Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Moment-based Dirac Mixture Approximation of Circular Densities

Uwe D. Hanebeck * **and Anders Lindquist** **

* *Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany (e-mail: Uwe.Hanebeck@ieee.org).*
** *Department of Automation, Shanghai Jiao Tong University, Shanghai, China, and CIAM and ACCESS Centers, Royal Institute of Technology, Stockholm, Sweden (e-mail: alq@kth.se).*

Abstract: Given a circular probability density function, called the true probability density function, the goal is to find a Dirac mixture approximation based on some circular moments of the true density. When keeping the locations of the Dirac points fixed, but almost arbitrarily located, we are applying recent results on the circulant rational covariance extension problem to the problem of calculating the weights. For the case of simultaneously calculating optimal locations, additional constraints have to be deduced from the given density. For that purpose, a distance measure for the deviation of the Dirac mixture approximation from the true density is derived, which then is minimized while considering the moment conditions as constraints. The method is based on progressive numerical minimization, converges quickly and gives well-distributed Dirac mixtures that fulfill the constraints, i.e., have the desired circular moments.

Keywords: Circular densities, Dirac mixture approximation, moment problem.

1. INTRODUCTION

Periodic quantities occur in a variety of applications (Mardia and Jupp, 2009). In the scalar case, they can always be mapped to the circumference of a circle with appropriate radius. When the error in these quantities is large, the topology cannot be ignored and the uncertainty is described by means of circular probability density functions, which leads to so-called circular statistics. Based on this description, systematic estimation, filtering, and data fusion can be performed. Non-periodic quantities will be called *linear* in the following.

Although the primary uncertainties often can be described by simple continuous circular probability density functions such as the von Mises distribution, see Sec. 6, or the wrapped Normal distribution (Mardia and Jupp, 2009), processing these distributions quickly becomes difficult (Azmani et al., 2009; Stienne et al., 2014). Hence, a discrete approximation is often useful, which could either mean discrete probabilities over a regular lattice or irregularly spaced samples. In the latter case, we can distinguish between randomly and deterministically placed samples. In this paper, we focus on deterministically placed samples, which we will call Dirac mixture approximation, and which either are equally weighted or equipped with weights (that sum to one).

Replacing a continuous circular probability density function by a Dirac mixture approximation can be posed as an approximation problem. Typically, either the resolution, i.e., the number of discrete samples on the circle, is prespecified or some measure of the approximation quality is given. The parameters of the Dirac mixture approximation are then calculated in such a way that the discrete approximation matches some properties of the given continuous density, e.g., its moments.

Moment problems have a long history in mathematics, going back to Chebyshev, Markov, and Lyapunov (for a reference, see, e.g., Krein and Nudelman (1977)). In the setting of this paper with probability densities on the unit circle, one needs to consider (truncated) trigonometric moment problems: given a finite sequence $(c_0, c_1, \ldots, c_N)$ of numbers, find a positive measure $d\mu$ such that

$$\int_{-\pi}^{\pi} e^{in\theta} \,\mathrm{d}\mu = c_n \ , \quad n = 0, 1, \ldots, N \ . \tag{1}$$

(Since $c_0 = 1$, there are N nontrivial moments.) If the measure is absolutely continuous, there is a probability density p such that $d\mu = p(\theta)d\theta$. In this paper, we shall consider the approximation of such a probability measure by a μ that is a staircase function. One can formally regard such a measure as one with a probability density

$$p(\theta) = \sum_{j=1}^{L} w_j \, \delta(\theta - \hat{\theta}_j) \ , \tag{2}$$

where w_j, $j = 1, \ldots, L$, are positive weights adding up to one, and where $\hat{\theta}_j$, $j = 1, \ldots, L$, are the Dirac locations. This is the formalism we shall use in this paper.

For noncircular/linear quantities, moment-based Dirac mixture approximations have been proposed in the context of the Linear Regression Kalman Filter (LRKF), see Lefebvre et al. (2005). As these are Gaussian filters, the Dirac mixture approximation is performed for Gaussian densities. Typical examples are the Unscented Kalman Filter (UKF) in Julier and Uhlmann (2004), its scaled version in Julier (2002), and its higher-order generalization in Tenne and Singh (2003). A generalization to an arbitrary number of deterministic samples that are placed along the coordinate axes is introduced in Huber

and Hanebeck (2008). Also for linear quantities, systematic Dirac mixture approximation based on distance measures have been proposed for the case of scalar continuous densities (Schrempf et al., 2006a,b). An algorithm for sequentially increasing the number of components is given in Hanebeck and Schrempf (2007) and applied to recursive nonlinear prediction in Schrempf and Hanebeck (2007). Systematic Dirac mixture approximations of arbitrary multi-dimensional Gaussian densities are calculated in Hanebeck et al. (2009). A more efficient method for the case of standard normal distributions with a subsequent transformation is given in Gilitschenski and Hanebeck (2013). In Hanebeck (2014), a faster version of Hanebeck et al. (2009) has been introduced that gives slightly suboptimal results. The method does not rely on comparing the probability masses on all scales as in Hanebeck et al. (2009). Instead, repulsion kernels are introduced to assemble an induced kernel density and perform the comparison of the given density with its Dirac mixture approximation.

For circular quantities, especially the von Mises distribution and the wrapped Normal distribution, a first approach to Dirac mixture approximation in the spirit of the UKF is introduced in Kurz et al. (2013b). The approach is based on matching the first circular moment and the locations of three Dirac components are calculated. The resulting approximation has already been applied to sensor scheduling based on bearings-only measurements (Gilitschenski et al. (2013)). The results are also used to perform recursive circular filtering for tracking an object constrained to an arbitrary one-dimensional manifold in Kurz et al. (2013a),

This paper focuses on Dirac mixture approximations with an arbitrary number of, say L, components that match a given number of, say $N \leq L$, circular moments of a given circular probability density function. We start with Dirac mixture approximations with prespecified component locations, where the weights are determined based on the theory of circulant rational covariance extension developed by Lindquist and Picci (2013), which in turn generalizes previous results on reciprocal processes in Krener (1986); Levy et al. (1990); Carli et al. (2011); Carli and Georgiou (2011) and on covariance extension with rationality constraints in Georgiou (1987); Byrnes et al. (1995, 1999). Subsequently, we consider additionally optimizing the component locations, as we desire to adapt the density of Dirac components to the height of the given continuous probability density function. This gives another L parameters. For the under-determined case, i.e., the number of constraints resulting from the given circular moments is less than the number of parameters, we minimize a measure of deviation of the Dirac mixture approximation from the given circular probability density function while considering the moment conditions as constraints. This is performed in analogy to the ideas proposed for linear quantities in Hanebeck (2014).

The paper is structured as follows. A detailed formulation of the considered approximation problem is given in Sec. 2. The resulting moment problem is formulated in Sec. 3. The case of a Dirac mixture approximation with fixed locations is treated in Sec. 4. For the case of a smaller amount of moments constraints compared to the number of parameters of the Dirac mixture approximation, a distance measure between the Dirac mixture approximation and the given density is derived in Sec. 5 and the resulting constrained optimization problem is given. An example of approximating circular probability density functions is given in Sec. 6, where a von Mises distribution is approximated. Conclusion are given in Sec. 7.

2. PROBLEM FORMULATION

We consider periodic probability density functions $p : \mathbb{R} \to \mathbb{R}_+$ with $p(\theta) = p(\theta + c)$ for $c \in \mathbb{R}_+$. When $c = 2\pi$, $p(\theta)$ is called a circular density and it is sufficient to consider a domain Γ that is some interval of length 2π. Without loss of generality we will use $\Gamma = [-\pi, \pi)$.

2.1 Circular Dirac mixtures

A circular Dirac mixture density $p(\theta)$ with L Dirac components is given by (2) with positive weights, i.e., $w_i > 0$ for $i = 1, \ldots, L$, adding up to one and locations $\hat{\theta}_i \in \Gamma$ for $i = 1, \ldots, L$. The symbol $\delta(.)$ denotes the Dirac delta distribution.

2.2 Circular moments

The circular moments of a probability density function $p(\theta)$ are given by

$$\mathrm{E}_p\left\{e^{in\theta}\right\} = \int_{-\pi}^{\pi} e^{in\theta}\, p(\theta)\, \mathrm{d}\theta \tag{3}$$

for $n \in \mathbb{N}$. Obviously, the circular moments are the coefficients of the Fourier series expansion of the probability density function $p(\theta)$.

For the circular Dirac mixture in (2), the circular moments are given by

$$\mathrm{E}_p\left\{e^{in\theta}\right\} = \sum_{j=1}^{L} w_j \int_{-\pi}^{\pi} e^{in\theta}\, \delta(\theta - \hat{\theta}_j)\, \mathrm{d}\theta \ . \tag{4}$$

Using the sifting property of the Dirac delta distribution gives

$$\mathrm{E}_p\left\{e^{in\theta}\right\} = \sum_{j=1}^{L} w_j\, e^{in\hat{\theta}_j} \ . \tag{5}$$

2.3 Approximation problem

Given a circular density $\tilde{p}(\theta)$, we would like to find a Dirac mixture approximation $p(\theta)$ as in (2) that has exactly the same first N circular moments

$$c_n = \int_{-\pi}^{\pi} e^{in\theta}\, \tilde{p}(\theta)\, \mathrm{d}\theta, \quad n = 1, 2, \ldots, N \ . \tag{6}$$

Clearly, $c_0 = 1$ is fixed.

Remark 1. Many interesting variations of this approximation problem exist. First of all, N arbitrary moments could be considered. Second, it would also make sense to consider more given moments than there are degrees of freedom in the approximating density. In that case, the moments cannot be exactly attained. Instead, some error between the given moments and the approximate moments has to be minimized.

3. MOMENT PROBLEM

We will now formalize the problem of constructing a Dirac mixture approximation for a given density in such a way that the two densities exactly share the first N circular moments. First of all, we focus on symmetric densities, which simplifies the notation. The results can then easily be generalized to non-symmetric densities. Without loss of generality, it is sufficient to consider densities symmetric to $\theta = 0$ during the approximation, as the resulting Dirac

mixture approximation can later be shifted to arbitrary center locations by shifting each component accordingly.

For symmetric densities $p(\theta)$, the imaginary parts of the circular moments are zero, so that we are left with real moments

$$\mathrm{E}_p\{\cos(n\,\theta)\} = \int_{-\pi}^{\pi} \cos(n\,\theta)\, p(\theta)\,\mathrm{d}\theta \tag{7}$$

that are from now on used as the given moments $c_n = \mathrm{E}_p\{\cos(n\,\theta)\}$.

For symmetric Dirac mixtures, we have to distinguish two cases, namely even and odd numbers of components. Here we focus on an even number of components, so that the circular Dirac mixture centered around $\theta = 0$ can be written as

$$p(\theta) = \sum_{j=1}^{L/2} w_j \cdot \Big(\delta(\theta+\hat{\theta}_j) + \delta(\theta-\hat{\theta}_j)\Big) \tag{8}$$

with $\hat{\theta}_j \in [0,\pi]$ for $j = 1,\ldots,L/2$. Its moments are given by

$$\mathrm{E}_p\{\cos(n\,\theta)\} = \sum_{j=1}^{L/2} 2\,w_j \cos(n\,\hat{\theta}_j)\ . \tag{9}$$

We are given N moments c_n for $n = 1,\ldots,N$ that we would like to match with a symmetric Dirac mixture approximation with L components as in (8). When weights and locations are free parameters, we have $L-1$ parameters available for matching the given N moments. The parameters are collected in a parameter vector $\boldsymbol{\eta}$ given by

$$\boldsymbol{\eta} = \big[w_1,\ldots,w_{L/2},\hat{\theta}_1,\ldots,\hat{\theta}_{L/2}\big]^T \tag{10}$$

and we now parametrize the circular Dirac mixture in (8) as $p(\theta) = p(\theta,\boldsymbol{\eta})$. The parameter vector is confined to the domain $\mathcal{S} = \mathbb{R}_+^{L/2} \times \Gamma^{L/2}$ with $2\,w_1 + \ldots + 2\,w_{L/2} = 1$.

In Sec. 4, we fix $\hat{\theta}_1,\ldots,\hat{\theta}_L$ in advance and just determine the weights $w_1,\ldots,w_L$. With this method, the moment equations should actually be underdetermined leaving room for tuning parameters.

4. APPLYING THE CIRCULANT RATIONAL TRIGONOMETRIC MOMENT PROBLEM

In this section we apply some recent results in Lindquist and Picci (2013) to the approximation problem of Sec. 2. (In that paper the density p is a spectral density, but the formalism remains the same.) Then, we need to fix the locations of the Dirac points *a priori*, but by choosing many equidistant points on the discrete circle and constraining the weights to be zero at certain points, we can achieve an almost arbitrary selection of the locations of the Dirac points.

To this end, set $\zeta_1 := e^{i\Delta}$, where $\Delta := 2\pi/L$, and define the the discrete variable ζ taking the L values $\zeta_k := \zeta_1^k = e^{ik\Delta}$ running counter-clockwise on the discrete unit circle $\mathbb{T}_L$. In particular, $\zeta_{L-k} = \zeta_k$. Then, for a sufficiently large $L > 2N$, we consider the Dirac delta distributions in (2) with $\hat{\theta}_j := (j-1)\Delta$ leading to a moment problem

$$\sum_{j=0}^{L-1} \zeta_j^n\, w_{j+1} = c_n, \quad n = 1,2,\ldots,N\ , \tag{11}$$

where $c_1, c_2, \ldots, c_N$ are the moments defined in (6) of the given continuous probability density $\tilde{p}$ to be matched, and some of the weights w_j are constrained to be zero. We also define the zeroth moment c_0, which clearly is one, i.e., $c_0 = 1$.

4.1 Dual cones

Consider the class of trigonometric polynomials

$$Q(e^{i\theta}) = \sum_{k=-N}^{N} q_k e^{-ik\theta}\ , \quad q_{-k} = q_k\ , \tag{12}$$

where $\mathbf{q} := [q_0, q_1, \ldots, q_N]^T \in \mathbb{R}^{N+1}$ with $q_0 > 0$. Let $\mathfrak{P}_+(L)$ be the cone of all such $(N+1)$-vectors for which

$$Q(\zeta_j) \geq 0\ , \quad j = 1,2,\ldots,L\ . \tag{13}$$

Then, setting $\hat{\mathbf{q}} = [Q(\zeta_0), Q(\zeta_1), \ldots, Q(\zeta_{L-1})]^T$, we have

$$\hat{\mathbf{q}} = \mathbf{F}\mathbf{q}\ , \tag{14}$$

where

$$\mathbf{F} = \begin{bmatrix} 1 & 2\cos(\hat{\theta}_1) & 2\cos(2\hat{\theta}_1) & \cdots & 2\cos(N\hat{\theta}_1) \\ 1 & 2\cos(\hat{\theta}_2) & 2\cos(2\hat{\theta}_2) & \cdots & 2\cos(N\hat{\theta}_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2\cos(\hat{\theta}_L) & 2\cos(2\hat{\theta}_L) & \cdots & 2\cos(N\hat{\theta}_L) \end{bmatrix}\ , \tag{15}$$

since

$$Q(e^{i\theta}) = q_0 + 2\sum_{k=1}^{N} q_k \cos(k\theta)\ .$$

Consequently, the vectors $\mathbf{q} \in \mathfrak{P}_+(L)$ are precisely the $\mathbf{q}$ satisfying $\mathbf{F}\mathbf{q} \geq 0$. However, $\hat{\mathbf{q}} \neq 0$ since $q_0 > 0$. We also define the interior $\mathring{\mathfrak{P}}_+(L)$ of $\mathfrak{P}_+(L)$. This is an open cone which requires all $\hat{q}_j := Q(\zeta_j)$, $j = 0,1,\ldots,L-1$, to be positive.

The $L \times (N+1)$ matrix $\mathbf{F}$ is a Vandermonde matrix and, as $L > 2N$, it has full row rank. It follows from the theory of discrete Fourier transforms and is easy to check that

$$\mathbf{F}^T\mathbf{F} = L\mathbf{I}\ , \tag{16}$$

where $\mathbf{I}$ is the $(N+1)\times(N+1)$ identity. Hence

$$\mathbf{q} = \frac{1}{L}\mathbf{F}^T\hat{\mathbf{q}}\ . \tag{17}$$

Next define the dual cone $\mathfrak{C}_+(L)$ of all

$$\mathbf{c} := [c_0, c_1, \ldots, c_N]^T \in \mathbb{R}^{N+1} \quad \text{with} \quad c_0 = 1 \tag{18}$$

such that

$$\mathbf{c}^T\mathbf{q} \geq 0 \quad \text{for all } \mathbf{q} \in \mathfrak{P}_+(L)\ , \tag{19}$$

and let the open cone $\mathring{\mathfrak{C}}_+(L)$ be its interior. Using Plancherel's Theorem for discrete Fourier transforms, we have

$$\mathbf{c}^T\mathbf{q} = \frac{1}{L}\sum_{j=0}^{L-1} C(\zeta_j)Q(\zeta_j)\ , \tag{20}$$

where

$$C(e^{i\theta}) = \sum_{n=-N}^{N} c_n e^{in\theta}\ , \tag{21}$$

yielding an alternative formulation of (19). Since the moments of a periodic probability density have a positive Toeplitz matrix, the following proposition follows from Proposition 6 in Lindquist and Picci (2013).

Proposition 2. Let $\mathbf{c} = [c_1, c_2, \ldots, c_N]^T$ be the moments (6). Then there is an integer L_0 so that $\mathbf{c} \in \mathring{\mathfrak{C}}_+(L)$ for all $L \geq L_0$.

4.2 Point-mass approximation by convex optimization

We are now in the position to reformulate Theorems 3, 4, and 7 in Lindquist and Picci (2013) in the present setting.

Theorem 3. (Lindquist and Picci (2013)). Let $\mathbf{c} \in \mathring{\mathfrak{C}}_+(L)$. Then, for each $P \in \mathfrak{P}_+(L)$, there is a unique $\hat{Q} \in \mathfrak{P}_+(L)$ such that

$$p_L(\zeta) = \frac{P(\zeta)}{\hat{Q}(\zeta)} \tag{22}$$

satisfies the moment condition

$$\sum_{j=0}^{L-1} \zeta_j^k p_L(\zeta_j) = c_n \ , \quad n = 0, 1, \dots, N \ . \tag{23}$$

If $\hat{Q}$ has zeros on the unit circle, then P has the same zeros, and there is cancellation of zeros in (22). Moreover, $\hat{Q}$ is the unique minimum of the strictly convex functional

$$\mathbb{J}_P(\mathbf{q}) = \sum_{j=0}^{L-1} \left[C(\zeta_j) Q(\zeta_j) - P(\zeta_j) \log Q(\zeta_j) \right] \ , \tag{24}$$

where the variable Q is given by (12) and C by (21). Moreover,

$$\lim_{L\to\infty} p_L(\zeta) = \frac{P(\zeta)}{\tilde{Q}(\zeta)} \ , \tag{25}$$

where $\tilde{Q}$ is the unique minimum of the functional

$$\mathbf{J}_P(Q) = \int_{-\pi}^{\pi} \left[C(e^{i\theta}) Q(e^{i\theta}) - P(e^{i\theta}) \log Q(e^{i\theta}) \right] d\theta \ . \tag{26}$$

It is shown in Lindquist and Picci (2013) that the problem to minimize (24) is the dual problem of the primal problem to maximize the generalized entropy gain

$$\mathbb{I}_P(p) = \sum_{j=0}^{L-1} P(\zeta_j) \log p(\zeta_j) \tag{27}$$

subject to the moment condition

$$\sum_{j=1}^{L-1} \zeta_j^n p(\zeta_j) = c_n \ , \quad n = 1, 2, \dots, N \ .$$

The optimal solution of the primal problem is precisely (22), where Q is the optimal solution to the dual problem to minimize (24). Choosing $P = 1$, we obtain the *maximum-entropy solution.*

Remark 4. It easy to see that maximizing $\mathbb{I}_P$ is equivalent to minimizing the Kullback-Leibler divergence

$$\hat{\mathbb{D}}(P\|p) = \sum_{j=0}^{L-1} P(\zeta_j) \log \left[\frac{P(\zeta_j)}{p(\zeta_j)} \right] \tag{28}$$

with respect to p. Likewise, the problem to minimize (26) is the dual of the problem to minimize the continuous Kullback-Leibler divergence

$$\mathbb{D}(P\|p) = \int_{-\pi}^{\pi} P(e^{i\theta}) \log \left[\frac{P(e^{i\theta})}{p(e^{i\theta})} \right] d\theta \tag{29}$$

subject to the moment constraints in (6). (See, e.g., Georgiou and Lindquist (2003).)

If we are content with equidistantly placed Dirac points, we now immediately have a procedure to determine a solution to the approximation problem of Sec. 2. In fact, applying Theorem 3, we obtain the solution

$$\boldsymbol{\eta} = [w_1, \dots, w_L, \hat{\theta}_1, \dots, \hat{\theta}_L]^T \ , \tag{30}$$

where

$$w_j = \frac{P(\zeta_{j-1})}{Q(\zeta_{j-1})}, \quad \hat{\theta}_j = (j-1)\Delta \ . \tag{31}$$

In particular, we can use the maximum entropy solution setting $P = 1$.

4.3 A generalization of Theorem 3

Inspecting the proof of Theorem 3 in Lindquist and Picci (2013) and modifying it along the lines of Byrnes and Lindquist (2003, 2006), it is trivial to see that P need not belong to $\mathfrak{P}_+(N)$ but could, for example, be an arbitrary probability density.

Theorem 5. Let f be an arbitrary continuous probability density. Then, replacing P by f in Theorem 3, all statements of that theorem remain true. Moreover, Q_L tends to a limit Q_∞ as $L \to \infty$ and

$$\lim_{L\to\infty} p_L(\zeta) = p_\infty(\zeta) := \frac{f(\zeta)}{Q_\infty(\zeta)} \ . \tag{32}$$

The limit density p_∞ satisfies the moment conditions (6).

Proof. Exchanging P for f, a trivial modification of the proofs of Theorems 3, 4, and 7 in Lindquist and Picci (2013) proves the theorem.

Corollary 6. Let $\tilde{p}$ be a probability density with the moments $c_1, c_2, \dots, c_N$, and set $P := \tilde{p}$ everywhere in the statement of Theorem 3. Then

$$\lim_{L\to\infty} p_L(\zeta) = \tilde{p}(\zeta) \ . \tag{33}$$

Proof. The limit of p_L as $L \to \infty$ is obtained from (32) setting $f := \tilde{p}$. It remains to show that $Q_\infty \equiv 1$ in this case so that $p_\infty = \tilde{p}$, which will be done by means of the primal optimization problems of Remark 4. However, since p_∞ satisfies the moment conditions (6), the minimum of $\mathbb{D}(\tilde{p}\|p) \geq 0$ is zero, and hence the unique minimizer of the Kullback-Leibler divergence $\mathbb{D}(\tilde{p}\|p)$ subject to the moment constraints (6) is precisely $\tilde{p}$. Consequently, the optimal dual solution is $Q_\infty \equiv 1$, as claimed.

4.4 Determining nonequidistant circular Dirac mixtures

As explained in the introduction, in general we would like to place the Dirac points denser where the probability mass is larger and more sparse where it is small. Then the procedure above needs to be modified.

To this end, let M be the unique integer such that $L = 2M$ when L is even and $L = 2M + 1$ when L is odd, and let $\mathfrak{R}_+$ be the class of probability densities

$$R(e^{i\theta}) = \sum_{k=-M}^{M} r_k e^{ik\theta}, \quad r_{-k} = r_k \tag{34}$$

that are nonnegative on the discrete unit circle. By Theorem 5, we can replace $P \in \mathfrak{P}_+$ by $R \in \mathfrak{R}_+$ in the statement of Theorem 3. We now choose L sufficiently large and remove Dirac points by setting $R(\zeta_j) = 0$ in points where no point mass is desired. Solving the optimization problem of Theorem 3, we then obtain a solution (30) with $\hat{\theta}_k$ removed whenever $R(e^{i\hat{\theta}_k}) = 0$.

To construct such an R we choose a vector $\hat{\mathbf{r}} \in \mathbb{R}^{M+1}$, such that $\hat{r}_j = R(\zeta_j)$, $j = 0, 1, \dots, M$, in the following manner. Choose $\hat{r}_j = 0$ in positions where no Dirac point is needed and take the other $\hat{r}_j$ to be positive tuning parameters.

We fix only $M+1$ values for R on the circle, rather than L, since the symmetry condition $\zeta_{L-k} = \bar{\zeta}_k$ will then prescribe values for R in the remaining $L-M-1$ points. Then, analogously to (17), we obtain

$$\mathbf{r} := \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_M \end{bmatrix} = \frac{1}{M+1} \tilde{\mathbf{F}}^T \hat{\mathbf{r}} \ , \tag{35}$$

where $\tilde{\mathbf{F}}$ is the $(M+1) \times (M+1)$ matrix

$$\tilde{\mathbf{F}} = \begin{bmatrix} 1 & 2\cos(\hat{\theta}_1) & \cdots & 2\cos(M\hat{\theta}_1) \\ 1 & 2\cos(\hat{\theta}_2) & \cdots & 2\cos(M\hat{\theta}_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2\cos(\hat{\theta}_{M+1}) & \cdots & 2\cos(M\hat{\theta}_{M+1}) \end{bmatrix} \ . \tag{36}$$

Since $\tilde{\mathbf{F}}$ is a square Vandermonde matrix, it is nonsingular and

$$\tilde{\mathbf{F}}\tilde{\mathbf{F}}^T = \tilde{\mathbf{F}}^T\tilde{\mathbf{F}} = (M+1)\mathbf{I} \ , \tag{37}$$

where $\mathbf{I}$ is the $(M+1)\times(M+1)$ identity, so there is a one-one correspondence between $\mathbf{r}$ and $\hat{\mathbf{r}}$. Therefore inserting (35) into (34) yields a suitable R.

5. DISTANCE MEASURE

In this section, we assume that more Dirac components are available than required for satisfying the given moment constraints. This redundancy is exploited by continuously placing the Dirac components in such a way that the resulting circular Dirac mixture $p(\theta)$ is as close to the given continuous density $\tilde{p}(\theta)$ as possible. Closeness between the two densities, the given continuous density $\tilde{p}(\theta)$ and its approximation $p(\theta)$, is quantified by a distance measure.

For the *direct* comparison of continuous densities and Dirac mixture densities, typical distance measures such as the Kullback-Leibler distance or squared integral distances are not well defined (Hanebeck and Klumpp, 2008). An alternative in the case of linear quantities is to compare cumulative distributions instead of densities. Doing so is, however, complicated in the circular case. In this paper, we adapt the method proposed for linear quantities in Hanebeck (2014) to the circular case. This approach is inspired by blue-noise sampling methods, e.g., see Schlömer et al. (2011). Here, we use kernels for describing the mass of individual Dirac components as in Fattal (2011) instead of discs.

The key idea of this paper is the following. Individual Dirac components of the circular Dirac mixture $p(\theta)$ are characterized by repulsion kernels. A repulsion kernel is an appropriate kernel function representing the spread of probability mass that is required to approximate the given density at a given location. The individual repulsion kernels are summed up to form an induced kernel density, which is then a continuous representation of the circular Dirac mixture $p(\theta)$. For obtaining a distance measure, the deviation between the given density and the induced kernel density is considered. In this paper, we use a squared integral distance, which is minimized by means of an optimization method in order to find the parameters of the circular Dirac mixture $p(\theta)$.

5.1 Using von Mises distribution kernels

For characterizing each Dirac component j located at $\hat{\theta}_j$, we use a von Mises distribution normalized to the height of the given density at the location of the Dirac component. For that purpose, we take an unnormalized von Mises distribution (see Sec. 6) according to $e^{\kappa_j \cos(\theta-\hat{\theta}_j)}$ located at $\hat{\theta}_j$. Multiplying it with the height of the given density at the Dirac component location $\tilde{p}(\hat{\theta}_j)$ gives the desired kernel functions

$$k_j(\theta, w_j, \hat{\theta}_j) = \tilde{p}(\hat{\theta}_j) \, \frac{e^{\kappa_j \cos(\theta-\hat{\theta}_j)}}{e^{\kappa_j}} \tag{38}$$

for $j = 1, \ldots, N$, where κ_j is a spread parameter to be determined. Dividing it by the normalization term e^{κ_j} ensures that the height at the component location $\hat{\theta}_j$ is equal to the given density at that point. Please note that κ_j will depend upon the weight w_j.

The constants κ_j in (38) are determined by maintaining the mass in the Dirac component given by w_j, so we obtain

$$\int_{-\pi}^{\pi} k_j(\theta) \ \mathrm{d}\theta = w_j \ , \tag{39}$$

which gives

$$2\pi \, \tilde{p}(\hat{\theta}_j) \, e^{-\kappa_j} \, I_0(\kappa_j) = w_j \ , \tag{40}$$

where $I_0(.)$ is the modified Bessel function of the first kind and zeroth order. For obtaining κ_j, $j = 1, \ldots, N$, (40) is rewritten as

$$f(\kappa_j) = 2\pi \, e^{-\kappa_j} \, I_0(\kappa_j) - \frac{w_j}{\tilde{p}(\hat{\theta}_j)} \tag{41}$$

and we desire to find the zero κ_j^0 of the function, i.e., $f(\kappa_j^0) = 0$. For that purpose, a Newton–like method is used with

$$\frac{\partial \, f(\kappa_j)}{\partial \, \kappa_j} = 2\pi \, e^{-\kappa_j} \ (I_1(\kappa_j) - I_0(\kappa_j)) \ . \tag{42}$$

The resulting kernels have a larger width in low-density regions of the given continuous density, as these regions are characterized with few Dirac components per unit length. In high-density areas, the kernels have a smaller widths, which results in more Dirac components per unit length.

The density induced by the kernels in (38) is given by

$$k(\theta, \boldsymbol{\eta}) = \sum_{j=1}^{L} k_j(\theta, w_j, \hat{\theta}_j) \ . \tag{43}$$

This induced kernel density $k(\theta, \boldsymbol{\eta})$ is now used to characterize the circular Dirac mixture $p(\theta)$. We assume that when the mass distributions of $\tilde{p}(\theta)$ and the circular Dirac mixture $p(\theta)$ are similar, then $k(\theta)$ is close to $\tilde{p}(\theta)$ on Γ and we consider the densities $\tilde{p}(\theta)$ and $p(\theta)$ to be close. Following this argument, $k(\theta)$ is used instead of the circular Dirac mixture $p(\theta, \boldsymbol{\eta})$ to define a distance measure $D(\boldsymbol{\eta})$ between $\tilde{p}(\theta)$ and $p(\theta, \boldsymbol{\eta})$ on Γ.

Here, we use a squared integral distance measure given by

$$D(\boldsymbol{\eta}) = \int_{\Gamma} \left(\tilde{p}(\theta) - k(\theta, \boldsymbol{\eta})\right)^2 \mathrm{d}\theta \ . \tag{44}$$

As the final optimization problem, we now desire to minimize the distance measure in (44) while maintaining the moment constraints, i.e.,

$$\begin{aligned} \boldsymbol{\eta}_{\text{opt}} &= \arg\min_{\boldsymbol{\eta} \in \mathcal{S}} d(\boldsymbol{\eta}) \\ \text{s.t. } \mathrm{E}_p\left\{\cos(n\,\theta)\right\} &= c_n \text{ for } n = 1, \ldots, N \ . \end{aligned} \tag{45}$$

Minimization of the distance measure under the moment constraints is nonlinear and nonconvex. Applying standard

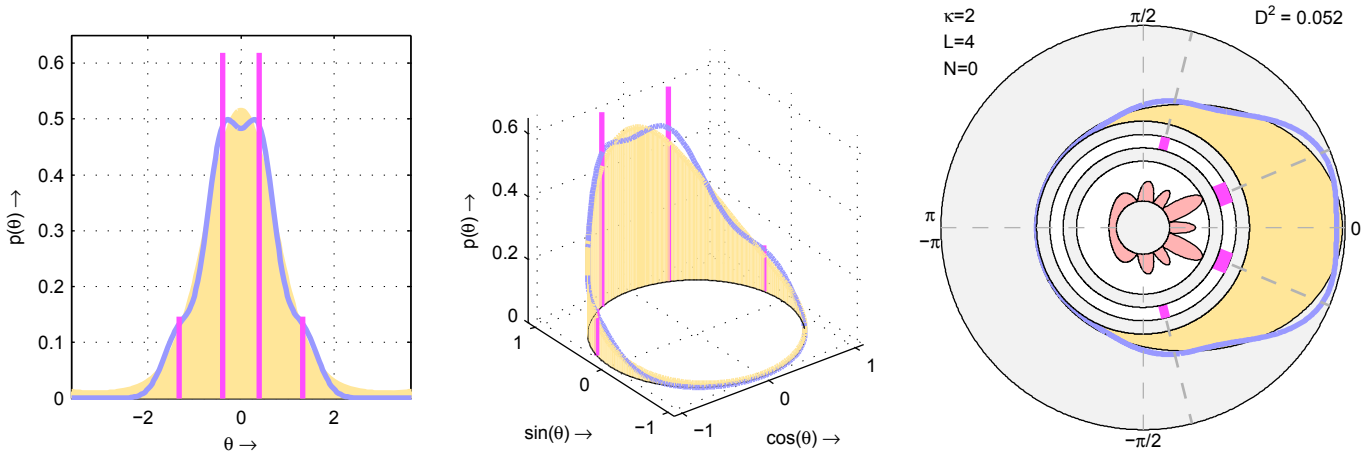

Fig. 1. Graphical visualizations for a symmetric Dirac mixture approximation of a von Mises distribution with $L = 4$ components and no moment constraints (N=0) for $\kappa = 2$. (Left) Standard plot. (Middle) Circular plot in three dimensions. (Right) Circular plot in two dimensions.

numerical minimization methods to (45) will most likely end up in local minima and the results depend on the starting values for the parameter vector. As this is undesired, we propose the use of a homotopy continuation method (Allgower and Georg, 2003). In doing so, we start with a density with a known approximation and gradually approach the desired density. For that purpose, a progression parameter γ is defined for parametrizing the given density as

$$\tilde{p}(\theta, \gamma) = \frac{\tilde{p}^{\gamma}(\theta)}{\int_{-\pi}^{\pi} \tilde{p}^{\gamma}(\theta)\, \mathrm{d}\theta} \quad . \tag{46}$$

We start with $\gamma = 0$, which gives us a uniform density on the circle, i.e., we have

$$\tilde{p}(\theta, \gamma = 0) = \frac{1}{2\pi} \quad . \tag{47}$$

The Dirac mixture approximation for this density is known and given by

$$w_j = \frac{1}{L} \text{ and } \theta_j = \frac{\pi}{L}(2j - L - 1) \tag{48}$$

for j=1, . . . , L.

We now progressively increase the parameter γ and for every γ solve the optimization problem by using the result from the previous step as starting values, which ensures tracking the desired solution, see Alg. 1. No predictor is used. As a corrector in line 7, we use the function `fmincon` for constrained optimization in our MATLAB implementation. A maximum amount of iterations is prespecified and success is reported (SuccessFlag = 1) when the optimum for the next value of γ can be found within these iterations. Otherwise, SuccessFlag = 0.

6. AN APPROXIMATION EXAMPLE

As an example of a continuous circular density, we consider the von Mises distribution (Mardia and Jupp, 2009) given by

$$p(\theta, \mu, \tilde{\kappa}) = \frac{1}{2\pi I_0(\tilde{\kappa})} e^{\tilde{\kappa}\, \cos(\theta - \mu)} \quad , \tag{49}$$

where μ is the location parameter, $\tilde{\kappa}$ is the concentration parameter, and we have $\mu \in \Gamma$ and $\tilde{\kappa} > 0$. $I_0(\cdot)$ is the modified Bessel function of the first kind and zeroth order.

The circular moments of the von Mises density are given by

$$\mathrm{E}_p\left\{e^{in\theta}\right\} = \frac{I_n(\tilde{\kappa})}{I_0(\tilde{\kappa})}\, e^{in\mu} \tag{50}$$

for $n \in \mathbb{N}$. $I_n(\cdot)$ is the modified Bessel function of the first kind and nth order.

The given circular probability density function $\tilde{p}(\theta)$ is approximated with a circular Dirac mixture $p(\theta, \boldsymbol{\eta})$ defined according to (8). Both weights w_j and locations $\hat{\theta}_j$ are optimized for $j = 1, \ldots, L$.

The approximation problem can now be simplified by noting that it is sufficient to consider a specific fixed location value μ. Without loss of generality, we use $\mu = 0$. The Dirac mixture approximation calculated for this location value can later be shifted to arbitrary location values by shifting each component accordingly. As a result, the true circular moments are given by

$$c_n = \frac{I_n(\tilde{\kappa})}{I_0(\tilde{\kappa})} \tag{51}$$

for $n \in \mathbb{N}$.

In the case of von Mises distributions, the progressive parametrization is given by

$$\tilde{p}(\theta, \gamma) = \frac{1}{2\pi I_0(\gamma\tilde{\kappa})} e^{\gamma\tilde{\kappa}\, \cos(\theta - \mu)} \quad . \tag{52}$$

For a von Mises distribution with $\tilde{\kappa} = 2$, a symmetric Dirac mixture approximation with $L = 4$ components and no moment constraints (N=0) is visualized with three different types of plots in Fig. 1. The given density is always shown in yellow, the induced kernel density corresponding to its Dirac mixture approximation in blue. On the left side, a standard linear plot of the periodic probability density functions is shown. The middle plot shows a circular plot in three dimensions. The right graph shows a dedicated circular plot in two dimensions to compactly represent the results of the simulations. The outermost ring shows the given circular probability density function in yellow and the induced kernel density corresponding to its Dirac mixture approximation in blue. The middle ring shows the locations of the Dirac components in purple, where the width is proportional to the associated weight.

Dirac mixture approximation

Input : Given density $\tilde{p}$, number of Dirac components L, number of moments N to maintain

Output : Dirac mixture approximation with weights w_j and locations $\hat{\theta}_j$, $j = 1, \ldots, L$ collected in parameter $\boldsymbol{\eta}$ according to (10)

```
// Initialize progression step counter
PC = 0;
// Initialize progression parameter γ
γ = 0;
// Initialize Δγ
Δγ = ε (some small positive number);
// Parameters for in-/decreasing Δγ
Down = 0.5 , Up = 1.5;
// Initial parameter vector for γ = 0
Initialize η with w_j and θ̂_j, j = 1, ..., L from (48);
while γ < 1 do
    // Try correcting values for increased γ
    [η_tmp, SuccessFlag] = Corrector(η, γ + Δγ);
    if Correction step successful? (SuccessFlag==1?) then
        // Make trial update the temporary estimate
        η = η_tmp;
        // Increment γ
        γ = γ + Δγ;
        // Increase step size
        Δγ = Up · Δγ;
        // Increment progression step counter PC
        PC = PC + 1;
    else
        // Decrease step size
        Δγ = Down · Δγ;
    end
    // Limit γ to [0, 1]
    if γ + Δγ > 1 then
        Δγ = 1 − γ;
    end
end
```

Algorithm 1: Dirac mixture approximation of given continuous circular probability density function.

The innermost ring shows the error between the given circular probability density function and the induced kernel density corresponding to its Dirac mixture approximation normalized to one in red.

Circular plots of the Dirac mixture approximation of a von Mises distribution with $L = 6$ and $L = 8$ components and no moment constraints (N=0) for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$ are shown in Fig. 2 and Fig. 3, respectively. It is obvious that for a larger number of components L, the error between the given density and its Dirac mixture approximation decreases. For $N = L - 2$ moment constraints, circular plots of the Dirac mixture approximation of a von Mises distribution with $L = 6$ and $L = 8$ components are shown for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$ in Fig. 4 and Fig. 5, respectively. Here, the errors are always a little larger than in the unconstrained case. However, in all cases, the Dirac mixture approximations are well-distributed and approximation quality is very good.

7. CONCLUSION

The first proposed method for calculating the weights for a Dirac mixture approximation with fixed component locations has not been simulated so far. Simulations for the second proposed approximation method are shown in the paper. It reliably provides well-distributed Dirac mixture approximations of given circular densities while exactly maintaining a set of predefined circular moments. It will be interesting to compare the performance of the two proposed methods. This will be the topic of a future study.

The next step is the generalization to higher-dimensional periodic manifolds such as tori or spheres.

ACKNOWLEDGEMENTS

The authors would like to thank Gerhard Kurz and Igor Gilitschenski from the Intelligent Sensor-Actuator-Systems Laboratory (ISAS) at the Karlsruhe Institute of Technology (KIT) for their helpful comments.

REFERENCES

Allgower, E.L. and Georg, K. (2003). *Introduction to Numerical Continuation Methods*. SIAM.

Azmani, M., Reboul, S., Choquel, J.B., and Benjelloun, M. (2009). A recursive fusion filter for angular data. In *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 882–887. doi: 10.1109/ROBIO.2009.5420492.

Byrnes, C.I. and Lindquist, A. (2006). The generalized moment problem with complexity constraint. *Integral Equations and Operator Theory*, 56, 163–180.

Byrnes, C., Gusev, S., and Lindquist, A. (1999). A convex optimization approach to the rational covariance extension problem. *SIAM J. Control and Optmimization*, 37, 211–229.

Byrnes, C. and Lindquist, A. (2003). A convex optimization approach to generalized moment problems. In K. Hashimoto, Y. Oishi, and Y. Yamamoto (eds.), *Control and Modeling of Complex Systems: Cybernetics in the 21st Century: Festschrift in Honor of Hidenori Kimura on the Occasion of his 60th Birthday*, 3–21. Birkhauser.

Byrnes, C., Lindquist, A., Gusev, S., and Matveev, A. (1995). A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Automatic Control*, 40, 1841–1857.

Carli, F., Ferrante, A., Pavon, M., and Picci, G. (2011). A maximum entropy solution of the covariance extension problem for reciprocal processes. *IEEE Trans. Automatic Control*, 56, 1999–2012.

Carli, F.P. and Georgiou, T. (2011). On the covariance completion problem under a circulant structure. *IEEE Trans. Automatic Control*, 56(1), 918 –922.

Fattal, R. (2011). Blue-noise point sampling using kernel density model. In *ACM SIGGRAPH 2011 papers*, SIGGRAPH '11, 48:1–48:12. ACM, New York, NY, USA. doi:10.1145/1964921.1964943. URL `http://doi.acm.org/10.1145/1964921.1964943`.

Georgiou, T.T. and Lindquist, A. (2003). Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Information Theory*, 49, 2910–2917.

Georgiou, T. (1987). Realization of power spectra from partial covariances. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 35, 438–449.

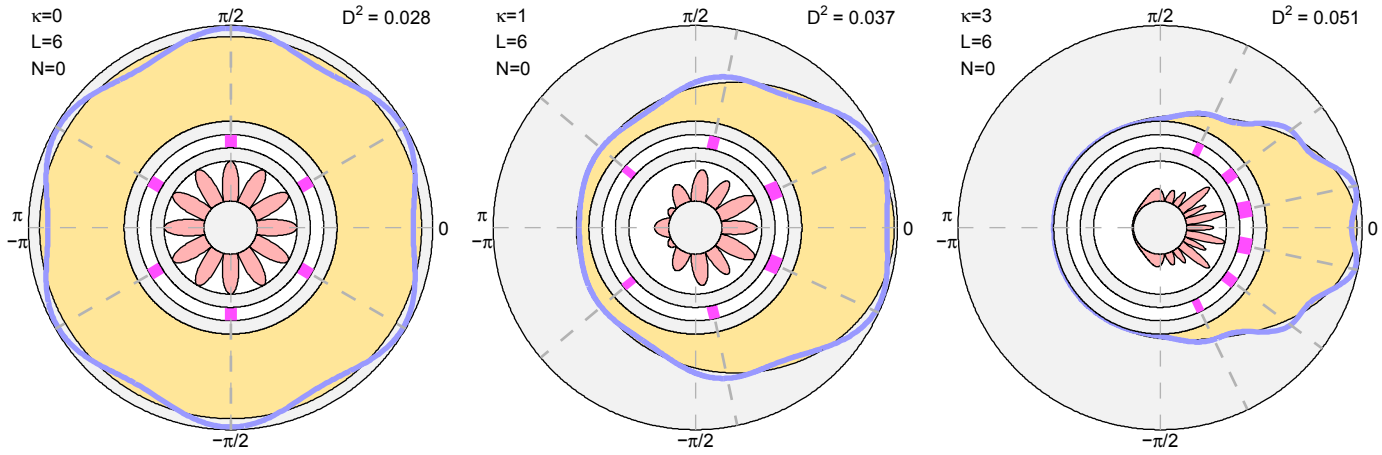

Fig. 2. Circular plot of Dirac mixture approximation of von Mises distribution with $L = 6$ components and no moment constraints (N=0) for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$.

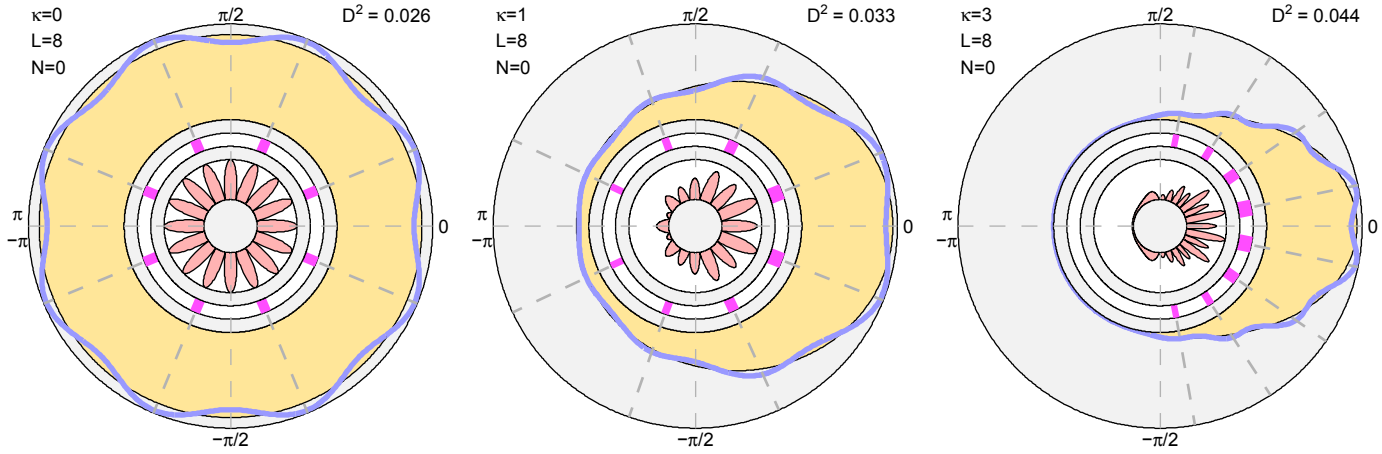

Fig. 3. Circular plot of Dirac mixture approximation of von Mises distribution with $L = 8$ components and no moment constraints (N=0) for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$.

Gilitschenski, I. and Hanebeck, U.D. (2013). Efficient Deterministic Dirac Mixture Approximation. In *Proceedings of the 2013 American Control Conference (ACC 2013)*. Washington D. C., USA.

Gilitschenski, I., Kurz, G., and Hanebeck, U.D. (2013). Circular Statistics in Bearings-only Sensor Scheduling. In *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*. Istanbul, Turkey.

Hanebeck, U.D. (2014). Kernel-based Deterministic Blue-noise Sampling of Arbitrary Probability Density Functions. In *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS 2014)*. Princeton, New Jersey, USA.

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 Schrempf, O.C. (2007). Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions. In *Proceedings of the 2007 IEEE Conference on Decision and Control (CDC 2007)*, 3065–3071. New Orleans, Louisiana, USA.

Huber, M.F. and Hanebeck, U.D. (2008). Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation. In *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, volume 17. Seoul, Republic of Korea.

Julier, S.J. (2002). The scaled unscented transformation. In *Proceedings of the 2002 IEEE American Control Conference (ACC 2002)*, volume 6, 4555– 4559. Anchorage, Alaska, USA. doi:10.1109/ACC.2002.1025369.

Julier, S.J. and Uhlmann, J.K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3), 401– 422. doi:10.1109/JPROC.2003.823141.

Krein, M. and Nudelman, A. (1977). *The Markov Moment Problem and Extremal Problems*. American Mathematical Society, Providence, Rhode Island.

Krener, A. (1986). Reciprocal processes and the stochastic realization problem for acausal systems. In C.I. Byrnes and A. Lindquist (eds.), *Modelling, Identification and Robust Control*, 197–211. North-Holland, Amsterdam.

Kurz, G., Faion, F., and Hanebeck, U.D. (2013a). Constrained Object Tracking on Compact One-dimensional Manifolds Based on Directional Statistics. In *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)*. Montbeliard, France.

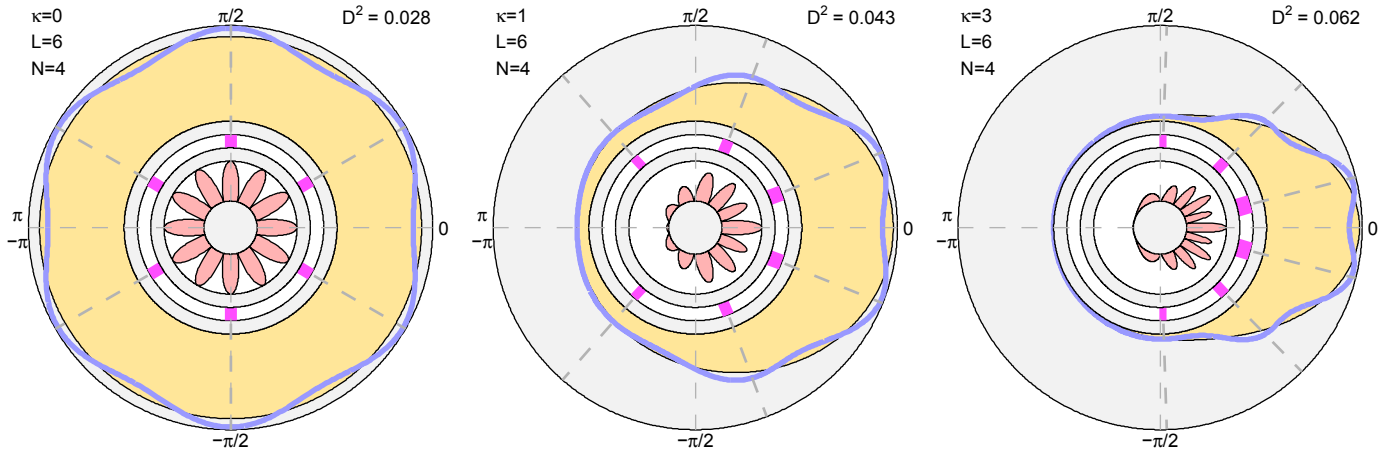

Fig. 4. Circular plot of Dirac mixture approximation of von Mises distribution with $L = 6$ components and $N = 4$ moment constraints for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$.

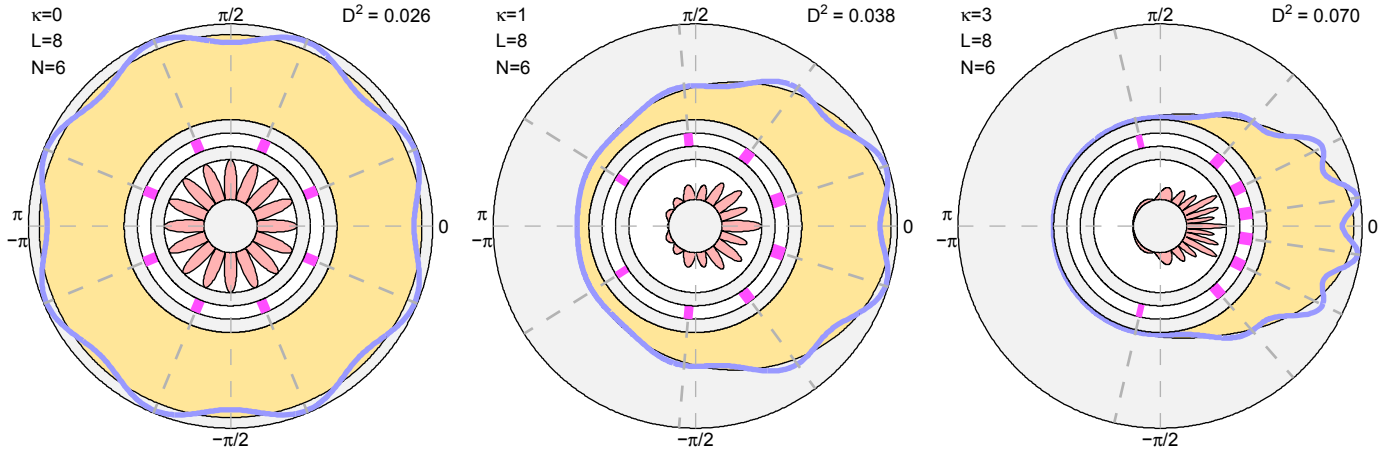

Fig. 5. Circular plot of Dirac mixture approximation of von Mises distribution with $L = 8$ components and $N = 6$ moment constraints for (left) $\kappa = 0$, (middle) $\kappa = 1$, (right) $\kappa = 3$.

Kurz, G., Gilitschenski, I., and Hanebeck, U.D. (2013b). Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions. In *Proceedings of the 2013 American Control Conference (ACC 2013)*. Washington D. C., USA.

Lefebvre, T., Bruyninckx, H., and De Schutter, J. (2005). The linear regression kalman filter. In *Nonlinear Kalman Filtering for Force-Controlled Robot Tasks*, volume 19 of *Springer Tracts in Advanced Robotics*.

Levy, B.C., Frezza, R., and Krener, A.J. (1990). Modeling and estimation of discrete-time gaussian reciprocal processes. *IEEE Trans. Automatic Control*, 35(9), 1013–1023.

Lindquist, A. and Picci, G. (2013). The circulant rational covariance extension problem: The complete solution. *IEEE Trans. Automatic Control*, 58(11), 2848–2861.

Mardia, K.V. and Jupp, P.E. (2009). *Directional Statistics*. John Wiley & Sons.

Schlömer, T., Heck, D., and Deussen, O. (2011). Farthest-point optimized point sets with maximized minimum distance. In *Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics*, HPG '11, 135–142. ACM, New York, NY, USA. doi:10.1145/2018323.2018345. URL `http://doi.acm.org/10.1145/2018323.2018345`.

Schrempf, O.C., Brunn, D., and Hanebeck, U.D. (2006a). Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering. In *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*. San Diego, California, USA.

Schrempf, O.C., Brunn, D., and Hanebeck, U.D. (2006b). Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance. In *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, 512–517. Heidelberg, Germany.

Schrempf, O.C. and Hanebeck, U.D. (2007). Recursive Prediction of Stochastic Nonlinear Systems Based on Optimal Dirac Mixture Approximations. In *Proceedings of the 2007 American Control Conference (ACC 2007)*, 1768–1774. New York, New York, USA.

Stienne, G., Reboul, S., Azmani, M., Choquel, J., and Benjelloun, M. (2014). A multi-temporal multi-sensor circular fusion filter. *Information Fusion*, 18, 86–100. doi:10.1016/j.inffus.2013.05.012.

Tenne, D. and Singh, T. (2003). The higher order unscented filter. In *Proceedings of the 2003 IEEE American Control Conference (ACC 2003)*, volume 3, 2441–2446. Denver, Colorado, USA. doi:10.1109/ACC.2003.1243441.