

# Circular Discrete Reapproximation

Kailai Li, Florian Pfaff, and Uwe D. Hanebeck

Intelligent Sensor-Actuator-Systems Laboratory (ISAS)

Institute for Anthropomatics and Robotics

Karlsruhe Institute of Technology (KIT), Germany

{kailai.li, pfaff, uwe.hanebeck}@kit.edu

**Abstract**—We present a novel nonparametric scheme for modeling circular random variables. For that, the circular Cramér–von Mises distance (CCvMD) is proposed to measure the statistical divergence between two circular discrete models based on a smooth characterization of the localized cumulative distribution. Given a set of weighted samples from empirical data, the underlying unknown distribution is then reapproximated by another sample set of configurable size and dispersion-adaptive layout in the sense of least CCvMD. Built upon the proposed circular discrete reapproximation (CDR), a new method is introduced for density estimation with von Mises mixtures. Moreover, the CDR scheme is extended to topological spaces composing the unit circle and Euclidean space of arbitrary dimension, and a new regression model for random circular vector fields is proposed based thereon. We provide case studies using synthetic and real-world data from wind climatology. Numerical results validate the efficacy of proposed approaches with promising potential of outperforming competitive methods.

**Index Terms**—State estimation and inference, directional statistics, nonparametric probabilistic modeling, regression model

## I. INTRODUCTION

Random variables in circular domains appear in various application scenarios ranging among human-machine interaction, sound source localization, meteorology, bioinformatics, and ecology [1]–[5]. Performing state estimation and inference in this scope depends heavily on the representation fidelity and efficiency of the deployed circular probabilistic models.

High-fidelity modeling of uncertainty on the unit circle is nontrivial. Common probabilistic models in Euclidean space (e.g., normal distribution) cannot be directly applied to circular domain due to its periodic and nonlinear topological structure. Conventional strategy interprets directional uncertainty in a locally linearized space based on local perturbation assumption. In many related scenarios, however, this strategy is likely to be invalidated by highly uncertain or dynamic systems [6].

In consideration of the aforementioned issues, recent effort has been dedicated to applying topology-aware probabilistic models to directional estimation and inference, in particular, by using distributions from directional statistics [7]. Unlike adapted models from Euclidean space, they are inherently defined on circular or (hyper-)spherical manifolds. Popular probabilistic models on the unit circle  $\mathbb{S}^1$  include the wrapped normal distribution, the von Mises distribution, and the Bingham distribution [8]. Various statistical methods, e.g., deterministic sampling and parameter estimation, have been established to facilitate tackling related tasks, in particular, recursive Bayesian estimation of directional random variables [9].

Though directional statistics provides a theoretically sound means for modeling directional uncertainty, there still exist several limitations. For instance, most probabilistic density functions (PDFs) are defined on directional manifolds of simple structures (e.g., circles or spheres). In real-world tasks, however, random variables often belong to domains with extra complex topological structures, such as antipodal symmetry (on unit quaternion manifold) and composition with other Euclidean or non-Euclidean space [10], [11]. The former can be handled by certain PDFs with antipodally symmetric dispersion (e.g., Bingham or Watson distributions). The latter refers to *composite directional domains*, where choices of off-the-shelf parametric models are fairly limited w.r.t. topological variety.

One major theoretical obstacle therein lies in the interpretation of correlation across domain components, for which solutions are often domain-specific. On the torus  $\mathbb{S}^1 \times \mathbb{S}^1$ , the bivariate wrapped normal distribution and the bivariate von Mises distribution were exploited for recursive estimation [12]. An antipodally symmetric distribution was proposed on  $\mathbb{S}^1 \times \mathbb{R}^2$  in [13] for estimating planar transformations using the dual quaternion representation [14]. Based on hyperspherical parallel transport, a topology-aware modeling scheme was introduced in [11] on the manifold of unit dual quaternions representing generic rigid transformations.

PDFs from directional statistics have parametric forms, which may largely violate the uncertain nature of randomness with multi-modal or complex dispersion. Given that directional domains are compact and bounded, one solution is to discretize the state space using a grid of a certain layout. Consequently, the underlying uncertainty can be approximated by Dirac mixtures or piecewise constant distributions at grid points [15]–[18]. An extension to the composite directional domain  $\mathbb{S}^1 \times \mathbb{R}^2$  was provided in [19] via Rao–Blackwellization for planar motion estimation. Due to the fixed grid layout, high-fidelity approximation of arbitrary dispersion can only be achieved by increasing grid resolutions, which may lead to runtime and memory inefficiency in common estimation and inference tasks [16]. Such issues also arise when exploiting discrete models based on random samples [20], [21], although their adaptiveness allows for handling uncertainty of arbitrary form.

One promising methodology to improve the representation efficiency of discrete models is *reapproximation*: Given a large set of weighted samples for density approximation, a smaller set of weighted samples is produced with deterministic and

dispersion-adaptive locations to reapproximate the underlying distribution. This methodology was first introduced in [22] on the planar dual quaternion manifold followed by a variant on arbitrary dimensions of unit (hyper-)spheres  $\mathbb{S}^{d-1}$  ( $d \geq 3$ ) in [23]. Both variants were applied to on-manifold recursive estimation of corresponding directional quantities, delivering superior performance over filtering schemes using random samples.

### Contributions

Following the generic design of on-manifold discrete reapproximation, we introduce its variant on the unit circle  $\mathbb{S}^1$  for efficient discrete probabilistic modeling of arbitrary circular distributions (Sec. II). The proposed circular discrete reapproximation (CDR) method further enables a new approach for circular density estimation using von Mises mixtures (Sec. III). Moreover, we extend the CDR method to composite circular domains  $\mathbb{S}^1 \times \mathbb{R}^d$  incorporating Euclidean space of arbitrary dimensions, leading to a novel discrete regression model on random vector field of circular quantities (Sec. IV). We validate the efficacy of the proposed methods with case studies using synthetic and real-world data. A few advantages over competitive methods are shown.

## II. DISCRETE REAPPROXIMATION ON THE UNIT CIRCLE

The on-manifold discrete reapproximation methodology was pioneered by the work in [24] for sample reduction in Euclidean space. Later, a theoretically sound and generic design for discrete modeling on directional manifolds was progressively established through [22] and [23]. In this section, we customize this methodology to the circular domain. We first introduce the so-called circular localized cumulative distribution with a topology-aware kernel for smooth characterization of Dirac mixtures on  $\mathbb{S}^1$ . Then, a circular variant of the Cramér–von Mises distance (CCvMD) is proposed to quantify the statistical divergence between two discrete models on  $\mathbb{S}^1$ . Given a set of weighted samples as the *source*, a *template* discrete model of fewer and dispersion-adaptive supports is produced in the sense of least CCvMD.

### A. Circular Localized Cumulative Distribution

Given a large set of samples  $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_k\}_{k=1}^{\hat{n}} \subset \mathbb{S}^1 \subset \mathbb{R}^2$  and their weights  $\hat{\mathbf{W}} = \{\hat{w}_k\}_{k=1}^{\hat{n}}$  holding  $\sum_{k=1}^{\hat{n}} \hat{w}_k = 1$ , the underlying unknown distribution can be approximated by a discrete model in the form of Dirac mixture

$$\hat{f}(\underline{x}) = \sum_{k=1}^{\hat{n}} \hat{w}_k \delta(\underline{x} - \hat{\mathbf{x}}_k). \quad (1)$$

$\delta(\cdot)$  denotes the Dirac delta function. Such a discrete approximation using large amount of raw, e.g., random, samples lacks representation efficiency for most empirical inference tasks. Thus, we take the discrete model in (1) as the *source* and aim to produce a *template* Dirac mixture with fewer supports at more representative locations  $\{\underline{x}_i\}_{i=1}^n \subset \mathbb{S}^1 \subset \mathbb{R}^2$  following

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

Here,  $\{w_i\}_{i=1}^n$  are the template sample weights that sum to one, i.e.,  $\sum_{i=1}^n w_i = 1$ .

Comparing two circular discrete models is theoretically infeasible using standard distance measures because they do not share common supports. To achieve smooth characterization using discrete models, we establish the localized cumulative distribution [24] on the unit circle defined as below.

**Definition 1** Suppose a random variable  $\underline{x} \in \mathbb{S}^1$  follows a PDF  $f : \mathbb{S}^1 \rightarrow \mathbb{R}^+$ . The circular localized cumulative distribution (CLCD) of  $f(\underline{x})$  is defined as

$$\mathcal{F}(\underline{\alpha}, \tau) = \int_{\mathbb{S}^1} f(\underline{x}) \kappa(\underline{x}; \underline{\alpha}, \tau) d\underline{x}, \quad (3)$$

with  $\kappa(\underline{x}; \underline{\alpha}, \tau) = \exp(\tau \underline{\alpha}^\top \underline{x})$  being the kernel located at  $\underline{\alpha} \in \mathbb{S}^1$  of concentration  $\tau > 0$ .

By applying the definition above to (1) and (2), we derive source and template CLCDs as

$$\begin{aligned} \hat{\mathcal{F}}(\underline{\alpha}, \tau) &= \int_{\mathbb{S}^1} \sum_{k=1}^{\hat{n}} \hat{w}_k \delta(\underline{x} - \hat{\mathbf{x}}_k) \kappa(\underline{x}; \underline{\alpha}, \tau) d\underline{x} \\ &= \sum_{k=1}^{\hat{n}} \hat{w}_k \exp(\tau \underline{\alpha}^\top \hat{\mathbf{x}}_k) \quad \text{and} \end{aligned} \quad (4)$$

$$\begin{aligned} \mathcal{F}(\underline{\alpha}, \tau) &= \int_{\mathbb{S}^1} \sum_{i=1}^n w_i \delta(\underline{x} - \underline{x}_i) \kappa(\underline{x}; \underline{\alpha}, \tau) d\underline{x} \\ &= \sum_{i=1}^n w_i \exp(\tau \underline{\alpha}^\top \underline{x}_i), \end{aligned} \quad (5)$$

respectively. As shown above, the kernel function takes the form of the unnormalized von Mises distribution, leading to a topology-aware and smooth characterization of the underlying distribution.

### B. Circular Cramér–von Mises Distance

The next step is to measure the statistical divergence between two circular discrete models with their CLCDs. For that, we adapt the Cramér–von Mises distance to the unit circle in a similar fashion to [23] as follows.

**Definition 2** Suppose  $\mathcal{F}$  and  $\hat{\mathcal{F}}$  are CLCDs of two discrete models  $f$  and  $\hat{f}$  on  $\mathbb{S}^1$ , respectively, their circular Cramér–von Mises distance (CCvMD) is

$$\mathcal{D}(\mathcal{F}, \hat{\mathcal{F}}) = \int_{\mathbb{R}^+} \mathcal{V}(\tau) \int_{\mathbb{S}^1} (\mathcal{F}(\underline{\alpha}, \tau) - \hat{\mathcal{F}}(\underline{\alpha}, \tau))^2 d\underline{\alpha} d\tau, \quad (6)$$

with  $\mathcal{V}(\tau) = \exp(-\epsilon\tau)$  being the weighting function.

Given the template discrete model of size  $n$  and weights  $\{w_i\}_{i=1}^n$  in (2), sample locations  $\mathbf{X} = \{\underline{x}_i\}_{i=1}^n$  can be obtained by minimizing its CCvMD to the source, i.e.,

$$\mathbf{X}^* = \arg \min_{\mathbf{X} \subset \mathbb{S}^1} \mathcal{D}(\mathcal{F}, \hat{\mathcal{F}}), \quad (7)$$

with  $\hat{\mathcal{F}}$  and  $\mathcal{F}$  being the source and template CLCDs in (4) and (5), respectively. We decompose the objective function in (7) into

$$\mathcal{D}(\mathcal{F}, \hat{\mathcal{F}}) =: \mathcal{D}_1(\mathcal{F}, \mathcal{F}) - 2\mathcal{D}_2(\mathcal{F}, \hat{\mathcal{F}}) + \mathcal{D}_3(\hat{\mathcal{F}}, \hat{\mathcal{F}}),$$

with  $\mathcal{D}_1$ ,  $\mathcal{D}_2$  and  $\mathcal{D}_3$  following the general expression of  $\int_{\mathbb{R}_+} \mathcal{V}(\tau) \int_{\mathbb{S}^1} \mathcal{F}_1(\underline{\alpha}, \tau) \mathcal{F}_2(\underline{\alpha}, \tau) d\underline{\alpha} d\tau$ . Here,  $\mathcal{F}_1$  and  $\mathcal{F}_2$  denote CLCDs of two arbitrary circular discrete models. Thus, the last term  $\mathcal{D}_3(\hat{\mathcal{F}}, \hat{\mathcal{F}})$  is constant given the source model and does not contribute to optimization. The rest terms follow

$$\begin{aligned} \mathcal{D}_1(\mathcal{F}, \mathcal{F}) &= \sum_{i=1}^n \sum_{r=1}^n w_i w_r \mathcal{Q}(\underline{x}_i, \underline{x}_r), \\ \mathcal{D}_2(\mathcal{F}, \hat{\mathcal{F}}) &= \sum_{i=1}^n \sum_{k=1}^{\hat{n}} w_i \hat{w}_k \mathcal{Q}(\underline{x}_i, \hat{\underline{x}}_k), \end{aligned} \quad (8)$$

and substitute  $\mathcal{Q}$  is computed according to the formula

$$\mathcal{Q}(\underline{u}, \underline{v}) = \int_{\mathbb{R}_+} \mathcal{V}(\tau) \int_{\mathbb{S}^1} \kappa(\underline{u}; \underline{\alpha}, \tau) \kappa(\underline{v}; \underline{\alpha}, \tau) d\underline{\alpha} d\tau, \quad (9)$$

where kernel values are measured at arbitrary sample locations  $\underline{u}, \underline{v} \in \mathbb{S}^1$ . By incorporating the von Mises-like kernel defined in (3) and the weighting function in (6), we then obtain

$$\begin{aligned} \mathcal{Q}(\underline{u}, \underline{v}) &= \int_{\mathbb{R}_+} \mathcal{V}(\tau) \int_{\mathbb{S}^1} \exp(\tau \underline{\alpha}^\top (\underline{u} + \underline{v})) d\underline{\alpha} d\tau \\ &= 2\pi \int_{\mathbb{R}_+} \exp(-\epsilon\tau) \mathcal{I}_0(\tau(2 + 2\underline{u}^\top \underline{v})) d\tau. \end{aligned}$$

Integration over kernel locations induces the normalization constant of the von Mises distribution, and  $\mathcal{I}_0$  denotes the modified Bessel function of the first kind and zeroth order. Further, the integral over kernel concentration  $\tau$  refers to the Laplace transform of function  $\mathcal{I}_0(2 + 2\underline{u}^\top \underline{v})$ . Following the formula in [25, Sec. 17.13.109], we have

$$\mathcal{Q}(\underline{u}, \underline{v}) = 2\pi(\epsilon^2 - 2(1 + \underline{u}^\top \underline{v}))^{-1/2} \quad (10)$$

under condition  $\epsilon > 2$ . Then, the CCvMD terms in (8) are expressed as

$$\begin{aligned} \mathcal{D}_1 &= 2\pi \sum_{i=1}^n \sum_{r=1}^n w_i w_r (\epsilon^2 - 2(1 + \underline{x}_i^\top \underline{x}_r))^{-1/2}, \\ \mathcal{D}_2 &= 2\pi \sum_{i=1}^n \sum_{k=1}^{\hat{n}} w_i \hat{w}_k (\epsilon^2 - 2(1 + \underline{x}_i^\top \hat{\underline{x}}_k))^{-1/2}. \end{aligned} \quad (11)$$

Note computations above only depend on the relative distance respecting the arc length over all pairs of samples. Therefore, the CCvMD provides a symmetric and unique measure for quantifying the statistical divergence between two discrete models adaptively to manifold geometry. Meanwhile, it leads to a closed form that is smooth w.r.t. sample locations. This further facilitates our optimization-based reapproximation.

### C. Implementation

Similar to the reapproximation scheme for hyperspheres in [23], we exploit Riemannian trust-region (RTR) method [26] to solve the optimization problem in (7) given its smooth and real-valued objective. More specifically, we concatenate the desired template sample locations  $\mathbb{X} = \{\underline{x}_i\}_{i=1}^n$  column-wise into a matrix, which belongs to the oblique manifold  $\mathbb{OB}(2, n) \subset \mathbb{R}^{2 \times n}$ . Implementation of the RTR is taken from Manopt [27]. Riemannian optimization handles manifold constraints explicitly without modification of the objective (e.g., through incorporating Lagrange multipliers). Further, symbolic forms of the objective's gradients and Hessians in the ambient space  $\mathbb{R}^{2 \times n}$  are often preferable for achieving fast and stable convergence. This is possible thanks to the closed-form expressions in (11), and essential derivatives are provided in Appendix A. We showcase the proposed CDR technique in the following example.

**Case Study 1** We synthesize a complex distribution on the unit circle by mixing parametric models from directional statistics. Its PDF is given by  $f(\underline{x}) = 0.2 \cdot f_{\text{VMM}}(\underline{x}) + 0.8 \cdot f_{\text{BM}}(\underline{x})$ , with  $\underline{x} \in \mathbb{S}^1$ .  $f_{\text{VMM}}$  and  $f_{\text{BM}}$  are von Mises mixture and Bingham mixture distributions defined as

$$\begin{aligned} f_{\text{VMM}}(\underline{x}) &= \sum_{i=1}^4 1/4 \cdot f_{\text{VM}}(\underline{x}; \theta_i, \lambda_i) \quad \text{and} \\ f_{\text{BM}}(\underline{x}) &= \sum_{i=1}^2 1/2 \cdot f_{\text{B}}(\underline{x}; \mathbf{R}(\phi_i), \mathbf{Z}_i), \end{aligned}$$

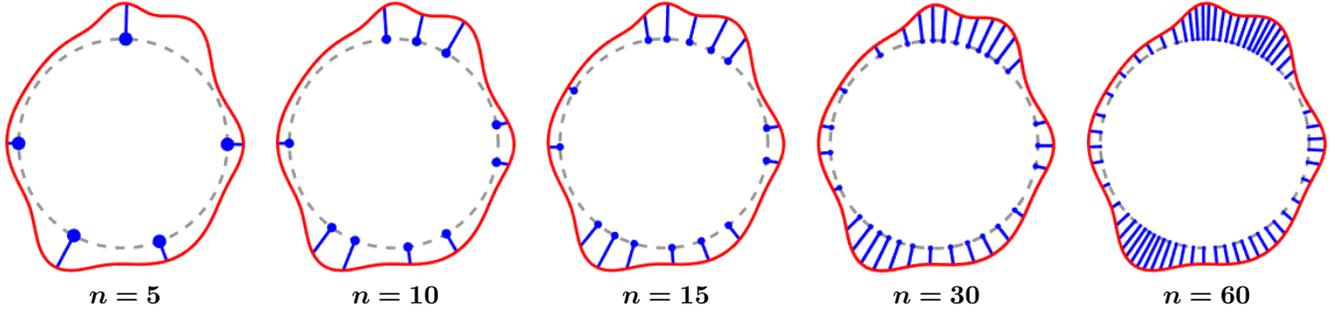
respectively. The von Mises components are parameterized with means  $\{\theta_i\}_{i=1}^4 = \{-\pi/4, \pi/2, \pi, 0\}$  and concentrations  $\{\lambda_i\}_{i=1}^4 = \{10, 50, 10, 20\}$ . According to the definition of the Bingham distribution [28], we set up the Bingham components with two-dimensional rotation matrices  $\mathbf{R}(\theta_i)$  through angles  $\{\phi_i\}_{i=1}^2 = \{-\pi/6, \pi/12\}$  and diagonal matrices  $\{\mathbf{Z}_i\}_{i=1}^2 = \{\text{diag}(-10, 0), \text{diag}(-2, 0)\}$ . The former gives the mode, and the latter controls the dispersion concentration. We draw  $\hat{n} = 10000$  random samples from the underlying distribution, and reapproximate it to  $n \in \{5, 10, 15, 30, 60\}$  template samples of equal weights via CDR.

Shown in Fig. 1, the proposed method yields template discrete models with supports of deterministic and dispersion-adaptive locations, which are much more representative and efficient compared with random samples.

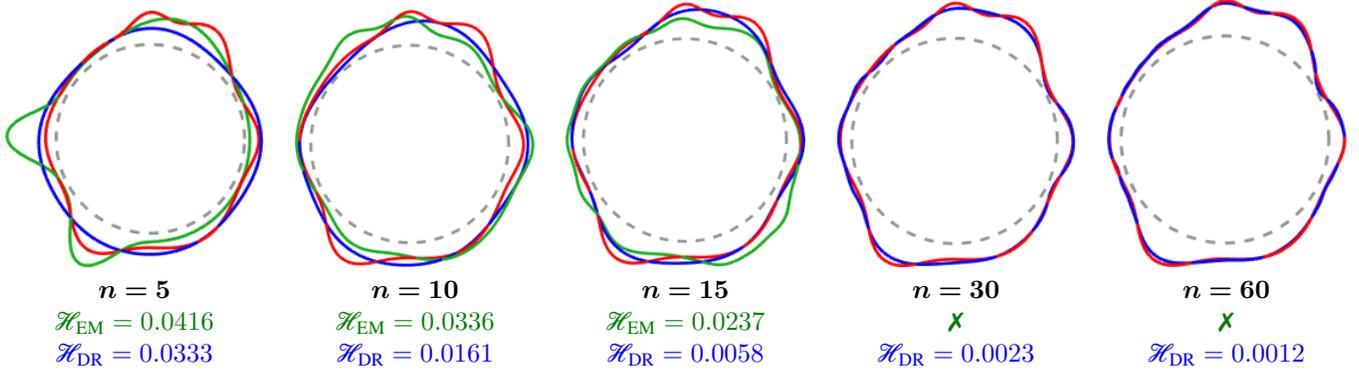
As validated in Case 1, the proposed CDR method can largely improve the representation efficiency of discrete models approximating arbitrary distributions. A quantitative test of such improvement can be conducted in the scenario of recursive circular estimation following the generic reapproximation discrete filtering in [23] (evaluated for hyperspheres). In this paper, we do not repeat it due to resemblance, and the reader may refer to the evaluation in [23] for more insight.

### III. CDR-BASED DENSITY ESTIMATION

The proposed CDR method can further be exploited for fitting a von Mises mixture to the source sample set via maximum likelihood estimation (MLE). For that, we configure the template samples  $\mathbb{X} = \{\underline{x}_i\}_{i=1}^n$  to be equally weighted for



**Figure 1:** Circular discrete reapproximation using random samples for Case 1. Red curves denote the underlying circular distribution, and template discrete models given by CDR are depicted by blue dots.



**Figure 2:** Circular density estimation based on discrete reapproximation (DR) and expectation maximum (EM) for Case 2. The proposed two-stage approach delivers von Mises mixtures of higher fidelity (blue) compared with the ones given by EM (green) w.r.t. Hellinger distances  $\mathcal{H}$  to the ground truth (red).

CDR. Then, each component of the targeted von Mises mixture is placed on each template sample with one joint concentration parameter  $\zeta$ , namely,

$$\begin{aligned}
 f_{VMM}(\mathbf{x}; \mathbf{X}, \zeta) &= \sum_{i=1}^n \frac{1}{n} f_{VM}(\mathbf{x}; \mathbf{x}_i, \zeta) \\
 &= \sum_{i=1}^n \frac{1}{2\pi n \mathcal{I}_0(\zeta)} \exp(\zeta \mathbf{x}_i^\top \mathbf{x}).
 \end{aligned}$$

The next step is to determine the optimal  $\zeta$  in the sense of MLE w.r.t. the source samples  $\{\hat{\mathbf{x}}_k\}_{k=1}^{\hat{n}}$  and their weights  $\{\hat{w}_k\}_{k=1}^{\hat{n}}$ , namely,

$$\begin{aligned}
 \zeta^* &= \arg \max_{\zeta \geq 0} \left\{ \sum_{k=1}^{\hat{n}} \hat{w}_k \ln \left( f_{VMM}(\hat{\mathbf{x}}_k; \mathbf{X}, \zeta) \right) \right\} \\
 &=: \arg \max_{\zeta \geq 0} \{ \mathcal{Y}(\zeta) \}.
 \end{aligned} \tag{12}$$

The objective function  $\mathcal{Y}(\zeta)$  is formulated in the form of log-likelihood and can further be derived as

$$\begin{aligned}
 \mathcal{Y}(\zeta) &= \sum_{k=1}^{\hat{n}} \hat{w}_k \ln \left( \sum_{i=1}^n \frac{1}{\mathcal{I}_0(\zeta)} \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k) \right) \\
 &= -\ln(\mathcal{I}_0(\zeta)) + \sum_{k=1}^{\hat{n}} \hat{w}_k \ln \left( \sum_{i=1}^n \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k) \right).
 \end{aligned}$$

To solve the MLE problem in (12), we need to find the zero of the objective's first derivative  $\mathcal{Y}'(\zeta)$ , which follows

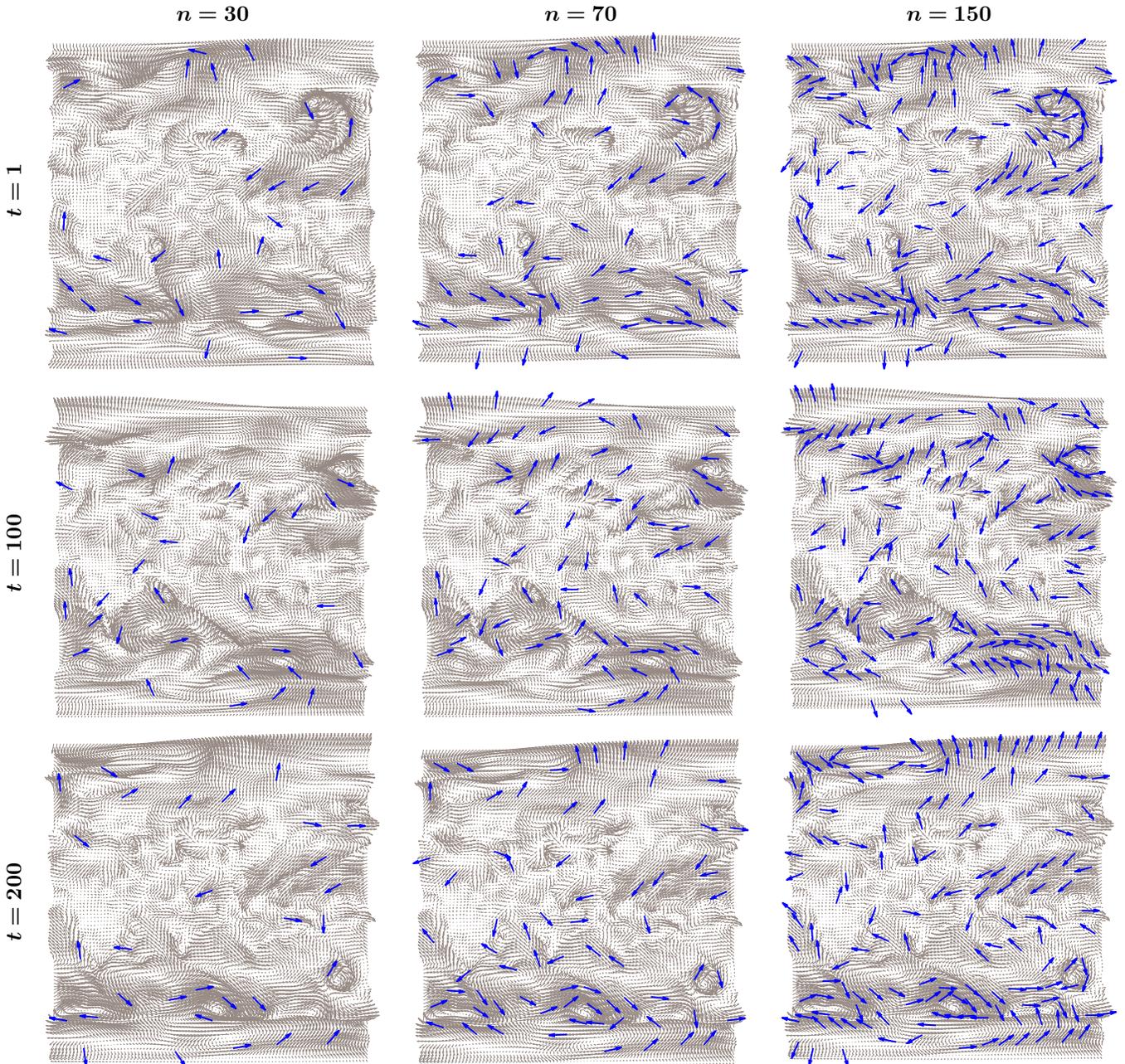
$$\mathcal{Y}'(\zeta) = \sum_{k=1}^{\hat{n}} \left( \hat{w}_k \frac{\sum_{i=1}^n (\mathbf{x}_i^\top \hat{\mathbf{x}}_k) \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)}{\sum_{i=1}^n \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)} \right) - \frac{\mathcal{I}'_0(\zeta)}{\mathcal{I}_0(\zeta)}. \tag{13}$$

We substitute the last term with  $\mathcal{A}(\zeta) = \mathcal{I}'_0(\zeta)/\mathcal{I}_0(\zeta)$ , and the numerator goes to  $\mathcal{I}_1(\zeta)$  according to the recurrence relations for derivative of modified Bessel function of the first kind [29]. Thus, we obtain  $\mathcal{A}(\zeta) = \mathcal{I}_1(\zeta)/\mathcal{I}_0(\zeta)$  in the form of Bessel function ratio [7].

Finding the zero of  $\mathcal{Y}'(\zeta)$  can only be done numerically. For that, we deploy the Newton's method with the second derivative of objective  $\mathcal{Y}(\zeta)$  given by

$$\begin{aligned}
 \mathcal{Y}''(\zeta) &= \sum_{k=1}^{\hat{n}} \left( \hat{w}_k \frac{\sum_{i=1}^n (\mathbf{x}_i^\top \hat{\mathbf{x}}_k)^2 \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)}{\sum_{i=1}^n \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)} \right) - \\
 &\quad \sum_{k=1}^{\hat{n}} \hat{w}_k \left( \frac{\sum_{i=1}^n (\mathbf{x}_i^\top \hat{\mathbf{x}}_k) \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)}{\sum_{i=1}^n \exp(\zeta \mathbf{x}_i^\top \hat{\mathbf{x}}_k)} \right)^2 - \mathcal{A}'(\zeta).
 \end{aligned} \tag{14}$$

For the last term in the formula above, we exploit  $\mathcal{A}'(\zeta) = 1 - \mathcal{A}(\zeta)^2 - \mathcal{A}(\zeta)/\zeta$  given by [7]. Then, an  $m$ -th Newton's step can be established following  $\zeta_m = \zeta_{m-1} - \mathcal{Y}'(\zeta_{m-1})/\mathcal{Y}''(\zeta_{m-1})$ , with  $\mathcal{Y}'$  and  $\mathcal{Y}''$  shown in (13) and (14), respectively. We provide the following case study for



**Figure 3:** ECDR-based wind field modeling for Case 3. Raw data of wind fields are exploited as source for reapproximation and are denoted by grey arrows with lengths proportional to wind speeds. Template wind fields from ECDR are in blue.

quantitative validation of the proposed circular density estimation procedure.

**Case Study 2** We stick to the scenario from Case 1 with the same underlying distribution and source discrete model. The proposed density estimation is performed based upon the template discrete model given by CDR in the former case as shown in Fig. 1. For comparison with the state of the art of circular density estimation, we deploy the approach in [30]

using its open-source implementation<sup>1</sup>. It fits a von Mises mixture given its number of components  $n$  to random samples on  $\mathbb{S}^1$  via expectation maximum (EM) and yields components of individual concentrations. Moreover, the approach only allows for matching to equally-weighted samples. To quantify the estimation quality of the two approaches, we compute the Hellinger distance of the fitted continuous density to the underlying ground truth [8]. As shown in Fig. 2, given same

<sup>1</sup><https://github.com/chrschy/mvmdist>

component quantities  $n \in \{5, 10, 15, 30, 60\}$ , the proposed CDR-based approach delivers better fitting accuracy compared with the EM. While the latter crashes when component quantity increases, our scheme produces refined continuous models (lower Hellinger distances).

The case study above characterizes that the proposed two-stage scheme using discrete reapproximation provides a superior alternative to estimating circular densities over standard methodology. CDR to the template samples of equal weights first fixes locations of the von Mises components in a dispersion-adaptive manner. This alleviates the model complexity for the second step of density fitting, because only the joint concentration parameter  $\zeta$  is to be determined.

#### IV. EXTENSIONS TO COMPOSITE DIRECTIONAL DOMAIN

Many tasks in ecological or climatological statistics refer to modeling and inference on random circular vector field, where data points incorporate both directions (circular) and locations (Euclidean). The following sections introduce the essential building blocks for extending the proposed CDR scheme from  $\mathbb{S}^1$  to domains composing both the unit circle and Euclidean space of arbitrary dimension.

##### A. Extended CDR on $\mathbb{S}^1 \times \mathbb{R}^d$

Following the generic design of discrete reapproximation in Sec. II, we now extend CDR to  $\mathbb{S}^1 \times \mathbb{R}^d$ . We denote random variables thereon as  $\underline{\mathbf{x}} = [\underline{\mathbf{x}}_j^\top, \underline{\mathbf{x}}_r^\top]^\top$ , with  $\underline{\mathbf{x}}_j \in \mathbb{S}^1$  and  $\underline{\mathbf{x}}_r \in \mathbb{R}^d$  being circular and Euclidean, respectively. We redesign the kernel in Definition 1 w.r.t. the manifold structure now to be  $\kappa(\underline{\mathbf{x}}; \underline{\alpha}, \tau) = \kappa_j(\underline{\mathbf{x}}_j; \underline{\alpha}_j, \tau) \kappa_r(\underline{\mathbf{x}}_r; \underline{\alpha}_r, \tau)$ , with

$$\begin{aligned} \kappa_j(\underline{\mathbf{x}}_j; \underline{\alpha}_j, \tau) &= \exp(\tau \underline{\alpha}_j^\top \underline{\mathbf{x}}_j) \quad \text{and} \\ \kappa_r(\underline{\mathbf{x}}_r; \underline{\alpha}_r, \tau) &= \exp(-\tau (\underline{\mathbf{x}}_r - \underline{\alpha}_r)^\top (\underline{\mathbf{x}}_r - \underline{\alpha}_r)) \end{aligned} \quad (15)$$

denoting the circular and Euclidean components, respectively. In order to model the correlation between domains  $\mathbb{S}^1$  and  $\mathbb{R}^d$ , the two kernel components with locations at  $\underline{\alpha}_j \in \mathbb{S}^1$  and  $\underline{\alpha}_r \in \mathbb{R}^d$  share the same concentration parameter  $\tau$ . Correspondingly, we adjust the weighting function in Definition 2 to be  $\mathcal{V}(\tau) = \tau^{d/2} \exp(-\epsilon\tau)$ . The double integrals over kernel locations and concentrations in (9) can be adapted to  $\mathbb{S}^1 \times \mathbb{R}^d$  by following a similar procedure in Sec. II-B with reference to [22] and Appendix D of [31]. We then obtain

$$\mathcal{Q}(\underline{u}, \underline{v}) = 4\pi (\pi/2)^{d/2} ((\delta_r + 2\epsilon)^2 - 8(1 + \delta_j))^{-1/2}, \quad (16)$$

with  $\underline{u}, \underline{v}$  being arbitrary on  $\mathbb{S}^1 \times \mathbb{R}^d \subset \mathbb{R}^{d+2}$ .  $\delta_j = \underline{u}_j^\top \underline{v}_j$  and  $\delta_r = \|\underline{u}_r - \underline{v}_r\|^2$  are topology-aware distance measures for the circular and Euclidean components, respectively.

Solving the optimization problem for extended CDR is performed on the matrix manifold  $\mathbb{O}\mathbb{B}(2, n) \times \mathbb{R}^{d \times n}$  using the Riemannian trust-region, for which pointwise gradients and Hessians of the objective in the ambient space  $\mathbb{R}^{2+d}$  are desired. Due to space limit and their complicated derivations (particularly for Hessians across the component domains), we do not provide the formulae in this paper. However, a unified implementation for all on-manifold discrete reapproximation

methods will be open-sourced together with upcoming publications. The following case study shows validations on the extended circular discrete reapproximation (ECDR) scheme.

**Case Study 3** We exploit wind climatology data<sup>2</sup> for discrete modeling of random circular vector field. In this data set, wind direction and speed of contiguous United States are given in a resolution of approximately  $1.9 \times 1.9$  degrees of latitude and longitude. The wind field was recorded in a temporally continuous manner with more than 1000 snapshots per year. We exploit the raw data recorded at stamps  $t \in \{1, 100, 200\}$  in the year of 2020 and clip the wind fields into  $94 \times 94$  squares as the source. Further, we exploit the wind speed at each sample location as the weight (after normalization).  $n = \{30, 70, 150\}$  template samples of equal weights are deployed for reapproximation using ECDR specified on  $\mathbb{S}^1 \times \mathbb{R}^2$ . Shown in Fig. 3, the proposed approach is efficacious over various parameterizations and real-world data, and the obtained template samples model the underlying random vector fields with representative layouts. This validation scenario is potentially of interest for wind farm planning, where locations and facing directions of wind turbines are to be configured for efficient power generation.

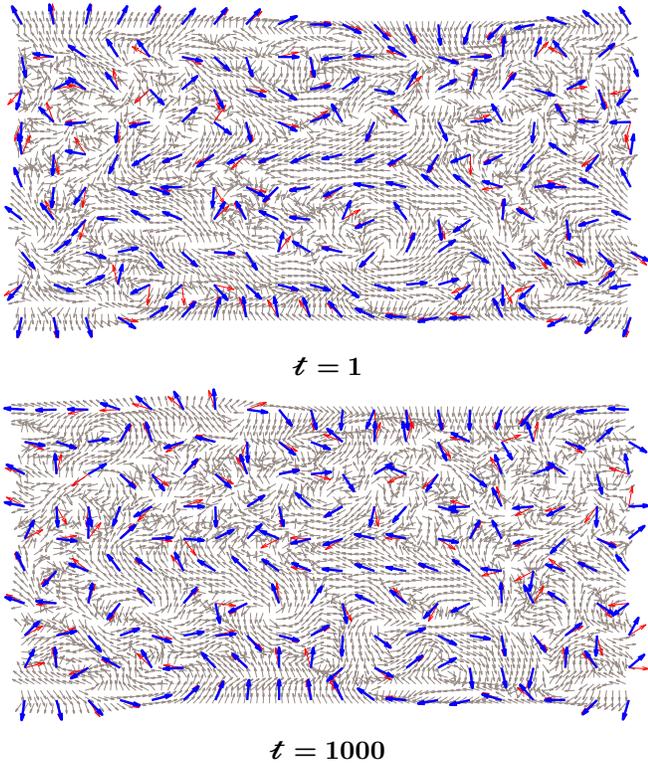
##### B. Regression on Circular Vector Fields

The proposed ECDR method further enables a novel regression model on random circular vector field. Given a training set  $\hat{\mathbf{X}} = \{[\hat{\mathbf{x}}_{k,j}^\top, \hat{\mathbf{x}}_{k,r}^\top]^\top\}_{k=1}^{\hat{n}} \subset \mathbb{S}^1 \times \mathbb{R}^d$  as the source and inquiry locations  $\{\underline{\mathbf{x}}_{i,j}\}_{i=1}^n \subset \mathbb{R}^d$  for the template set, the inquired circular quantities  $\{\underline{\mathbf{x}}_{i,j}\}_{i=1}^n \subset \mathbb{S}^1$  can be predicted via discrete reapproximation. Since the Euclidean components of the template samples are fixed, the optimization takes the ECDR objective, but is confined to the unit circle as in CDR.

We validate the ECDR-based regression model based on the wind field in the year of 2021 from the same data set mentioned in Case 3. The whole wind field of  $94 \times 192$  locations is downsampled with a ratio of 0.5 and used as training set. Among the rest of data points, another  $n = 200$  spatially uniform locations are selected for inquiry. All samples (in source and template sets) are equally weighted. Shown in Fig. 4 at time stamps  $t \in \{1, 1000\}$ , the ECDR-based regression model produces effective predictions of wind directions at the inquired locations.

Furthermore, we reset the number of inquiries to be  $n = 50$  and perform wind direction prediction over the whole year of 2021 at an interval of 10 time stamps. For comparing ECDR regression with state of the art, we apply two Gaussian processes (GPs) to modeling the two components of wind directions individually over the field [32] (implementation from Matlab). Normalization of the two individually predicted components is needed for prediction. One obvious drawback of this strategy lies in the ignorance of correlation between the two direction components. Multioutput GPs might

<sup>2</sup><https://www.ncdc.noaa.gov/societal-impacts/wind/>



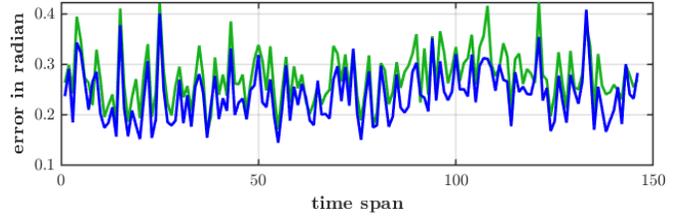
**Figure 4:** ECDR-based wind field regression. Given training sets (grey), inquired wind directions at  $n = 200$  locations are predicted (blue) and compared with the ground truth (red).

be applicable, however, no off-the-shelf scheme with circular output domain exists for our reference.

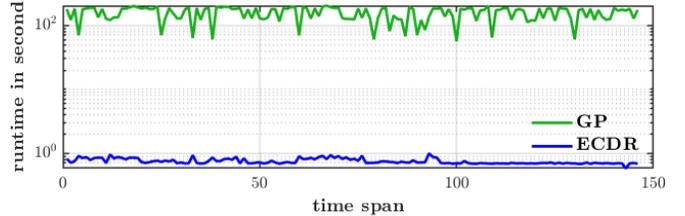
Fig. 5 shows evaluation results w.r.t. prediction accuracy (average error in radian of all inquiries) and runtime efficiency over the time span for the two regression models. The proposed ECDR-based model delivers better regression quality and substantially shorter runtime (over two orders of magnitude). It considers the correlation across the domain components and has a comparably simpler computational design (with only one single optimization using symbolic gradients and Hessians). However, the standard GP scheme suffers from large computational cost due to its well-known cubic complexity.

## V. CONCLUSIONS

In this paper, we propose the circular discrete reapproximation (CDR), a novel probabilistic scheme for discrete modeling and inference of random variables on the unit circle. The new circular Cramér–von Mises distance (CCvMD) is proposed to quantify the statistical divergence between two circular discrete models in a topology-aware manner. Given a large set of weighted samples as source, the CDR scheme reapproximates the underlying unknown distribution with template samples of configurable quantities and dispersion-adaptive locations. The basic CDR scheme facilitates circular density estimation using von Mises mixtures and is extended to reapproximation



(A) Accuracy given by the evaluated regression models.



(B) Runtime given by the evaluated regression models.

**Figure 5:** Evaluation on wind field regression using proposed ECDR (blue) and standard GP (green).

on the composite directional domain incorporating Euclidean space of arbitrary dimension. Both methods are novel and the latter directly enables a novel regression model on random circular vector field. For improvement, the parameter  $\epsilon$  in the weighting function of Definition 2 should be considered as a hyperparameter and needs to be automatically tuned. Methods for uncertainty quantification of ECDR-based regression model should also be developed. A unified implementation of the proposed approaches will be open-sourced with upcoming publications, and we expect more extensive evaluations of the proposed approaches in various real-world scenarios.

## APPENDIX

### A. Gradients and Hessians of CCvMD in Ambient Space

Deriving gradients and Hessians of the CCvMD terms in (8) is essentially done for the expression in (10) w.r.t. sample locations. For that, we reformulate (10) into

$$\mathcal{Q}(\delta) = 2\pi(\epsilon^2 - 2(1 + \delta))^{-1/2},$$

with  $\delta = \underline{u}^\top \underline{v}$  denoting the distance between two circular locations. The derivative w.r.t.  $\delta$  then follows

$$\frac{d\mathcal{Q}(\delta)}{d\delta} = 2\pi(\epsilon^2 - 2(1 + \delta))^{-3/2} = \frac{\mathcal{Q}(\delta)}{\epsilon^2 - 2(1 + \delta)}, \quad (17)$$

and we obtain the first derivative of  $\mathcal{Q}$  w.r.t. location  $\underline{u}$  as

$$\frac{\partial \mathcal{Q}(\underline{u}, \underline{v})}{\partial \underline{u}} = \frac{d\mathcal{Q}(\delta)}{d\delta} \frac{\partial \delta}{\partial \underline{u}} = \frac{d\mathcal{Q}(\delta)}{d\delta} \underline{v} =: \mathcal{P}(\underline{u}, \underline{v}) \underline{v}, \quad (18)$$

with  $\mathcal{P}(\underline{u}, \underline{v})$  substituting (17). Further, the second derivatives of  $\mathcal{Q}(\underline{u}, \underline{v})$  w.r.t. sample locations can be derived as

$$\begin{aligned} \frac{\partial^2 \mathcal{Q}(\underline{u}, \underline{v})}{\partial \underline{u} \partial \underline{u}^\top} &= \frac{3\mathcal{P}(\underline{u}, \underline{v})}{\epsilon^2 - 2(1 + \underline{u}^\top \underline{v})} \underline{v} \underline{v}^\top \quad \text{and} \\ \frac{\partial^2 \mathcal{Q}(\underline{u}, \underline{v})}{\partial \underline{u} \partial \underline{v}^\top} &= \left( \frac{3}{\epsilon^2 - 2(1 + \underline{u}^\top \underline{v})} \underline{u} \underline{v}^\top + \mathbf{I}_{2 \times 2} \right) \mathcal{P}(\underline{u}, \underline{v}). \end{aligned} \quad (19)$$

Deploying (18) and (19) pointwise to the expressions in (8) then delivers the desired gradients and Hessians of the objective in (7) for optimization.

#### ACKNOWLEDGMENT

This work is partially funded by the Helmholtz AI Cooperation Unit within the scope of the project Ubiquitous Spatio-Temporal Learning for Future Mobility (ULearn4Mobility).

#### REFERENCES

- [1] R. Senanayake and F. Ramos, "Directional Grid Maps: Modeling Multimodal Angular Uncertainty in Dynamic Environments," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, Oct. 2018.
- [2] Y. Ban, X. Alameda-Pineda, C. Evers, and R. Horaud, "Tracking Multiple Audio Sources with the von Mises Distribution and Variational EM," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 798–802, 2019.
- [3] M. N. Lang, L. Schlosser, T. Hothorn, G. J. Mayr, R. Stauffer, and A. Zeileis, "Circular Regression Trees and Forests with an Application to Probabilistic Wind Direction Forecasting," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 69, no. 5, pp. 1357–1374, 2020.
- [4] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A. W. Nelson, A. Bridgland, *et al.*, "Improved Protein Structure Prediction Using Potentials from Deep Learning," *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.
- [5] J. A. Carta, C. Bueno, and P. Ramírez, "Statistical Modelling of Directional Wind Speeds Using Mixtures of von Mises Distributions: Case Study," *Energy Conversion and Management*, vol. 49, no. 5, pp. 897–907, 2008.
- [6] Á. F. García-Fernández, F. Tronarp, and S. Särkkä, "Gaussian Target Tracking with Direction-of-Arrival von Mises–Fisher Measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 2960–2972, 2019.
- [7] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2009, vol. 494.
- [8] G. Kurz, I. Gilitschenski, F. Pfaff, L. Drude, U. D. Hanebeck, R. Haeb-Umbach, and R. Y. Siegwart, "Directional Statistics and Filtering Using libDirectional," *Journal of Statistical Software*, May 2019.
- [9] K. Li, F. Pfaff, and U. D. Hanebeck, "Progressive von Mises–Fisher Filtering Using Isotropic Sample Sets for Nonlinear Hyperspherical Estimation," *Sensors*, Apr. 2021.
- [10] K. Li, D. Frisch, S. Radtke, B. Noack, and U. D. Hanebeck, "Wavefront Orientation Estimation Based on Progressive Bingham Filtering," in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2018)*, Oct. 2018.
- [11] K. Li, F. Pfaff, and U. D. Hanebeck, "Geometry-Driven Stochastic Modeling of SE(3) States Based on Dual Quaternion Representation," in *Proceedings of the 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019)*, Taipei, Republic of China, May 2019.
- [12] G. Kurz and U. D. Hanebeck, "Deterministic Sampling on the Torus for Bivariate Circular Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 530–534, Feb. 2017.
- [13] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "A New Probability Distribution for Simultaneous Representation of Uncertain Position and Orientation," in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, July 2014.
- [14] K. Li, G. Kurz, L. Bernreiter, and U. D. Hanebeck, "Nonlinear Progressive Filtering for SE(2) Estimation," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, July 2018.
- [15] F. Pfaff, K. Li, and U. D. Hanebeck, "The Spherical Grid Filter for Nonlinear Estimation on the Unit Sphere," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.
- [16] K. Li, F. Pfaff, and U. D. Hanebeck, "Grid-Based Quaternion Filter for SO(3) Estimation," in *Proceedings of the 2020 European Control Conference (ECC 2020)*, Virtual, May 2020.
- [17] F. Pfaff, K. Li, and U. D. Hanebeck, "Estimating Correlated Angles Using the Hypertoroidal Grid Filter," in *Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2020)*, Virtual, Sept. 2020.
- [18] K. Li, F. Pfaff, and U. D. Hanebeck, "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, July 2019.
- [19] G. Kurz, F. Pfaff, and U. D. Hanebeck, "Application of Discrete Recursive Bayesian Estimation on Intervals and the Unit Circle to Filtering on SE(2)," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1197–1206, Mar. 2018.
- [20] K. Li, F. Pfaff, and U. D. Hanebeck, "Hyperspherical Unscented Particle Filter for Nonlinear Orientation Estimation," in *Proceedings of the 1st Virtual IFAC World Congress (IFAC-V 2020)*, July 2020.
- [21] —, "Unscented Dual Quaternion Particle Filter for SE(3) Estimation," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 647–652, Apr. 2021.
- [22] —, "Dual Quaternion Sample Reduction for SE(2) Estimation," in *Proceedings of the 23rd International Conference on Information Fusion (Fusion 2020)*, Virtual, July 2020.
- [23] —, "Hyperspherical Dirac Mixture Reapproximation," *arXiv preprint arXiv:2110.10411*, Oct. 2021.
- [24] U. D. Hanebeck, "Optimal Reduction of Multivariate Dirac Mixture Densities," *at – Automatisierungstechnik, Oldenbourg Verlag*, vol. 63, no. 4, pp. 265–278, Apr. 2015.
- [25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. Academic press, 2014.
- [26] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.
- [27] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab Toolbox for Optimization on Manifolds," *Journal of Machine Learning Research*, vol. 15, pp. 1455–1459, 2014.
- [28] K. Li, D. Frisch, B. Noack, and U. D. Hanebeck, "Geometry-Driven Deterministic Sampling for Nonlinear Bingham Filtering," in *Proceedings of the 2019 European Control Conference (ECC 2019)*, Naples, Italy, June 2019.
- [29] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the Unit Hypersphere Using von Mises–Fisher Distributions," *Journal of Machine Learning Research*, vol. 6, no. Sep, pp. 1345–1382, 2005.
- [30] W.-L. Hung, S.-J. Chang-Chien, and M.-S. Yang, "Self-Updating Clustering Algorithm for Estimating the Parameters in Mixtures of von Mises Distributions," *Journal of Applied Statistics*, vol. 39, no. 10, pp. 2259–2274, 2012.
- [31] R. P. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House Norwood, MA, USA, 2007, vol. 685.
- [32] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.