Recursive Estimation of Orientation Based on the Bingham Distribution

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

[*]Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute of Anthropomatics
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—**Directional estimation is a common problem in many tracking applications. Traditional filters such as the Kalman filter perform poorly in a directional setting because they fail to take the periodic nature of the problem into account. We present a recursive filter for directional data based on the Bingham distribution in two dimensions. The proposed filter can be applied to circular filtering problems with 180 degree symmetry, i.e., rotations by 180 degrees cannot be distinguished. It is easily implemented using standard numerical techniques and is suitable for real-time applications. The presented approach is extensible to quaternions, which allow tracking arbitrary three-dimensional orientations. We evaluate our filter in a challenging scenario and compare it to a traditional Kalman filtering approach.**

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

I. Introduction

Many estimation problems involve the task of estimating angular values. These problems include, but are not limited to, estimating the pose or orientation of objects. For example, 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 reduces to estimating an axis. This can be thought of as estimation of a directionless orientation or estimation with $180°$ symmetry. All these estimation problems share the need for processing angular or directional data, which differs in many ways from the classical Euclidean setting. First, periodicity needs to be taken into account. This is especially important for measurement updates around 0, respectively 2π. Second, directional quantities do not lie in a vector space. Thus, there is no equivalent to a classical linear model, as there are no linear mappings.

In many applications, even simple estimation problems involving angular data are often considered as traditional linear or nonlinear estimation problems and handled with classical techniques such as the Kalman Filter [1], the extended Kalman Filter (EKF), or the unscented Kalman Filter (UKF) [2]. In a circular setting, most traditional 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 problem 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 filter may even diverge.

Classical strategies to avoid these problems in an angular setting involve an "intelligent" repositioning of measurements 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 [3]. There are also approaches that use operators on a manifold to provide a local approximation of a vector space [4]. While these approaches yield feasible results, they still suffer from ignoring the true geometry of circular data within their probabilistic models, which are usually based on assuming a 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 [5]. For angular data, this is not the case. Choosing a circular distribution for describing uncertainty can offer better results.

In this paper, we consider the use of the Bingham distribution [6] for recursive estimation of orientation. The Bingham distribution is defined on the hypersphere of arbitrary dimension, so it can be applied to problems of different dimensionality. Here, we focus on the two-dimensional case and apply our results to axis estimation. To the best of our knowledge, this is the first published attempt to create a recursive filter based on the Bingham distribution.

The presented methods can also be applied to the four-dimensional case, which would allow the representation of unit quaternions. Unit quaternions could then be used to estimate the full 3D orientation of an object. It is well known that Quaternions avoid the singularities present in other

representations such as Euler angles. Their only downsides are the fact that they must remain normalized and the property that there are two quaternions for every rotation (q and $-q$). Both of these issues can elegantly be overcome by use of the Bingham distribution, since it is by definition restricted to the hypersphere and is $180°$ symmetric.

This paper is structured as follows. First, we present an overview of previous work in the area of directional statistics and angular estimation (Sec. II). Then, we introduce our key idea in Sec. III. In Sec. IV, we give a detailed introduction to the Bingham distribution and we derive the necessary operations, which we will need to create a recursive Bingham filter. Based on these prerequisites, we introduce our filter in Sec. V. We have carried out an evaluation in simulations, which is presented in Sec. VI. Finally, we conclude this work in Sec. VII.

II. RELATED WORK

Directional statistics is a subdiscipline of statistics, which focuses on dealing with directional data. Classical results in directional statistics are summed up in the books by Mardia and Jupp [7] and by Jammalamadaka and Sengupta [8]. Directional statistics differs from traditional statistics by the fact that random variables located on manifolds (for example the circle or the sphere) are considered rather than random variables located in vector spaces (typically $\mathbb{R}^d$).

There is a broad range of research for investigating the 2D orientation, e. g., the work by Krindis et al. [9]. A recursive filter based on the von Mises distribution for estimating the orientation on the $SO(2)$ was presented in [10]. Later, a nonlinear filter based on von Mises and wrapped normal distributions was presented in [11]. 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. Later, it will be compared to the proposed filter.

In 1974, Bingham proposed his distribution in [6]. Further work on the Bingham distribution has been done by Kent [12], [13] as well as Jupp and Mardia [14]. So far, there have only been a few applications of the Bingham distribution, for example in geology [15], [16]. In 2011, Glover used the Bingham distribution for a Monte Carlo based pose estimation [17]. Glover also released a library called libbingham [18] that includes implementations of some of the methods discussed in Sec. IV. 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.

III. KEY IDEA OF THE BINGHAM FILTER

The goal of this paper is the derivation of a recursive filter based on the Bingham distribution. Rather than relying on the traditional Gaussian distribution, we chose to represent all occurring probability densities as Bingham. The Bingham distribution is defined on the hypersphere and is antipodally symmetric, which makes it interesting for applications in angular estimation with inherent $180°$ symmetry and for

problems where $180°$ symmetry occurs as a result of parameterization, e.g., in the case of quaternions. Although we restrict ourselves to the two-dimensional case in this paper, we would like to emphasize that most of presented methods are easily 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 traditional estimation problem in $\mathbb{R}^d$ with additive noise, this involves a convolution with the noise density. We provide a suitable analogon on the hypersphere, which we call *composition*. Since Bingham distributions are not closed under compositions, 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.

IV. BINGHAM DISTRIBUTION

The Bingham distribution appears naturally when a d-dimensional normal random vector $\underline{\mathbf{x}}$ with $\mathrm{E}(\underline{\mathbf{x}}) = \underline{0}$ is conditioned on $||\underline{\mathbf{x}}|| = 1$ [19]. In the following, we will introduce the Bingham distribution and 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.

A. Probability Density Function

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)*

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

of a Bingham distribution [6] is given by

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

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

As Bingham showed, 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 $z_1 \leq \cdots \leq z_{d-1}$. This representation allows us to obtain the mode of the distribution very easily by taking the last column of $\mathbf{M}$.

The pdf is antipodally symmetric, i. e., $f(\underline{x}) = f(-\underline{x})$ holds for all $x \in S_{d-1}$. Consequently, the Bingham distribution is invariant to rotations by $180°$. Examples of the pdf for two dimensions ($d = 2$) are shown in Fig. 1 and Fig. 2. The
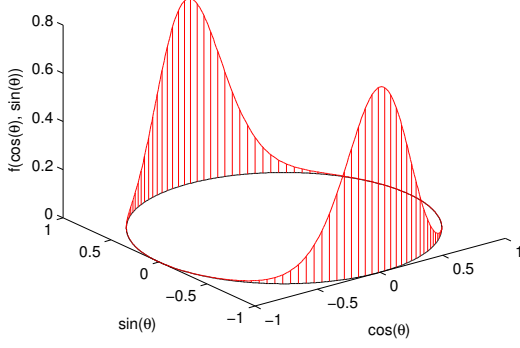
Fig. 1: Bingham pdf with $\mathbf{M} = \mathbf{I}_{2 \times 2}$ and $\mathbf{Z} = \mathrm{diag}(-8, 0)$ as a 3D plot. This corresponds to a standard deviation of $16°$.
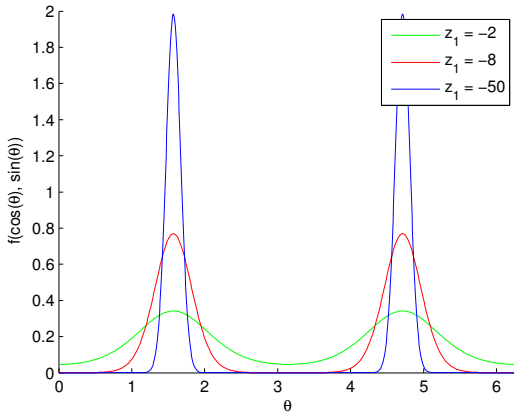

Fig. 2: Bingham pdf with $\mathbf{M} = \mathbf{I}_{2 \times 2}$ for different values of $\mathbf{Z} = \mathrm{diag}(z_1, 0)$ and $x = (\cos(\theta), \sin(\theta))^T$. These values for z_1 correspond to standard deviations of approximately $36°$, $16°$, and $6°$, respectively.

Bingham distribution is very similar to a Gaussian if and only if the uncertainty is small. This can be seen in Fig. 3, which shows the Kullback-Leibler divergence between a Bingham pdf and a corresponding Gaussian pdf.

B. Normalization Constant

The normalization constant can be calculated with the help of the hypergeometric function of a matrix argument [20], [21], [22]. It is given by

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

where $|S_{d-1}|$ is the surface area of the d-sphere and ${}_1F_1(\cdot, \cdot, \cdot)$ is the hypergeometric function of matrix argument. In the two-dimensional case ($d = 2$), this reduces to

$$F = 2\pi \cdot {}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & 0 \end{pmatrix}\right) = 2\pi \cdot {}_1F_1\left(\frac{1}{2}, 1, z_1\right) ,$$

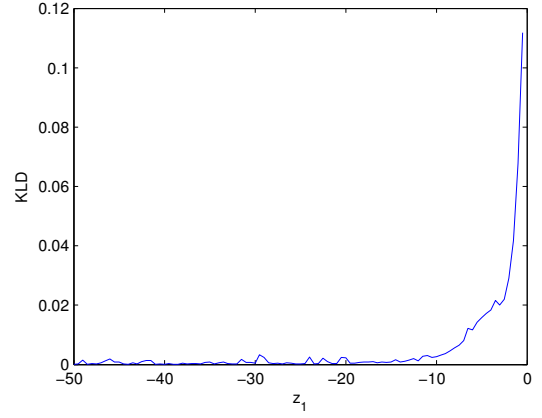so it is sufficient to compute the hypergeometric function of a scalar argument, which is described in [23].

Fig. 3: Kullback-Leibler divergence on the interval $[0, \pi]$ between a Bingham pdf with $\mathbf{M} = \mathbf{I}_{2 \times 2}$, $\mathbf{Z} = \mathrm{diag}(z_1, 0)$ 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. Note that this curve is smooth in reality and that the jitter is caused by the fact that a Monte Carlo method was used to create this plot.

C. 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

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

and

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

Then

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

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

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. $\qquad\square$

D. Estimation of Parameters

Estimating parameters for the Bingham distribution is not only motivated by the need to estimate noise parameters from samples. It also plays a crucial role in the prediction process when computing the composition of two Bingham random

vectors. 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 its covariance 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 originate from an arbitrary distribution on the hypersphere).

The maximum likelihood estimate for the parameters $(\mathbf{M}, \mathbf{Z})$ of a Bingham distribution can be obtained as described in [6]. $\mathbf{M}$ can be obtained as the matrix of eigenvectors of the covariance $\mathbf{S}$ with eigenvalues $\omega_1 \le \omega_2$. In other words, $\mathbf{M}$ can be found as the eigendecomposition of $\mathbf{S} = \mathbf{M} \cdot \mathrm{diag}(\omega_1, \omega_2) \cdot \mathbf{M}^T$. To calculate $\mathbf{Z}$, the equations

$$\frac{\frac{\partial}{\partial z_i} {}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\right)}{{}_1F_1(\frac{1}{2}, 1, z_1)} = \omega_i, \quad i = 1, 2$$

have to be solved under the constraint $z_2 = 0$, which is justified by the argumentation above and used to simplify the computation. This operation is performed numerically.

Conversely, for a given a Bingham distribution $(\mathbf{M}, \mathbf{Z})$, the covariance matrix can be calculated according to

$$\begin{aligned} \mathbf{S} &= \mathbf{M} \cdot \mathrm{diag}(\omega_1, \omega_2) \cdot \mathbf{M}^T \\ &= \mathbf{M} \cdot \mathrm{diag}\left(\frac{1}{F}\frac{\partial F}{\partial z_1}, \frac{1}{F}\frac{\partial F}{\partial z_2}\right) \cdot \mathbf{M}^T \\ &= \sum_{i=1}^{2} \frac{1}{F}\frac{\partial F}{\partial z_i} \mathbf{M}(:,i)\mathbf{M}(:,i)^T , \end{aligned}$$

where $\mathbf{M}(:,i)$ refers to the i-th column of $\mathbf{M}$ [18]. Thus, any Bingham distribution is uniquely defined by its covariance matrix and vice versa. The following Lemma simplifies the computation of partial derivatives of a confluent hypergeometric function of a 2×2 matrix argument, which is used in computation of the covariance matrix as derived above.

Lemma 2. *For $d = 2$, the partial derivatives*

$$\frac{\partial}{\partial z_i} {}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\right) , \quad i = 1, 2$$

can be reduced to hypergeometric functions of scalar argument.

Proof. See Appendix A. □

E. Composition

Now, we want to derive the composition of Bingham distributed random variables, which is the directional analogue to adding random variables. This operation can, for example, be used to disturb an uncertain Bingham-distributed system state with Bingham-distributed noise, similar to using a convolution to disturb a probability distribution on $\mathbb{R}$ with additive noise. First, we define a composition of individual points on the hypersphere S_{d-1}, which we then use to derive the composition of Bingham distributions.

The composition of two Bingham distributions depends on the interpretation of the unit vectors, for example as complex numbers or quaternions. We assume that a composition function

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

is given. The function $\oplus$ has to be compatible with $180°$ degree symmetry, i.e.,

$$\pm(x \oplus y) = \pm((-x) \oplus y) = \pm(x \oplus (-y)) = \pm((-x) \oplus (-y))$$

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.

In the complex case, we interpret $S_1 \subset \mathbb{R}^2$ as unit vectors in $\mathbb{C}$, 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. *The composition function $\oplus$ is defined to be complex multiplication, i.e.,*

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \oplus \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 y_1 - x_2 y_2 \\ x_1 y_2 + x_2 y_1 \end{pmatrix}$$

analogous to

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

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) is the complex conjugate $\pm(x_1, -x_2)$.

Unfortunately, the Bingham distribution is not closed under this kind of composition. That is, the resulting random vector is not Bingham distributed. Thus, we propose a technique to approximate a Bingham distribution to the composed random vector. 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 and mark it with an asterisk.

Lemma 3. *Let $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ be Bingham distributions with covariance matrices*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ * & a_{22} \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ * & b_{22} \end{pmatrix} ,$$

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

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

of the composition is given by

$$c_{11} = a_{11}b_{11} - 2a_{12}b_{12} + a_{22}b_{22} \ ,$$
$$c_{12} = a_{11}b_{12} - a_{12}b_{22} + a_{12}b_{11} - a_{22}b_{12} \ ,$$
$$c_{22} = a_{11}b_{22} + 2a_{12}b_{12} + a_{22}b_{11} \ .$$

Proof. See Appendix B. □

Based on $\mathbf{C}$, maximum likelihood estimation 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, because the solution of the equation involving the hypergeometric function is the only part which 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 [24], [25].

V. FILTER IMPLEMENTATION

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

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

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

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

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 $\underline{w}_k$ and $\underline{v}_k$ can include a constant offset. For example, $\underline{w}_k$ could include a known angular velocity. Alternatively, to avoid dealing with biased noise distributions, a rotation may be applied to $\underline{x}_k$ first and unbiased 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 $(\mathbf{M}_k^v, \mathbf{Z}_k^v)$.

A. Prediction

The prediction can be calculated according to

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

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

B. Update

Given a measurement $\hat{\underline{z}}$, we can calculate the updated distribution according to Bayes' rule

$$f(\mathbf{M}_k, \mathbf{Z}_k | \hat{\underline{z}}) = c \cdot f(\hat{\underline{z}} | \mathbf{M}_k, \mathbf{Z}_k) \cdot f(\mathbf{M}_k, \mathbf{Z}_k)$$

with some normalization constant c, which yields the update procedure

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

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

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

```
/* calculate covariance matrices A, B */
```
$\mathbf{A} = \sum_{i=1}^d \frac{1}{F} \frac{\partial F}{\partial z_i} \mathbf{M}_k^e(:,i) \mathbf{M}_k^e(:,i)^T;$
$\mathbf{B} = \sum_{i=1}^d \frac{1}{F} \frac{\partial F}{\partial z_i} \mathbf{M}_k^w(:,i) \mathbf{M}_k^w(:,i)^T;$
```
/* calculate C according to Lemma 3 */
```
$c_{11} = a_{11}b_{11} - 2a_{12}b_{12} + a_{22}b_{22};$
$c_{12} = a_{11}b_{12} - a_{22}b_{12} - a_{12}b_{22} + a_{12}b_{11};$
$c_{22} = a_{11}b_{22} + 2a_{12}b_{12} + a_{22}b_{11};$
$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix};$
```
/* calculate M_{k+1}^p, Z_{k+1}^p based on C */
```
$\mathbf{M}_{k+1}^p, \mathbf{Z}_{k+1}^p \leftarrow \text{MLE}(\mathbf{C});$

Fig. 4: Algorithm for prediction 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                      */
```
$\mathbf{M} \leftarrow (\bar{\hat{\underline{z}}} \oplus \mathbf{M}_k^v);$
```
/* multiply with prior distribution */
```
$(\mathbf{M}_k^e, \mathbf{Z}_k^e) \leftarrow \text{multiply}((\mathbf{M}, \mathbf{Z}_k^v), (\mathbf{M}_k^p, \mathbf{Z}_k^p));$

Fig. 5: Algorithm for update step.

with $\mathbf{M} = (\bar{\hat{\underline{z}}} \oplus \mathbf{M}_k^v)$, where $\bar{\underline{a}}$ indicates the complex conjugate of $\underline{a}$ and $\oplus$ is evaluated for each column of $\mathbf{M}_k^v$.

VI. EVALUATION

The proposed filter was evaluated in simulations. In this section, all angles are given in radians unless specified differently.

For comparison, we implemented a one-dimensional Kalman filter [1]. A traditional one-dimensional Kalman filter has two issues when confronted with our situation. First, it does not take the circular nature of the problem into account. Second, it does not handle $180°$ symmetry. We can circumvent both issues by restricting the estimate x_k according to $0 \le x_k \le \pi$ and by shifting the measurement, so that $|x_k - \hat{z}_k| \le \frac{\pi}{2}$ is satisfied.

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 $(0,1)^T$

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

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

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

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

$$\mathbf{M}_k^v = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{Z}_k^v = \begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix} \ .$$

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

$$x_0^e = \operatorname{atan2}(\operatorname{mode}(\mathbf{M}_0^e)) = \operatorname{atan2}(1,0) = \frac{\pi}{2}$$

and the noise means are

$$\mu_k^w = \mu_k^v = \operatorname{atan2}(0,1) = 0 \ .$$

The covariance matrices for the Kalman filter are obtained by sampling the Bingham noise parameters and calculating the empirical covariance from the samples. This yields

$$C_0^e = 0.5956, \quad C_k^w = 0.0027, \quad C_k^v = 0.2836 \ ,$$

which is equivalent to standard deviations of $44°$ for the first time step, $3°$ for the system noise and $30°$ for the measurement noise.

We simulate the system for a duration of $k_{\max} = 100$ time steps. An example run is depicted in Fig. 6. In addition to the mode of the estimate, we plot the 95% confidence interval, which is equivalent to the 2σ bounds in the case of the Kalman filter.

For evaluation, we consider the angular RMSE which is given by

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

with angular error

$$e_k = \min(\measuredangle(\underline{x}_k^{\text{true}}, \operatorname{mode}(\mathbf{M}_k^e)), \pi - \measuredangle(\underline{x}_k^{\text{true}}, \operatorname{mode}(\mathbf{M}_k^e)))$$

at time step k. Obviously, $0 \leq e_k \leq \frac{\pi}{2}$ holds, which is consistent with our assumption of $180°$ symmetry.

The presented results are based on 1000 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, our non-optimized implementation in MATLAB needs approximately $60\,\mathrm{ms}$ for one time step (prediction and update), which could be significantly improved by a faster evaluation of the hypergeometric function. In Fig. 8, we plot the error of our filter against the error of the Kalman filter for all runs. The proposed filter outperforms the Kalman filter in most cases, which is also true for the mean angular error in every time step as shown in Fig. 7. In particular, the significantly faster rate of convergence of the proposed filter is evident. This superiority is due to the reasons listed in the introduction. The use of the Gaussian distribution, which does not consider the problem geometry leads to suboptimal results
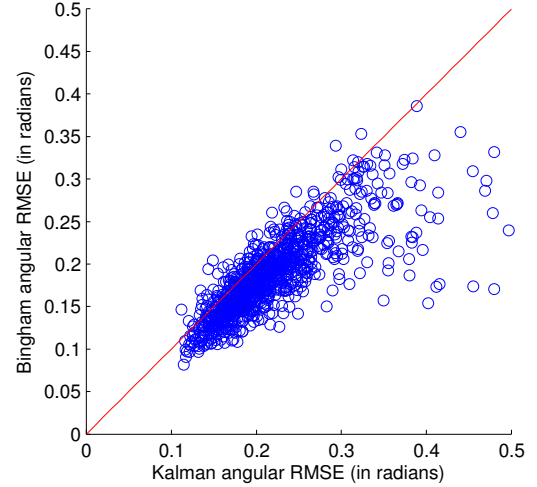

Fig. 8: Results of 1000 Monte Carlo runs. Each sample represents one run. Samples below the red line indicate that the proposed filter has performed better, samples above the red line indicate that the Kalman filter has performed better.

compared to the proposed approach based on the Bingham distribution.

In Fig. 7, we also show a comparison with a filter based on the wrapped normal distribution (denoted WN), which we previously published in [11] and modified for the $180°$ case. This modification was done by a simple rescaling technique mapping the angles between $0°$ and $180°$ to the full circle. The angular error of both filters is almost indistinguishable. However, unlike the proposed filter based on the Bingham distribution, the previously published filter cannot easily be generalized to higher dimensions.

VII. Conclusion

We have presented a recursive filter based on the Bingham distribution. It can be applied to circular estimation problems with $180°$ symmetry. Our simulations have shown the superiority of the presented approach compared to the traditional solution of modifying a Kalman filter for the circular setting.

Future work will focus on recursive 3D pose estimation using Bingham distribution. This can be achieved by applying the presented methods in the four-dimensional case for estimating quaternions. 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.
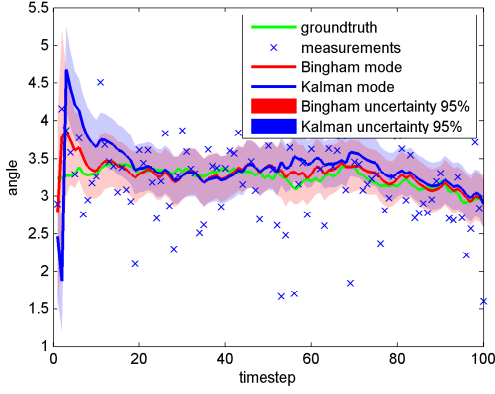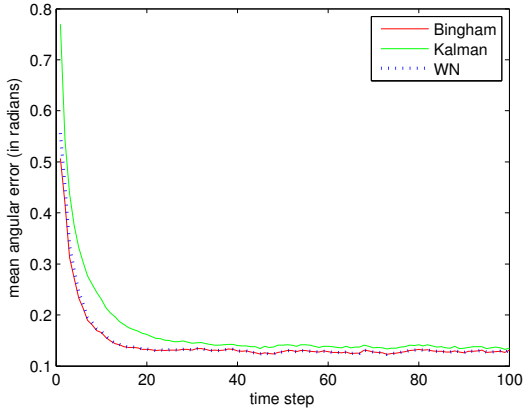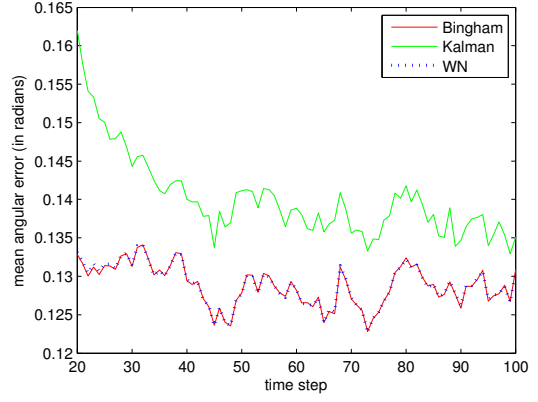
(a) Ground truth and estimate.
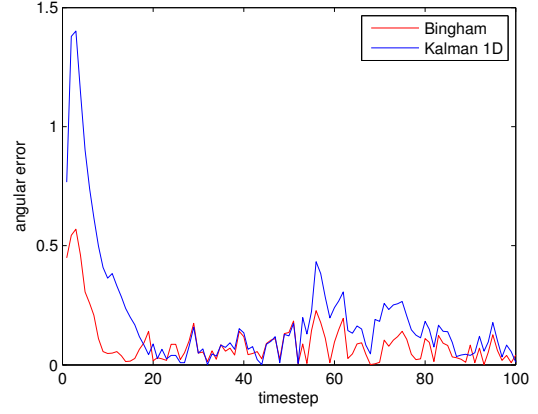
(b) Angular error.

Fig. 6: An example run of Bingham and Kalman filter.

(a) Time steps $1 \leq k \leq 100$.

(b) Time steps $20 \leq k \leq 100$.

Fig. 7: Results of 1000 Monte Carlo runs. We show the mean error at every time step across all runs for the proposed Bingham filter, a Kalman [1] filter and a filter based on the wrapped normal (WN) distribution [11]. Because the initial error is large as a result of the poor initial estimate, we show two plots of different time intervals.

APPENDIX

A. Proof of Lemma 2.

Proof. We use the identities

$${}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\right) = \exp(z_2) \cdot {}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right)$$

and

$$\frac{\partial}{\partial z} {}_1F_1(a, b, z) = \frac{a}{b} {}_1F_1(a+1, b+1, z) .$$

The partial derivative with respect to z_1 is given by

$$\begin{aligned} \frac{\partial}{\partial z_1} \frac{F}{|S_{d-1}|} &= \frac{\partial}{\partial z_1} {}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\right) \\ &= \exp(z_2) \frac{\partial}{\partial z_1} {}_1F_1\left(\frac{1}{2}, 1, z1 - z2\right) \\ &= \exp(z_2) \frac{1}{2} {}_1F_1\left(\frac{3}{2}, 2, z1 - z2\right) . \end{aligned}$$

The partial derivative with respect to z_2 is calculated in a similar fashion. We calculate

$$\begin{aligned} \frac{\partial}{\partial z_2} \frac{F}{|S_{d-1}|} &= \frac{\partial}{\partial z_2} {}_1F_1\left(\frac{1}{2}, 1, \begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\right) \\ &= \frac{\partial}{\partial z_2}\left(\exp(z_2) {}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right)\right) \\ &= \exp(z_2) {}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right) \\ &\quad + \exp(z_2) \frac{\partial}{\partial z_2} {}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right) \\ &= \exp(z_2) {}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right) \\ &\quad - \exp(z_2) \frac{1}{2} {}_1F_1\left(\frac{3}{2}, 2, z_1 - z_2\right) . \end{aligned}$$

and obtain

$$\frac{\partial}{\partial z_2} \frac{F}{|S_{d-1}|} = \exp(z_2)\left({}_1F_1\left(\frac{1}{2}, 1, z_1 - z_2\right)\right.$$
$$\left. - \frac{1}{2}{}_1F_1\left(\frac{3}{2}, 2, z_1 - z_2\right)\right)$$

$\square$

B. Proof of Lemma 3.

Proof. The covariance of the composition

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

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

$$c_{11} = \mathrm{Var}(x_1 y_1 - x_2 y_2)$$
$$= \mathrm{E}((x_1 y_1 - x_2 y_2)^2) - (\mathrm{E}(x_1 y_1 - x_2 y_2))^2$$
$$= E(x_1^2 y_1^2 - 2 x_1 y_1 x_2 y_2 + x_2^2 y_2^2)$$
$$\quad - (\mathrm{E}(x_1 y_1) - \mathrm{E}(x_2 y_2))^2 \tag{1}$$
$$= \mathrm{E}(x_1^2)\,\mathrm{E}(y_1^2) - 2\,\mathrm{E}(x_1 x_2)\,\mathrm{E}(y_1 y_2) + \mathrm{E}(x_2^2)\,\mathrm{E}(y_2^2) \tag{2}$$
$$\quad - (\underbrace{\mathrm{E}(x_1)}_{0}\underbrace{\mathrm{E}(y_1)}_{0} - \underbrace{\mathrm{E}(x_2)}_{0}\underbrace{\mathrm{E}(y_2)}_{0})^2 \tag{3}$$
$$= a_{11} b_{11} - 2 a_{12} b_{12} + a_{22} b_{22} \ .$$

We use independence of $\underline{x}$ and $\underline{y}$ in (1), linearity of the expectation value in (2) and symmetry of the Bingham in (3). Analogously we calculate

$$c_{22} = \mathrm{Var}(x_1 y_2 - x_2 y_1)$$
$$= \mathrm{E}((x_1 y_2 - x_2 y_1)^2) - (\mathrm{E}(x_1 y_2 - x_2 y_1))^2$$
$$= E(x_1^2 y_2^2 - 2 x_1 y_1 x_2 y_2 + x_2^2 y_1^2)$$
$$\quad - (\mathrm{E}(x_1 y_2) - \mathrm{E}(x_2 y_1))^2$$
$$= \mathrm{E}(x_1^2)\,\mathrm{E}(y_2^2) - 2\,\mathrm{E}(x_1 x_2)\,\mathrm{E}(y_1 y_2) + \mathrm{E}(x_2^2)\,\mathrm{E}(y_1^2)$$
$$\quad - (\underbrace{\mathrm{E}(x_1)}_{0}\underbrace{\mathrm{E}(y_2)}_{0} - \underbrace{\mathrm{E}(x_2)}_{0}\underbrace{\mathrm{E}(y_1)}_{0})^2$$
$$= a_{11} b_{22} - 2 a_{12} b_{12} + a_{22} b_{11} \ .$$

The off-diagonal entry can be calculated similarly

$$c_{12} = \mathrm{Cov}(x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1)$$
$$= \mathrm{E}((x_1 y_1 - x_2 y_2) \cdot (x_1 y_2 + x_2 y_1))$$
$$\quad - \mathrm{E}(x_1 y_1 - x_2 y_2) \cdot \mathrm{E}(x_1 y_2 + x_2 y_1)$$
$$= \mathrm{E}(x_1^2 y_1 y_2 - x_1 x_2 y_2^2 + x_1 x_2 y_1^2 - x_2^2 y_1 y_2)$$
$$\quad - (\mathrm{E}(x_1)\,\mathrm{E}(y_1) - \mathrm{E}(x_2)\,\mathrm{E}(y_2))$$
$$\quad \cdot (\mathrm{E}(x_1)\,\mathrm{E}(y_2) + \mathrm{E}(x_2)\,\mathrm{E}(y_1))$$
$$= a_{11} b_{12} - a_{12} b_{22} + a_{12} b_{11} - a_{22} b_{12} \ .$$

Because $\mathbf{C}$ is a symmetrical matrix, this concludes the proof of Lemma 3. $\square$

REFERENCES

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[2] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, mar 2004.

[3] S. Julier and J. LaViola, "On kalman filtering with nonlinear equality constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2774–2784, 2007.

[4] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds," *Information Fusion*, vol. 14, no. 1, pp. 57 – 77, 2013.

[5] A. N. Shiryaev, *Probability*, 2nd ed. Springer, 1995.

[6] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, Nov. 1974.

[7] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Wiley, 1999.

[8] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*. World Scientific Pub Co Inc, 2001.

[9] S. Krinidis and V. Chatzis, "Frequency-based object orientation and scaling determination," in *2006 IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings*, May 2006, p. 4 pp.

[10] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A recursive fusion filter for angular data," in *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2009, pp. 882 –887.

[11] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive nonlinear filtering for angular data based on circular distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013) (to appear)*, Washington D.C., USA, Jun. 2013.

[12] J. T. Kent, "Asymptotic expansions for the bingham distribution," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 36 (2), pp. 139–144, 1987.

[13] ——, "The complex bingham distribution and shape analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285–299, 1994.

[14] P. E. Jupp and K. V. Mardia, "Maximum likelihood estimators for the matrix von mises-fisher and bingham distributions," *Annals of Statistics*, vol. 7 (3), pp. 599–606, 1979.

[15] K. Mardia, "Directional statistics in geosciences," *Communications in Statistics - Theory and Methods*, vol. 10, no. 15, pp. 1523–1543, 1981.

[16] K. Kunze and H. Schaeben, "The bingham distribution of quaternions and its spherical radon transform in texture analysis," *Mathematical Geology*, vol. 36, pp. 917–943, 2004.

[17] J. Glover, R. Rusu, and G. Bradski, "Monte carlo pose estimation with quaternion kernels and the bingham distribution," in *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, Jun. 2011.

[18] J. Glover, "libbingham bingham statistics library," 2013. [Online]. Available: http://code.google.com/p/bingham/

[19] A. Kume and A. T. A. Wood, "Saddlepoint approximations for the bingham and fisher-bingham normalising constants," *Biometrika*, vol. 92, no. 2, pp. 465–476, 2005.

[20] C. S. Herz, "Bessel functions of matrix argument," *Annals of Mathematics*, vol. 61, no. 3, pp. 474–523, 1955.

[21] P. Koev and A. Edelman, "The efficient evaluation of the hypergeometric function of a matrix argument," *Math. Comp.*, vol. 75, pp. 833–846, 2006.

[22] R. Muirhead, *Aspects of multivariate statistical theory*, ser. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley, 1982.

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth dover printing, tenth gpo printing ed. New York: Dover, 1964.

[24] Y. Luke, *Algorithms for the computation of mathematical functions*, ser. Computer science and applied mathematics. Academic Press, 1977.

[25] K. E. Muller, "Computing the confluent hypergeometric function, m(a,b,x)," *Numerische Mathematik*, vol. 90, no. 1, pp. 179–196, 2001.