vol. 26, no. 3, pp. 1083–1102, Jun. 1998.

[33] J. Avery, *Hyperspherical Harmonics: Applications in Quantum Theory*, 1st ed., ser. Reidel Texts in the Mathematical Sciences.   Springer, 1989, vol. 5.

[34] I. S. Abramson, "On bandwidth variation in kernel estimates - a square root law," *Annals Of Statistics*, vol. 10, no. 4, pp. 1217–1223, 1982.

[35] N. Oudjane and C. Musso, "L^2-density estimation with negative kernels," *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pp. 34–39, Sept. 2005.

[36] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal Of The American Statistical Association*, vol. 91, no. 433, pp. 401–407, Mar. 1996.

[37] S. T. Chiu, "Bandwidth selection for kernel density-estimation," *Annals Of Statistics*, vol. 19, no. 4, pp. 1883–1905, Dec. 1991.

[38] P. Hall, G. S. Watson, and J. Cabrera, "Kernel density-estimation with spherical data," *Biometrika*, vol. 74, no. 4, pp. 751–762, Dec. 1987.

[39] J. Klemelä, "Estimation of densities and derivatives of densities with directional data," *Journal Of Multivariate Analysis*, vol. 73, no. 1, pp. 18–40, Apr. 2000.

[40] Z. D. Bai, C. R. Rao, and L. C. Zhao, "Kernel estimators of density-function of directional-data," *Journal Of Multivariate Analysis*, vol. 27, no. 1, pp. 24–39, Oct. 1988.

[41] M. P. Wand and M. C. Jones, *Kernel Smoothing (Monographs on Statistics and Applied Probability)*.   Chapman & Hall/CRC, December 1994.

[42] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*.   Chapman & Hall/CRC, April 1986.

[43] P. J. Kostelec and D. N. Rockmore, "S2kit: A lite version of spharmonickit," 2004. [Online]. Available: www.cs.dartmouth.edu/~geelong/sphere/

[44] F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signals," *The Journal of the Acoustical Society of America*, vol. 57, pp. S35–S35, Apr. 1975. [Online]. Available: http://link.aip.org/link/?JAS/57/S35/2

[45] N. Sugamura and F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signal and its statistical properties," *Trans. Inst. Electron. Commun. Eng. Japan*, vol. J64-A, pp. 323–340, 1981.

[46] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds.   Academic Press, Inc., 2007.

Ivan Dokmanić (S'03) received the Dipl. ing. degree in electrical engineering from the University of Zagreb in 2007. He is currently working toward his Ph.D. degree at the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb.

His current research interests include efficient signal representations, digital audio processing and signal processing for bioinformatics.

Davor Petrinović (M'96) was born in Croatia in 1965. He received the Dipl. ing. degree in electrical engineering from the Faculty of Electrical Engineering (currently, the Faculty of Electrical Engineering and Computing), University of Zagreb, Croatia, in 1988 and the M.Sc. and Dr.Sc. degree in electrical engineering from the same institution in 1996 and 1999, respectively. In 2005, he was appointed an Associate Professor at the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb. In 2000-2001, he was a Fulbright Post-doctoral Scholar at the SCL Laboratory, University of California, Santa Barbara, and a Visiting Researcher at Sound and Image Processing Lab, School of Electrical Engineering, KTH, Stockholm, Sweden in 2005-06.

His current research interests include speech and audio modeling, processing and coding.