Hyperspherical Deterministic Sampling Based on Riemannian Geometry for Improved Nonlinear Bingham Filtering

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@kit.edu, florian.pfaff@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—We present a novel geometry-driven scheme for generating equally weighted deterministic samples of Bingham distributions in arbitrary dimensions. Unlike existing approaches, our method provides flexibility in the sampling size with samples satisfying requirements of the unscented transform while approximating higher-order moments of the Bingham distribution. This is done by first using Dirac mixture approximation as a sampling scheme on the tangent plane at the mode with respect to the Bingham density via gnomonic projection. Subsequently, the tangent sigma points are retracted backwards to the hypersphere, after which an on-manifold moment correction is performed via Riemannian optimization. The proposed approach is further applied to quaternion Bingham filtering for recursive orientation estimations. Evaluation results show that the geometry-adaptive sampling scheme gives better tracking accuracy and robustness for nonlinear orientation estimations.**

I. Introduction

Recursive orientation estimation plays a crucial rule in various applications such as robotic pose estimation [1]–[3], localization and mapping [4] as well as multilateration [5], etc. However, due to the inherently periodic and nonlinear structure of the special orthogonal group SO(3), robust and accurate orientation estimations are nontrivial. Moreover, various options for parameterizing orientations exist. The Euler angles, for instance, enable minimal representations. However, they suffer from ambiguity caused by gimbal lock. The well-known 3×3 rotation matrices eliminate the ambiguity via over-parameterization, but introduce a high degree of redundancy (9 elements used for representing 3 DoF), potentially causing numerical instabilities. In contrast, unit quaternions parameterize the spatial orientation without ambiguity and with only one degree of redundancy. Thus, they have been employed for pose representation in various robotic tracking tasks [6], [7].

As unit quaternions are located on the hypersphere $\mathbb{S}^3$, conventional quaternion filters for orientation estimations rely on local linearizations of the nonlinear manifold (e.g., via Lie algebra) [7]. However, this is based on the assumption of local perturbations and might be error-prone for fast rotations under large uncertainty and nonlinearity. By utilizing distributions from directional statistics [8], recent efforts were dedicated to model the uncertainty of unit quaternions directly on the hypersphere without local linearization. For instance, the Bingham distribution on $\mathbb{S}^3$ inherently enables the on-manifold modeling and filtering of unit quaternions for orientation estimations [9]–[11].

However, filtering approaches using the Bingham distribution usually rely on sampling-approximation schemes, in which the approximation is based on the second-order moment matching [9], [12]. Compared to random sampling schemes, deterministic approaches guarantee reproducible results and are in general more efficient [13]. Taking the idea of the unscented transform (UT) [14], the conventional unscented quaternion filter uses sigma points that approximate the Bingham distribution up to the second order. However, the fixed and limited sample size[1] might pose risks for performing filtering under large uncertainty and nonlinearity due to the sample degeneration issue [12]. Therefore, we need more samples drawn adaptively according to the information geometry of the underlying density while satisfying requirements of the UT for moment matching. Moreover, samples with uniform weights are preferred as they can equally contribute during the nonlinear filtering.

In order to flexibly generate a given numbers of samples in the sense of information geometry [15] for nonlinear filtering, efforts were first devoted to Gaussian distributions in the Euclidean space through Dirac mixture approximation (DMA) [16]–[18]. The basic methodology here is to minimize a certain distance measure between the Dirac mixture density described by the samples and the underlying density under the constraint of second-order moment. The sampling scheme is then formulated as a constrained optimization problem and solved by using standard optimizers (e.g., quasi-Newton method). In [18], samples were drawn in the principal axes for DMA by minimizing the Cramér–von Mises distance. In [19], a modified Cramér–von Mises distance adapted to the localized cumulative distribution (LCD) was proposed for the DMA of multivariate Gaussian distributions. Here, the deterministic samples are not restricted to the principal axes but can be drawn in the entire domain.

[1] In the context of unscented orientation filtering, 3 and 5 samples are drawn on $\mathbb{S}^1$ (planar case) and $\mathbb{S}^3$ (spatial case), respectively.

Driven by hyperspherical geometry, the first geometry-adaptive sampling scheme for the Bingham distribution was proposed in [20]. Though improved accuracy and robustness have been shown for nonlinear orientation filtering, the samples are still drawn in the principal directions. In this paper, we propose a novel hyperspherical sampling scheme to better approximate higher-order moments of the Bingham distribution. Here, we apply the LCD-based Dirac mixture approximation as a sampling scheme on the tangent plane at the mode with respect to the underlying Bingham density via the gnomonic projection. The tangent sigma points are then retracted to the manifold followed by a moment correction of the second order to satisfy requirements of the UT. The moment correction is formulated as a manifold optimization problem and solved by the Riemannian steepest descent algorithm [21]. More specifically, we summarize our main contributions as follows:

- A generic deterministic sampling scheme is proposed for Bingham distributions in arbitrary dimensions with equally weighted samples of flexible size.
- The samples drawn on the hypersphere (not only in principal directions) satisfy requirements of the unscented transform and give approximation to higher-order moments of the Bingham distribution.
- By using the proposed sampling scheme, better accuracy and robustness are shown in nonlinear Bingham filtering for orientation estimations.

The remainder of the paper is structured as follows. In Sec. II, preliminaries about quaternion-based orientation representation and the Bingham distribution as well as the LCD-based DMA are introduced. In Sec. III, we propose the novel deterministic sampling approach for the Bingham distribution based on Riemannian geometry. The proposed sampling scheme is further evaluated for nonlinear Bingham filtering in Sec. IV. The work is finally concluded in Sec. V.

II. Preliminaries

A. Unit Quaternion Parameterization for Spatial Orientations

By convention, orientation states belonging to the special orthogonal group SO(3) can be represented by the unit quaternion as follows

$$\mathbf{x} = \left[\, \cos(\theta/2),\, \underline{u}^\top \sin(\theta/2) \,\right]^\top \,, \tag{1}$$

with the unit vector $\underline{u}$ being the rotation axis and θ the rotation angle. Any vector $\underline{v} \in \mathbb{R}^3$ can be rotated around axis $\underline{u}$ of degree θ by (1) according to

$$\underline{v}' = \mathbf{x} \otimes \underline{v} \otimes \mathbf{x}^* \,, \tag{2}$$

with $\otimes$ denoting the Hamilton product and $\mathbf{x}^*$ the conjugate of $\mathbf{x}$ [5], [22], namely[2] $\mathbf{x}^* = \mathrm{diag}\,(1, -1, -1, -1)\,\mathbf{x}$. Unit quaternions are thus naturally located on the hypersphere $\mathbb{S}^3 \subset \mathbb{R}^4$ because of the unit norm constraint. Furthermore, two antipodal quaternions on the hypersphere, i.e., $\mathbf{x}$ and $-\mathbf{x}$, denote the same rotation as shown in (2). Therefore, a density f for stochastically modeling uncertain unit quaternions should be antipodally symmetric on the hypersphere, such that $f(\mathbf{x}) = f(-\mathbf{x})$.

[2] $\mathrm{diag}(\cdot)$ maps each entry to the diagonal element of the diagonal matrix.

B. Bingham Distribution

Intuitively, the Bingham distribution on $\mathbb{S}^n$ can be derived by constraining a zero-mean Gaussian distribution of $\mathbb{R}^{n+1}$ on the hypersphere followed by a re-normalization. Traditionally, it is defined as follows

$$f_{\mathcal{B}}(\mathbf{x}; \mathbf{Z}, \mathbf{M}) = \frac{1}{N(\mathbf{Z})} \exp\left(\mathbf{x}^\top \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top \mathbf{x}\right) , \; \mathbf{x} \in \mathbb{S}^n \,, \tag{3}$$

with the diagonal matrix $\mathbf{Z}$ determining the concentration as well as the normalization constant $N(\mathbf{Z})$, and the real orthogonal matrix $\mathbf{M}$ controlling the orientation of the distribution on the hypersphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$. The parameters $\mathbf{Z}$ and $\mathbf{M}$ can be derived via eigendecomposition of a negative semidefinite matrix $\mathbf{C}_{\mathcal{B}} \in \mathbb{R}^{(n+1)\times(n+1)}$. Afterwards, elements of $\mathbf{Z}$ are by convention adjusted in ascending order, namely $\mathbf{Z} = \mathrm{diag}(z_1, \dots, z_n, z_0)$, with $z_1 < \dots < z_n < z_0 \leq 0$, followed by a reordering of the column vectors in $\mathbf{M}$ accordingly. The mode of the Bingham distribution is thus denoted by the last column of $\mathbf{M}$ as it corresponds to the largest eigenvalue in the space $\mathbb{R}^{n+1}$. Furthermore, the Bingham distribution is inherently antipodally symmetric on $\mathbb{S}^n$ since $f_{\mathcal{B}}(\mathbf{x}) = f_{\mathcal{B}}(-\mathbf{x})$. For orientation estimations using unit quaternions, the Bingham distribution on $\mathbb{S}^3$ and $\mathbb{S}^1$ can therefore be employed for stochastic modeling uncertain spatial and planar orientations, respectively.

C. Dirac Mixture Approximation for Multivariate Gaussian Distribution Based on the Localized Cumulative Distribution

In [19], [23], multivariate Gaussian distributions were approximated with the following Dirac mixture

$$f(\mathbf{x}) = \sum_{j=1}^{m} w_i \cdot \delta(\mathbf{x} - \mathbf{x}_j) \,, \tag{4}$$

with $\delta(\cdot)$ being the Dirac delta function. Here, $\mathbf{x}_j$ denotes the location of each Dirac component and $w_i \in [0, 1]$ the weighting factor of each sample (for uniformly weighted samples $w_i = 1/m$). For an arbitrarily given probability density function $g : \mathbb{R}^n \to \mathbb{R}_+$, its localized cumulative distribution (LCD) is defined as follows

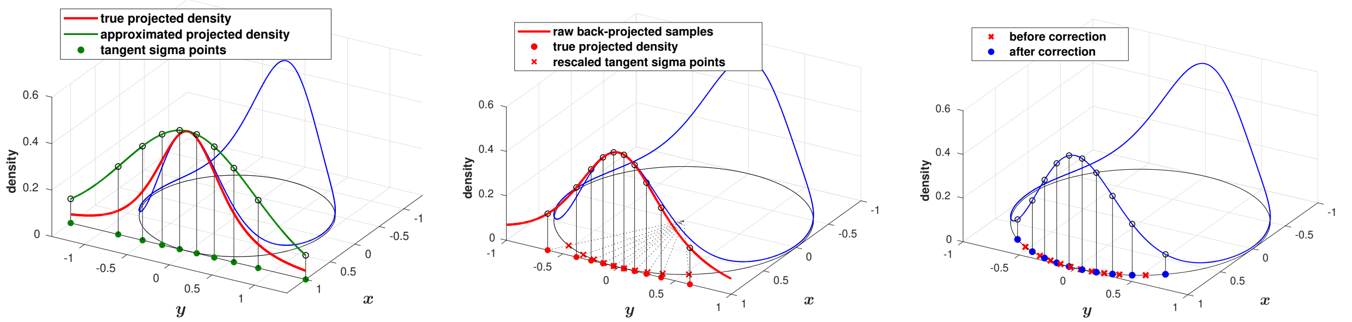
$$\mathcal{F}(\mathbf{r}, \mathbf{b}) = \int_{\mathbb{R}^n} g(\mathbf{x}) \cdot \kappa(\mathbf{x} - \mathbf{r}, \mathbf{b}) \, \mathrm{d}\mathbf{x} \,, \tag{5}$$

with $\kappa(\mathbf{x} - \mathbf{r}, \mathbf{b})$ denoting a suitable kernel (symmetric and integrable) that locally measures the density at $\mathbf{m}$ in a range of $\mathbf{b}$. In [19], [24], axis-wise separable kernels were used, i.e.,

$$\kappa(\mathbf{x} - \mathbf{r}, \mathbf{b}) = \prod_{i=1}^{n} \kappa(x_i - r_i, b_i) \,,$$

with each kernel suggested to be of Gaussian type with isotropic dispersion ($\mathbf{b}$ can be written as a scalar value), thus

$$\kappa(\mathbf{x} - \mathbf{r}, b) = \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2} \frac{(x_i - r_i)^2}{b^2} \right\} .$$



(a) (b) (c)

Fig. 1: Illustration of the proposed deterministic sampling approach on unit circle $\mathbb{S}^1$. Here, a Bingham distribution with $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 5)$ is used and $m = 10$ samples are drawn with the following procedure: (a) On-tangent-plane deterministic sampling using the LCD-based DMA to the approximated projected density (Alg. 1, line 5), (b) gnomonic retraction of the tangent sigma points after rescaling (Alg. 1, line 7-8), and (c) on-manifold second-order moment correction (Alg. 1, line 10). The proposed RSS approach can therefore guarantee the first two moments while approximating higher-order moments of the Bingham distribution.

Based on the LCD defined in (5), a modified Cramér–von Mises distance is proposed to measure the distance between the Dirac mixture $f(\mathbf{x})$ given in (4) and the underlying Gaussian distribution $\hat{f}(\mathbf{x})$. This distance measure is given in the following form

$$\mathcal{D} = \int_{\mathbb{R}+} w(b) \int_{\mathbb{R}^n} \left(\hat{\mathcal{F}}(\mathbf{r}, b) - \mathcal{F}(\mathbf{r}, b) \right) \mathrm{d}\mathbf{r} \, \mathrm{d}b \,, \qquad (6)$$

with $\hat{\mathcal{F}}(\mathbf{m}, b)$ and $\mathcal{F}(\mathbf{m}, b)$ being the LCD of the underlying Gaussian distribution and the Dirac mixture, respectively. Here, $w(b)$ is a weighting function determined by the range factor for the kernel and is given by

$$w(b) = \begin{cases} b^{1-n} \,, & b \in [\, 0, b_{\max} \,] \\ 0 \,, & \text{elsewhere} \end{cases} .$$

As proposed in [19], the Dirac mixture approximation to a multivariate Gaussian distribution can thus be formulated as a constrained optimization problem, where the distance measure in (6) is minimized and the second-order moment constraint should be satisfied.

The aforementioned LCD-based DMA for multivariate Gaussian distribution can be efficiently done with closed-form gradient and error term in either an online or offline manner [19]. In this way, deterministic samples of arbitrary size that satisfy the second-order moment constraint and approximate the higher-order moments can be generated. For nonlinear recursive estimation in the Euclidean space, the LCD-based adaptive sampling scheme has facilitated the development of trackers with better accuracy and higher robustness [25].

III. Riemannian Spherical Sampling (RSS)

For the DMA-based sampling approach, a proper kernel function should be defined on the domain of the distribution, based on which a certain error metric between the Dirac mixture and the underlying density can be formulated and minimized [17], [19]. However, this scheme cannot be trivially applied on manifolds. The kernel function used to evaluate the density in the Euclidean space is normally an isotropic Gaussian. On the hypersphere, the Gaussian kernel should be modified according to the geometric structure of the nonlinear domain. A Bingham or von Mises–Fisher kernel can be employed for this purpose. However, the subsequent error metric formulation normally refers to an integral on the nonlinear domain. For Bingham distributions, this boils down to an integral over the hypersphere, which leads to a similar or even larger computational cost than calculating the normalization constant [26]. This is, obviously, against the spirit of pursuing good efficiency via deterministic sampling for recursive estimations.

In order to alleviate the computational burden, we exploit hyperspherical geometry to enable the LCD-based DMA on the tangent plane at the Bingham mode. As pointed out in [20], projections or the logarithm map can be employed to derive the on-tangent-plane density from the underlying hypersphere. We hereby apply the gnomonic projection [27], as it results in unbounded domain of the projected Bingham density on the tangent plane. This is coherent to the definition of the LCD and paves the way to retract tangent sigma points back to the hypersphere.

A. Gnomonic Projection

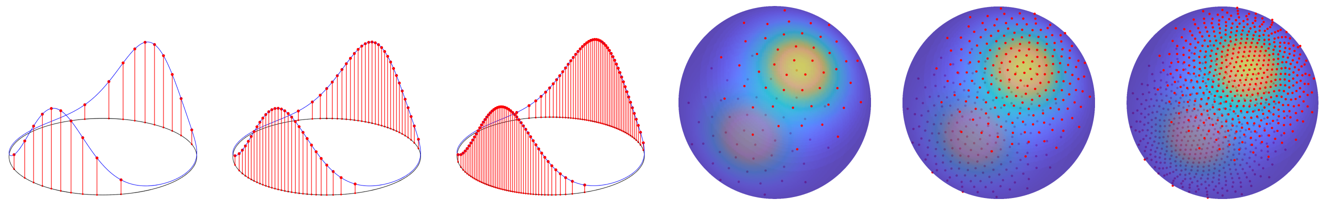
As the parameter matrix $\mathbf{M}$ of the Bingham distribution in (3) is orthogonal, the first n columns of matrix $\mathbf{M}$ provide an orthonormal basis of the tangent plane at the mode of the Bingham distribution, namely

$$\mathbb{T}_\mathbf{m}\mathbb{S}^n = \mathrm{span}\{\mathbf{m}_1, \dots, \mathbf{m}_n\} \,,$$

with $\mathbf{M} = [\, \mathbf{m}_1, \dots, \mathbf{m}_n, \mathbf{m} \,] \in \mathbb{R}^{(n+1)\times(n+1)}$ and $\mathbf{m}$ being the mode. For better readability, we denote the tangent local basis as $\mathbf{E} = [\, \mathbf{m}_1, \dots, \mathbf{m}_n \,] \in \mathbb{R}^{(n+1)\times n}$ such that $\mathbf{M} = [\, \mathbf{E}\,, \mathbf{m} \,]$. For any $\mathbf{x}_\mathrm{t} \in \mathbb{T}_\mathbf{m}\mathbb{S}^n \subset \mathbb{R}^n$, its coordinate expressed in the local basis $\mathbf{E}$ is thus

$$\tilde{\mathbf{x}}_\mathrm{t} = \mathbf{E}^\top \mathbf{x}_\mathrm{t} \,. \qquad (7)$$

For any $\mathbf{x} \in \mathbb{S}^n$, it can be mapped from the hypersphere to the point $\mathbf{x}_\mathrm{t}$ on the tangent plane via the gnomonic projection



$\mathcal{E} = 2.7 \times 10^{-13}$, $m = 10$ | $\mathcal{E} = 4.0 \times 10^{-12}$, $m = 30$ | $\mathcal{E} = 1.4 \times 10^{-12}$, $m = 50$ | $\mathcal{E} = 4.2 \times 10^{-13}$, $m = 50$ | $\mathcal{E} = 1.0 \times 10^{-11}$, $m = 200$ | $\mathcal{E} = 6.0 \times 10^{-11}$, $m = 500$

(a) Bingham distribution on $\mathbb{S}^1$ (left) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 5)$ and on $\mathbb{S}^2$ (right) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 5, 5)$.

$\mathcal{E} = 8.6 \times 10^{-15}$, $m = 10$ | $\mathcal{E} = 1.7 \times 10^{-11}$, $m = 30$ | $\mathcal{E} = 2.4 \times 10^{-11}$, $m = 50$ | $\mathcal{E} = 2.8 \times 10^{-12}$, $m = 50$ | $\mathcal{E} = 4.6 \times 10^{-11}$, $m = 200$ | $\mathcal{E} = 2.1 \times 10^{-11}$, $m = 500$

(b) Bingham distribution on $\mathbb{S}^1$ (left) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 10)$ and on $\mathbb{S}^2$ (right) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 5, 10)$.

$\mathcal{E} = 5.7 \times 10^{-13}$, $m = 10$ | $\mathcal{E} = 5.8 \times 10^{-12}$, $m = 30$ | $\mathcal{E} = 5.7 \times 10^{-12}$, $m = 50$ | $\mathcal{E} = 1.9 \times 10^{-11}$, $m = 50$ | $\mathcal{E} = 3.2 \times 10^{-11}$, $m = 200$ | $\mathcal{E} = 4.2 \times 10^{-10}$, $m = 500$

(c) Bingham distribution on $\mathbb{S}^1$ (left) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 30)$ and on $\mathbb{S}^2$ (right) with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 5, 40)$.

Fig. 2: Deterministic sampling results given by the proposed approach for different parameter configurations.

$\mathcal{P}_\mathbf{m} : \mathbb{S}^n \to \mathbb{T}_\mathbf{m}\mathbb{S}^n$ [20]. Its inverse operation is essentially a retraction that project $\mathbf{x}_\mathrm{t}$ from the tangent plane back to the hypersphere, namely $\mathcal{R}_\mathbf{m} : \mathbb{T}_\mathbf{m}\mathbb{S}^n \to \mathbb{S}^n$, according to

$$\mathcal{R}_\mathbf{m}(\mathbf{x}_\mathrm{t}) = \frac{\mathbf{x}_\mathrm{t} + \mathbf{m}}{\sqrt{1 + \|\mathbf{x}_\mathrm{t}\|^2}} \,. \tag{8}$$

B. On-Tangent-Plane Approximation to On-Manifold Density

In order to perform LCD-based DMA on the tangent plane at the mode, we use the gnomonic projection to map the density from the hypersphere to the tangent space at the mode. We therefore substitute $\mathbf{x}$ in (3) by its projected point via the gnomonic retraction given in (8), such that the projected Bingham can be derived as follows

$$\begin{aligned} f(\mathbf{x}_\mathrm{t}) &= f_\mathcal{B}(\mathcal{R}_\mathbf{m}(\mathbf{x}_\mathrm{t})) \\ &= \frac{1}{N(\mathbf{Z})} \exp\left\{ \frac{(\mathbf{x}_\mathrm{t} + \mathbf{m})^\top \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top (\mathbf{x}_\mathrm{t} + \mathbf{m})}{1 + \|\mathbf{x}_\mathrm{t}\|^2} \right\} \\ &= \frac{1}{N(\mathbf{Z})} \exp\left\{ \frac{\mathbf{x}_\mathrm{t}^\top \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^\top \mathbf{x}_\mathrm{t}}{1 + \|\mathbf{x}_\mathrm{t}\|^2} \right\} . \end{aligned}$$

As the last column of $\mathbf{M}$ is orthogonal to the tangent plane $\mathbb{T}_\mathbf{m}\mathbb{S}^n$ and the last diagonal element of $\mathbf{Z}$ is set to be zero, we can reformulate the Bingham density under gnomonic projection with respect to the local coordinate on $\mathbb{T}_\mathbf{m}\mathbb{S}^n$ into the following form

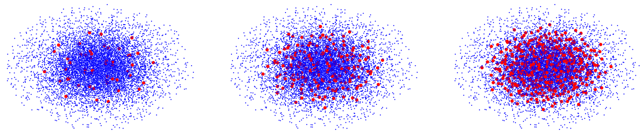
$$\begin{aligned} & f(\tilde{\mathbf{x}}_\mathrm{t}) = \frac{1}{\beta} \exp\left\{ \frac{\tilde{\mathbf{x}}_\mathrm{t}^\top \, \mathbf{Z}_\mathrm{t} \, \tilde{\mathbf{x}}_\mathrm{t}}{1 + \|\tilde{\mathbf{x}}_\mathrm{t}\|^2} \right\} , \quad \text{with} \\ & \tilde{\mathbf{x}}_\mathrm{t} = \mathbf{E}^\top \mathbf{x}_\mathrm{t} \in \mathbb{R}^n \,, \quad \mathbf{Z}_\mathrm{t} = \mathrm{diag}(z_1, \dots, z_n) \in \mathbb{R}^{n \times n} \,. \end{aligned} \tag{9}$$

Here, $\beta = N(\mathbf{Z})$ denotes the Bingham normalization constant.

The function in (9) provides the density value at $\tilde{\mathbf{x}}_\mathrm{t}$ according to the underlying Bingham on the hypersphere via the gnomonic projection. However, to perform deterministic sampling using the LCD-based DMA approach introduced in Sec. II-C, the density should be converted to or approximated by a Gaussian distribution. Essentially, the denominator in (9) denotes the squared cosine of the angle induced by the projection line, namely $\cos(\alpha)^2 = 1/(1 + \|\tilde{\mathbf{x}}_\mathrm{t}\|)^2$, with $\alpha = \mathbf{x}^\top \mathbf{m}$. By applying Taylor expansion to α, the density can be approximated by $f(\tilde{\mathbf{x}}_\mathrm{t}) \approx \frac{1}{\beta} \exp\left\{ \tilde{\mathbf{x}}_\mathrm{t}^\top \, \mathbf{Z}_\mathrm{t} \, \tilde{\mathbf{x}}_\mathrm{t} \right\}$. This density can be further converted into an isotropic zero-mean Gaussian by imposing the normalization constant β to be equal to the covariance

$$\mathbf{\Omega} = \frac{\beta^{2/n}}{2\pi} \mathbf{I} \,, \tag{10}$$

such that $\beta = \sqrt{\det(2\pi\mathbf{\Omega})}$. We can therefore perform the LCD-based sampling approach introduced in Sec. II-C to



$\mathcal{E} = 7.7 \times 10^{-13}$ $\quad m = 50$ $\qquad \mathcal{E} = 2.6 \times 10^{-13}$ $\quad m = 200$ $\qquad \mathcal{E} = 1.2 \times 10^{-13}$ $\quad m = 500$

Fig. 3: Samples (red dots) given by the proposed approach for the Bingham distribution on $\mathbb{S}^3$ with $\mathbf{C}_\mathcal{B} = \mathrm{diag}(1, 5, 5, 5)$. We draw random samples (blue dots) to illustrate the density. For the purpose of visualization, both the deterministic and random samples are mapped to the tangent plane $\mathbb{T}_\mathbf{m}\mathbb{S}^3$ via the spherical logarithm map [28].

generate sigma points $\boldsymbol{\sigma}_\mathbf{u}$ from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ through the following substitution

$$\mathbf{u} = \sqrt{-\frac{\beta^{2/n}}{\pi} \mathbf{Z}_\mathrm{t}}\, \tilde{\mathbf{x}}_\mathrm{t} \,. \tag{11}$$

C. On-Manifold Moment Correction

To obtain deterministic samples on the hypersphere $\mathbb{S}^n$, the sigma points $\{\boldsymbol{\sigma}_j^\mathrm{t}\}_{j=1}^m$ generated with respect to the local basis on $\mathbb{T}_\mathbf{m}\mathbb{S}^n$ should be first converted to global coordinates and then undergo the gnomonic retraction

$$\boldsymbol{\sigma}_j = \mathcal{R}_\mathbf{m}(\mathbf{E}\,\boldsymbol{\sigma}_j^\mathrm{t}) = \frac{\mathbf{E}\,\boldsymbol{\sigma}_j^\mathrm{t} + \mathbf{m}}{\sqrt{1 + \|\boldsymbol{\sigma}_j^\mathrm{t}\|^2}} \,. \tag{12}$$

However, these raw retracted points do not necessarily satisfy the requirements of the unscented transform. Here, the first moment (zero) can be easily maintained by mirroring the samples to the opposite hemi-hypersphere. The second-order moment, however, should be explicitly corrected. In [25], [29], the moment correction was done by applying a Cholesky decomposition to the covariance. This correction method works efficiently for the Euclidean space but cannot be trivially applied to nonlinear domains. In the case of Bingham sampling, the samples should always be located on the hypersphere $\mathbb{S}^n$ before and after the correction. The transformation given by the Cholesky decomposition, however, cannot guarantee the additional constraints imposed by the curved manifold structure. We therefore propose to formulate the moment correction as the following optimization problem

$$\begin{aligned} \{\boldsymbol{\sigma}_j^*\}_{j=1}^m &= \underset{\{\boldsymbol{\sigma}_j\}_{j=1}^m}{\arg\min} \left\| \frac{1}{m} \boldsymbol{\sigma}_j \boldsymbol{\sigma}_j^\top - \mathrm{cov}_\mathcal{B} \right\|_\mathrm{F}^2 , \\ \text{s.t.} \quad & \boldsymbol{\sigma}_j \in \mathbb{S}^n \,, j = 1, \dots, m \,, \end{aligned} \tag{13}$$

where the distance between the covariance of the sigma points and the Bingham second-order moment is minimized in the sense of least Frobenius norm.

The objective function in (13) is non-convex and normally has multiple global minima especially for large sample sizes. Applying standard solvers for constrained optimization problems (e.g., with a Lagrange multiplier) usually shows non-ideal convergence behavior. We therefore employ the Riemannian optimization approach [21], in which the manifold structure is explicitly considered in each iteration. As Riemannian optimizers are geometry-aware, a better convergence speed and robustness can be expected [21].

Moreover, to let the samples after moment correction maintain the higher-order shape information given by the LCD-based DMA as much as possible, the location change of each sample due to the correction should also be restricted. We thus recommend to use the Riemannian steepest descent (RSD) algorithm with raw retracted sigma points as the initialization. Another solution to overcome convergence to a global minimum of large distortion is to add a penalty term to (13), such that

$$\mathcal{E}(\{\boldsymbol{\sigma}_j\}_{j=1}^n) = \left\| \frac{1}{m} \boldsymbol{\sigma}_j \boldsymbol{\sigma}_j^\top - \mathrm{cov}_\mathcal{B} \right\|_\mathrm{F}^2 + \frac{\lambda}{2\,m} \sum_{j=1}^m \boldsymbol{\sigma}_j^\top \boldsymbol{\sigma}_j^\mathrm{ini} \,.$$

The second term in the equation above denotes a scaled (by λ) average location change on the hypersphere compared with the initialization. For this penalty term, the distance metric from von Mises-Fisher distribution is used[3]. Fig. 5 further illustrates the effect given by different combinations of objective function formulations and Riemannian optimizers. Here, the Bingham distribution is parameterized by $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 8, 70)$ and $m = 150$ samples are first drawn on the tangent plane $\mathbb{T}_\mathbf{m}\mathbb{S}^2$ using the LCD-based DMA introduced in (III-B). All the combinations converge to numerically global minima defined in (13). When using the Riemannian trust region approach for on-manifold moment correction, a penalty term is needed in the objective function. However, this is not necessary when applying the RSD, as samples are updated with limited location changes in each iteration, where only the nearest local minimum is to be reached given the gradient descent direction.

Algorithm 1 Riemannian Spherical Sampling

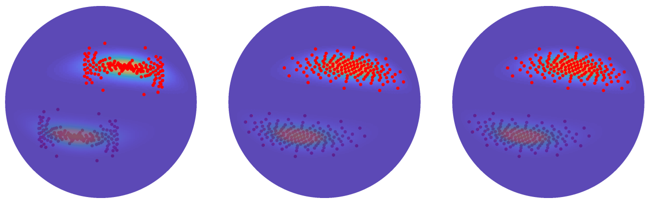
procedure RSS $(f_\mathrm{B}\,, m)$
- $n \leftarrow$ getHypersphereDimension $(f_\mathcal{B})$;
- $\beta \leftarrow$ getNormalizatonConstant $(f_\mathcal{B})$;
- $\mathbf{E} \leftarrow$ getLocalCoord $(f_\mathcal{B})$;
- $\boldsymbol{\Omega} \leftarrow \frac{\beta^{2/n}}{2\pi} \mathbf{I}$; // see (10)
- $\{\mathbf{u}_j\}_{j=1}^m \leftarrow$ LCD-DMA $(\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})\,, m)$;
- **for** $j = 1$ to m **do**
  - $\boldsymbol{\sigma}_j^\mathrm{t} \leftarrow -\sqrt{\frac{\pi}{\beta^{2/n}} \mathbf{Z}_\mathrm{t}^{-1}}\, \mathbf{u}_j$; // see (11)
  - $\boldsymbol{\sigma}_j = \mathcal{R}_\mathbf{m}(\mathbf{E}\,\boldsymbol{\sigma}_j^\mathrm{t})$; // see (12)
- **end for**
- $\{\boldsymbol{\sigma}_j\}_{j=1}^m \leftarrow$ momentCorrection $(\{\boldsymbol{\sigma}_j\}_{j=1}^m)$; // see (13)
- **return** $\{\boldsymbol{\sigma}_j\}_{j=1}^m$

end procedure

D. Implementation

The proposed sampling scheme is summarized in Alg. 1 and illustrated in Fig. 1 with an example on the unit circle $\mathbb{S}^1$. We employ *The Nonlinear Estimation Toolbox* [30] for efficient on-tangent-plane deterministic sampling via the LCD-based DMA

[3]Obviously, other metrics (e.g., Euclidean distance) can be also employed.



(a) $\mathcal{E} = 2.8 \times 10^{-10}$ (b) $\mathcal{E} = 1.5 \times 10^{-9}$ (c) $\mathcal{E} = 1.6 \times 10^{-10}$

Fig. 5: Illustration of moment correction approaches using different combinations of objective functions and Riemannian optimizers: (a) trust region method without penalty, (b) trust region method for penalized objective function, and (c) steepest descent without penalty. The residual value $\mathcal{E}$ after correction is defined in (13).

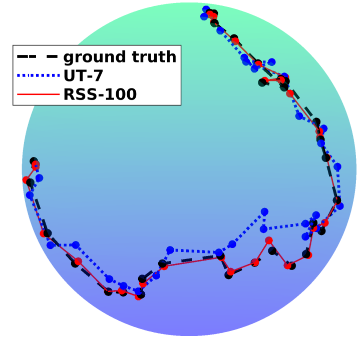


Fig. 6: A representative result of the orientation estimation given by Bingham filters using the sampling approach of the RSS and the one only satisfying the requirement of UT [12]. Here, 100 deterministic samples are drawn by the RSS, whereas the UT-based sampling approach has a fixed sample size of 7. The estimates are visualized as trajectories on the unit sphere $\mathbb{S}^2$.

introduced in Sec. II-C. The on-manifold moment correction is performed by employing the manifold optimization toolbox *Manopt* [31], where the RSD algorithm is used with retracted tangent sigma points as the initialization. In this way, the optimization-based correction can quickly converge to a global minimum that leads to little distortion of higher-order moments. We use the following example to show the performance of the proposed Riemannian spherical sampling (RSS) approach under different configurations.

Example III.1 *We apply the RSS for Bingham distributions of different dimensions with various parameter matrices $\mathbf{C}_\mathcal{B}$ introduced in Sec. II-B and sample size m as follows.*

- $\mathbb{S}^1$*: parameter $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, a)$ with $a = 5, 10, 30$ and sample size $m = 10, 30, 50$.*
- $\mathbb{S}^2$*: parameter $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 5, a)$ with $a = 5, 10, 30$ and sample size $m = 50, 200, 500$.*
- $\mathbb{S}^3$*: parameter $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 5, 5, 5)$ and sample size $m = 50, 200, 500$.*

Fig. 2 shows the sampling results given by the proposed approach on $\mathbb{S}^1$ and $\mathbb{S}^2$ and results for $\mathbb{S}^3$ are shown in Fig. 3. For varying dimensionality and parameter configurations, the proposed sampling approach can generate deterministic samples maintaining the second-order moment (denoted by $\mathcal{E}$) while being adaptive to the higher-order moments of the Bingham distribution.

IV. Evaluation

We integrate the proposed deterministic sampling approach into the nonlinear Bingham filter for orientation estimation. Here, the system model is set up as

$$\mathbf{x}_k = \mathbf{x}_{k-1} \otimes \mathbf{u}_{k-1} \otimes \mathbf{w}_{k-1} \,, \tag{14}$$

with $\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k \in \mathbb{S}^3$ denoting system state, input and noise represented by unit quaternions, respectively. The system noise is assumed to be Bingham-distributed with a mode at $[1, 0, 0, 0]^\top$. The measurement model is given as

$$\underline{z}_k = \mathbf{x}_k \otimes \underline{n}_0 \otimes \mathbf{x}_k^* + \underline{v}_k \,, \tag{15}$$

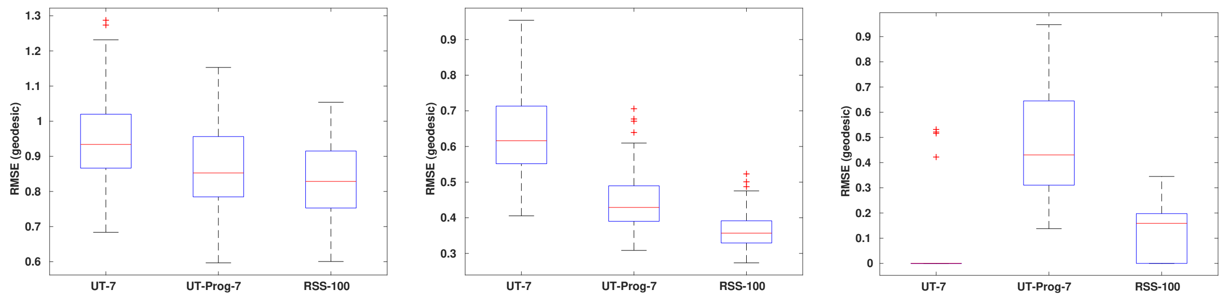
which is the orientation $\underline{z}_k$ rotated from the initialization $\underline{n}_0$. We assume the noise $\underline{v}_k \in \mathbb{R}^3$ to be additive and follow a zero-mean Gaussian distribution, i.e., $\underline{v}_k \sim \mathcal{N}(\underline{0}, \mathbf{\Omega}_{\underline{v}})$.

As proposed in [5], [12], for quaternion Bingham filtering, the prediction step can be performed via the composition of Bingham distributions. For the non-identity measurement model in (15), the posterior is normally approximated from the prior samples that are reweighted according to the likelihoods

$$f(\underline{z}_k | \mathbf{x}_k) = f_{\underline{v}_k}(\underline{z}_k - \mathbf{x}_k \otimes \underline{n}_0 \otimes \mathbf{x}_k^*) \,.$$

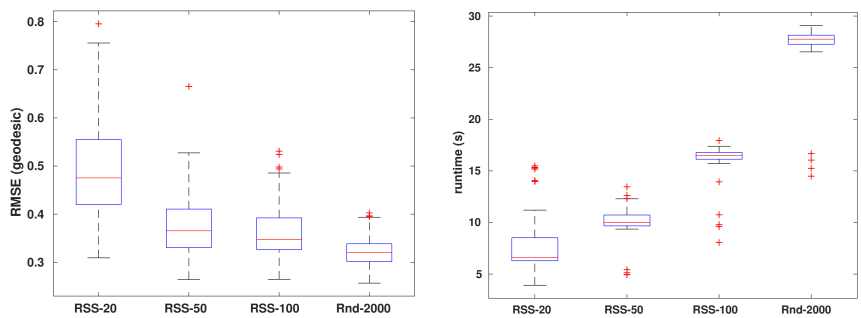
It can therefore be risky to rely on a small number of samples for performing the update step. For instance, the UT-based approach from [12] can only generate seven samples, which can easily degenerate after the reweighting. A typical method to overcome the sample degeneration issue is to apply the progressive update step [5], [33], which gradually fuse the measurement with the prior. However, the existing progressive Bingham filter [5] still relies on the UT-based deterministic sampling approach, leading to an ineffective progression step as the non-mode samples can also easily degenerate. For the following evaluation cases, we apply the proposed Riemannian spherical sampling (RSS) approach in a simple sampling-approximation Bingham filtering scheme (without progressive update) for spatial orientation estimation. For each of the cases, we perform 100 Monte Carlo runs of 50 time steps.

First, we evaluate the proposed RSS-based Bingham filtering in comparison with the unscented Bingham filter (UBF) from [12] and the progressive UBF from [5]. Here, the system noise $\mathbf{w}$ is Bingham-distributed with the parameter $\mathbf{C}_\mathcal{B} = -\mathrm{diag}(1, 10, 10, 10)$. The additive measurement noise follows zero-mean Gaussian distributions with different uncertainty levels of $\mathbf{\Omega}_{\underline{v}} = a \times \mathrm{diag}(1, 10, 10)$, where $a = 10^{-1}$, 10^{-2} and 10^{-4}. We use $m = 100$ deterministic samples for the RSS-based filter. As shown in Fig. 7, for all levels of measurement uncertainties the proposed approach shows better tracking accuracy than the other two existing approaches. Specifically, for low measurement noise, where samples are prone to degenerate due to the peaky likelihood function, the UT-based



(a) large noise (b) medium noise (c) low noise

Fig. 7: Evaluation of the proposed deterministic sampling approach for quaternion-based orientation estimation under different measurement noise levels (shown by box plot of MATLAB). We compare the Bingham filter (BF) based on the RSS (100 deterministic samples, shown as RSS-100) with the unscented BF [12] (7 deterministic samples satisfying requirements of the unscented transform (UT), shown as UT-7) and the UT-based progressive BF [5] (shown as UT-Prog-7). The proposed deterministic sampling approach enables the most robust and accurate tracking performance using the non-progressive filtering scheme with only a sampling-approximation-based update step.



(a) tracking accuracy (b) runtime

Fig. 8: Evaluation of Bingham filtering for orientation estimation using the proposed deterministic sampling approach with different sample sizes (shown by box plot of MATLAB). The proposed RSS shows improved tracking accuracy and higher runtime with increasing sample size. Compared with the Bingham particle filter [32] (using 2000 random samples, shown as Rnd-2000), the proposed RSS enables good combination of tracking accuracy and efficiency.

approach almost totally failed with several outliers showing large error. However, the Bingham filter based on the proposed sampling approach works robustly and most accurately. Fig. 6 further shows an example of the visualized tracking result given by Bingham filters based on different sampling approaches.

Second, we evaluate the proposed RSS-based Bingham filtering with different sample sizes of $m = 20, 50, 100$ under medium measurement noise. Additionally, we employ a naively implemented particle filter based on 2000 random samples [32] as a reference. Fig. 8 shows that the RSS-based Bingham filter improves tracking accuracy with larger size of deterministic samples. Due to the online optimization-based sampling scheme, the RSS-based Bingham filter shows worse efficiency with larger sample size. Compared to the Bingham particle filter with 2000 samples, the RSS-based Bingham filter with 100 deterministic samples gives accurate enough results with much less runtime.

V. Conclusion

In this paper, we introduced a novel deterministic sampling approach that can be applied to Bingham distributions in arbitrary dimensions for any given sample size. Unlike existing sampling schemes, the proposed approach satisfies requirements of the unscented transform (preserve the first two moments) and approximates the higher-order moments of the Bingham distribution with samples drawn on the hypersphere. Therefore, our novel approach enables improved nonlinear Bingham filtering for quaternion-based orientation estimations.

However, there is still much potential to exploit for the presented work. The proposed sampling scheme gives DMA to an on-tangent-plane Gaussian distribution, which is an approximation of the true projected Bingham density. It is thus valuable to further investigate the difference between drawing samples from the approximated density and the true one. Moreover, by applying the proposed sampling approach to the Bingham-based pose filtering frameworks [4], [6], [34]–[36], a better performance regarding the tracking accuracy, efficiency, and robustness can be expected.

Acknowledgment

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

References

[1] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, “Iterated Extended Kalman Filter Based Visual-Inertial Odometry Using Direct Photometric Feedback,” *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[2] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, 2018.

[3] F. Pfaff, K. Li, and U. D. Hanebeck, "Fourier Filters, Grid Filters, and the Fourier-Interpreted Grid Filter," in *Proceedings of the 22nd International Conference on Information Fusion (Fusion 2019)*, Ottawa, Canada, July 2019.

[4] S. Bultmann, K. Li, and U. D. Hanebeck, "Stereo Visual SLAM Based on Unscented Dual Quaternion Filtering," in *Proceedings of the 22st International Conference on Information Fusion (Fusion 2019)*, June 2019.

[5] 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.

[6] K. Li, G. Kurz, L. Bernreiter, and U. D. Hanebeck, "Simultaneous Localization and Mapping Using a Novel Dual Quaternion Particle Filter," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, Jul. 2018.

[7] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-Manifold Preintegration for Real-Time Visual–Inertial Odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[8] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2009, vol. 494.

[9] J. M. Glover, "The Quaternion Bingham Distribution, 3D Object Detection, and Dynamic Manipulation," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[10] J. Glover and L. P. Kaelbling, "Tracking the Spin on a Ping Pong Ball with the Quaternion Bingham Filter," in *2014 IEEE International Conference on Robotics and Automation (ICRA 2014)*, May 2014, pp. 4133–4140.

[11] T. Birdal, U. Şimşekli, M. O. Eken, and S. Ilic, "Bayesian Pose Graph Optimization via Bingham Distributions and Tempered Geodesic MCMC," *arXiv preprint arXiv:1805.12279*, 2018.

[12] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented Orientation Estimation Based on the Bingham Distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, Jan. 2016.

[13] G. Kurz, I. Gilitschenski, S. J. Julier, and U. D. Hanebeck, "Recursive Estimation of Orientation Based on the Bingham Distribution," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.

[14] S. J. Julier and J. K. Uhlmann, "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.

[15] S.-i. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Soc., 2007, vol. 191.

[16] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "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, Dec. 2006.

[17] U. D. Hanebeck, "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, Mar. 2014.

[18] M. F. Huber and U. D. Hanebeck, "Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation," in *Proceedings of the 17th IFAC World Congress (IFAC 2008)*, vol. 17, no. 2, Seoul, Republic of Korea, Jul. 2008.

[19] U. D. Hanebeck, M. F. Huber, and V. Klumpp, "Dirac Mixture Approximation of Multivariate Gaussian Densities," in *Proceedings of the 2009 IEEE Conference on Decision and Control (CDC 2009)*, Shanghai, China, Dec. 2009.

[20] 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)*, June 2019.

[21] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2009.

[22] W. R. Hamilton, "On Quaternions; Or on a New System of Imaginaries in Algebra," *Philosophical Magazine Series 3*, vol. 25, no. 163, pp. 10–13, 1844.

[23] U. D. Hanebeck, "Optimal Reduction of Multivariate Dirac Mixture Densities," *at - Automatisierungstechnik, Oldenbourg Verlag*, vol. 63, no. 4, pp. 265–278, Apr. 2015.

[24] I. Gilitschenski and U. D. Hanebeck, "Efficient Deterministic Dirac Mixture Approximation," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[25] J. Steinbring and U. D. Hanebeck, "LRKF Revisited: The Smart Sampling Kalman Filter (S2KF)," *Journal of Advances in Information Fusion*, vol. 9, no. 2, pp. 106–123, Dec. 2014.

[26] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Efficient Bingham Filtering based on Saddlepoint Approximations," in *Proceedings of the 2014 IEEE International Conference on Multisensor Fusion and Information Integration (MFI 2014)*, Beijing, China, Sep. 2014.

[27] P.-A. Absil and J. Malick, "Projection-Like Retractions on Matrix Manifolds," *SIAM Journal on Optimization*, vol. 22, no. 1, pp. 135–158, 2012.

[28] S. Hauberg, "Directional Statistics with the Spherical Normal Distribution," in *Proceedings of the 21st International Conference on Information Fusion (Fusion 2018)*, Cambridge, United Kingdom, Jul. 2018.

[29] G. Kurz and U. D. Hanebeck, "Linear Regression Kalman Filtering Based on Hyperspherical Deterministic Sampling," in *Proceedings of the 56th IEEE Conference on Decision and Control (CDC 2017)*, Melbourne, Australia, Dec. 2017.

[30] J. Steinbring, "Nonlinear Estimation Toolbox." [Online]. Available: https://bitbucket.org/nonlinearestimation/toolbox

[31] 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. [Online]. Available: http://www.manopt.org

[32] J. T. Kent, A. M. Ganeiber, and K. V. Mardia, "A New Method to Simulate the Bingham and Related Distributions in Directional Data Analysis with Applications," *arXiv preprint arXiv:1310.8110*, 2013.

[33] 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, Jul. 2018.

[34] 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)*, May 2019.

[35] F. Pfaff, K. Li, and U. D. Hanebeck, "Association Likelihoods for Directional Estimation," in *Proceedings of the 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019)*, May 2019.

[36] S. Radtke, K. Li, B. Noack, and U. D. Hanebeck, "Comparative Study of Track-to-Track Fusion Methods for Cooperative Tracking with Bearing-Only Measurements," in *Proceedings of the 2019 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2019)*, May 2019.