

# Gaussian Process on the Product of Directional Manifolds

Ziyu Cao and Kailai Li

**Abstract**— We present a principled study on defining Gaussian processes (GPs) with inputs on the product of directional manifolds. A circular kernel is first presented according to the von Mises distribution. Based thereon, the hypertoroidal von Mises (HvM) kernel is proposed to establish GPs on hypertori with consideration of correlated circular components. The proposed HvM kernel is demonstrated with multi-output GP regression for learning vector-valued functions on hypertori using the intrinsic coregionalization model. Analytic derivatives for hyperparameter optimization are provided for runtime-critical applications. For evaluation, we synthesize a ranging-based sensor network and employ the HvM-based GPs for data-driven recursive localization. Numerical results show that the HvM-based GP achieves superior tracking accuracy compared to parametric model and GPs of conventional kernel designs.

## I. INTRODUCTION

Directional variables appear ubiquitously in control-related scenarios including signal processing, object tracking, and computer vision, among others [1]–[6]. Typical examples of directional manifolds include the unit circle  $\mathbb{S}^1 \subset \mathbb{R}^2$  and hyperspheres  $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ , which are by nature nonlinear and periodic. Many real-world applications, such as ground/aerial locomotion, robotic manipulation, and angle-of-arrival-based tracking, involve stochastic systems composing multiple directional inputs. Gaussian processes (GPs) have emerged as a powerful statistical tool for inferential modeling of uncertain dynamical systems [7], where dependence between system outputs at different inputs is captured by the similarity across the input domain quantified by the kernel. This paper focuses on the study of a representative case, namely GP modeling of vector-valued functions defined on the hypertorus  $\mathbb{T}^3 := \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^6$  ( $\times$  denotes the Cartesian product).

The key challenge in addressing the considered problem is to design a suitable kernel quantifying the similarity between hypertoroidal data points adaptively to the underlying manifold structure. The investigation can be decomposed into two sub-problems. First, each circular component is inherently periodic and nonlinear, which cannot be simply handled by conventional kernels defined based on Euclidean distances. Second, it is essential to properly interpret the correlations among the circular components for informative similarity quantification. The former challenge can be overcome by transforming common periodic kernels w.r.t. the unit circle  $\mathbb{S}^1$  [7]. The latter refers to constructing kernels on a manifold composing multiple domains through the Cartesian product. In this regard, most existing solutions employ a straightforward strategy of multiplying the kernels belonging to each domain component, which overlooks the correlational

interpretation between them and may result in information loss for inferential modeling [8].

The major contribution of the presented work is a principled approach on establishing Gaussian processes on the product of directional manifolds. We propose the hypertoroidal von Mises (HvM) kernel for handling input data on hypertori with consideration of correlated circular components in a manifold-adaptive manner. Closed-form derivatives are provided for efficient hyperparameter optimization. Further, we synthesize an evaluation scenario for recursive tracking in sensor networks, where an unknown range sensing model is learned using angle-of-arrival data. Compared with parametric and GP modeling using common kernel designs, the proposed HvM-based GP enables superior tracking accuracy and robustness. We open-source our code under [github.com/ASIG-X/hyperToroidalGP](https://github.com/ASIG-X/hyperToroidalGP).

The remainder of the paper is organized as follows. We summarize related works in Sec. II. Preliminaries on GP modeling are introduced in Sec. III, after which the HvM kernel is proposed in Sec. IV. We elaborate the evaluation in Sec. V, and the work is concluded in Sec. VI.

## II. RELATED WORKS

*GPs on Riemannian Manifolds:* In [9], GP kernels on Riemannian manifolds (RMs) have been derived implicitly by solving stochastic partial differential equations incorporating the Laplace–Beltrami operator. An improvement on scalability has been achieved in [10] via spectral theory for Matérn kernels on RMs. The approach was further extended in [11] to model tangential vector fields defined on RMs through projective geometry. These implicit strategies are generally applicable for developing RM-adaptive kernels, however, suffer from high theoretical complexity and computational expense, such as the eigendecomposition of Laplace–Beltrami operators, leading to a considerable gap for their usage in engineering practices. Alternatively, manifold-adaptive kernels can be explicitly designed w.r.t. the underlying manifold structure. This can lead to conciser procedures of establishing GPs on RMs. However, simply replacing the distance metric with the geodesic to generalize Gaussian kernels on RMs for non-Euclidean inputs does not guarantee positive definiteness. In this regard, Laplacian kernels are eligible if and only if the distance metric is conditionally negative definite [12]. For GP modeling on domains with complex but known structures, basic kernels of each domain component are typically combined through summation or multiplication [8]. In [13], GPs have been established on cylindrical surface ( $\mathbb{S}^1 \times \mathbb{R}$ ) for spatiotemporal modeling of meteorological fields, where the kernel is obtained by multiplying a sinusoidal and

The authors are with the Department of Electrical Engineering, Linköping University, Sweden. E-Mails: {ziyu.cao, kailai.li}@liu.se

a Laplacian kernel of  $\ell^1$ -norm. In [14], various GP kernels have been investigated on manifolds of unit quaternions and unit dual quaternions based on hyperspherical geometry on  $\mathbb{S}^3$  for spatial orientation/pose inferences.

*Directional Statistics:* Distributions from directional statistics have provided valuable references for quantifying similarities between directional data points [15]. Popular choices include the wrapped normal and the von Mises distributions on the unit circle, and the von Mises–Fisher distribution on hyperspheres [16]–[18]. For modeling directional quantities of antipodal symmetry, e.g., unit quaternions on  $\mathbb{S}^3 \subset \mathbb{R}^4$ , the Bingham distribution is commonly used and has been widely exploited in reasoning uncertain spatial orientations [19], [20]. As for distributions on composite directional domains, one important aspect is to interpret correlation across manifold components [21], [22]. Former works were shown on the torus  $\mathbb{S}^1 \times \mathbb{S}^1$  for temporal modeling of correlated wind directions [23] and on the manifold of unit dual quaternions for modeling uncertain spatial poses [24]–[26].

### III. GAUSSIAN PROCESS REGRESSION AND INFERENCE

#### A. Gaussian Process Modeling

A Gaussian process is a collection of random variables on a certain domain. Any finite set of these random variables follows a multivariate Gaussian distribution. We denote a GP with scalar output as  $\mathbf{r}(\underline{x}) \sim \mathcal{GP}(m(\underline{x}), \mathcal{K}(\underline{x}, \underline{x}'))$ , with  $m(\underline{x})$  and  $\mathcal{K}(\underline{x}, \underline{x}')$  being the mean and covariance functions of  $\mathbf{r}(\underline{x})$ , respectively [7]. Suppose there is an arbitrary function that is observed under uncertainty following  $z = r(\underline{x}) + \epsilon$  with noise  $\epsilon \sim \mathcal{N}(0, \sigma_r^2)$ . Given a set of  $n$  input locations and corresponding observations corrupted by noise,  $\{(\underline{x}_{\bullet,i}, z_i)\}_{i=1}^n$ , a GP yields a posterior of the function value at test locations  $\{\underline{x}_{\circ,i}\}_{i=1}^m$  in the form of a multivariate Gaussian distribution

$$\mathbf{r}_{\circ} | \{(\underline{x}_{\bullet,i}, z_i)\}_{i=1}^n, \{\underline{x}_{\circ,i}\}_{i=1}^m \sim \mathcal{N}(\hat{\mathbf{r}}_{\circ}, \mathbf{C}_{\circ}),$$

with the mean and covariance given by

$$\hat{\mathbf{r}}_{\circ} = \mathbf{K}_{\bullet\bullet} \mathbf{K}^{-1} \underline{z} \quad \text{and} \quad \mathbf{C}_{\circ} = \mathbf{K}_{\circ\circ} - \mathbf{K}_{\bullet\circ} \mathbf{K}^{-1} \mathbf{K}_{\circ\bullet}, \quad (1)$$

respectively [7], where  $\mathbf{K} = \mathbf{K}_{\bullet\bullet} + \sigma_r^2 \mathbf{I}_n$ . Element in matrix  $\mathbf{K}_{\bullet\bullet} \in \mathbb{R}^{n \times m}$  at entry  $(i, j)$  is the covariance given by the kernel evaluated at training and test locations  $(\underline{x}_{\bullet,i}, \underline{x}_{\circ,j})$ . Matrices  $\mathbf{K}_{\bullet\bullet} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{K}_{\circ\circ} \in \mathbb{R}^{m \times m}$  can be obtained analogously. Vector  $\underline{z} = [z_1, \dots, z_n]^T$  collects all observations, and  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$  is an identity matrix. Further, the vector of predicted observations at test locations follows the distribution  $\mathcal{N}(\hat{\mathbf{r}}_{\circ}, \mathbf{C}_{\circ} + \sigma_r^2 \mathbf{I}_m)$ .

#### B. Multi-Output Gaussian Processes

For modeling vector-valued functions expressed as  $\underline{z} = \mathbf{r}(\underline{x}) + \underline{\epsilon} \in \mathbb{R}^d$ , where  $\underline{\epsilon} \sim \mathcal{N}(\mathbf{0}_d, \mathbf{R})$  is the noise term and  $\mathbf{R} = \text{diag}(\sigma_{r,1}^2, \dots, \sigma_{r,d}^2)$  its associated covariances, multi-output Gaussian processes can be utilized. It takes the form  $\mathbf{r}(\underline{x}) \sim \mathcal{GP}(m(\underline{x}), \mathcal{K}(\underline{x}, \underline{x}'))$ , with  $m(\underline{x}) \in \mathbb{R}^d$  being a vector-valued function collecting the mean of each dimension and  $\mathcal{K}(\underline{x}, \underline{x}')$  the matrix-valued covariance function. To

obtain a valid  $\mathcal{K}(\underline{x}, \underline{x}')$ , one popular method is the intrinsic coregionalization model. This leads to

$$\mathcal{K}(\underline{x}, \underline{x}') = \mathcal{K}(\underline{x}, \underline{x}') \mathbf{B}, \quad \text{with} \quad \mathbf{B} \in \mathbb{R}^{d \times d} \quad (2)$$

being a coregionalization matrix that is positive semidefinite [27]. Given uncertain observations  $\{z_i\}_{i=1}^n$  collected at training locations  $\{\underline{x}_{\bullet,i}\}_{i=1}^n$ , the posterior at a test location  $\underline{x}_{\circ}$  follows

$$\mathbf{r}_{\circ} | \{(\underline{x}_{\bullet,i}, z_i)\}_{i=1}^n, \underline{x}_{\circ} \sim \mathcal{N}(\hat{\mathbf{r}}_{\circ}, \mathbf{C}_{\circ}).$$

Here, the mean and covariance are calculated according to the general form in (1), with components newly defined as  $\mathbf{K}_{\bullet\bullet} = \mathbf{B} \otimes \mathbf{K}_{\bullet\bullet}$ ,  $\mathbf{K}_{\circ\circ} = \mathcal{K}(\underline{x}_{\circ}, \underline{x}_{\circ}) \mathbf{B}$  and  $\mathbf{K} = \mathbf{B} \otimes \mathbf{K}_{\bullet\bullet} + \mathbf{R} \otimes \mathbf{I}_n$ . Note that  $\mathbf{K}_{\bullet\bullet}$  still denotes the kernel matrix at training locations as given in (1). Furthermore, vector  $\underline{z}$  is obtained through the vectorization  $\underline{z} = \text{vec}([\underline{z}_1, \dots, \underline{z}_n]^T) \in \mathbb{R}^{nd}$ , and vector  $\mathbf{k}_{\bullet\bullet} = [\mathcal{K}(\underline{x}_{\circ}, \underline{x}_{\bullet,1}), \dots, \mathcal{K}(\underline{x}_{\circ}, \underline{x}_{\bullet,n})]$ . Similarly to the scalar case, the predictive distribution of the observation at  $\underline{x}_{\circ}$  is obtained as  $\underline{z}_{\circ} \sim \mathcal{N}(\hat{\mathbf{r}}_{\circ}, \mathbf{C}_{\circ} + \mathbf{R})$ .

### IV. GAUSSIAN PROCESSES ON THE HYPERTORUS

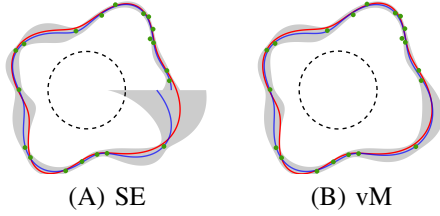
#### A. A Circular Kernel Based on the von Mises Distribution

We now consider a scalar-valued function  $r(\underline{x})$  with inputs defined on the unit circle, namely,  $r : \mathbb{S}^1 \rightarrow \mathbb{R}$  for GP modeling. In order to quantify the similarity between two circular inputs  $\underline{u}, \underline{v} \in \mathbb{S}^1$ , we formulate a kernel  $\mathcal{K}_{\text{vM}}(\underline{u}, \underline{v}) = \omega^2 \exp(\lambda \underline{u}^T \underline{v})$  based on the von Mises distribution, with  $\omega$  controlling signal variance and  $\lambda > 0$  the concentration. Note that this kernel can be referred to as a reformulation of the periodic kernel with angular inputs [7, Eq. 4.31]. To showcase its functionality within the GP framework compared with kernels on Euclidean domains, we demonstrate the following case study.

**Case Study 1** We synthesize a function on the unit circle by mixing three von Mises distributions and one Bingham distribution, with each component configured with individual parameters. The function is observed via  $z = \frac{1}{3} \sum_{i=1}^3 f_{\text{vM}}^i(\underline{x}) + f_{\text{B}}(\underline{x}) + \epsilon$ , with  $\epsilon \sim \mathcal{N}(0, 0.0025)$  being additive noise. We embed the squared exponential (SE) kernel (distance metric  $d = \theta - \theta'$ , with  $\theta$  and  $\theta'$  denoting the angular positions of  $\underline{u}$  and  $\underline{v}$ , respectively) and the proposed von Mises (vM) kernel into the same GP regression scheme as introduced in Sec. III-A. Shown in Fig. 1-(A), the SE kernel produces discontinuous curves due to the aperiodic distance metric of Euclidean geometry. In contrast, the proposed vM kernel quantifies periodic similarity adaptively to circular geometry, inducing identical posteriors with period of  $2\pi$  as plotted in Fig. 1-(B).

#### B. The Hypertoroidal von Mises Kernel

We now aim to establish GPs on the hypertorus  $\mathbb{T}^3 \subset \mathbb{R}^6$ . For constructing kernels on the Cartesian product of multiple unit circles, a common strategy is to multiply kernels on each circular component [8]. However, this only considers the similarity within each unit circle and neglects any potential



**Fig. 1:** GP regression using different kernels on the circular function in Case 1. Given the same training set (green dots), the vM kernel produces a geometry-adaptive posterior (with blue curves and gray areas denoting means and variances, respectively), whereas the SE kernel fails.

correlations in between. To achieve informative GP modeling on hypertori, it is crucial to design a manifold-adaptive kernel with consideration of correlated circular components. For that, we define the hypertoroidal von Mises (HvM) kernel

$$\mathcal{K}_{\text{HvM}}(\underline{u}, \underline{v}) = \omega^2 \exp(\underline{\lambda}^\top \underline{d}(\underline{u}, \underline{v}) + \underline{d}(\underline{u}, \underline{v})^\top \underline{\Lambda} \underline{d}(\underline{u}, \underline{v})), \quad (3)$$

with  $\underline{u}, \underline{v} \in \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$  being a pair of hypertoroidal inputs. We express them component-wise w.r.t. each circle as  $\underline{u} = [(\underline{u}^1)^\top, (\underline{u}^2)^\top, (\underline{u}^3)^\top]^\top$  and  $\underline{v} = [(\underline{v}^1)^\top, (\underline{v}^2)^\top, (\underline{v}^3)^\top]^\top$ . The metric  $\underline{d}$  is defined as

$$\underline{d}(\underline{u}, \underline{v}) = [(\underline{u}^1)^\top \underline{v}^1, (\underline{u}^2)^\top \underline{v}^2, (\underline{u}^3)^\top \underline{v}^3]^\top,$$

which measures distances between data points via inner products on each circular component. This is an analogous construction to the von Mises kernel given in Sec. IV-A, and inherently guarantees the symmetry of function (3) and the periodic nature of the hypertoroidal manifold.

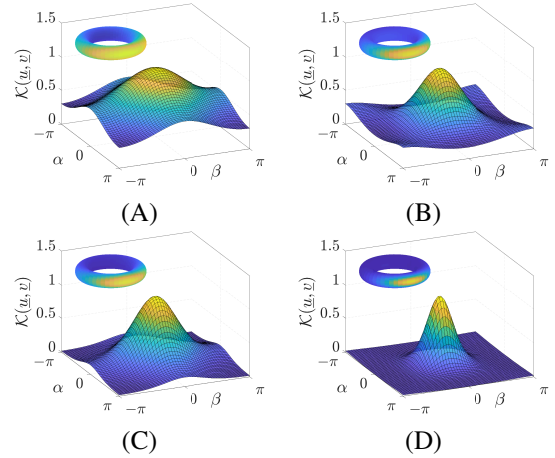
The proposed HvM kernel incorporates three hyperparameters, i.e., the  $\omega \in \mathbb{R}$ ,  $\underline{\lambda} \in \mathbb{R}^3$ , and  $\underline{\Lambda} \in \mathbb{R}^{3 \times 3}$ . Similar to the von Mises kernel in Sec. IV-A,  $\omega$  controls signal variance. The exponent of the kernel consists of a linear term and a quadratic term w.r.t. the distance metric  $\underline{d}$ . In the linear term,  $\underline{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^\top$  is a nonnegative vector and interprets concentrations on each circular component. In the second term,  $\underline{\Lambda}$  serves as a weighting matrix in the quadratic formulation and is defined as

$$\underline{\Lambda} = \begin{bmatrix} 0 & a_1 & a_3 \\ a_1 & 0 & a_2 \\ a_3 & a_2 & 0 \end{bmatrix},$$

with all elements being nonnegative. It is specifically designed to capture the correlations between the three pairs of circular components through the quadratic term. It allows for more informative similarity quantification of hypertoroidal data compared to simple products of circular kernels. To showcase this efficacy, we provide the following case study.

**Case Study 2** We configure the proposed kernel in (3) w.r.t. the torus  $\mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^4$  based on four sets of parameters  $\{(\omega_i, \underline{\lambda}_i, \underline{\Lambda}_i)\}_{i=1}^4$ . These are as follows

- $\omega_1 = \omega_2 = \omega_3 = \omega_4 = 1$ ,
- $\underline{\lambda}_1 = \underline{\lambda}_2 = [0.3, 0.3]^\top$ ,  $\underline{\lambda}_3 = \underline{\lambda}_4 = [1, 1]^\top$ , and



**Fig. 2:** Proposed HvM kernels (normalized function values) configured in Case 2 with mappings to tori.

$$\bullet \quad \underline{\Lambda}_1 = \underline{\Lambda}_3 = \mathbf{0}_{2 \times 2}, \quad \underline{\Lambda}_2 = \begin{bmatrix} 0 & 0.3 \\ 0.3 & 0 \end{bmatrix}, \quad \underline{\Lambda}_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

We evaluate the proposed kernel with one input fixed at  $\underline{u} = [1, 0, 1, 0]^\top$ , which corresponds to zero angular positions on each circle, and the other one given by

$$\underline{v} = [\cos(\alpha), \sin(\alpha), \cos(\beta), \sin(\beta)]^\top$$

w.r.t. angles  $(\alpha, \beta)$ . Shown in Fig. 2-(A) to (D), the resulted kernels are plotted corresponding to the parameter sets 1 to 4. In all cases, the kernel indicates the highest similarity at  $(\alpha, \beta) = (0, 0)$ , namely,  $\underline{u} = \underline{v}$ . When  $\underline{\Lambda} = \mathbf{0}_{2 \times 2}$ , as illustrated by Fig. 2-(A) and (C), the proposed hypertoroidal kernel degenerates to a simple multiplication of von Mises kernels, which disregards the correlation between the two circular components. In contrast, the nonzero off-diagonal elements in  $\underline{\Lambda}_2$  and  $\underline{\Lambda}_4$  enable correlation interpretation between circles, leading to more distinguishable similarity quantification of function values as plotted in Fig. 2-(B) and (D), respectively.

### C. Hyperparameter Optimization

Varying the free parameters, e.g., those in the kernels, has a considerable impact on the performance of GP regression and inference. In general, these hyperparameters  $\underline{\theta}$  can be obtained via maximum likelihood estimation following  $\underline{\theta}^* = \arg \max_{\underline{\theta}} \mathcal{F}(\underline{\theta})$ . The objective is derived in the form of log marginal likelihood

$$\begin{aligned} \mathcal{F}(\underline{\theta}) &= 2 \log p(\underline{z} | \{\underline{x}_{\bullet, i}\}_{i=1}^n, \underline{\theta}) \\ &= -\underline{z}^\top \mathbf{K}^{-1} \underline{z} - \log |\mathbf{K}| - n \log(2\pi), \end{aligned} \quad (4)$$

with  $\mathbf{K}$  and  $\underline{z}$  specified according to Sec. III. In practice, the nonlinear optimization problem above is solved numerically with the kernel matrix evaluated in each iteration. The gradient of the objective w.r.t. the  $i$ -th element in  $\underline{\theta}$  takes the following general form

$$\frac{\partial \mathcal{F}(\underline{\theta})}{\partial \theta_i} = \underline{z}^\top \mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \theta_i} \mathbf{K}^{-1} \underline{z} - \text{tr} \left( \mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \theta_i} \right), \quad (5)$$

with  $\text{tr}$  denoting the trace of a matrix.

As for multi-output GPs using the proposed hypertoroidal von Mises kernel, the hyperparameters can be collected into  $\theta = [\omega, \underline{\lambda}^\top, \underline{a}^\top, \underline{b}^\top, \underline{\sigma}_r^\top]^\top$ . The first three components,  $\omega$ ,  $\underline{\lambda}$ , and  $\underline{a} = [a_1, a_2, a_3]^\top$  are free parameters in (3).  $\underline{b} = \text{vec}(\mathbf{B})$  denotes the vectorized coregionalization matrix in (2). And  $\underline{\sigma}_r = [\sigma_{r,1}, \dots, \sigma_{r,d}]^\top$  indicates observation noise variance on each output dimension according to Sec. III-B.

Computing (5) boils down to deriving the derivative of  $\mathbf{K}(\theta)$  w.r.t. hyperparameters in  $\theta$ , which we express element-wise now for the ease of exposition. The derivative of the kernel matrix w.r.t. the signal deviation  $\omega$  follows

$$\frac{\partial \mathbf{K}}{\partial \omega} = \mathbf{B} \otimes \frac{\partial \mathbf{K}_{\bullet\bullet}}{\partial \omega} = \frac{2}{\omega} \mathbf{B} \otimes \mathbf{K}_{\bullet\bullet},$$

where  $\mathbf{K}_{\bullet\bullet}$  denotes the kernel matrix evaluated at training sets as introduced in Sec. III-B. Further, the derivative of  $\mathbf{K}$  w.r.t. each element in the concentration vector  $\underline{\lambda}$  follows

$$\frac{\partial \mathbf{K}}{\partial \lambda_s} = \mathbf{B} \otimes \frac{\partial \mathbf{K}_{\bullet\bullet}}{\partial \lambda_s} = \mathbf{B} \otimes (\mathbf{K}_{\bullet\bullet} \odot \mathbf{D}_{\bullet\bullet}^s),$$

with  $s \in \{1, 2, 3\}$  being the index of the  $s$ -th circular component of the hypertorus. Correspondingly, elements in  $\mathbf{D}_{\bullet\bullet}^s \in \mathbb{R}^{n \times n}$  follow  $(\mathbf{D}_{\bullet\bullet}^s)_{ij} := (\underline{x}_{\bullet,i}^s)^\top \underline{x}_{\bullet,j}^s$ , which interprets distance between training inputs on each circle, and  $\odot$  denotes the Hadamard product. As for differentiation w.r.t. the nonzero elements in matrix  $\mathbf{A}$ , we have

$$\frac{\partial \mathbf{K}}{\partial a_s} = \mathbf{B} \otimes \frac{\partial \mathbf{K}_{\bullet\bullet}}{\partial a_s} = 2\mathbf{B} \otimes (\mathbf{K}_{\bullet\bullet} \odot \mathbf{D}_{\bullet\bullet}^s \odot \mathbf{D}_{\bullet\bullet}^{s \bmod 3 + 1}),$$

with  $s \in \{1, 2, 3\}$ . The derivative of  $\mathbf{K}$  w.r.t. the  $s$ -th element in the vectorized coregionalization matrix takes the form

$$\frac{\partial \mathbf{K}}{\partial b_s} = \frac{\partial (\mathbf{B} \otimes \mathbf{K}_{\bullet\bullet})}{\partial b_s} = \frac{\partial \mathbf{B}}{\partial b_s} \otimes \mathbf{K}_{\bullet\bullet} = \mathbf{E}_{ij} \otimes \mathbf{K}_{\bullet\bullet},$$

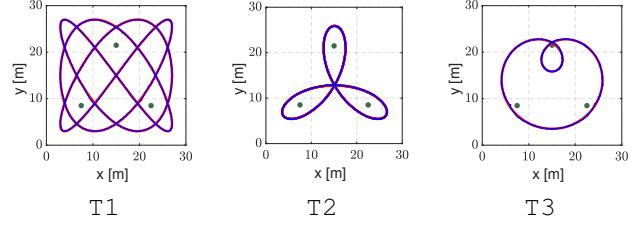
with  $\mathbf{E}_{ij} \in \mathbb{R}^{d \times d}$  denoting a matrix unit. Here,  $b_s = (\mathbf{B})_{ij}$ , with  $s = d(j-1) + i \in \{1, \dots, d^2\}$ . The derivative of  $\mathbf{K}$  w.r.t. observation deviation  $\sigma_{r,s}$  of each output domain  $s \in \{1, \dots, d\}$  can be derived in a similar fashion. It follows

$$\frac{\partial \mathbf{K}}{\partial \sigma_{r,s}} = \frac{\partial (\mathbf{R} \otimes \mathbf{I}_n)}{\partial \sigma_{r,s}} = 2\sigma_{r,s} \mathbf{E}_{ss} \otimes \mathbf{I}_n,$$

where  $\mathbf{E}_{ss} \in \mathbb{R}^{d \times d}$  is a matrix unit. In practice, we utilize the trust-regions method from Manopt [28] for maximizing the objective in (4).

## V. EVALUATION

We now evaluate the proposed HvM-based GPs within particle filtering for data-driven recursive tracking in sensor networks. Here, an unknown range sensing model is learned on the hypertorus, which consists of multiple angle-of-arrival data as input.



**Fig. 3:** Considered trajectories (red) and estimates (blue) given by GP-based particle filtering using the proposed HvM kernel. Noise level is set at  $\xi = 0.01$  (corresponding to sequences of  $\mathbb{S}1$ ). Green dots denote reference points.

### A. Scenario Setup

We set up a two-dimensional area of  $30 \times 30 \text{ m}^2$ , where a mobile agent is tracked while moving along different trajectories. We have the following system dynamics according to a random walk

$$\underline{x}_{t+1} = \underline{x}_t + \underline{w}_t, \quad (6)$$

with  $\underline{x}_t, \underline{x}_{t+1} \in \mathbb{R}^2$  denoting positions and  $\underline{w}_t \in \mathbb{R}^2$  the process noise following  $\underline{w}_t \sim \mathcal{N}(\mathbf{0}_2, \mathbf{Q}_t)$ . A range sensor, e.g., of ultrasonic or ultra-wideband modalities, is equipped onboard the platform observing distances to three reference points of coordinates  $\{\underline{L}_s^r\}_{s=1}^3 \subset \mathbb{R}^2$  under uncertainty. This can be formulated as

$$\underline{z}_t = \underline{h}(\underline{x}_t) + \underline{\delta}_t + \underline{v}_t, \quad (7)$$

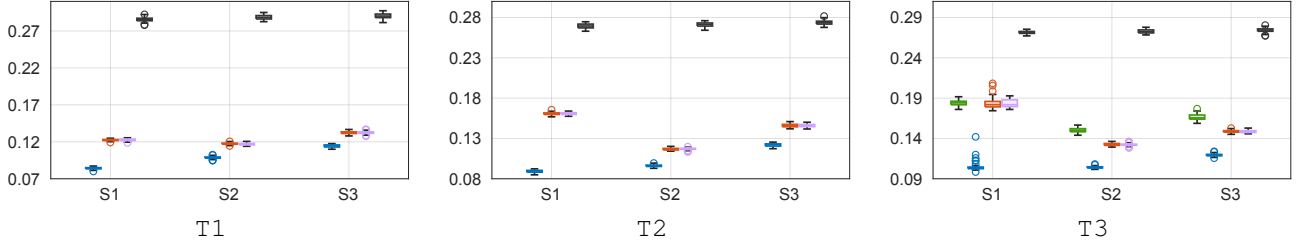
where  $\underline{z}_t \in \mathbb{R}^3$  and  $\underline{v}_t \in \mathbb{R}^3$  denote the range measurement and corresponding noise, respectively, and  $\underline{v}_t \sim \mathcal{N}(\mathbf{0}_3, \mathbf{R}_t)$ . The observation function is given by  $\underline{h}(\underline{x}_t) := [\|\underline{L}_1^r - \underline{x}_t\|, \|\underline{L}_2^r - \underline{x}_t\|, \|\underline{L}_3^r - \underline{x}_t\|]^\top$ , which generates noise-free range readings of position state  $\underline{x}_t$  w.r.t. the three reference points. Moreover, a time-varying offset  $\underline{\delta}_t \in \mathbb{R}^3$  is added to the range signal to include possible interference in sensor networks, such as clock drift or signal reflection [29]. For the sake of demonstration, we assume that it is proportional to the range, namely,  $\underline{\delta}_t = c \cdot \underline{h}(\underline{x}_t)$ , with the ratio  $c = 0.05$ .

### B. Multi-Output GPs on Hypertorus

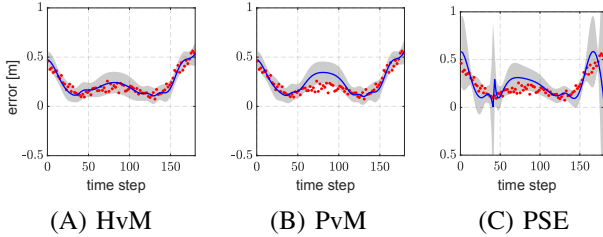
Suppose that the offset and uncertainty in range observations are unknown. However, we are able to obtain accurate angle-of-arrival (AoA) readings from each reference point w.r.t.  $24 \times 10$  grid points  $\{\underline{x}_i\}_{i=1}^{n=240}$  that are uniformly spaced on the  $xy$ -plane. Meanwhile, range measurements given by the onboard sensor are collected, inducing a training set of  $\{(\underline{\alpha}_i, \underline{z}_i)\}_{i=1}^n$ , where the range  $\underline{z}_i \in \mathbb{R}^3$ . The input  $\underline{\alpha}_i = [(\underline{\alpha}_i^1)^\top, (\underline{\alpha}_i^2)^\top, (\underline{\alpha}_i^3)^\top]^\top \in \mathbb{T}^3 \subset \mathbb{R}^6$  incorporates the AoA signal of each reference point w.r.t. grid point  $\underline{x}_i$ , which can be obtained via

$$\underline{\alpha}_i^s = [(\underline{L}_s^r - \underline{x}_i)_x, (\underline{L}_s^r - \underline{x}_i)_y]^\top / \|\underline{L}_s^r - \underline{x}_i\| \in \mathbb{S}^1, \quad (8)$$

with  $s \in \{1, 2, 3\}$ . Such a scenario occurs commonly when there exists no positioning system (e.g., camera networks) covering the whole tracking space, whereas certain portable devices such as total stations can be easily deployed to provide



**Fig. 4:** Results of particle filtering using parametric and different GP-based reweighting schemes in Sec. V-D. RMSEs of the APEs are plotted for the trajectories T1 – T3 using `boxchart` of default setting in `Matlab`. The proposed HvM-based GP (blue) reweighting consistently enables superior tracking accuracy compared to those based on PvM (red), and PPRD (lilac) kernels constructed via kernel product, whereas the parametric model (black) produces inferior performance with a considerable margin. The PSE kernel (green) leads to tracking failures in most sequences due to its nonperiodic definition.



**Fig. 5:** Range measurements (red) and predictions (means and variances plotted by blue curves and gray areas, respectively) w.r.t. ground truth over T2 of noise level S2. The reference point on the top in Fig. 3 is selected for demonstration.

accurate relative angles between two locations [30]. Based on the training set, we exploit the proposed HvM-based vector-valued GPs to learn the unknown range sensing model as introduced in Sec. IV. According to Sec. III-B, we reformulate the measurement model in (7) into  $\mathbf{z}_t = \mathbf{r}(\alpha_t) + \mathbf{v}_t$ , with  $\alpha_t \in \mathbb{T}^3$  incorporating the AoA inputs and  $\mathbf{v}_t$  the uncertainty in the output. Given the scenario setup in Sec. V-A, the noise-free range observation follows  $\mathbf{r}(\alpha_t) = \mathbf{h}(\mathbf{x}_t) + \delta_t = (1+c)\mathbf{h}(\mathbf{x}_t)$ .

### C. GP-Based Particle Filtering

We apply particle filtering to track the mobile agent set up in Sec. V-A [31]. Given a particle  $\hat{\mathbf{x}}_{t|t-1}$  drawn at timestamp  $t$  from the prior estimate (particle index is omitted for brevity), we first compute the corresponding hypertoroidal state  $\hat{\alpha}_{t|t-1} \in \mathbb{T}^3$  following (8). Here, each circular component is given by  $\hat{\alpha}_{t|t-1}^s = [(\mathbf{l}_s^x - \hat{\mathbf{x}}_{t|t-1})_x, (\mathbf{l}_s^x - \hat{\mathbf{x}}_{t|t-1})_y]^\top / \|\mathbf{l}_s^x - \hat{\mathbf{x}}_{t|t-1}\|$  w.r.t. each reference point indexed by  $s \in \{1, 2, 3\}$ . We further inquire the pre-trained GPs (using training data  $\{(\alpha_i, \mathbf{z}_i)\}_{i=1}^n$ ) at  $\hat{\alpha}_{t|t-1}$  to predict corresponding range measurement distribution, i.e.,  $\mathbf{z}_t | \{(\alpha_i, \mathbf{z}_i)\}_{i=1}^n, \hat{\alpha}_{t|t-1} \sim \mathcal{N}(\hat{\mathbf{r}}_{t|t-1}, \mathbf{C}_{t|t-1} + \mathbf{R})$ , where  $\hat{\mathbf{r}}_{t|t-1}$  and  $(\mathbf{C}_{t|t-1} + \mathbf{R})$  are posterior mean and covariance given by the GP regression in Sec. III-B, respectively. Based thereon, the likelihood function is directly obtained for reweighting the prior particle  $\hat{\mathbf{x}}_{t|t-1}$  given current measurement  $\mathbf{z}_t$ . Afterward, particles are updated via resampling.

### D. Evaluation and Results

We equip the GP-based particle filter in Sec. V-C with the proposed HvM kernel. For comparison, we instrument the

product of squared exponential kernels (PSE), the product of periodic kernels (PPRD), and the product of von Mises kernels (PvM) on the hypertorus through multiplication of the corresponding circular kernels [8]. GPs based on different kernels are trained using the same data set collected as introduced in Sec. V-B. Furthermore, we approximate range measurement noise with a three-dimensional Gaussian distribution using the same training data to provide likelihood functions for a parametric form. We set the process noise in (6) with covariance  $\mathbf{Q}_t = \text{diag}(0.16, 0.16)$  and the measurement noise in (7) with covariance  $\mathbf{R}_t = \xi^2 \mathbf{I}_3$ , where  $\xi \in \{0.01, 0.03, 0.05\}$  controls the level of uncertainty. We deploy 100 particles to all particle filters for 1000 time steps and perform evaluation based on 100 Monte Carlo runs for different noise levels and trajectories shown in Fig. 3. Subsequently, we compute the absolute position error (APE) to quantify the tracking accuracy.

Shown in Fig. 4, we collect the root-mean-square error (RMSE) of APEs given by the particle filters configured above on sequences of varying trajectories (T1-T3) and measurement noise levels (S1-S3 w.r.t.  $\xi$ ). The proposed HvM kernel delivers superior tracking accuracy of particle filtering via GP-based reweighting over the PPRD, PvM, PSE kernels, and the parametric modeling method. Exemplary runs of HvM-GP-based particle filtering on these three trajectories are shown in Fig. 3 under the noise level of  $\xi = 0.01$ . Almost identical accuracies are achieved by the PPRD and PvM kernels, which suggests that the von Mises kernel can be interpreted as a reformulation of the common periodic kernel. Moreover, results from these two kernels show better accuracy compared with the parametric method, whereas the PSE kernel fails tracking on most sequences due to disregard of periodic nature of hypertori.

To bring more insights to the functionality of the proposed HvM kernel on particle reweighting, we demonstrate in Fig. 5 range measurement distributions predicted by GPs using HvM, PvM, and PSE kernels over time. The proposed HvM-GP predicts more data-adaptive range distributions than PvM-GP, and the PSE kernel induces discontinuity when periodicity in inputs takes effect. The parametric method fails at modeling the time-varying noise pattern, which can be interpreted in a straightforward way (thus not plotted).

## VI. CONCLUSIONS

We provide a novel study on establishing GPs on the product of directional manifolds. Based on the circular kernel following the form of the von Mises distribution, a novel hypertoroidal von Mises (HvM) kernel has been proposed on the hypertorus  $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \subset \mathbb{R}^6$  in a manifold-adaptive manner. It captures information not only within each circular component but also in the correlated region, leading to a more distinct similarity quantification of hypertoroidal data compared to conventional strategies such as kernel multiplication. In consideration of potential usage in runtime-critical scenarios, derivatives of the marginal likelihood w.r.t. the hyperparameters have been provided to facilitate efficient GP modeling. The proposed HvM-based GP has been evaluated for data-driven recursive localization in ranging-based sensor networks. Simulations have shown that it delivers superior tracking performance over the parametric model and GPs based on kernel multiplications.

For future work, the efficiency of the proposed method can be improved by reducing the computational complexity, e.g., via sparse GPs [32]. We will also present the proof on the positive definiteness of the proposed kernel in a follow-up work. In addition, the proposed HvM-GP is to be exploited in real-world applications for data-driven modeling of complex stochastic systems with multiple angular inputs [33]. Also, the HvM kernel can be extended to higher-dimensional spaces, enabling its application in more extensive scenarios.

## ACKNOWLEDGMENT

This work was supported in part by the German Research Foundation (DFG) under grant HA 3789/25-1, in part by the Swedish Research Council under grant Scalable Kalman Filters, and in part by ZENITH of Linköping University under grant CASIIS. The work was partially conducted at Karlsruhe Institute of Technology (KIT), Germany.

## REFERENCES

- [1] N. Ito, S. Araki, and T. Nakatani, "Complex angular central Gaussian mixture model for directional statistics in mask-based microphone array signal processing," in *European Signal Processing Conference*, Aug. 2016, pp. 1153–1157.
- [2] J. Glover and L. P. Kaelbling, "Tracking the spin on a ping pong ball with the quaternion Bingham filter," in *IEEE International Conference on Robotics and Automation*, June 2014, pp. 4133–4140.
- [3] X. Zhe, S. Chen, and H. Yan, "Directional statistics-based deep metric learning for image classification and retrieval," *Pattern Recognition*, vol. 93, pp. 113–123, 2019.
- [4] H. Möls, K. Li, and U. D. Hanebeck, "Highly parallelizable plane extraction for organized point clouds using spherical convex hulls," in *IEEE International Conference on Robotics and Automation*, May 2020.
- [5] K. Li, F. Pfaff, and U. D. Hanebeck, "Grid-based quaternion filter for SO(3) estimation," in *European Control Conference*, May 2020.
- [6] —, "Circular discrete reapproximation," in *International Conference on Information Fusion*, July 2022.
- [7] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [8] D. Duvenaud, "Automatic model construction with Gaussian processes," Ph.D. dissertation, University of Cambridge, 2014.
- [9] F. Lindgren, H. Rue, and J. Lindt, "An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 4, pp. 423–498, 2011.
- [10] V. Borovitskiy, A. Terenin, P. Mostowsky, and P. D. Marc, "Matérn Gaussian processes on Riemannian manifolds," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 426–12 437, 2020.
- [11] M. Hutchinson, A. Terenin, V. Borovitskiy, S. Takao, Y. Teh, and M. Deisenroth, "Vector-valued Gaussian processes on Riemannian manifolds via gauge independent projected kernels," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17 160–17 169, 2021.
- [12] A. Feragen, F. Lauze, and S. Hauberg, "Geodesic exponential kernels: When curvature and linearity conflict," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2015, pp. 3032–3042.
- [13] A. Majumdar, "Gaussian processes on the support of cylindrical surfaces, with application to periodic spatio-temporal data," *Journal of Statistical Planning and Inference*, vol. 153, pp. 27–41, 2014.
- [14] M. Lang, O. Dunkley, and S. Hirche, "Gaussian process kernels for rotations and 6D rigid body motions," in *IEEE International Conference on Robotics and Automation*, June 2014, pp. 5165–5170.
- [15] K. V. Mardia, P. E. Jupp, and K. Mardia, *Directional Statistics*. John Wiley & Sons, 2000.
- [16] R. Fisher, "Dispersion on a sphere," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 217, no. 1130, pp. 295–305, 1953.
- [17] K. Li, F. Pfaff, and U. D. Hanebeck, "Progressive von Mises–Fisher filtering using isotropic sample sets for nonlinear hyperspherical estimation," *Sensors*, vol. 21, no. 9, p. 2991, 2021.
- [18] —, "Nonlinear von Mises–Fisher filtering based on isotropic deterministic sampling," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sept. 2020.
- [19] I. Gilitschenski, R. Sahoo, W. Schwarting, A. Amini, S. Karaman, and D. Rus, "Deep orientation uncertainty learning based on a Bingham loss," in *International Conference on Learning Representations*, Apr. 2020.
- [20] K. Li, D. Frisch, B. Noack, and U. D. Hanebeck, "Geometry-driven deterministic sampling for nonlinear Bingham filtering," in *European Control Conference*, June 2019.
- [21] K. Li, "On-manifold recursive Bayesian estimation for directional domains," Ph.D. dissertation, Karlsruhe Institute of Technology, 2022.
- [22] K. Li, F. Pfaff, and U. D. Hanebeck, "Dual quaternion sample reduction for SE(2) estimation," in *International Conference on Information Fusion*, July 2020.
- [23] G. Kurz and U. D. Hanebeck, "Toroidal information fusion based on the bivariate von Mises distribution," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sept. 2015, pp. 309–315.
- [24] S. Bultmann, K. Li, and U. D. Hanebeck, "Stereo visual SLAM based on unscented dual quaternion filtering," in *International Conference on Information Fusion*, July 2019, pp. 1–8.
- [25] K. Li, F. Pfaff, and U. D. Hanebeck, "Unscented dual quaternion particle filter for SE(3) estimation," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 647–652, 2021.
- [26] —, "Geometry-driven stochastic modeling of SE(3) states based on dual quaternion representation," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, May 2019.
- [27] M. A. Alvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012.
- [28] 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, no. 42, pp. 1455–1459, 2014.
- [29] M. Kok, J. D. Hol, and T. B. Schön, "Indoor positioning using ultrawideband and inertial measurements," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1293–1303, 2015.
- [30] Y. Oshman and P. Davidson, "Optimization of observer trajectories for bearings-only target localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 892–902, 1999.
- [31] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forsell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, 2002.
- [32] M. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *International Conference on Artificial Intelligence and Statistics*, Apr. 2009, pp. 567–574.
- [33] M. Cai, M. Hasanbeig, S. Xiao, A. Abate, and Z. Kan, "Modular deep reinforcement learning for continuous motion planning with temporal logic," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7973–7980, 2021.