Recursive Bingham Filter for Directional Estimation Involving 180 Degree Symmetry

Gerhard Kurz*, **Igor Gilitschenski***, **Simon Julier**†, and **Uwe D. Hanebeck***

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

†Virtual Environments and Computer Graphics Group
Department of Computer Science
University College London (UCL), United Kingdom
s.julier@cs.ucl.ac.uk

***Abstract*—This work considers filtering of uncertain data defined on periodic domains, particularly the circle and the manifold of orientations in 3D space. Filters based on the Kalman filter perform poorly in this directional setting as they fail to take the structure of the underlying manifold into account. We present a recursive filter based on the Bingham distribution, which is defined on the considered domains. The proposed filter can be applied to circular filtering problems with 180 degree symmetry and to estimation of orientations in three dimensional space. It is easily implemented using standard numerical techniques and suitable for real-time applications. We evaluate our filter in a challenging scenario and compare it to a Kalman filtering approach adapted to the particular setting.**

***Index Terms*—angular quantities, circular data, directional statistics, recursive filtering**

This is an extended version of the paper *Recursive Estimation of Orientation Based on the Bingham Distribution* published at the *16th International Conference on Information Fusion (Fusion 2013)*, which received the *Best Student Paper Award, First Runner-Up*.

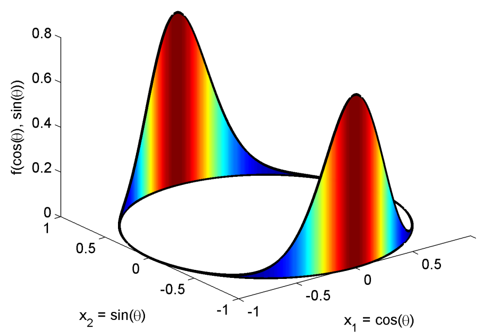

Fig. 1: Bingham probability density function with $\mathbf{M} = \mathbf{I}_{2\times 2}$ and $\mathbf{Z} = \operatorname{diag}(-8, 0)$ as a 3D plot. This corresponds to a standard deviation of $16°$.

1. Introduction

Tracking cars, ships, or airplanes may involve estimation of their current orientation or heading. Furthermore, many applications in the area of robotics or augmented reality depend on reliable estimation of the pose of certain objects. When estimating the orientation of two-way roads or relative angles of two unlabeled targets, the estimation task can be thought of as estimation of a directionless orientation. Thus, the estimation task reduces to estimating the alignment of an axis, i.e., estimation with $180°$ symmetry.

All these estimation problems share the need for processing angular or directional data, which differs in many ways from the linear setting. First, periodicity of the underlying manifold needs to be taken into account. Second, directional quantities do not lie in a vector space. Thus, there is no equivalent to a linear model, as there are no linear mappings. These problems become particularly significant for high uncertainties, e.g., as a result of poor initialization, inaccurate sensors such as magnetometers, or sparse measurements causing a large integration error.

In many applications, even simple estimation problems involving angular data are often considered as linear or nonlinear estimation problems on linear domains and handled with techniques such as the Kalman Filter [19], the Extended Kalman Filter (EKF), or the Unscented Kalman Filter (UKF) [17]. In a circular setting, most approaches to filtering suffer from assuming a Gaussian probability density at a certain point. They fail to take into account the periodic nature of the underlying domain and assume a (linear) vector space instead of a curved manifold. This shortcoming can cause poor results, in particular when the angular uncertainty is large. In certain cases, the filters may even diverge.

Strategies to avoid these problems in an angular setting involve an "intelligent" repositioning of measurements (typically by multiples of π) or even discarding certain undesired measurements. Sometimes, nonlinear equality constraints have to be fulfilled, for example, unit length of a vector, which makes it necessary to inflate the covariance [16]. There are also approaches that use operators on a manifold to provide a local approximation of a vector space [13]. While these approaches yield reasonable results in some circumstances, they still suffer from ignoring the true geometry of circular data within their probabilistic models, which are usually based on assuming normally distributed noise. This assumption is often motivated by the Central Limit Theorem, i.e., the limit distribution of a normalized sum of i.i.d. random variables with finite variance is normally distributed [42]. However, this motivation does not apply to uncertain data from a periodic domain. Thus, choosing a circular distribution for describing uncertainty can offer better results.

In this paper, we consider the use of the Bingham distribution [5] (see Fig. 1) for recursive estimation. The Bingham distribution is defined on the hypersphere of arbitrary dimension. Here, we focus on the cases of two- and four-dimensional Bingham distributed random vectors and apply our results to angular estimation with $180°$ symmetry and estimating orientation in 3D space.

Estimating orientation is achieved by using unit quaternions to represent the full 3D orientation of an object. It is well known that quaternions avoid the singularities present in other representations such as Euler angles [25]. Their only downsides are the fact that they must remain normalized and the property that the quaternions $\underline{q}$ and $-\underline{q}$ represent the same orientation. Both of these issues can elegantly be resolved by use of the Bingham distribution, since it is by definition restricted to the hypersphere and is $180°$ symmetric.

This work extends our results on Bingham filtering [32] and the first-order quaternion Bingham filter proposed in [11] in several ways. First of all, we present a relationship between the two-dimensional Bingham distribution and the von Mises distribution and we show how to exploit it to obtain a more efficient way of computing the normalization constant and its derivatives. Furthermore, we show a relation to the von Mises-Fisher distribution, which can be used to speed up parameter estimation and moment matching procedures in an important special case. In that situation, we avoid the need for precomputed lookup tables. This is of considerable interest because the computation of the normalization constant plays a crucial role for the performance of the Bingham filter. Finally, we perform a more thorough evaluation of both two- and four-dimensional scenarios using different types of noise distributions and different degrees of uncertainty.

This paper is structured as follows. First, we present an overview of previous work in the area of directional statistics and angular estimation (Sec. 2). Then, we introduce our key idea in Sec. 3. In Sec. 4, we give a detailed introduction to the Bingham distribution and in Sec. 5, we derive the necessary operations needed to create a recursive Bingham filter. Based on these prerequisites, we introduce our filter in Sec. 6. We have carried out an evaluation in simulations, which is presented in Sec. 7. Finally, we conclude this work in Sec. 8.

2. Related Work

Directional statistics is a subdiscipline of statistics, which focuses on dealing with directional data. That is, it considers random variables which are constrained to lie on manifolds (for example the circle or the sphere) rather than random variables located in $d-$dimensional vector spaces (typically $\mathbb{R}^d$). Classical results in directional statistics are summed up in the books by Mardia and Jupp [37] and by Jammalamadaka and Sengupta [15]. Probability distributions on the unit sphere are described in more detail in [6].

There is a broad range of research investigating the two-dimensional orientation estimation. A recursive filter based on the von Mises distribution for estimating the orientation on the $SO(2)$ was presented in [3], [45]. It has been applied to GPS phase estimation problems [44]. Furthermore, a nonlinear filter based on von Mises and wrapped normal distributions was presented in [30], [31]. This filter takes advantage of the fact that wrapped normal distributions are closed under convolution and the fact that von Mises distributions are closed under Bayesian inference. This filter has also been applied to constrained object tracking [29].

In 1974, Bingham published the special case for three dimensions of his distribution in [5], which he originally developed in his PhD thesis [4]. Further work on the Bingham distribution has been done by Kent [21], [22] as well as Jupp and Mardia [18], [35]. So far, there have been a few applications of the Bingham distribution, for example in geology [36], [28], [33].

Antone published some results on a maximum likelihood approach for Bingham-based pose estimation in 2001 [2]. However, this method was limited to offline applications. In 2011, Glover used the Bingham distribution for a Monte Carlo based pose estimation [10], which he later generalized into a quaternion-based recursive filter [11] and applied it to tracking the spin of a ping pong ball [12]. Glover also released a library called `libbingham` [9] that includes C and MATLAB implementations of some of the methods discussed in Sec. 4. It should be noted that our implementation is not based on `libbingham`. Our implementation calculates the normalization constant online, whereas `libbingham` relies on values that have been precomputed offline. In the case of a two-dimensional Bingham-distributed random vector, the computation of the normalization constant of the corresponding probability density function reduces to the evaluation of Bessel functions. In higher dimensions, a saddlepoint approximation can be used [26].

In 2013, we proposed a recursive Bingham filter for 2D axis estimation [32], which serves as a foundation for this paper. We also published a nonlinear generalization to the quaternion case in [8].

3. Key Idea of the Bingham Filter

In this paper, we derive a recursive filter based on the Bingham distribution for two- and four-dimensional random vectors of unit length, because they can be used to represent orientations on the plane and in three-dimensional space.

Rather than relying on approximations involving the Gaussian distribution, we chose to represent all occurring probability densities as Bingham distributions. The Bingham distribution is defined on the hypersphere and is antipodally symmetric. Our use of the Bingham distribution is motivated by its convenient representation of hyperspherical random vectors, its relationship to the Gaussian distribution, and a maximum entropy property [35]. Although we restrict ourselves to the two- and four-dimensional cases in this paper, we would like to emphasize that some of the presented methods can easily be generalized to higher dimensions.

In order to derive a recursive filter, we need to be able to perform two operations. First, we need to calculate the predicted state at the next time step from the current state and the system noise affecting the state. In a recursive estimation problem in $\mathbb{R}^d$ with additive noise, this involves a convolution with the noise density. We provide a suitable analogue on the hypersphere in order to account for the *composition* of uncertain rotations. Since Bingham distributions are not closed under this operation, we present an approximate solution to this problem based on matching covariance matrices.

Second, we need to perform a Bayes update. As usual, this requires the *multiplication* of the prior density with the likelihood density. We prove that Bingham distributions are closed under multiplication and show how to obtain the posterior density.

4. Bingham Distribution

In this section, we lay out the Bingham distribution and the fundamental operations that we use to develop the filter and discuss its relation to several other distributions. The Bingham distribution on the hypersphere naturally appears when a d-dimensional normal random vector $\underline{x}$ with $\mathrm{E}(\underline{x}) = \underline{0}$ is conditioned on $||\underline{x}|| = 1$ [26]. One of the main challenges when dealing with the Bingham distribution is the calculation of its normalization constant, so we discuss this issue in some detail.

4.1 Probability Density Function

As a consequence of the motivation above, it can be seen that the Bingham probability density function (pdf) looks exactly like its Gaussian counterpart except for the normalization constant. Furthermore, the parameter matrix of the Bingham distribution appearing in the exponential (which is the inverse covariance matrix in the Gaussian case) is usually decomposed into an orthogonal and a diagonal matrix, which yields an intuitive interpretation of the matrices. This results in the following definition.

Definition 1. *Let $S_{d-1} = \{\underline{x} \in \mathbb{R}^d : ||\underline{x}|| = 1\} \subset \mathbb{R}^d$ be the unit hypersphere in $\mathbb{R}^d$. The probability density function (pdf)*

$$f : S_{d-1} \to \mathbb{R} \tag{1}$$

of a Bingham distribution [5] is given by

$$f(\underline{x}) = \frac{1}{F} \cdot \exp(\underline{x}^T \mathbf{M}\,\mathbf{Z}\,\mathbf{M}^T \underline{x}) \ , \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{d\times d}$ is an orthogonal matrix[1] *describing the orientation, $\mathbf{Z} = \mathrm{diag}(z_1, \dots z_{d-1}, 0) \in \mathbb{R}^{d\times d}$ with $z_1 \le \dots \le z_{d-1} \le 0$ is the concentration matrix, and F is a normalization constant.*

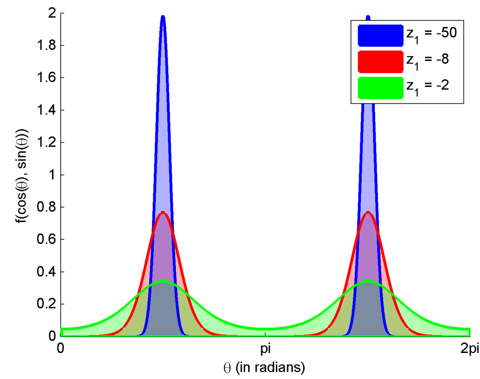

Fig. 2: Bingham probability density function with $\mathbf{M} = \mathbf{I}_{2\times 2}$ for different values of $\mathbf{Z} = \mathrm{diag}(z_1, 0)$ and $\underline{x} = (\cos(\theta), \sin(\theta))^T$. These values for z_1 correspond to standard deviations of approximately 6°, 16°, and 36°, respectively.

As Bingham showed [5], adding a multiple of the identity matrix $\mathbf{I}_{d\times d}$ to $\mathbf{Z}$ does not change the distribution. Thus, we conveniently force the last entry of $\mathbf{Z}$ to be zero. Because it is possible to swap columns of $\mathbf{M}$ and the according diagonal entries in $\mathbf{Z}$ without changing the distribution, we can enforce $z_1 \le \dots \le z_{d-1}$.

The probability density function is antipodally symmetric, i.e., $f(\underline{x}) = f(-\underline{x})$ holds for all $\underline{x} \in S_{d-1}$. Consequently, the Bingham distribution is invariant to rotations by 180°. Examples for two dimensions ($d = 2$) are shown in Fig. 1 and Fig. 2. Examples for three dimensions ($d = 3$) are shown in Fig. 3. The relation of the Bingham distribution to certain other distributions is discussed the appendix.

It deserves to mention that some authors use slightly different parameterizations of the Bingham distribution. In particular, the rightmost column of $\mathbf{M}$ is sometimes omitted [11], because it is, up to sign, uniquely determined by being a unit vector that is orthogonal to the other columns of $\mathbf{M}$. As a result of antipodal symmetry, the sign can safely be ignored. Still, we prefer to include the entire matrix $\mathbf{M}$ because this representation allows us to obtain the mode of the distribution very easily by taking the last column of $\mathbf{M}$.

4.2 Normalization Constant

The normalization constant of the Bingham distribution is difficult to calculate, which constitutes one of the most significant challenges when dealing with the Bingham distribution.

[1] An orthogonal matrix $\mathbf{M}$ fulfills the equation $\mathbf{M}\,\mathbf{M}^T = \mathbf{M}^T\,\mathbf{M} = \mathbf{I}_{d\times d}$.

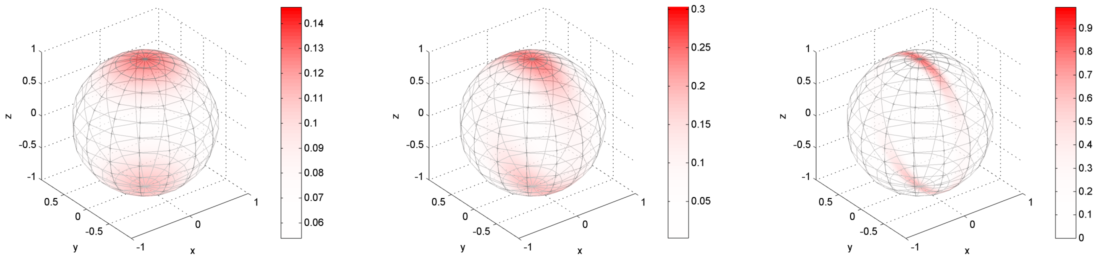

Fig. 3: Bingham pdf with $\mathbf{M} = \mathbf{I}_{3\times 3}$ for values of $\mathbf{Z} = \mathrm{diag}(-1, -1, 0)$, $\mathbf{Z} = \mathrm{diag}(-5, -1, 0)$, and $\mathbf{Z} = \mathrm{diag}(-50, -1, 0)$.

Because

$$F = \int_{S_{d-1}} \exp(\underline{x}^T \mathbf{M} \mathbf{Z} \mathbf{M}^T \underline{x}) d\underline{x} \tag{3}$$

$$= \int_{S_{d-1}} \exp(\underline{x}^T \mathbf{Z} \underline{x}) d\underline{x} \ , \tag{4}$$

the normalization constant does not depend on $\mathbf{M}$. It can be calculated with the help of the hypergeometric function of a matrix argument [14], [23], [40] according to

$$F := |S_{d-1}| \cdot {}_1F_1\left(\frac{1}{2}, \frac{d}{2}, \mathbf{Z}\right) \ , \tag{5}$$

where

$$|S_{d-1}| = \frac{2 \cdot \pi^{n/2}}{\Gamma(N/2)} \tag{6}$$

is the surface area of the d-sphere and ${}_1F_1(\cdot, \cdot, \cdot)$ is the hypergeometric function of matrix argument. In the d-dimensional case, this reduces to

$$F = |S_{d-1}| \cdot {}_1F_1\left(\frac{1}{2}, \frac{d}{2}, \begin{bmatrix} z_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & z_{d-1} & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix}\right) \tag{7}$$

$$= |S_{d-1}| \cdot {}_1F_1\left(\frac{1}{2}, \frac{d}{2}, \begin{bmatrix} z_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & z_{d-1} \end{bmatrix}\right) \ , \tag{8}$$

so it is sufficient to compute the hypergeometric function a diagonal matrix of size $(d-1) \times (d-1)$. If $d = 2$, this is a hypergeometric function of a scalar argument, which is described in [1]. We will later show how to further reduce this to a Bessel function for $d = 2$.

A number of algorithms for computing the hypergeometric function have been proposed, for example saddle point-approximations [26], a series of Jack functions [23], and holonomic gradient descent [24]. Glover has suggested the formula [11, (8)]

$$F = 2\sqrt{\pi} \sum_{\alpha_1=0}^{\infty} \cdots \sum_{\alpha_{d-1}=0}^{\infty} \frac{\prod_{i=1}^{d-1} \Gamma(\alpha_i + \frac{1}{2}) \frac{z_i^{\alpha_i}}{\alpha_i!}}{\Gamma\left(\frac{d}{2} + \sum_{i=1}^{d-1} \alpha_i\right)} \tag{9}$$

which should only be evaluated for positive $z_1, \dots, z_{d-1}$ to avoid a numerically unstable alternating series[2]. Because of the computational complexity involved, `libbingham` [9] provides a precomputed lookup table and linear interpolation is used at runtime to obtain an approximate value. The technique of precomputed tables has previously been used by Mardia et al. for the maximum likelihood estimate, which involves the normalization constant [38].

To allow for online calculation of the normalization constant, we use Bessel functions for $d = 2$ and the saddle-point approximation by Kume et al. [26] for $d > 2$. The derivatives of the normalization constant, which are required for the maximum likelihood estimation procedure, can be calculated according to [27].

5. Operations on the Bingham Distribution

In this section, we derive the formulas for multiplication of two Bingham probability density functions. Furthermore, we will present a method for computing the composition of two Bingham-distributed random variables, which is analogous to the addition of real random variables.

5.1 Multiplication

For two given Bingham densities, we want to obtain their product. This product is used for Bayesian inference involving Bingham densities. The result presented below yields a convenient way to calculate the product of Bingham densities.

Lemma 1. *Bingham densities are closed under multiplication with renormalization.*

Proof. Consider two Bingham densities

$$f_1(\underline{x}) = F_1 \cdot \exp(\underline{x}^T \mathbf{M}_1 \, \mathbf{Z}_1 \, \mathbf{M}_1^T \underline{x}) \tag{10}$$

and

$$f_2(\underline{x}) = F_2 \cdot \exp(\underline{x}^T \mathbf{M}_2 \, \mathbf{Z}_2 \, \mathbf{M}_2^T \underline{x}) \ . \tag{11}$$

Then

$$f_1(\underline{x}) \cdot f_2(\underline{x}) = F_1 F_2 \cdot \exp(\underline{x}^T \underbrace{\left(\mathbf{M}_1 \mathbf{Z}_1 \mathbf{M}_1^T + \mathbf{M}_2 \mathbf{Z}_2 \mathbf{M}_2^T \right)}_{=:\mathbf{C}} \underline{x})$$

$$\propto F \cdot \exp(\underline{x}^T \mathbf{M} \, \mathbf{Z} \, \mathbf{M}^T \underline{x}) \tag{12}$$

[2] This can easily be achieved by adding a multiple of the identity matrix to the concentration matrix $\mathbf{Z}$.

with F as the new normalization constant after renormalization, $\mathbf{M}$ are the unit eigenvectors of $\mathbf{C}$, $\mathbf{D}$ has the eigenvalues of $\mathbf{C}$ on the diagonal (sorted in ascending order) and $\mathbf{Z} = \mathbf{D} - \mathbf{D}_{dd}\mathbf{I}_{d\times d}$ where $\mathbf{D}_{dd}$ refers to the bottom right entry of $\mathbf{D}$, i. e., the largest eigenvalue. □

5.2 Estimation of Bingham Distribution Parameters

Estimating parameters for the Bingham distribution is not only motivated by the need to estimate distribution parameters of the process noise. It also plays a crucial role in the prediction process when computing the composition of two Bingham random vectors and reapproximating a Bingham distribution. This procedure is based on matching covariance matrices. Be aware that although the Bingham distribution is only defined on S_{d-1}, we can still compute the covariance matrix of a Bingham-distributed random vector $\underline{x} \in S_{d-1}$ according to $\mathbf{S} = \mathrm{E}(\underline{x} \cdot \underline{x}^T)$ in $\mathbb{R}^d$. Thus, we will present both the computation of the covariance matrix of a Bingham distributed random vector and the computation of parameters for a Bingham distribution with a given covariance (which could correspond to an arbitrary distribution on the hypersphere).

The maximum-likelihood estimate for the parameters $(\mathbf{M}, \mathbf{Z})$ of a Bingham distribution can be obtained from given or empirical moments (in the case of given samples) as described in [5]. $\mathbf{M}$ can be obtained as the matrix of eigenvectors of the covariance $\mathbf{S}$ with eigenvalues $\omega_1 \leq \cdots \leq \omega_d$. In other words, $\mathbf{M}$ can be found as the eigendecomposition of

$$\mathbf{S} = \mathbf{M} \cdot \mathrm{diag}(\omega_1, \ldots, \omega_d) \cdot \mathbf{M}^T \ . \tag{13}$$

To calculate $\mathbf{Z}$, the equations

$$\frac{\frac{\partial}{\partial z_i} {}_1F_1\left(\frac{1}{2}, 1, \mathrm{diag}(z_1, \ldots, z_d)\right)}{{}_1F_1(\frac{1}{2}, 1, \mathrm{diag}(z_1, \ldots, z_d))} = \omega_i \ , \quad i = 1, \ldots, d \tag{14}$$

have to be solved under the constraint $z_d = 0$, which is justified by the argumentation above and used to simplify the computation. The actual computation is performed numerically. In our case, the `fsolve` routine from Matlab was used, which utilizes a trust region method for solving nonlinear equations.

Conversely, for a given Bingham$(\mathbf{M}, \mathbf{Z})$-distributed random vector $\underline{x} \in S_{d-1}$, the covariance matrix can be calculated according to

$$\mathrm{E}(\underline{x} \cdot \underline{x}^T) = \mathbf{M} \cdot \mathrm{diag}(\omega_1, \ldots, \omega_d) \cdot \mathbf{M}^T \tag{15}$$

$$= \mathbf{M} \cdot \mathrm{diag}\left(\frac{1}{F}\frac{\partial F}{\partial z_1}, \ldots, \frac{1}{F}\frac{\partial F}{\partial z_d}\right) \cdot \mathbf{M}^T \ . \tag{16}$$

Thus, the underlying distribution parameters of a Bingham distributed random vector are uniquely defined by its covariance matrix and vice versa. However, it is important to note that this covariance matrix is usually not the same as the covariance matrix of a Gaussian random vector which was conditioned to one in order to obtain the Bingham distribution.

Remark 1. *For $d = 2$, there is an interesting relation of the covariance matrix to the circular (or trigonometric) moments*

$$m_n = \int_0^{2\pi} \exp(inx) f(x) dx \in \mathbb{C} \ , \quad i^2 = -1 \tag{17}$$

that are commonly used for circular distributions. A Bingham distribution with $\mathbf{M} = \mathbf{I}_{2\times 2}$ and $\underline{x} = [\cos(\theta), \sin(\theta)]^T$ has covariance

$$\mathbf{S} = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix} = \begin{bmatrix} \mathrm{E}(x_1^2) & 0 \\ 0 & \mathrm{E}(x_2^2) \end{bmatrix} \tag{18}$$

$$= \begin{bmatrix} \mathrm{E}(\cos(\theta)^2) & 0 \\ 0 & \mathrm{E}(\sin(\theta)^2) \end{bmatrix} \ , \tag{19}$$

i.e., $\omega_1 = \mathrm{Re}\, m_2$ and $\omega_2 = \mathrm{Im}\, m_2$.

5.3 Composition

Now, we want to derive the composition of Bingham distributed random vectors, which is the directional analogue to addition of random vectors in a linear space. Thus, the density of the random vector resulting from this operation is the directional analogue to the convolution in linear space. First, we define a composition of individual points on the hypersphere S_{d-1}, which we then use to derive the composition of Bingham distributed random vectors. We consider a composition function

$$\oplus : S_{d-1} \times S_{d-1} \to S_{d-1} \ , \tag{20}$$

where $\oplus$ has to be compatible with 180° degree symmetry, i.e.,

$$x \oplus y = \pm((-x) \oplus y) \tag{21}$$

$$= \pm(x \oplus (-y)) \tag{22}$$

$$= \pm((-x) \oplus (-y)) \tag{23}$$

for all $x, y \in S_{d-1}$. Furthermore, we require the quotient $(S_{d-1}/\{\pm 1\}, \oplus)$ to have an algebraic group structure. This guarantees associativity, the existence of an identity element, and the existence of inverse elements.

Remark 2. *It has been shown that the only hyperspheres admitting a topological group structure are S_0, S_1, and S_3 [39]. Because S_0 only consists of two elements, S_1 and S_3 (i.e., $d = 2$ and $d = 4$) are the only two relevant hyperspheres. This structure is necessary to obtain a suitable composition operation.*

For this reason, we only consider the cases $d = 2$ and $d = 4$ from now on. These two cases are of practical interest as they conveniently allow the representation of two-dimensional axes and three-dimensional orientations via quaternions.

5.3.1 Two-dimensional case: For $d = 2$, we interpret $S_1 \subset \mathbb{R}^2$ as elements in $\mathbb{C}$ of unit length, where the first dimension is the real part and the second dimension the imaginary part. In this interpretation, the Bingham distributions can be understood as a distribution on a subset of the complex plane, namely the unit circle.

Definition 2. *For $d = 2$, the composition function $\oplus$ is defined to be complex multiplication, i.e.,*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \oplus \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 y_1 - x_2 y_2 \\ x_1 y_2 + x_2 y_1 \end{bmatrix} \tag{24}$$

analogous to

$$(x_1 + ix_2) \cdot (y_1 + iy_2) = (x_1 y_1 - x_2 y_2) + i(x_1 y_2 + x_2 y_1) \ . \quad (25)$$

Since we only consider unit vectors, the composition $\oplus$ is equivalent to adding the angles of both complex numbers when they are represented in polar form. The identity element is ± 1 and the inverse element for $(x_1, x_2)^T$ is the complex conjugate $\pm(x_1, -x_2)^T$.

Unfortunately, the Bingham distribution is not closed under this kind of composition. That is, the resulting random vector is no longer Bingham distributed (see Lemma 3). Thus, we propose a technique to approximate the composed random vector with a Bingham distribution. The composition of two Bingham distributions $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ is calculated by considering the composition of their covariance matrices $\mathbf{A}, \mathbf{B}$ and estimating the parameters of $f_{\mathbf{C}}$ based on the resulting covariance matrix. Composition of covariance matrices can be derived from the composition of random vectors. Note that since covariance matrices are always symmetric, we can ignore the bottom left entry in our notation and mark it with an asterisk.

Lemma 2. *Let $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ be Bingham distributions with covariance matrices*

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ * & a_{22} \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ * & b_{22} \end{bmatrix} \ , \quad (26)$$

respectively. Let $\underline{x}, \underline{y} \in S_1 \subset \mathbb{R}^2$ be independent random vectors distributed according to $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$. Then the covariance

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ * & c_{22} \end{bmatrix} := \mathrm{Cov}(\underline{x} \oplus \underline{y}) \quad (27)$$

of the composition is given by

$$c_{11} = a_{11} b_{11} - 2 a_{12} b_{12} + a_{22} b_{22} \ , \quad (28)$$

$$c_{12} = a_{11} b_{12} - a_{12} b_{22} + a_{12} b_{11} - a_{22} b_{12} \ , \quad (29)$$

$$c_{22} = a_{11} b_{22} + 2 a_{12} b_{12} + a_{22} b_{11} \ . \quad (30)$$

Proof. See Appendix E. □

Based on $\mathbf{C}$, the maximum likelihood estimate is used to obtain the parameters $\mathbf{M}$ and $\mathbf{Z}$ of the uniquely defined Bingham distribution with covariance $\mathbf{C}$ as described above. This computation can be done in an efficient way, even though the solution of the equation involving the hypergeometric function is not given in closed form. This does not present a limitation to the proposed algorithm, because there are many efficient ways for the computation of the confluent hypergeometric function of a scalar argument [34], [41].

5.3.2 Four-dimensional case: In the four-dimensional case ($d = 4$), we interpret $S_3 \subset \mathbb{R}^4$ as unit quaternions in $\mathbb{H}$ [25]. A quaternion $\underline{q} = [q_1, q_2, q_3, q_4]^T$ consists of the real part q_1 and imaginary parts q_2, q_3, q_4. It is written as

$$\mathbf{q} = q_1 + q_2\, i + q_3\, j + q_4\, k \ , \quad (31)$$

where $i^2 = j^2 = k^2 = ijk = -1$ are the imaginary units. A rotation in $SO(3)$ with rotation axis $[v_1, v_2, v_3]^T \in S_2$ and rotation angle $\phi \in [0, 2\pi)$ can be represented as the quaternion

$$\mathbf{q} = \cos(\phi/2) + \sin(\phi/2)(v_1\, i + v_2\, j + v_3\, k) \quad (32)$$

and applied to a vector $\underline{w} = [w_1, w_2, w_3] \in \mathbb{R}^3$ according to

$$\mathbf{w}^{rot} = \mathbf{q}(0 + w_1\, i + w_2\, j + w_3\, k)\bar{\mathbf{q}} \ . \quad (33)$$

Here, $\bar{\mathbf{q}} = q_1 - q_2 i - q_3 j - q_4 k$ denotes the conjugate of $\mathbf{q}$ and $\mathbf{w}^{rot}$ quaternion containing the rotated vector encoded as the factors of the quaternion basis elements i, j, and k.

Definition 3. *For $d = 4$, the composition function $\oplus$ is defined to be quaternion multiplication, i.e.,*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \oplus \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4 \\ x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3 \\ x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2 \\ x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1 \end{bmatrix} \ , \quad (34)$$

sometimes also referred to as Hamilton product.

This definition corresponds to the composition of rotations. The identity element is $\pm[1, 0, 0, 0]^T$ and the inverse element is given by the quaternion conjugate as given above.

Lemma 3. *For all nontrivial hyperspheres that allow a topological group structure ($d = 2$ and $d = 4$, see Remark 2), the Bingham distribution is not closed under composition of random variables.*

Proof. We prove this Lemma by computing the true distribution of the (Hamilton-) product $\oplus$ of two Bingham distributed random vectors $\underline{x} \sim f_x(\cdot)$ and $\underline{y} \sim f_y(\cdot)$ with respective parameter matrices $\mathbf{M}_x$, $\mathbf{M}_y$, $\mathbf{Z}_x$, and $\mathbf{Z}_y$. The true density $f(\cdot)$ of $\underline{x} \oplus \underline{y}$ can be expressed in terms of the densities of $\underline{x}$ and $\underline{y}$ by

$$f(\underline{z}) = \int_{S_{d-1}} f_x(\underline{z} \oplus \underline{a}^{-1}) f_y(\underline{a}) d\underline{a} \ . \quad (35)$$

Inversion of unit quaternions and complex numbers of unit length can both be obtained by conjugation. Furthermore, complex numbers and quaternions can both be represented by matrices. This can be used to construct a matrix $\mathbf{Q}_z$ such that $\underline{z} \oplus \underline{a}^{-1} = \mathbf{Q}_z \underline{a}$. Thus, we obtain

$$f(\underline{z}) = \int_{S_{d-1}} f_x(\mathbf{Q}_z \underline{a}) f_y(\underline{a}) da \quad (36)$$

$$\propto \int_{S_{d-1}} \exp(\underline{a}^T \mathbf{Q}_z^T \mathbf{M}_x \mathbf{Z}_x \mathbf{M}_x^T \mathbf{Q}_z \underline{a} + \underline{a}^T \mathbf{M}_y \mathbf{Z}_y \mathbf{M}_y^T \underline{a}) d\underline{a} \quad (37)$$

$$\propto \int_{S_{d-1}} \exp(\underline{a}^T (\mathbf{Q}_z^T \mathbf{M}_x \mathbf{Z}_x \mathbf{M}_x^T \mathbf{Q}_z + \mathbf{M}_y \mathbf{Z}_y \mathbf{M}_y^T) \underline{a}) d\underline{a} \ .$$

Computation of the integral yields a rescaled hypergeometric function of matrix argument. Therefore, the random variable $\underline{x} \oplus \underline{y}$ does not follow a Bingham distribution. □

Lemma 4. *Let $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ be Bingham distributions with covariance matrices*

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ * & a_{22} & a_{23} & a_{24} \\ * & * & a_{33} & a_{34} \\ * & * & * & a_{44} \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ * & b_{22} & b_{23} & b_{24} \\ * & * & b_{33} & b_{34} \\ * & * & * & b_{44} \end{bmatrix},$$

respectively. Let $\underline{x}, \underline{y} \in S_3 \subset \mathbb{R}^4$ be independent random vectors distributed according to $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$. Then the covariance matrix

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ * & c_{22} & c_{23} & c_{24} \\ * & * & c_{33} & c_{34} \\ * & * & * & c_{44} \end{bmatrix} = \mathrm{Cov}(\underline{x} \oplus \underline{y}) \tag{38}$$

of the composition is given by

$$c_{ij} = \mathrm{E}((\underline{x} \oplus \underline{y})_i \cdot (\underline{x} \oplus \underline{y})_j) \ , \quad i, j = 1, \ldots, 4 \ . \tag{39}$$

Proof. Analogous to Lemma 2. The complete formula for c_{ij} is given in [11, A.9.2]. □

6. Filter Implementation

The techniques presented in the preceding section can be applied to derive a recursive filter based on the Bingham distribution. The system model is given by

$$\underline{x}_{k+1} = \underline{x}_k \oplus \underline{w}_k \ , \tag{40}$$

where $\underline{w}_k$ is Bingham-distributed noise. The measurement model is given by

$$\underline{z}_k = \underline{x}_k \oplus \underline{v}_k \ , \tag{41}$$

where $\underline{v}_k$ is Bingham-distributed noise and $\underline{x}_k$ is an uncertain Bingham-distributed system state. Intuitively, this means that both system and measurement model are the identity disturbed by Bingham-distributed noise. Note that the modes of the distributions of $\underline{w}_k$ and $\underline{v}_k$ can be chosen to include a constant offset. This can be thought of as a directional equivalent to non-zero noise in the linear setting. For example, the mode of $\underline{w}_k$ can be chosen such that it represents a known angular velocity or a given control input. Alternatively, to avoid dealing with nonzero-mean noise distributions, a rotation may be applied to $\underline{x}_k$ first and zero-mean noise added subsequently.

The predicted and estimated distributions at time k are described by their parameter matrices $(\mathbf{M}_k^p, \mathbf{Z}_k^p)$ and $(\mathbf{M}_k^e, \mathbf{Z}_k^e)$, respectively. The noise distributions at time k are described by $(\mathbf{M}_k^w, \mathbf{Z}_k^w)$ and $(\mathbf{M}_k^v, \mathbf{Z}_k^v)$.

Algorithm 1: Algorithm for prediction step.

Input: estimate $\mathbf{M}_k^e, \mathbf{Z}_k^e$, noise $\mathbf{M}_k^w, \mathbf{Z}_k^w$

Output: prediction $\mathbf{M}_{k+1}^p, \mathbf{Z}_{k+1}^p$

/* obtain covariance matrices $\mathbf{A}, \mathbf{B}$ */
$\mathbf{A} \leftarrow \mathbf{M}_k^e \cdot \mathrm{diag}\left(\frac{1}{F}\frac{\partial F}{\partial z_1}, \ldots, \frac{1}{F}\frac{\partial F}{\partial z_d}\right) \cdot (\mathbf{M}_k^e)^T$;
$\mathbf{B} \leftarrow \mathbf{M}_k^w \cdot \mathrm{diag}\left(\frac{1}{F}\frac{\partial F}{\partial z_1}, \ldots, \frac{1}{F}\frac{\partial F}{\partial z_d}\right) \cdot (\mathbf{M}_k^w)^T$;
/* obtain $\mathbf{C}$ with to Lemma 2 or 4 */
$c_{ij} \leftarrow \mathrm{E}((\underline{x} \oplus \underline{y})_i \cdot (\underline{x} \oplus \underline{y})_j), \quad i, j = 1, \ldots, d$;
$\mathbf{C} = (c_{ij})_{ij}$;
/* obtain $\mathbf{M}_{k+1}^p, \mathbf{Z}_{k+1}^p$ based on $\mathbf{C}$ */
$\mathbf{M}_{k+1}^p, \mathbf{Z}_{k+1}^p \leftarrow \mathrm{MLE}(\mathbf{C})$;

6.1 Prediction Step

The prediction can be calculated with the Chapman-Kolmogorov-equation

$$f_p(\underline{x}_{k+1}) \tag{42}$$
$$= \int_{S_{d-1}} f(\underline{x}_{k+1} | \underline{x}_k) f_e(\underline{x}_k) d\underline{x}_k \tag{43}$$
$$= \int_{S_{d-1}} \int_{S_{d-1}} f(\underline{x}_{k+1}, \underline{w}_k | \underline{x}_k) d\underline{w}_k f_e(\underline{x}_k) d\underline{x}_k \tag{44}$$
$$= \int_{S_{d-1}} \int_{S_{d-1}} f(\underline{x}_{k+1}, | \underline{w}_k, \underline{x}_k) f_w(\underline{w}_k) d\underline{w}_k f_e(\underline{x}_k) d\underline{x}_k \tag{45}$$
$$= \int_{S_{d-1}} \int_{S_{d-1}} \delta(\underline{w}_k - (\underline{x}_k^{-1} \oplus \underline{x}_{k+1})) f_w(\underline{w}_k) d\underline{w}_k f_e(\underline{x}_k) d\underline{x}_k$$
$$= \int_{S_{d-1}} f_w(\underline{x}_k^{-1} \oplus \underline{x}_{k+1}) f_e(\underline{x}_k) d\underline{x}_k \ . \tag{46}$$

This yields

$$(\mathbf{M}_{k+1}^p, \mathbf{Z}_{k+1}^p) = \mathrm{composition}((\mathbf{M}_k^e, \mathbf{Z}_k^e), (\mathbf{M}_k^w, \mathbf{Z}_k^w)) \ , \tag{47}$$

which uses the previously introduced composition operation to disturb the estimate with the system noise.

6.2 Measurement Update

Given a measurement $\hat{\underline{z}}_k$, we can calculate the updated density $\hat{f}$ of $\underline{x}_k$ given $\underline{z}_k$ from the density f_v of $\underline{v}_k$ and the prior density f_x of $\underline{x}_k$. This is performed using the transformation theorem for densities and Bayes' rule

$$\hat{f}(\underline{a}) \propto f_v(\underline{a}^{-1} \oplus \hat{\underline{z}}) \cdot f_x(\underline{a}) \ . \tag{48}$$

First, we make use of the fact that negation corresponds to conjugation for quaternions and complex numbers of unit length. Thus, we have $\underline{a}^{-1} \oplus \hat{\underline{z}} = \mathbf{D}(\hat{\underline{z}}^{-1} \oplus \underline{a})$ with $\mathbf{D} = \mathrm{diag}(1, -1)$ for $d = 2$ and $\mathbf{D} = \mathrm{diag}(1, -1, -1, -1)$. As in our proof of Lemma 7, we can use a matrix representation $\mathbf{Q}_{\hat{\underline{z}}^{-1}}$ of $\hat{\underline{z}}^{-1}$ such that $\hat{\underline{z}}^{-1} \oplus \underline{a} = \mathbf{Q}_{\hat{\underline{z}}^{-1}} \underline{a}$. Thus, we obtain

$$f_v(\underline{a}^{-1} \oplus \hat{\underline{z}}) = f_v(\mathbf{D} \cdot \mathbf{Q}_{\hat{\underline{z}}^{-1}} \underline{a}) \ . \tag{49}$$

Algorithm 2: Algorithm for update step.

Input: prediction $\mathbf{M}_k^p, \mathbf{Z}_k^p$, noise $\mathbf{M}_k^v, \mathbf{Z}_k^v$, measurement $\hat{\underline{z}}_k$

Output: estimate $\mathbf{M}_k^e, \mathbf{Z}_k^e$

```
/* rotate noise according to
   measurement                          */
```
$\mathbf{M} \leftarrow \hat{\underline{z}} \oplus (\mathbf{D}\mathbf{M}_k^v)$;
```
/* multiply with prior distribution     */
```
$(\mathbf{M}_k^e, \mathbf{Z}_k^e) \leftarrow \text{multiply}((\mathbf{M}, \mathbf{Z}_k^v)), (\mathbf{M}_k^p, \mathbf{Z}_k^p))$;

This yields

$$
\begin{aligned}
& f_v(\mathbf{D} \cdot \mathbf{Q}_{\hat{\underline{z}}^{-1}} \cdot \underline{a}) && (50)\\
& \quad \propto \exp(\underline{a}^T \, \mathbf{Q}_{\hat{\underline{z}}^{-1}}^T \, \mathbf{D}^T \, \mathbf{M}_k^v \, \mathbf{Z}_k^v (\mathbf{M}_k^v)^T \, \mathbf{D} \, \mathbf{Q}_{\hat{\underline{z}}^{-1}} \underline{a}) && (51)\\
& \quad = \exp(\underline{a}^T \, \mathbf{Q}_{\hat{\underline{z}}} \, \mathbf{D} \, \mathbf{M}_k^v \, \mathbf{Z}_k^v (\mathbf{M}_k^v)^T \, \mathbf{D} \, \mathbf{Q}_{\hat{\underline{z}}^{-1}} \underline{a}) \; . && (52)
\end{aligned}
$$

The last identity is due to $\mathbf{D}^T = \mathbf{D}$ and the fact that the transpose of the usual matrix representations of complex numbers and quaternions corresponds to the representation of their conjugates.

Finally, the parameters of the resulting Bingham distribution are obtained by

$$
(\mathbf{M}_k^e, \mathbf{Z}_k^e) = \text{multiply}((\mathbf{M}, \mathbf{Z}_k^v), (\mathbf{M}_k^p, \mathbf{Z}_k^p)) \qquad (53)
$$

with $\mathbf{M} = (\hat{\underline{z}} \oplus (\mathbf{D}\,\mathbf{M}_k^v))$, where $\oplus$ is evaluated for each column of $\mathbf{D}\,\mathbf{M}_k^v$ and "multiply" denotes the procedure outlined in Sec. 5-A. This operation can be performed solely on the Bingham parameters and does not involve the calculation of normalization constants (see Algorithm 2).

7. Evaluation

The proposed filter was evaluated in simulations for both the 2D and 4D cases. In this section, all angles are given in radians unless specified differently.

For comparison, we implemented modified Kalman filters with two- and four-dimensional state vectors [19]. In order to deal with axial estimates, we introduce two modifications:

1) We mirror the estimate $\hat{\underline{z}} \leftarrow -\hat{\underline{z}}$ if the angle between prediction and measurement $\angle(\underline{x}_k^p, \hat{\underline{z}}) > \pi/2$.
2) We normalize the estimate $\underline{x}_k^e$ after each update step $\underline{x}_k^e \leftarrow \frac{\underline{x}_k^e}{||\underline{x}_k^e||}$.

It should be noted that in two-dimensional scenarios, a comparison to a Kalman filter with a scalar state is also possible. We previously performed this simulation in [32] and showed that the Bingham filter is superior to Kalman filter with scalar state in the considered scenario.

7.1 Two-Dimensional Case

In our example, we consider the estimation of an axis in robotics. This could be the axis of a symmetric rotor blade or any robotic joint with 180° symmetry. We use the initial estimate with mode $(0, 1)^T$

$$
\mathbf{M}_0^e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{Z}_0^e = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \; , \qquad (54)
$$

the system noise with mode $(1, 0)^T$

$$
\mathbf{M}_k^w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{Z}_k^w = \begin{pmatrix} -200 & 0 \\ 0 & 0 \end{pmatrix} \; , \qquad (55)
$$

and the measurement noise with mode $(1, 0)^T$

$$
\mathbf{M}_k^v = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{Z}_k^v = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} \; . \qquad (56)
$$

The true initial state is given by $(1, 0)^T$, i.e., the initial estimate with mode $(0, 1)^T$ is very poor.

To calculate the covariance matrices for the Kalman filter we fit a Gaussian to one of the two Bingham modes by means of numerical integration, i.e.,

$$
\begin{aligned}
C = & \int_{\alpha_m - \pi/2}^{\alpha_m + \pi/2} f([\cos(\phi), \sin(\phi)]^T) && (57)\\
& \cdot \begin{bmatrix} (\cos(\phi) - m_1)^2 & (\cos(\phi) - m_1)(\sin(\phi) - m_2) \\ * & (\sin(\phi) - m_2)^2 \end{bmatrix} d\phi \; ,
\end{aligned}
$$

where $(m_1, m_2)^T$ is a mode of the Bingham distribution and $\alpha_m = \text{atan2}(m_2, m_1)$. The original Bingham distribution and the resulting Gaussian are illustrated in Fig. 6. We obtain the parameters

$$
C_0^e = \begin{bmatrix} 3.8 \times 10^{-1} & 0 \\ 0 & 1.5 \times 10^{-1} \end{bmatrix} \; , \qquad (58)
$$

$$
C_k^w = \begin{bmatrix} 4.7 \times 10^{-6} & 0 \\ 0 & 2.5 \times 10^{-3} \end{bmatrix} \; , \qquad (59)
$$

$$
C_k^v = \begin{bmatrix} 8.8 \times 10^{-2} & 0 \\ 0 & 2.8 \times 10^{-1} \end{bmatrix} \; , \qquad (60)
$$

which is equivalent to angular standard deviations of 43.9° for the first time step, 2.9° for the system noise and 36.3° for the measurement noise.

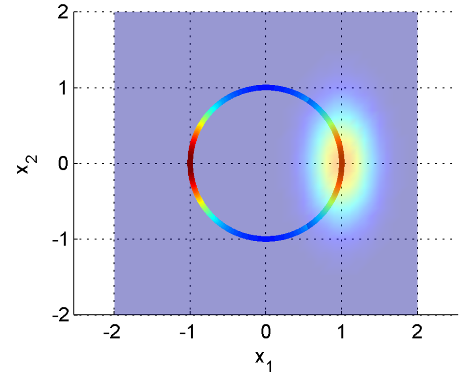

Fig. 6: The Bingham density with parameters $\mathbf{M}_k^v, \mathbf{Z}_k^v$ (on the circle) and a Gaussian (in the plane) fitted to one of the modes with the mean located at the mode and covariance computed according to (57).

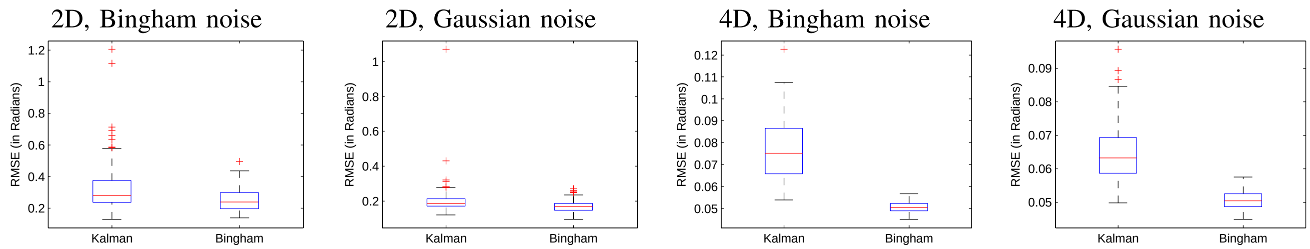

Fig. 4: RMSE from 100 Monte Carlo runs.

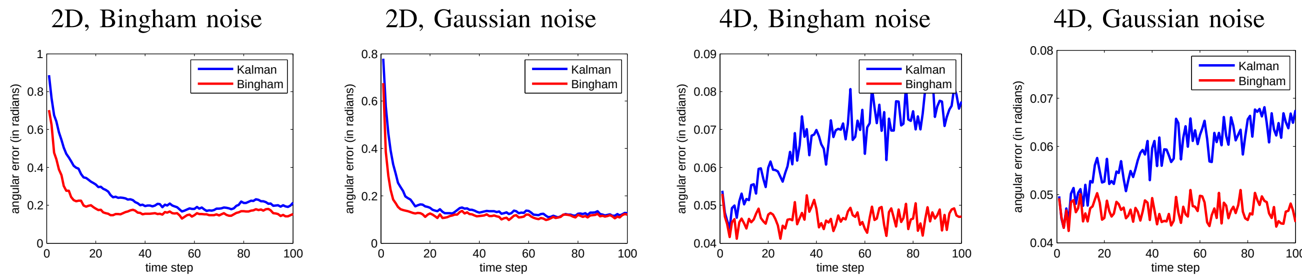

Fig. 5: Average error over time from 100 Monte Carlo runs.

7.2 Four-Dimensional Case

For the quaternion case, we use the initial estimate with mode $(0, 0, 0, 1)^T$

$$\mathbf{M}_0^e = \mathbf{I}_{4\times 4}, \quad \mathbf{Z}_0^e = \mathrm{diag}(-1, -1, -1, 0) \ , \tag{61}$$

the system noise with mode $(1, 0, 0, 0)^T$

$$\mathbf{M}_k^w = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} , \tag{62}$$

$$\mathbf{Z}_k^w = \mathrm{diag}(-200, -200, -2, 0) \ , \tag{63}$$

and the measurement noise with mode $(1, 0, 0, 0)^T$

$$\mathbf{M}_k^v = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} , \tag{64}$$

$$\mathbf{Z}_k^v = \mathrm{diag}(-500, -500, -500, 0) \ . \tag{65}$$

The true initial state is $(0, 1, 0, 0)^T$, i.e., the initial estimate with mode $(0, 0, 0, 1)^T$ is very poor. It should be noted that the system noise is not isotropic, because the uncertainty is significantly higher in the third dimension than in the first two.

We converted the Bingham noise parameters to Gaussians analogous to the two-dimensional case.

7.3 Results

We simulate the system for a duration of $k_{\max} = 100$ time steps. For evaluation, we consider the angular RMSE given by

$$\sqrt{\frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} (e_k)^2} \tag{66}$$

with angular error

$$e_k = \min\Big(\measuredangle(\underline{x}_k^{\text{true}}, \mathrm{mode}(\mathbf{M}_k^e)), \tag{67}$$

$$\pi - \measuredangle(\underline{x}_k^{\text{true}}, \mathrm{mode}(\mathbf{M}_k^e)) \Big) \tag{68}$$

at time step k. Obviously, $0 \leq e_k \leq \frac{\pi}{2}$ holds, which is consistent with our assumption of $180°$ symmetry. This error measure can be used in the two- and the four-dimensional setting. As we have shown in [8], the angle between two quaternions in four-dimensional space is proportional to the angle of the corresponding rotation between the two orientations in three dimensions, so e_k is a reasonable measure for quaternions.

The presented results are based on 100 Monte Carlo runs. Even though our filter is computationally more demanding than a Kalman filter, it is still fast enough for real-time applications. On a standard laptop with an Intel Core i7-2640M CPU, our non-optimized implementation in MATLAB needs approximately 8 ms for one time step (prediction and update) in the two-dimensional case. In the four-dimensional case, we implemented the hypergeometric function in C, but the maximum likelihood estimation is written in MATLAB. The calculations fore one time step require 13 ms on our laptop.

We consider two different types of noise, Bingham and Gaussian. Even though Bingham distributed noise may be a more realistic assumption in a circular setting, we do not want to give the proposed filter an unfair advantage by comparing it to a filter with an incorrect noise assumption. In the cases of Gaussian noise, we obtain the parameters of the Gaussian as described in (57) and convert the resulting Gaussians back to Bingham distributions to account for any information that was lost in the conversion from Bingham to Gaussian.

The results for all considered scenarios are depicted in Fig. 4 and Fig. 5. It can be seen that the proposed filter outperforms the Kalman filter in all considered scenarios. Particularly, it outperforms the Kalman filter even if Gaussian noise is used. This is due to the fact that projecting the Gaussian noise to the unit sphere does not yield a Gaussian distribution, which makes the Kalman filter suboptimal. Furthermore, the Kalman filter does not consider the nonlinearity of the underlying domain. As expected, the advantage of using the Bingham filter is larger if the noise is following a Bingham distribution.

8. Conclusion

We have presented a recursive filter based on the Bingham distribution. It can be applied to angular estimation in the plane with $180°$ symmetry and to quaternion-based estimation of orientation of objects in three-dimensional space. Thus, it is relevant for a wide area of applications, particularly when uncertainties occur, for example as a result of cheap sensors or very limited prior knowledge.

We have evaluated the proposed approaches in very challenging settings involving large non-isotropic noise. Our simulations have shown the superiority of the presented approach compared to the solution based on an adapted Kalman filter for both the circular and the quaternion case. This is true no matter if the noise is distributed according to a Bingham or a Gaussian distribution. Furthermore, we have shown that the proposed algorithms are fast enough on a typical laptop to be used in real-time applications.

Open challenges include an efficient estimator of the Bingham parameters based on available data. This makes an efficient evaluation of the confluent hypergeometric function necessary. Furthermore, extensions to nonlinear measurement equations and the group of rigid body motions $SE(3)$ may be of interest.

Acknowledgment

This work was partially supported by grants from the German Research Foundation (DFG) within the Research Training Groups RTG 1194 “Self-organizing Sensor-Actuator-Networks” and RTG 1126 “Soft-tissue Surgery: New Computer-based Methods for the Future Workplace”.

Appendix A
Relation to Gaussian Distribution

The Bingham distribution is closely related to the widely used Gaussian distribution.

Definition 4. *The pdf of a multivariate Gaussian distribution in $\mathbb{R}^d$ is given by*

$$f^G(\underline{x}) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{\Sigma}}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \mathbf{\Sigma}^{-1}(\underline{x}-\underline{\mu})\right)$$

with mean $\underline{\mu} \in \mathbb{R}^d$ and positive definite covariance $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$.

If we require $\underline{\mu} = 0$ and restrict $\underline{x}$ to the unit hypersphere, i.e., $||\underline{x}|| = 1$, we have

$$f^G(\underline{x}) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{\Sigma}}} \exp\left(\underline{x}^T \left(-\frac{1}{2}\mathbf{\Sigma}^{-1}\right)\underline{x}\right) , \quad (69)$$

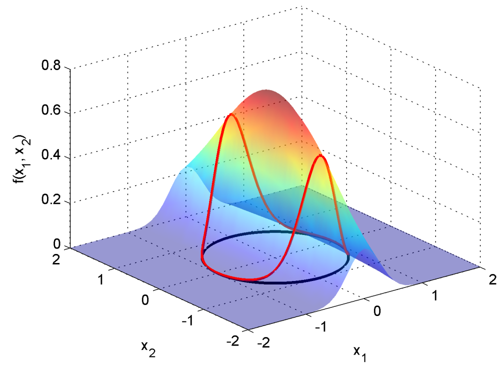

Fig. 7: A two-dimensional Gaussian distribution, which is restricted to the unit circle to obtain a two-dimensional Bingham distribution.

which is an unnormalized Bingham distribution with $\mathbf{MZM}^T = -\frac{1}{2}\mathbf{\Sigma}^{-1}$. Conversely, any Bingham distribution is a restricted Gaussian distribution with $\mathbf{\Sigma} = (-2\mathbf{MZM}^T)^{-1}$ if $\mathbf{MZM}^T$ is negative definite. This condition can always be fulfilled by adding a multiple of the identity matrix $\mathbf{I}_{d\times d}$ to $\mathbf{Z}$. Modifying $\mathbf{Z}$ in this way yields a different Gaussian distribution, but the values on the unit hypersphere stay the same, i.e., the Bingham distribution does not change. A graphical illustration of the relation between a Gaussian density and the corresponding Bingham resulting from conditioning the original Gaussian random vector to unit length is given in Fig. 7.

Due to local linear structure of the underlying manifold, each mode of the Bingham distribution defined on this manifold is very similar to a Gaussian of dimension $d-1$ if and only if the uncertainty is small. This can be seen in Fig. 8, which shows the Kullback-Leibler divergence

$$\int_0^\pi f([\cos(\theta), \sin(\theta)]^T) \log\left(\frac{f([\cos(\theta), \sin(\theta)]^T)}{f^G(\theta, \mu, \sigma)}\right) d\theta \quad (70)$$

between one mode of a Bingham pdf for $d = 2$ and a corresponding one-dimensional Gaussian pdf on the semicircle.

Appendix B
Relation to von Mises Distribution

The Bingham distribution for $d = 2$ is closely related to the von Mises distribution. We can exploit this fact at some points in this paper.

Definition 5. *A von Mises distribution [15] is given by the probability density function*

$$f^{VM}(\phi; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\phi - \mu)) \quad (71)$$

for $\phi \in [0, 2\pi)$, location parameter $\mu \in [0, 2\pi)$ and concentration parameter $\kappa > 0$, where $I_0(\kappa)$ is the modified Bessel function [1] of order 0.

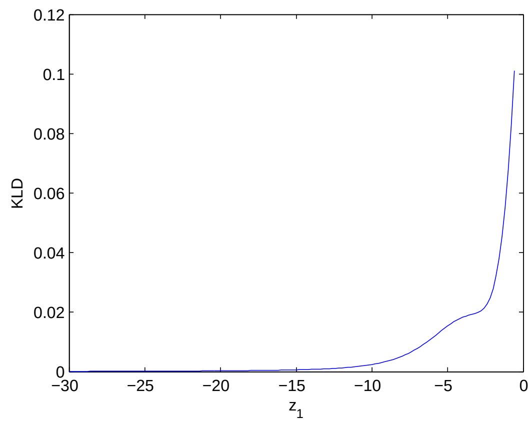

Fig. 8: Kullback-Leibler divergence on the interval $[0, \pi]$ between a Bingham pdf with $\mathbf{M} = \mathbf{I}_{2\times 2}$, $\mathbf{Z} = \text{diag}(z_1, 0)$ and a Gaussian pdf with equal mode and standard deviation. For small uncertainties ($z_1 < -15$, which corresponds to a standard deviation of about 11°), the Gaussian and Bingham distributions are almost indistinguishable. However, for large uncertainties, the Gaussian approximation becomes quite poor.

Based on this definition, we can show an interesting relation between Bingham and von Mises distributions [35].

Lemma 5. *For the circular case, every Bingham density is equal to a von Mises density rescaled to* $[0, \pi)$ *and repeated on* $[\pi, 2\pi)$.

Proof. We can reparameterize a Bingham distribution with $d = 2$ by substituting $\underline{x} = [\cos(\theta), \sin(\theta)]^T$ and

$$\mathbf{M} = \begin{bmatrix} -\sin(\nu) & \cos(\nu) \\ \cos(\nu) & \sin(\nu) \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} z_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{72}$$

to attain the von Mises distribution. With

$$\mathbf{MZM}^T = z_1 \begin{bmatrix} \sin^2(\nu) & -\cos(\nu)\sin(\nu) \\ -\cos(\nu)\sin(\nu) & \cos^2(\nu) \end{bmatrix}, \tag{73}$$

this yields the pdf

$$\begin{aligned} f(\theta) =& \frac{1}{F} \exp\left([\cos(\theta), \sin(\theta)]\mathbf{MZM}^T[\cos(\theta), \sin(\theta)]^T\right) \\ =& \frac{1}{F} \exp\left(z_1(\cos(\theta)\sin(\nu) - \sin(\theta)\cos(\nu))^2\right) \quad (74) \\ =& \frac{1}{F} \exp\left(z_1 \sin^2(\theta - \nu))\right) \quad (75) \end{aligned}$$

according to $\sin(a - b) = \sin(a)\cos(b) - \cos(a)\sin(b)$. Now we apply $\sin^2(a) = \frac{1}{2}(1 - \cos(2a))$ and get

$$f(\theta) = \frac{1}{F} \exp\left(\frac{z_1}{2}\right) \exp\left(-\frac{z_1}{2}\cos(2\theta - 2\nu)\right), \tag{76}$$

which exactly matches a von Mises distribution with $\phi = 2\theta, \mu = 2\nu$, and $\kappa = -\frac{z_1}{2}$ that has been repeated twice, i.e., $\theta \in [0, 2\pi)$ and $\phi \in [0, 4\pi)$. □

This property can be exploited to derive a formula for the normalization constant of the Bingham distribution.

Lemma 6. *For* $d = 2$*, the normalization constant is given by*

$$F = 2\pi \cdot I_0\left(\frac{z_1}{2}\right) \exp\left(\frac{z_1}{2}\right) \tag{77}$$

with derivatives

$$\frac{\partial}{\partial z_1} F = \pi \exp\left(\frac{z_1}{2}\right)\left(I_1\left(\frac{z_1}{2}\right) + I_0\left(\frac{z_1}{2}\right)\right) \tag{78}$$

$$\begin{aligned} \frac{\partial}{\partial z_2} F =& \pi \exp\left(\frac{z_1 + z_2}{2}\right) \\ & \cdot \left(I_0\left(\frac{z_1 - z_2}{2}\right) - I_1\left(\frac{z_1 - z_2}{2}\right)\right) \end{aligned} \tag{79}$$

Proof. In order to consider the derivative with respect to z_2, we first consider a Bingham density with arbitrary z_2, which yields

$$f(\underline{x}) = \frac{1}{F} \exp\left(\underline{x}^T \mathbf{M} \begin{bmatrix} z_1 & 0 \\ 0 & z_2 \end{bmatrix} \mathbf{M}^T \underline{x}\right) \tag{80}$$

$$\begin{aligned} &= \frac{1}{F} \exp\left(\underline{x}^T \mathbf{M} \begin{bmatrix} z_1 - z_2 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{M}^T \underline{x} + z_2 \cdot \underline{x}^T \mathbf{MM}^T \underline{x}\right) \\ &= \frac{\exp(z_2)}{F} \exp\left(\underline{x}^T \mathbf{M} \begin{bmatrix} z_1 - z_2 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{M}^T \underline{x}\right). \quad (81) \end{aligned}$$

We use the formula for the normalization constant of a von Mises distribution to obtain

$$\frac{\exp(z_2)}{F} = \frac{1}{2\pi I_0(\frac{z_1 - z_2}{2})}. \tag{82}$$

Solving this equation for F and substituting $z_2 = 0$ shows (77). The derivatives are calculated with [1, eq. 9.6.27]. □

Appendix C
Relation to von Mises-Fisher Distribution

The von Mises-Fisher distribution is a hyperspherical generalization of the von Mises distribution.

Definition 6. *A von Mises-Fisher distribution [7] is given by the pdf*

$$f^{VMF}(\underline{x}; \underline{\mu}, \kappa) = C_d(\kappa) \exp(\kappa \underline{\mu}^T \underline{x}) \tag{83}$$

with

$$C_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)} \tag{84}$$

for $\underline{x} \in S_{d-1}$*, location parameter* $\underline{\mu} \in S_{d-1}$ *and scalar concentration parameter* $\kappa > 0$*, where* $I_n(\kappa)$ *is the modified Bessel function [1] of order* n*.*

Unlike the Bingham distribution, the von Mises-Fisher distribution is unimodal and not antipodally symmetric, but radially symmetric around the axis of $\underline{\mu}$. We note that by use of hyperspherical coordinates, we can reformulate the pdf of the von Mises-Fisher distribution as

$$f^{\text{VMF}}(\cdot; \kappa) : [0, \pi] \to \mathbb{R}^+, \tag{85}$$

$$f^{\text{VMF}}(\phi; \kappa) = C_d(\kappa) \exp(\kappa \cos(\phi)) \sin^{d-1}(\phi), \tag{86}$$

$$\text{where } \phi = \angle(\underline{\mu}, \underline{x}). \tag{87}$$

The term $\sin^{d-1}(\phi)$ arises as a volume-correcting term when the substitution rule is applied. Using this definition, we can show an interesting relation between certain Bingham distributions and the von Mises-Fisher distribution.

Lemma 7. *For a Bingham distribution with $z_1 = \cdots = z_{d-1}$ with pdf $f(\cdot)$, we have the relation*

$$f^{VMF}(\theta;\kappa) = (2\cos(\theta))^{d-1} \cdot f(\theta) \tag{88}$$

to the von Mises-Fisher distribution.

Proof. We consider $\mathbf{Z} = \operatorname{diag}(z_1 \dots, z_1, 0)$ and $\mathbf{M} = [\dots|\underline{\mu}]$. From the Bingham pdf, we obtain

$$\begin{aligned} f(\underline{x}) =& \frac{1}{F}\exp(\underline{x}^T \mathbf{M}\mathbf{Z}\mathbf{M}^T \underline{x}) & (89)\\ =& \frac{1}{F}\exp(\underline{x}^T \mathbf{M}\operatorname{diag}(z_1 \dots, z_1, 0)\mathbf{M}^T \underline{x}) & (90)\\ =& \frac{1}{F}\exp(\underline{x}^T \mathbf{M}\operatorname{diag}(0 \dots, 0, -z_1)\mathbf{M}^T \underline{x} \\ & \quad + z_1 \underline{x}^T \mathbf{M}\mathbf{M}^T \underline{x}) & (91)\\ =& \frac{\exp(z_1)}{F}\exp(-z_1 \underline{x}^T \mathbf{M}\operatorname{diag}(0 \dots, 0, 1)\mathbf{M}^T \underline{x}) \ . & (92) \end{aligned}$$

We use the fact that the last column of $\mathbf{M}$ contains the mode $\underline{\mu}$ and obtain

$$\begin{aligned} f(\underline{x}) =& \frac{\exp(z_1)}{F}\exp(-z_1 \underline{x}^T \underline{\mu}\underline{\mu}^T \underline{x}) & (93)\\ =& \frac{\exp(z_1)}{F}\exp(-z_1 (\underline{\mu}^T \underline{x})^2) & (94)\\ =& \frac{\exp(z_1)}{F}\exp(-z_1 (\cos(\angle(\underline{x}, \underline{\mu}))^2) \ . & (95) \end{aligned}$$

By using the trigonometric identity $\cos^2(x) = (1+\cos(2x))/2$, we obtain

$$f(\underline{x}) = \frac{\exp(\frac{z_1}{2})}{F}\exp\left(-\frac{z_1}{2}\cos(2\angle(\underline{x}, \underline{\mu}))\right) \ . \tag{96}$$

Substitution of spherical coordinates as above yields the pdf $f : [0, \frac{\pi}{2}] \to \mathbb{R}^+$,

$$f(\theta) = \frac{\exp(\frac{z_1}{2})}{F}\exp\left(-\frac{z_1}{2}\cos(2\theta)\right)\sin^{d-1}(\theta) \ . \tag{97}$$

On the other hand, the von Mises-Fisher pdf can be stated as

$$f^{VMF}(\phi;\kappa) = C_d(\kappa)\exp(\kappa\cos(\phi))\sin^{d-1}(\phi) \ . \tag{98}$$

We set $\kappa = -\frac{z_1}{2}$ and $\phi = 2\theta$, which yields

$$\begin{aligned} & f^{VMF}(\theta;\kappa) & (99)\\ &= C_d\left(-\frac{z_1}{2}\right)\exp\left(-\frac{z_1}{2}\cos(2\theta)\right)\sin^{d-1}(2\theta) & (100)\\ &= \frac{\sin^{d-1}(2\theta)}{\sin^{d-1}(\theta)} \cdot f(\theta) = (2\cos(\theta))^{d-1} \cdot f(\theta) \ . & (101) \end{aligned}$$

□

This fact can be used to simplify the maximum likelihood estimation when the underlying samples are (or can be assumed to be) generated by an isotropic Bingham distribution, i.e., when the corresponding density is circularly symmetric around the modes If the samples are reweighted by a factor of $(2\cos(\theta))^{d-1}$ and their angle around the mean is doubled, a von Mises-Fisher maximum likelihood estimate can be performed to obtain κ and subsequently z_1. This can be advantageous, because the maximum likelihood estimate for a von Mises-Fisher distribution is computationally less demanding than for the Bingham distribution [43].

APPENDIX D
RELATION TO KENT DISTRIBUTION

Furthermore, it should be noted that the d-dimensional Bingham distribution is a special case of the d-dimensional Kent distribution [20]. The Kent distribution is also commonly referred to as the Fisher-Bingham distribution because it is a generalization of both the von Mises-Fisher and the Bingham distribution.

Definition 7. *The pdf of the Kent distribution is given by*

$$f(\underline{x}) \propto \exp(\kappa\underline{\mu}^T \underline{x} + \sum_{j=2}^{d} \beta_j (\underline{\gamma}_j^T \underline{x})^2) \ , \tag{102}$$

where $\underline{x} \in S_{d-1}$, and $\underline{\mu} \in S_{d-1}$ is the location parameter, $\kappa \geq 0$ is the concentration around μ, the directions $\underline{\gamma}_2, \dots, \underline{\gamma}_d \in S_{d-1}$ are orthogonal and have corresponding concentrations $\beta_2 \geq \cdots \geq \beta_d \in \mathbb{R}$.

It can be seen that for $\kappa = 0$, this yields a Bingham distribution. The vectors $\underline{\gamma}_2, \dots, \underline{\gamma}_d$ correspond to the $\mathbf{M}$ matrix and the coefficients $\beta_2, \dots, \beta_d$ correspond to the diagonal of the $\mathbf{Z}$ matrix. This fact allows the application of methods developed for the Kent distribution such as [24], [26] in conjunction with the Bingham distribution. For $\beta_2 = \cdots = \beta_d = 0$, the Kent distribution reduces to a von Mises-Fisher distribution.

APPENDIX E
PROOF OF LEMMA 2.

Proof. The covariance of the composition

$$\begin{aligned} \mathbf{C} &= \operatorname{Cov}(\underline{x} \oplus \underline{y}) & (103)\\ &= \operatorname{Cov}\left(\begin{pmatrix} x_1 y_1 - x_2 y_2 \\ x_1 y_2 + x_2 y_1 \end{pmatrix}\right) & (104)\\ &= \begin{pmatrix} \operatorname{Var}(x_1 y_1 - x_2 y_2) & \operatorname{Cov}(x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1) \\ * & \operatorname{Var}(x_1 y_2 + x_2 y_1) \end{pmatrix} \end{aligned}$$

can be obtained by calculating the matrix entries individually. For the first entry we get

$$\begin{aligned} c_{11} &= \operatorname{Var}(x_1 y_1 - x_2 y_2) & (105)\\ &= \operatorname{E}((x_1 y_1 - x_2 y_2)^2) - (\operatorname{E}(x_1 y_1 - x_2 y_2))^2 & (106)\\ &= \operatorname{E}(x_1^2 y_1^2 - 2 x_1 y_1 x_2 y_2 + x_2^2 y_2^2) \\ &\quad - (\operatorname{E}(x_1 y_1) - \operatorname{E}(x_2 y_2))^2 & (107)\\ &= \operatorname{E}(x_1^2)\operatorname{E}(y_1^2) - 2\operatorname{E}(x_1 x_2)\operatorname{E}(y_1 y_2) + \operatorname{E}(x_2^2)\operatorname{E}(y_2^2) & (108)\\ &\quad - (\underbrace{\operatorname{E}(x_1)}_{0}\underbrace{\operatorname{E}(y_1)}_{0} - \underbrace{\operatorname{E}(x_2)}_{0}\underbrace{\operatorname{E}(y_2)}_{0})^2 & (109)\\ &= a_{11} b_{11} - 2 a_{12} b_{12} + a_{22} b_{22} \ . & (110) \end{aligned}$$

We use independence of $\underline{x}$ and $\underline{y}$ in (107), linearity of the expectation value in (108), and symmetry of the Bingham in (109). Analogously we calculate

$$\begin{aligned}
c_{22} &= \mathrm{Var}(x_1 y_2 - x_2 y_1) && (111)\\
&= \mathrm{E}((x_1 y_2 - x_2 y_1)^2) - (\mathrm{E}(x_1 y_2 - x_2 y_1))^2 && (112)\\
&= \mathrm{E}(x_1^2 y_2^2 - 2 x_1 y_1 x_2 y_2 + x_2^2 y_1^2)\\
&\quad - (\mathrm{E}(x_1 y_2) - \mathrm{E}(x_2 y_1))^2 && (113)\\
&= \mathrm{E}(x_1^2)\,\mathrm{E}(y_2^2) - 2\,\mathrm{E}(x_1 x_2)\,\mathrm{E}(y_1 y_2) + \mathrm{E}(x_2^2)\,\mathrm{E}(y_1^2)\\
&\quad - (\underbrace{\mathrm{E}(x_1)}_{0}\underbrace{\mathrm{E}(y_2)}_{0} - \underbrace{\mathrm{E}(x_2)}_{0}\underbrace{\mathrm{E}(y_1)}_{0})^2 && (114)\\
&= a_{11} b_{22} - 2 a_{12} b_{12} + a_{22} b_{11}\ . && (115)
\end{aligned}$$

The off-diagonal entry can be calculated similarly

$$\begin{aligned}
c_{12} &= \mathrm{Cov}(x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1) && (116)\\
&= \mathrm{E}((x_1 y_1 - x_2 y_2) \cdot (x_1 y_2 + x_2 y_1))\\
&\quad - \mathrm{E}(x_1 y_1 - x_2 y_2) \cdot \mathrm{E}(x_1 y_2 + x_2 y_1) && (117)\\
&= \mathrm{E}(x_1^2 y_1 y_2 - x_1 x_2 y_2^2 + x_1 x_2 y_1^2 - x_2^2 y_1 y_2)\\
&\quad - (\mathrm{E}(x_1)\,\mathrm{E}(y_1) - \mathrm{E}(x_2)\,\mathrm{E}(y_2))\\
&\quad \cdot (\mathrm{E}(x_1)\,\mathrm{E}(y_2) + \mathrm{E}(x_2)\,\mathrm{E}(y_1)) && (118)\\
&= a_{11} b_{12} - a_{12} b_{22} + a_{12} b_{11} - a_{22} b_{12}\ . && (119)
\end{aligned}$$

Because $\mathbf{C}$ is a symmetrical matrix, this concludes the proof of Lemma 2. □

References

[1] ABRAMOWITZ, M., AND STEGUN, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. Dover, New York, 1972.

[2] ANTONE, M., AND TELLER, S. Scalable, Absolute Position Recovery for Omni-directional Image Networks. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (2001), vol. 1, pp. I–398–I–405.

[3] AZMANI, M., REBOUL, S., CHOQUEL, J.-B., AND BENJELLOUN, M. A Recursive Fusion Filter for Angular Data. In *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)* (2009), pp. 882–887.

[4] BINGHAM, C. *Distributions on the Sphere and on the Projective Plane*. PhD thesis, Yale University, 1964.

[5] BINGHAM, C. An Antipodally Symmetric Distribution on the Sphere. *The Annals of Statistics 2*, 6 (Nov. 1974), 1201–1225.

[6] FISHER, N. I., LEWIS, T., AND EMBLETON, B. J. *Statistical Analysis of Spherical Data*. Cambridge University Press, 1987.

[7] FISHER, R. Dispersion on a Sphere. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 217*, 1130 (1953), 295–305.

[8] GILITSCHENSKI, I., KURZ, G., JULIER, S. J., AND HANEBECK, U. D. Unscented Orientation Estimation Based on the Bingham Distribution. *arXiv preprint: Systems and Control (cs.SY)* (2013).

[9] GLOVER, J. libbingham Bingham Statistics Library, 2013.

[10] GLOVER, J., BRADSKI, G., AND RUSU, R. Monte Carlo Pose Estimation with Quaternion Kernels and the Bingham Distribution. In *Proceedings of Robotics: Science and Systems (RSS 2011)* (Los Angeles, USA, 2011).

[11] GLOVER, J., AND KAELBLING, L. P. Tracking 3-D Rotations with the Quaternion Bingham Filter. Tech. rep., MIT, Mar. 2013.

[12] GLOVER, J., AND KAELBLING, L. P. Tracking the Spin on a Ping Pong Ball with the Quaternion Bingham Filter. In *IEEE Conference on Robotics and Automation (ICRA)* (2014). To Appear.

[13] HERTZBERG, C., WAGNER, R., FRESE, U., AND SCHRÖDER, L. Integrating Generic Sensor Fusion Algorithms with Sound State Representations Through Encapsulation of Manifolds. *Information Fusion 14*, 1 (Jan. 2013), 57–77.

[14] HERZ, C. S. Bessel Functions of Matrix Argument. *Annals of Mathematics 61*, 3 (1955), 474–523.

[15] JAMMALAMADAKA, S. R., AND SENGUPTA, A. *Topics in Circular Statistics*. World Scientific Pub Co Inc, 2001.

[16] JULIER, S., AND LAVIOLA, J. On Kalman Filtering With Nonlinear Equality Constraints. *IEEE Transactions on Signal Processing 55*, 6 (2007), 2774–2784.

[17] JULIER, S. J., AND UHLMANN, J. K. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE 92*, 3 (Mar. 2004), 401–422.

[18] JUPP, P. E., AND MARDIA, K. V. Maximum Likelihood Estimators for the Matrix Von Mises-Fisher and Bingham Distributions. *Annals of Statistics 7 (3)* (1979), 599–606.

[19] KALMAN, R. E. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME Journal of Basic Engineering 82* (1960), 35–45.

[20] KENT, J. T. The Fisher-Bingham Distribution on the Sphere. *Journal of the Royal Statistical Society. Series B (Methodological) 44*, 1 (1982), 71–80.

[21] KENT, J. T. Asymptotic Expansions for the Bingham Distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 36 (2)* (1987), 139–144.

[22] KENT, J. T. The Complex Bingham Distribution and Shape Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* (1994), 285–299.

[23] KOEV, P., AND EDELMAN, A. The Efficient Evaluation of the Hypergeometric Function of a Matrix Argument. *Math. Comp. 75* (2006), 833–846.

[24] KOYAMA, T., NAKAYAMA, H., NISHIYAMA, K., AND TAKAYAMA, N. Holonomic Gradient Descent for the Fisher-Bingham Distribution on the d-dimensional Sphere. *ArXiv e-prints* (Jan. 2012).

[25] KUIPERS, J. B. *Quaternions and Rotation Sequences*, vol. 66. Princeton University Press, 2002.

[26] KUME, A., AND WOOD, A. T. A. Saddlepoint Approximations for the Bingham and Fisher-Bingham Normalising Constants. *Biometrika 92*, 2 (2005), 465–476.

[27] KUME, A., AND WOOD, A. T. A. On the Derivatives of the Normalising Constant of the Bingham Distribution. *Statistics & Probability Letters 77*, 8 (2007), 832–837.

[28] KUNZE, K., AND SCHAEBEN, H. The Bingham Distribution of Quaternions and Its Spherical Radon Transform in Texture Analysis. *Mathematical Geology 36* (2004), 917–943.

[29] KURZ, G., FAION, F., AND HANEBECK, U. D. Constrained Object Tracking on Compact One-dimensional Manifolds Based on Directional Statistics. In *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)* (Montbeliard, France, Oct. 2013).

[30] KURZ, G., GILITSCHENSKI, I., AND HANEBECK, U. D. Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions. In *Proceedings of the 2013 American Control Conference (ACC 2013)* (Washington D. C., USA, June 2013).

[31] KURZ, G., GILITSCHENSKI, I., AND HANEBECK, U. D. Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions. In *Proceedings of the 2014 American Control Conference (ACC 2014)* (Portland, Oregon, USA, June 2014).

[32] KURZ, G., GILITSCHENSKI, I., JULIER, S. J., AND HANEBECK, U. D. Recursive Estimation of Orientation Based on the Bingham Distribution. In *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)* (Istanbul, Turkey, July 2013).

[33] LOVE, J. J. *Encyclopedia of Geomagnetism & Paleomagnetism*. Springer, Dordrecht, The Netherlands, 2007, ch. Bingham statistics, pp. 45–47.

[34] LUKE, Y. *Algorithms for the Computation of Mathematical Functions*. Computer science and applied mathematics. Academic Press, 1977.

[35] MARDIA, K. Characterizations of Directional Distributions. In *A Modern Course on Statistical Distributions in Scientific Work*, vol. 17. Springer Netherlands, 1975, pp. 365–385.

[36] MARDIA, K. Directional Statistics in Geosciences. *Communications in Statistics - Theory and Methods 10*, 15 (1981), 1523–1543.

[37] MARDIA, K. V., AND JUPP, P. E. *Directional Statistics*, 1 ed. Wiley, 1999.

[38] MARDIA, K. V., AND ZEMROCH, P. J. Table of Maximum Likelihood Estimates for the Bingham Distribution. *Journal of Statistical Computation and Simulation 6*, 1 (1977), 29–34.

[39] MEGÍA, I. S.-M. Which Spheres Admit a Topological Group Structure? *Revista de la Academia de Ciencias Exactas, Físico-Químicas y Naturales de Zaragoza 62* (2007), 75–79.

[40] MUIRHEAD, R. J. *Aspects of Multivariate Statistical Theory*. Wiley, 1982.

[41] MULLER, K. E. Computing the Confluent Hypergeometric Function, M(a,b,x). *Numerische Mathematik 90*, 1 (2001), 179–196.

[42] SHIRYAEV, A. N. *Probability*, 2nd ed. Springer, 1995.

[43] SRA, S. A Short Note on Parameter Approximation for von Mises-Fisher Distributions: and a Fast Implementation of Is (x). *Computational Statistics 27*, 1 (2012), 177–190.

[44] STIENNE, G., REBOUL, S., AZMANI, M., CHOQUEL, J., AND BENJELLOUN, M. A Multi-sensor Circular Particle Filter Applied to the Fusion of the GPS-L2C channels. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on* (2011).

[45] STIENNE, G., REBOUL, S., AZMANI, M., CHOQUEL, J., AND BENJELLOUN, M. A Multi-temporal Multi-sensor Circular Fusion Filter. *Information Fusion 18* (July 2013), 86–100.