

Optimal Reduction of Dirac Mixture Densities on the 2-Sphere

Daniel Frisch, Kailai Li, and Uwe D. Hanebeck

*Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
daniel.frisch@ieee.org, kailai.li@kit.edu, uwe.hanebeck@ieee.org*

Abstract: This paper is concerned with optimal approximation of a given Dirac mixture density on the \mathbb{S}^2 manifold, i.e., a set of weighted samples located on the unit sphere, by an equally weighted Dirac mixture with a reduced number of components. The sample locations of the approximating density are calculated by minimizing a smooth global distance measure, a generalization of the well-known Cramér-von Mises Distance. First, the Localized Cumulative Distribution (LCD) together with the von Mises–Fisher kernel provides a continuous characterization of Dirac mixtures on the \mathbb{S}^2 manifold. Second, the L^2 norm of the difference of two LCDs is a unique and symmetric distance between the corresponding Dirac mixtures. Thereby we integrate over all possible kernel sizes instead of choosing one specific kernel size. The resulting approximation method facilitates various efficient nonlinear sample-based state estimation methods.

Keywords: Probability density function, function approximation, optimization problems, least-squares approximation, parameter estimation, Dirac mixture approximation, samples, manifolds, target tracking, system state estimation, Kalman filters, density approximation

1. INTRODUCTION

Context: Exact Bayesian state estimation for nonlinear systems is a nontrivial task, because analytic representations of the occurring densities can become arbitrarily complex over time, or closed-form expressions may not even exist. Therefore, *samples* instead of density functions are often used as stochastic state representation in applications of state estimation and control.

In the prediction step, samples can simply be inserted individually into any nonlinear generative state transition model. However, incorporating additive or non-additive noise multiplies the number of samples in every time step. To keep the algorithm tractable, the number of samples must be reduced for further processing, while maintaining as much information as possible. In the update step, one method is to re-weight state samples according to the measurement likelihood. Over time, repeated re-weighting will result in samples with vanishing weight, i.e., sample degeneration. Thus, individually weighted samples should be re-approximated by equally weighted samples for further processing. In this context, our method could provide a more faithful resampling than many existing bootstrap resamplings.

In general, wherever a cloud of many randomly drawn samples is employed, be it in assumed density filters or sequential Monte Carlo (SMC) state estimators, similar results can be achieved with far less *deterministically* chosen samples, see Fig. 1.

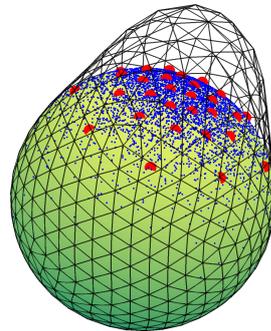


Fig. 1. A von Mises–Fisher density with $\kappa = 7$ (black mesh) is defined on a 2-sphere (green object). Initially, 5000 samples (blue points) are drawn randomly from this density. Using the proposed method, a reduction to 30 samples (red points) is performed.

State variables that describe directions or orientations, cannot be faithfully described in the Euclidean space \mathbb{R}^D . They require spaces with different topologies, for example the 2-sphere \mathbb{S}^2 . As a manifold, it is locally homeomorphic to \mathbb{R}^2 , so for small uncertainties, linearization of the space (“tangent plane”) at the point of interest is actually an adequate approximation. Of course, probability mass far apart from this linearization point will increasingly deteriorate estimation results.

Directional estimation problems occur in many practical problems, for example crystal orientation estimation using electron backscatter diffraction, Bingham et al. (2010), fiber tracking in biomedical image analysis, Zhang et al. (2009), or path planning in robotics, Sánchez and Latombe (2002).

* This work is supported by the German Research Foundation (DFG) under grant HA 3789/16-1.

State of Art: Measures to compare two *continuous* density functions are for example the Kullback-Leibler-divergence, Kullback and Leibler (1951), or simply the L_2 norm of the difference of the density functions. If at least one of the densities is discrete, i.e., a Dirac mixture density, and the samples are not equidistant, said distance measures cannot be applied anymore: the local *concentration* of samples must be involved in some way into the distance measure, in addition to the sample weight. The Cramér-von Mises distance as the L_2 norm of the difference of the *cumulative* distribution functions, see Darling (1957) and Choi and Bulgren (1968), is well suited to compare discrete densities on a one-dimensional domain. However for multivariate densities, the definition of the cumulative distribution is ambiguous, and the Cramér-von Mises distance cannot be employed directly.

Hanebeck and Klumpp (2008) proposed a *localized* cumulative distribution (LCD). Given a continuous or discrete density function, it provides a representation of that density that is unique and symmetric also in the multivariate case. The LCD of continuous or discrete densities is always continuous, so the L_2 norm of the difference of two LCDs can be calculated. This is called the *modified* Cramér-von Mises distance, and serves as an objective function to match two multivariate densities, regardless whether they are continuous or discrete. It has been used to vastly reduce numbers of samples while keeping most of the relevant information, see Eberhardt et al. (2010) and Hanebeck (2015). On the other hand, the LCD has been applied to obtain deterministic samples of given multivariate Gaussian densities in Hanebeck et al. (2009). Using these deterministic samples, some Gaussian filtering algorithms have been developed, for example the Smart Sampling Kalman Filter (S²KF) in Steinbring and Hanebeck (2013) and Steinbring et al. (2016), or progressive Gaussian filters in Steinbring and Hanebeck (2014) and Hanebeck and Pander (2016).

All of the publications mentioned so far applied the LCD in D -dimensional Euclidean space \mathbb{R}^D . Now the question arises whether the LCD can also be used for sample reduction on non-Euclidean manifolds. The first idea was proposed in Li et al. (2019a), where the LCD was applied in the Euclidean tangent space \mathbb{R}^{D-1} of a hypersphere S^{D-1} . Deterministic samples were placed on the principle axes only. This limitation was overcome in Li et al. (2019b), as samples were placed not only on the principle axes but in the entire tangent space of the hypersphere. The resulting Bingham filter can be seen as spherical equivalent to the S²KF. The linearization point for the tangent space was always placed at the mode of the spherical (Bingham) distribution, and various projection methods were used to transform samples between hypersphere and tangent space.

Finally we want to address some of the literature about non-parametric density function estimation on manifolds. Methods for spheres, Ruymgaart (1989), or more generally, closed manifolds, Hendriks (1990), have been proposed based on Fourier expansions, for example. Furthermore, there are Bayesian nonparametric density estimators based on mixtures of parametric densities like the vMF for spheres (Bhattacharya and Bhattacharya, 2012, Sec. 13.2), Bhattacharya and Oliver (2019). These methods allow to compute an adequate *continuous* representation of a given set of samples.

Contribution: We propose a definition of the LCD for DMDs directly on the \mathbb{S}^2 sphere. We choose the vMF density as kernel function and it will become obvious that it can be seen as a spherical equivalent to the Gaussian kernel. Furthermore, a modified Cramér-von Mises distance is formulated on the sphere. It provides a scalar distance measure between two spherical LCDs. With these tools, we propose a distance measure between two sets of samples on the sphere. This distance measure is then minimized using nonlinear optimization. Altogether, a deterministic method for optimal sample reduction on the \mathbb{S}^2 sphere is proposed, see Fig. 1 for an example.

Outline: In a first step, a distance measure between two DMDs, i.e., sets of samples, that are located on the unit sphere, is introduced. The numbers of samples are fixed beforehand, as well as the sample locations of one of the sets. In a second step, by minimizing the distance measure we propose a deterministic resampling method. Thereby it is possible to i) reduce the number of samples, and/or to ii) approximate weighted samples with unweighted ones. No local linearization of \mathbb{S}^2 is necessary to calculate the distance measure. The distance measure is given in closed form as a nonlinear function of the sample locations. The computational effort lies in numerical optimization of the distance measure. The unknown variables to be optimized are the sample locations of the smaller set of samples.

2. PROBLEM FORMULATION

Consider a given and fixed Dirac mixture density (DMD) $\tilde{f}(\underline{x})$, $\underline{x} \in \mathbb{S}^2 \subset \mathbb{R}^3$, with M components, given by

$$\tilde{f}(\underline{x}) = \sum_{j=1}^M w_j^y \delta(\underline{x} - \underline{y}_j) , \quad (1)$$

with positive weights w_j^y , i.e., $w_j^y > 0$ for $j \in \{1, 2, \dots, M\}$, and $\sum_{j=1}^M w_j^y = 1$. Samples are located at

$$\underline{y}_j = \begin{bmatrix} y_{j,x} \\ y_{j,y} \\ y_{j,z} \end{bmatrix} \quad \underline{y}_j = \begin{bmatrix} y_{j,\vartheta} \\ y_{j,\varphi} \end{bmatrix}$$

on the unit sphere, in Cartesian and spherical coordinates, respectively. This density is approximated by another DMD $f(\underline{x})$ with L components, given by

$$f(\underline{x}) = \sum_{i=1}^L w_i^x \delta(\underline{x} - \underline{x}_i) , \quad (2)$$

again with positive weights w_i^x that sum up to one. Sample locations on the unit sphere are

$$\underline{x}_i = \begin{bmatrix} x_{i,x} \\ x_{i,y} \\ x_{i,z} \end{bmatrix} \quad \underline{x}_i = \begin{bmatrix} x_{i,\vartheta} \\ x_{i,\varphi} \end{bmatrix}$$

in Cartesian and spherical coordinates, respectively. Typically, the approximating density $f(\underline{x})$ has much less components than the approximated density $\tilde{f}(\underline{x})$, i.e., $L \ll M$.

The goal is to select the location parameters \underline{x}_i , $i \in \{1, 2, \dots, L\}$ of the approximating density $f(\underline{x})$ such that a distance measure D between the true density $\tilde{f}(\underline{x})$ and its approximation $f(\underline{x})$ is minimized. L is assumed to be given and fixed. The corresponding weights w_i^x are assumed to be either given or set to be equal, i.e., $w_i^x = 1/L$.

The true DMD $\tilde{f}(\underline{x})$ might already have equally weighted components, so that the information is solely stored in the component locations. In this case, the goal of the approximation is a pure reduction of the number of components. On the other hand, the components of $\tilde{f}(\underline{x})$ might have different weights, e.g. as a result of weighting some prior DMD with a likelihood function in a Bayesian filtering setup. In that case, the approximation replaces an already weighted DMD by an equally weighted one.

3. LOCALIZED CUMULATIVE DISTRIBUTION

DMDs cannot easily be compared to each other, especially when they do not share a common support. Not the “vertical” difference of density values can be considered here, but the “horizontal” distance between samples and clusters of samples is essential. Therefore each DMD, $f(\underline{x})$, is first transformed to its LCD, $F(\underline{m}, b)$, which is a continuous characterization of the DMD. Intuitively, a kernel function $K(\underline{x}, \underline{m}, b)$ is placed on each component of the DMD. The kernel attains a maximum at $\underline{x} = \underline{m}$ (location parameter), and b controls the kernel size (scale parameter). We consider the LCD as *cumulative* distribution because we take b as an equivalent to the “upper integration limit” of the traditional scalar cumulative distribution $\int_{-\infty}^x f(t) dt$. Instead from $-\infty$, we integrate around \underline{m} in the LCD.

We propose to define the LCD $F(\underline{m}, b)$ for arbitrary densities $f(\underline{x})$ on \mathbb{S}^2 as

$$F(\underline{m}, b) = \int_{\mathbb{S}^2} f(\underline{x}) K(\underline{x}, \underline{m}, b) d\underline{x} .$$

In particular, the LCD $F(\underline{m}, b)$ of the approximating DMD $f(\underline{x})$ from (2) is

$$\begin{aligned} F(\underline{m}, b) &= \int_{\mathbb{S}^2} \overbrace{\left(\sum_{i=1}^L w_i^x \delta(\underline{x} - \underline{x}_i) \right)}^{f(\underline{x})} K(\underline{x}, \underline{m}, b) d\underline{x} \\ &= \sum_{i=1}^L w_i^x \int_{\mathbb{S}^2} \delta(\underline{x} - \underline{x}_i) K(\underline{x}, \underline{m}, b) d\underline{x} . \end{aligned} \quad (3)$$

With the Dirac delta distribution for spheres

$$\begin{aligned} \delta(\underline{x} - \underline{x}_i) &= \frac{\delta(x_\vartheta - x_{i,\vartheta}) \delta(x_\varphi - x_{i,\varphi})}{\sin(x_\vartheta)} , \\ d\underline{x} &= \sin(x_\vartheta) dx_\vartheta dx_\varphi , \end{aligned}$$

or by just using the general definition of the Dirac delta distribution for any manifold

$$\int_X f(\underline{x}) \delta(\underline{x} - \underline{x}_i) d\underline{x} = f(\underline{x}_i) ,$$

the LCD (3) becomes

$$F(\underline{m}, b) = \sum_{i=1}^L w_i^x K(\underline{x}_i, \underline{m}, b) . \quad (4)$$

Similarly, the LCD $\tilde{F}(\underline{m}, b)$ of the true density $\tilde{f}(\underline{x})$ (1) is

$$\tilde{F}(\underline{m}, b) = \sum_{j=1}^M w_j^y K(\underline{y}_j, \underline{m}, b) . \quad (5)$$

The LCDs $\tilde{F}(\underline{m}, b)$ and $F(\underline{m}, b)$ are continuous in \underline{m} and b and can thus be used to easily calculate a distance between $\tilde{f}(\underline{x})$ and $f(\underline{x})$.

4. A MODIFIED CRAMÉR-VON MISES DISTANCE

A straightforward distance measure D between two LCDs $\tilde{F}(\underline{m}, b)$ and $F(\underline{m}, b)$ is their L_2 norm. We also integrate this over b , thereby becoming independent of a specific choice of kernel width and instead using the information from all possible kernel widths,

$$\begin{aligned} D^2 &= \int_{\mathbb{R}_+} w(b) \int_{\mathbb{S}^2} (\tilde{F}(\underline{m}, b) - F(\underline{m}, b))^2 d\underline{m} db \quad (6) \\ &= D_{xx} - 2D_{xy} + D_{yy} \quad (7) \end{aligned}$$

with

$$\begin{aligned} D_{xx} &= \int_{\mathbb{R}_+} w(b) \int_{\mathbb{S}^2} F^2(\underline{m}, b) d\underline{m} db , \\ D_{xy} &= \int_{\mathbb{R}_+} w(b) \int_{\mathbb{S}^2} F(\underline{m}, b) \tilde{F}(\underline{m}, b) d\underline{m} db , \\ D_{yy} &= \int_{\mathbb{R}_+} w(b) \int_{\mathbb{S}^2} \tilde{F}^2(\underline{m}, b) d\underline{m} db . \end{aligned} \quad (8)$$

Inserting the definitions (4) and (5) of the particular LCDs, the products between any two LCDs yields

$$F^2(\underline{m}, b) = \sum_{i=1}^L \sum_{j=1}^L w_i^x w_j^x K(\underline{x}_i, \underline{m}, b) K(\underline{x}_j, \underline{m}, b) ,$$

$$F(\underline{m}, b) \tilde{F}(\underline{m}, b) = \sum_{i=1}^L \sum_{j=1}^M w_i^x w_j^y K(\underline{x}_i, \underline{m}, b) K(\underline{y}_j, \underline{m}, b) ,$$

$$\tilde{F}^2(\underline{m}, b) = \sum_{i=1}^M \sum_{j=1}^M w_i^y w_j^y K(\underline{y}_i, \underline{m}, b) K(\underline{y}_j, \underline{m}, b) .$$

The following integral $D_s(\underline{x}, \underline{y})$, depending on the locations \underline{x} and \underline{y} of one pair of samples,

$$D_s(\underline{x}, \underline{y}) = \int_{\mathbb{R}_+} w(b) \int_{\mathbb{S}^2} K(\underline{x}, \underline{m}, b) K(\underline{y}, \underline{m}, b) d\underline{m} db \quad (9)$$

conveniently encapsulates most of the complexity and allows for a very concise representation of the three main parts (7) of the modified Cramér-von Mises distance (7)

$$\begin{aligned} D_{xx} &= \sum_{i=1}^L \sum_{j=1}^L w_i^x w_j^x D_s(\underline{x}_i, \underline{x}_j) , \\ D_{xy} &= \sum_{i=1}^L \sum_{j=1}^M w_i^x w_j^y D_s(\underline{x}_i, \underline{y}_j) , \\ D_{yy} &= \sum_{i=1}^M \sum_{j=1}^M w_i^y w_j^y D_s(\underline{y}_i, \underline{y}_j) . \end{aligned}$$

Obviously, in order to get an analytic representation of the modified Cramér-von Mises distance (7), a closed form solution of the integral (9) must be available. The inner and outer part of the integral will be denoted separately as

$$D_{s,b}(\underline{x}, \underline{y}, b) = \int_{\mathbb{S}^2} K(\underline{x}, \underline{m}, b) K(\underline{y}, \underline{m}, b) d\underline{m} , \quad (10)$$

$$D_s(\underline{x}, \underline{y}) = \int_{\mathbb{R}_+} w(b) D_{s,b}(\underline{x}, \underline{y}, b) db . \quad (11)$$

5. SYMMETRIC KERNEL FUNCTIONS

Only symmetric kernel functions will be considered from now on. That means, the value of the kernel function $K(\underline{x}, \underline{y}, b)$ depends, except for b , only on the geodesic arc length $d(\underline{x}, \underline{y})$ between \underline{x} and \underline{y} , and not on their absolute location on the sphere,

$$K(\underline{x}, \underline{y}, b) = K(d(\underline{x}, \underline{y}), b) .$$

In this case, due to symmetry also the sample distance D_s will depend solely on the geodesic between the samples

$$\begin{aligned} D_s(\underline{x}, \underline{y}) &= D_s(d(\underline{x}, \underline{y})) , \\ D_{s,b}(\underline{x}, \underline{y}, b) &= D_{s,b}(d(\underline{x}, \underline{y}), b) . \end{aligned} \quad (12)$$

In order to obtain the geodesic of a pair of samples, the dot product is calculated first. The value of the dot product of two unit vectors $\underline{u}, \underline{v}$ is

$$\begin{aligned} d_c(\underline{u}, \underline{v}) &= \cos(d(\underline{u}, \underline{v})) \\ &= \underline{u}^\top \underline{v} \\ &= u_x v_x + u_y v_y + u_z v_z \end{aligned}$$

from their Cartesian representation $\underline{u} = [u_x, u_y, u_z]^\top$, and

$$d_c = \sin(u_\vartheta) \sin(v_\vartheta) \cos(u_\varphi - v_\varphi) + \cos(u_\vartheta) \cos(v_\vartheta)$$

from their spherical coordinates.

6. VON MISES-FISHER KERNEL

The LCD has previously been developed for densities in \mathbb{R}^N , where a Gaussian kernel

$$\begin{aligned} K(\underline{x}, \underline{m}, \sigma) &= \exp \left\{ -\frac{1}{2\sigma} (\underline{x} - \underline{m})^\top (\underline{x} - \underline{m}) \right\} , \\ K(d(\underline{x}, \underline{m}), \sigma) &= \exp \left\{ -\frac{1}{2\sigma} (d(\underline{x}, \underline{m}))^2 \right\} \end{aligned} \quad (13)$$

provided good results together with the weighting function $w(\sigma) = 1/\sigma$ in the two-dimensional case, see Hanebeck (2015).

Following (8), the integral over the product of two such Gaussian kernel functions and said weighting factor is

$$\begin{aligned} &\int_{\sigma=\sigma_1}^{\sigma_2} w(\sigma) K(d_1, \sigma) K(d_2, \sigma) d\sigma \\ &= \int_{\sigma=\sigma_1}^{\sigma_2} \underbrace{\frac{1}{\sigma}}_{w(\sigma)} \underbrace{\exp \left\{ -\frac{1}{2} \frac{d_1^2 + d_2^2}{\sigma^2} \right\}}_{K(d_1, \sigma) \cdot K(d_2, \sigma)} d\sigma . \end{aligned}$$

By substitution

$$b = \sigma^{-2} , \quad \sigma = b^{-\frac{1}{2}} , \quad \sigma' = -\frac{b^{-\frac{3}{2}}}{2} ,$$

this can be transformed into

$$\begin{aligned} &\int_{b=\sigma_1^{-2}}^{\sigma_2^{-2}} \frac{1}{b^{-\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \frac{d_1^2 + d_2^2}{(b^{-\frac{1}{2}})^2} \right\} \left(-\frac{b^{-\frac{3}{2}}}{2} \right) db \\ &= \frac{1}{2} \int_{b=\sigma_2^{-2}}^{\sigma_1^{-2}} \frac{1}{b} \exp \left\{ b \cdot \left(-\frac{d_1^2 + d_2^2}{2} \right) \right\} db . \end{aligned}$$

While the dot product in the exponent of (13) is a natural squared distance of points in Euclidean space, the cosine with its periodicity is tailored for spherical geometries.

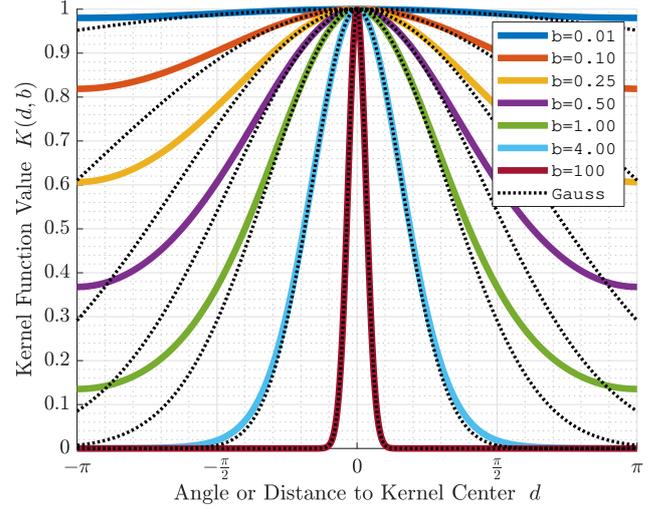


Fig. 2. Comparison of vMF kernel and Gaussian kernel. Colored lines denote the vMF kernel function value over the geodesic d to the kernel center. Individual colors represent various kernel widths b , as indicated in the legend. For each of these vMF kernel functions, the corresponding Gaussian kernel function ($\sigma = b^{-1/2}$) is plotted as a dashed black line. Note that both kernels are very similar near the kernel center, and the narrow kernels (small σ , large b) are similar everywhere.

Incorporating $(\cos(d) - 1)$ as the “spherical version” of $(-d^2/2)$,

$$-\frac{d^2}{2} \approx \cos(d) - 1 ,$$

where the approximation stands for second order Taylor polynomial, one obtains

$$\begin{aligned} &\int_{b=\sigma_1^{-2}}^{\sigma_2^{-2}} \frac{1}{b} e^{-2b} e^{b \cos(d_1)} e^{b \cos(d_2)} db \\ &= \int_{b=\sigma_1^{-2}}^{\sigma_2^{-2}} w(b) K(d_1, b) K(d_2, b) db , \end{aligned}$$

with the vMF kernel function

$$\begin{aligned} K(d, b) &= e^{b \cdot \cos(d)} \\ K(d(\underline{x}, \underline{m}), b) &= e^{b \cdot \cos(d(\underline{x}, \underline{m}))} \\ &= e^{b \cdot \underline{x}^\top \underline{m}} \end{aligned}$$

and the weighting factor

$$w(b) = \frac{1}{b} e^{-2b} . \quad (14)$$

Thus, the vMF kernel with $w(b) = b^{-1} e^{-2b}$ can be seen as the spherical equivalent to the Gaussian kernel with $w(\sigma) = 1/\sigma$. See also Fig. 2 for a visual comparison between Gaussian and vMF kernel. Note that the kernel function is scaled such that it has a maximum value of one, as opposed to the vMF distribution, which is normalized such that it integrates to one, see Fisher (1953).

Having established the choice of kernel function and weighting factor, the function $D_{s,b}(\underline{u}, \underline{v}, b)$ will be calculated by integrating the product of two vMF kernels on the sphere according to (10)

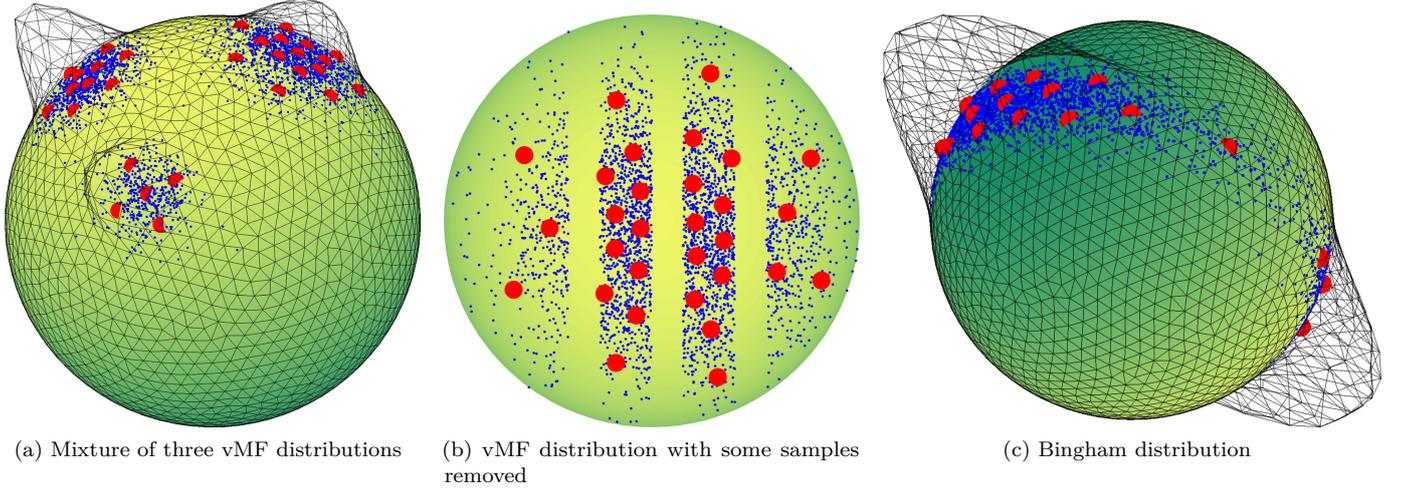


Fig. 3. Different examples for sample reduction on the 2-sphere. a) Mixture of three vMF distributions. 1000 random samples with $b = 50$ in the upper left, 1500 samples with $b = 30$ in the upper right, and 500 samples with $b = 70$ in the center, altogether approximated by 30 samples (red dots). b) vMF distribution with $b = 7$, where three strips of samples are removed. The remaining 3333 samples are approximated by 30 deterministic samples. c) Bingham distribution with $Z = \text{diag}(-40, -3, 0)$. 5000 stochastic samples are reduced to 30 deterministic samples.

$$\begin{aligned}
 D_{s,b}(\underline{u}, \underline{v}, b) &= \int_{\mathbb{S}^2} e^{b \cdot \underline{u}^\top \underline{m}} \cdot e^{b \cdot \underline{v}^\top \underline{m}} d\underline{m} \\
 &= \int_{\mathbb{S}^2} e^{b \cdot (\underline{u} + \underline{v})^\top \underline{m}} d\underline{m} . \quad (15)
 \end{aligned}$$

Due to symmetry, it is not important where \underline{u} and \underline{v} are located on the sphere, only their geodesic distance d has an influence on the result. One convenient choice of \underline{u} and \underline{v} with geodesic distance d between them is

$$\begin{aligned}
 \begin{bmatrix} u_\vartheta \\ u_\varphi \end{bmatrix} = \begin{bmatrix} d/2 \\ 0 \end{bmatrix} &\Leftrightarrow \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \begin{bmatrix} \sin(d/2) \\ 0 \\ \cos(d/2) \end{bmatrix} , \\
 \begin{bmatrix} v_\vartheta \\ v_\varphi \end{bmatrix} = \begin{bmatrix} d/2 \\ \pi \end{bmatrix} &\Leftrightarrow \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} -\sin(d/2) \\ 0 \\ \cos(d/2) \end{bmatrix} .
 \end{aligned}$$

The integration variable \underline{m} will be denoted as

$$\begin{bmatrix} m_\vartheta \\ m_\varphi \end{bmatrix} = \begin{bmatrix} \vartheta \\ \varphi \end{bmatrix} \Leftrightarrow \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} = \begin{bmatrix} \sin(\vartheta) \cos(\varphi) \\ \sin(\vartheta) \sin(\varphi) \\ \cos(\vartheta) \end{bmatrix} .$$

Then,

$$(\underline{u} + \underline{v})^\top \underline{m} = 2 \cos(d/2) \cos(\vartheta) ,$$

and the integral (15) becomes

$$\begin{aligned}
 D_{s,b}(d, b) &= \int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} e^{2b \cos(d/2) \cos(\vartheta)} \sin(\vartheta) d\vartheta d\varphi \\
 &= 2\pi \int_{\vartheta=0}^{\pi} e^{2b \cos(d/2) \cos(\vartheta)} \sin(\vartheta) d\vartheta \\
 &\quad | \quad u = \cos(\vartheta), \quad u' = -\sin(\vartheta) \\
 &= -2\pi \int_{u=1}^{-1} e^{2b \cos(d/2) u} du \\
 &= \frac{2\pi}{2b \cos(d/2)} \left[e^{2b \cos(d/2) u} \right]_{u=-1}^1 \\
 &= \frac{\pi}{b \cos(d/2)} \left(e^{2b \cos(d/2)} - e^{-2b \cos(d/2)} \right)
 \end{aligned}$$

$$= \frac{\pi}{b \cos(d/2)} 2 \sinh(2b \cos(d/2)) .$$

Multiplication with the weighting function (14) and integration (11) over b yields the sample distance function $D_s(d)$ (12)

$$D_s(d) = \frac{\pi}{\cos(d/2)} \int_{b_1}^{b_2} \frac{e^{-2b}}{b^2} 2 \sinh(2b \cos(d/2)) db .$$

Using the relationships from Zeidler (2013)

$$\begin{aligned}
 \frac{d}{dx} \text{Ei}(x) &= \frac{e^x}{x} , \\
 \int \frac{1}{b^2} e^{ab} db &= \begin{cases} a \text{Ei}(ab) - \frac{e^{ab}}{b} , & a \neq 0 , \\ -\frac{1}{b} , & a = 0 , \end{cases}
 \end{aligned}$$

the integral can be solved analytically. Then $D_s(d)$ for the vMF kernel calculates to

$$D_s(d) = \begin{cases} \pi \left[4 \text{Ei}(-4b) + \frac{1}{b} (e^{-4b} - 1) \right]_{b_1}^{b_2} , & d = 0 , \\ \frac{\pi}{\cos(d/2)} \left[c_1 \text{Ei}(c_1 b) - c_2 \text{Ei}(c_2 b) + \frac{1}{b} (e^{c_2 b} - e^{c_1 b}) \right]_{b_1}^{b_2} , & d \in (0, \pi) , \\ [4\pi \text{Ei}(-2b)]_{b_1}^{b_2} , & d = \pi , \end{cases}$$

where

$$\begin{aligned}
 c_1 &= +2 \cos(d/2) - 2 , \\
 c_2 &= -2 \cos(d/2) - 2 .
 \end{aligned}$$

Bounds of integration were set to $b_1 = 0.001$ and $b_2 \rightarrow \infty$.

7. EVALUATION

In order to demonstrate and visualize the performance of this method, various sets of samples were first randomly drawn by stochastic sampling, see blue points in Fig. 3. For the vMF distribution, the stochastic sampling method by Kurz and Hanebeck (2015) was used, and for the Bingham distribution Kent et al. (2013). In both cases,

the implementation in the `libDirectional` library, Kurz et al. (2015) and Kurz et al. (2019) was employed. Each DMD (1) consisting of $M = 3000 \dots 5000$ stochastic points \underline{y}_j was then compared to a DMD (2) of $L = 30$ points \underline{x}_i , that were as well randomly chosen as an initial guess, using the distance measure (6). The $(2 \cdot L)$ parameters $x_{i,\vartheta}$ and $x_{i,\varphi}$ were subsequently optimized using Matlab's `fminunc`. Analytical gradients were provided for better optimization performance. The code of the presented method will be published as part of `libDirectional` as well and can be found in the class `AbstractHypersphericalDistribution`.

8. CONCLUSION

We propose a distance measure between two sets of samples on the sphere. Minimizing this distance using off-the-shelf numerical optimization algorithms provides an efficient method for optimal sample reduction, deterministic sampling, and resampling of weighted samples by unweighted samples.

REFERENCES

- Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric Inference on Manifolds: With Applications to Shape Spaces*, volume 2. Cambridge University Press.
- Bhattacharya, R. and Oliver, R. (2019). Nonparametric Analysis of Non-Euclidean Data on Shapes and Images. *Sankhya A*, 81(1), 1–36.
- Bingham, M.A., Lograsso, B.K., and Laabs, F.C. (2010). A Statistical Analysis of the Variation in Measured Crystal Orientations obtained through Electron Backscatter Diffraction. *Ultramicroscopy*, 110(10), 1312 – 1319.
- Choi, K. and Bulgren, W.G. (1968). An Estimation Procedure for Mixtures of Distributions. *Journal of the Royal Statistical Society*.
- Darling, D.A. (1957). The Kolmogorov-Smirnov, Cramér-von Mises Tests. *The Annals of Mathematical Statistics*, 28(4), 823–838.
- Eberhardt, H., Klumpp, V., and Hanebeck, U.D. (2010). Optimal Dirac Approximation by Exploiting Independencies. In *Proceedings of the 2010 American Control Conference (ACC 2010)*. Baltimore, Maryland, USA.
- Fisher, R. (1953). Dispersion on a Sphere. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 217(1130), 295–305.
- Hanebeck, U.D. (2015). Optimal Reduction of Multivariate Dirac Mixture Densities. *at – Automatisierungstechnik, Oldenbourg Verlag*, 63(4), 265–278. doi:10.1515/ato-2015-0005.
- Hanebeck, U.D., Huber, M.F., and Klumpp, V. (2009). Dirac Mixture Approximation of Multivariate Gaussian Densities. In *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*. Shanghai, China.
- Hanebeck, U.D. and Klumpp, V. (2008). Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance. In *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)*, 33–39. Seoul, Republic of Korea.
- Hanebeck, U.D. and Pander, M. (2016). Progressive Bayesian Estimation with Deterministic Particles. In *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)*. Heidelberg, Germany.
- Hendriks, H. (1990). Nonparametric Estimation of a Probability Density on a Riemannian Manifold Using Fourier Expansions. *The Annals of Statistics*, 18(2), 832–849.
- Kent, J.T., Ganeiber, A.M., and Mardia, K.V. (2013). A new method to simulate the Bingham and related distributions in directional data analysis with applications. *arXiv preprint arXiv:1310.8110*.
- Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Kurz, G., Gilitschenski, I., Pfaff, F., and Drude, L. (2015). `libDirectional`. URL <https://github.com/libDirectional>.
- Kurz, G., Gilitschenski, I., Pfaff, F., Drude, L., Hanebeck, U.D., Haeb-Umbach, R., and Siegwart, R.Y. (2019). Directional Statistics and Filtering Using `libDirectional`. *Journal of Stat. Software*.
- Kurz, G. and Hanebeck, U.D. (2015). Stochastic Sampling of the Hyperspherical von Mises–Fisher Distribution Without Rejection Methods. In *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion (SDF 2015)*. Bonn, Germany.
- Li, K., Frisch, D., Noack, B., and Hanebeck, U.D. (2019a). Geometry-Driven Deterministic Sampling for Nonlinear Bingham Filtering. In *Proceedings of the 2019 European Control Conference (ECC 2019)*. Naples, Italy.
- Li, K., Pfaff, F., and Hanebeck, U.D. (2019b). Hyperspherical Deterministic Sampling Based on Riemannian Geometry for Improved Nonlinear Bingham Filtering. In *Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019)*. Ottawa, Canada.
- Ruymgaart, F. (1989). "strong uniform convergence of density estimators on spheres". *Journal of Statistical Planning and Inference*, 23(1), 45 – 52.
- Steinbring, J. and Hanebeck, U.D. (2013). S2KF: The Smart Sampling Kalman Filter. In *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*. Istanbul, Turkey.
- Steinbring, J. and Hanebeck, U.D. (2014). Progressive Gaussian Filtering Using Explicit Likelihoods. In *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*. Salamanca, Spain.
- Steinbring, J., Pander, M., and Hanebeck, U.D. (2016). The Smart Sampling Kalman Filter with Symmetric Samples. *Journal of Advances in Information Fusion*, 11(1), 71–90.
- Sánchez, G. and Latombe, J.C. (2002). On Delaying Collision Checking in PRM Planning: Application to Multi-Robot Coordination. *The International Journal of Robotics Research*, 21(1), 5–26.
- Zeidler, E. (2013). *Springer-Taschenbuch der Mathematik: Begründet von I.N. Bronstein und K.A. Semendjajew Weitergeführt von G. Grosche, V. Ziegler und D. Ziegler Herausgegeben von E. Zeidler*, 174. Springer Fachmedien Wiesbaden.
- Zhang, F., Hancock, E.R., Goodlett, C., and Gerig, G. (2009). Probabilistic White Matter Fiber Tracking using Particle Filtering and von Mises–Fisher Sampling. *Medical Image Analysis*, 13(1), 5 – 18.