# Unscented Orientation Estimation Based on the Bingham Distribution

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

*Autonomous Systems Laboratory (ASL)
Institute of Robotics and Intelligent Systems
Swiss Federal Institute of Technology (ETH) Zurich, Switzerland
igilitschenski@ethz.ch

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

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

***Abstract*—In this work, we develop a recursive filter to estimate orientation in 3D, represented by quaternions, using directional distributions. Many closed-form orientation estimation algorithms are based on traditional nonlinear filtering techniques, such as the extended Kalman filter (EKF) or the unscented Kalman filter (UKF). These approaches assume the uncertainties in the system state and measurements to be Gaussian-distributed. However, Gaussians cannot account for the periodic nature of the manifold of orientations and thus small angular errors have to be assumed and *ad hoc* fixes must be used. In this work, we develop computationally efficient recursive estimators that use the Bingham distribution. This distribution is defined on the hypersphere and is inherently more suitable for periodic problems. As a result, these algorithms are able to consistently estimate orientation even in the presence of large angular errors. Furthermore, handling of nontrivial system functions is performed using an entirely deterministic method which avoids any random sampling. A scheme reminiscent of the UKF is proposed for the nonlinear manifold of orientations. It is the first deterministic sampling scheme that truly reflects the nonlinear manifold of orientations.**

***Keywords*—*Directional Statistics, Nonlinear Filtering, Deterministic Sampling, Quaternions***

## I. Introduction

Accurate estimation of orientation is fundamental for many mobile systems, where precise orientation is required for tracking and control. High performance can be achieved in environments with an expensive sensor infrastructure. However, in many applications such as emergency response management or UAVs, systems must be low cost, lightweight, and thus tend to use sensors with higher noise and greater levels of dropout. Therefore, there is a need to perform orientation estimation in the presence of significant uncertainties.

In this work, we consider a stochastic filter that does not use the Gaussian distribution for describing the uncertain system state and uncertain measurements. Instead, it uses directional statistics for a better description of uncertainty in data defined on periodic domains such as the circle or the (hyper-)sphere. This promises better results because certain assumptions motivating Gaussian distribution based filtering approaches do not hold in a periodic setting. First, the Gaussian distribution is defined in the Euclidean space and thus does not consider periodicity. Second, Gaussians are preserved under linear operations. Unfortunately, there is no equivalent of a linear function on the manifold of orientations. Third, the central limit theorem usually motivates the use of Gaussians but it does not apply to certain manifolds such as the hypersphere $S^d$. Consequently, Gaussians are not a good approximation for uncertain quantities defined on periodic domains in the case of strong noise [1]. On a circle, for instance, the true limit distribution might be the wrapped normal distribution which arises by wrapping the density of a normal distribution around an interval of length $2\pi$. Thus, it differs in its shape from the classical Gaussian distribution.

### A. *Orientation Estimation using Directional Statistics*

There has been a lot of work on orientation estimation but almost all methods are based on the assumption that the uncertainty can be adequately represented by a Gaussian distribution. Thus, they are usually using the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) [2]. Although three parameters are sufficient to represent orientation, they often suffer from an ambiguity problem known as "gimbal lock". Therefore, many applications use quaternions [3], [4]. These represent uncertainty as a point on the surface of a four dimensional hypersphere. Moreover, current approaches use nonlinear projection [5] or other *ad hoc* approaches to push the state estimate back onto the surface of the hypersphere of unit quaternions. For properly representing uncertain orientations, we need to use an antipodally symmetric probability distribution defined on this hypersphere reflecting the fact that the unit quaternions $q$ and $-q$ represent the same rotation.

Thus, rather than using Gaussians with projection operations, we use a distribution that explicitly considers the structure of the nonlinear manifold. Specifically, we develop recursive estimators that use the Bingham Distribution (BD). The BD is an antipodally symmetric distribution defined on an $d$-dimensional hypersphere [6] and thus the proposed approach takes the periodicity of orientation into account. Therefore, BDs can be used to describe uncertainty on the group of rotations $SO(3)$ parametrized by quaternions.

Other directional distributions defined on the Hypersphere include the von Mises-Fisher distribution [7], which is a multi-dimensional generalization of the von Mises distribution. This unimodal distribution is particularly suitable for estimating directions in $\mathbb{R}^3$. However, the lack of antipodal symmetry makes quaternion-based orientation estimation infeasible. The (Dimroth-Scheidegger-)Watson distribution, independently introduced in [8], [9], introduces antipodal symmetry. However, this distribution is a special case of the more general BD considered in this work. The BD itself is generalized by the Fisher-Bingham distribution, which is not necessarily antipodally symmetric and thus not of interest in the considered scenario. Furthermore, the angular central Gaussian distribution [10] can also be considered for representing uncertain orientations.

The use of directional statistics in recursive filtering has recently been discussed for the estimation of angles [11]–[14]. A Monte Carlo pose estimation approach was presented in [15], where a BD was used for representing uncertainties. However, this possibly requires costly random sampling. In [13], we proposed a recursive Bingham filter on the circle. Our original filter considers system functions performing a predefined change of orientation, which is performed by choosing a suitable mode for the Bingham distributed noise term. This can be thought of as an equivalent to an identity system function in $\mathbb{R}^n$ with additive (possibly non-zero mean) stochastic noise. A quaternion equivalent to our Bingham-based estimation approach was independently developed in [16], where precomputed lookup tables are used for handling complicated computations involving the Bingham normalization constant. That approach also considers system models performing a predefined shift of orientation and additionally considers a model with variable velocities using a second-order Bingham filter. In most practical applications, system functions are usually more complicated than a predefined shift of orientation. Thus, it is important to efficiently consider these complex system models.

### B. Main Contribution

In order to take non-trivial system functions into account, we extend the existing work on Bingham filtering by proposing an equivalent of the unscented Kalman filter (UKF) for orientations based on the Bingham distribution. Thus, the system state is an orientation represented directly in terms of quaternions. The new filter consists of two consecutive steps. The *prediction step*, is performed by deterministically sampling the current system state and propagating the samples through the system function. Compared to random sampling, our propagation approach ensures reproducible results with a relatively small number of samples. After the propagation, system noise is imposed and a corresponding BD is found by moment matching. In the *measurement update step*, we assume noisy measurements of the system state. It makes use of the fact that the product of two BD densities is itself a rescaled BD density. Thus, the measurement update step can be performed in closed form.

The remainder of the paper is structured as follows. In Sec. II, we review the Bingham distribution and then propose a method for deterministic sampling in Sec. III. The presented approach can easily be generalized to higher dimensions. Our filter is presented in Sec. IV, where we make use of moment matching for quaternion multiplication in the prediction step and present a closed-form measurement update step. The proposed filter is compared to the UKF and the particle filter in Sec. V. Our work is concluded in Sec. VI.

## II. Bingham Distribution

The Bingham distribution (BD) naturally arises when conditioning a $d$-dimensional Gaussian distribution to the $d$-dimensional unit sphere (denoted by $S^{d-1}$). This distribution is of particular interest due to the following two facts. First, it offers a very natural way to characterize uncertainty over unit quaternions and can easily be used to describe uncertain orientations. Second, it is closed under Bayesian inference (i.e., the product of two BD densities is itself a rescaled BD density), making efficient algorithms for recursive filtering possible.

**Definition 1.** *The Bingham distribution is defined by its probability density function*

$$f(\underline{x}; \mathbf{M}, \mathbf{Z}) = \frac{1}{N(\mathbf{Z})} \exp(\underline{x}^\top \mathbf{M} \mathbf{Z} \mathbf{M}^\top \underline{x}) \ ,$$

*where $\mathbf{Z}$ is a diagonal matrix with increasing entries $z_1 \leq z_2 \leq .. \leq z_d$, $\mathbf{M}$ is an orthonormal matrix, and $N(\mathbf{Z})$ is a normalization constant. The notation $\underline{y} \sim \mathrm{Bingham}(\mathbf{M}, \mathbf{Z})$ is used to indicate that $\underline{y}$ is a Bingham distributed random vector with parameters $\mathbf{M}$ and $\mathbf{Z}$.*

Antipodal symmetry, i.e., $f(\underline{x}) = f(-\underline{x})$, follows immediately. Examples of BD densities are visualized in Fig. 1. Furthermore, the normalization constant is

$$N(\mathbf{Z}) = \int_{S^{d-1}} \exp\left(\underline{x}^\top \mathbf{Z} \underline{x}\right) \, \mathrm{d}\underline{x} \ .$$

The matrix $\mathbf{M}$ is not a part of the normalization constant, because $N(\mathbf{Z}) = N(\mathbf{M}\mathbf{Z}\mathbf{M}^\top)$ can be shown by choosing $\underline{y} = \mathbf{M}\underline{x}$ and applying the transformation theorem. It is also possible to represent the normalization constant in terms of a hypergeometric function of matrix argument ${}_1F_1(\cdot,\cdot,\cdot)$ [17], that is

$$N(\mathbf{Z}) = |S^{d-1}| \cdot \ {}_1F_1\left(\frac{1}{2}, \frac{d}{2}, \mathbf{Z}\right) \ ,$$

where $|S^{d-1}|$ is the surface area of the unit ball in $\mathbb{R}^d$. Computing the normalization constant is difficult and usually gets harder as the density becomes more peaked. Approaches to solve this problem include series expansions [18], saddle point approximations [19], holonomic gradient descent [20], and precomputed lookup tables [21]. The latter approach is used in this work, because it offers a fast evaluation and thus, makes an efficient implementation of the filter possible. This approach is also chosen for computing derivatives of the normalization constant, which will be needed in parameter estimation.

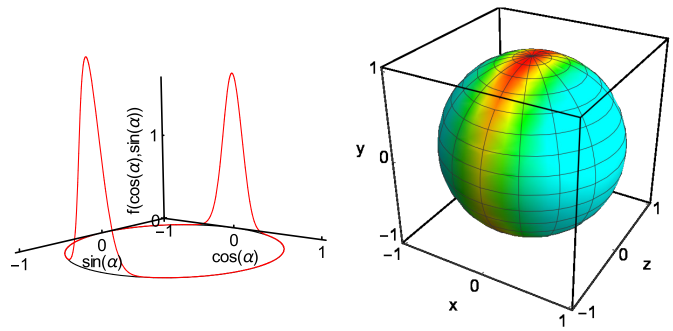



Fig. 1: Probability density functions for BDs on the circle and on the sphere.

The product of two BDs is again a (rescaled) BD. The proof is similar to the Gaussian case and presented in [13]. Furthermore, for a $\mathrm{Bingham}(\mathbf{M}, \mathbf{Z})$ distribution, the identity matrix $\mathbf{I}$, and all $c \in \mathbb{R}$,

$$\mathrm{Bingham}(\mathbf{M}, \mathbf{Z}) = \mathrm{Bingham}(\mathbf{M}, \mathbf{Z} + c\mathbf{I}) \ , \qquad (1)$$

which is a consequence of the facts that $\mathbf{M}$ is orthogonal, every possible value $\underline{x}$ has unit length, and $N(\mathbf{Z} + c\mathbf{I}) = N(\mathbf{Z}) \cdot \exp(c)$. Even though a Bingham distributed random vector $\underline{y}$ only takes values on the unit sphere, it is still possible to compute a classical (i.e., non-spherical) covariance matrix in $\mathbb{R}^d$, which is given by

$$\begin{aligned}\mathrm{Cov}(\underline{y}) &= \mathrm{E}(\underline{y}^2) - \underbrace{(\mathrm{E}(\underline{y}))^2}_{0} \\ &= \mathbf{M} \cdot \mathrm{diag}\left(\frac{\frac{\partial}{\partial z_1} N(\mathbf{Z})}{N(\mathbf{Z})}, \ldots, \frac{\frac{\partial}{\partial z_d} N(\mathbf{Z})}{N(\mathbf{Z})}\right) \cdot \mathbf{M}^\top\end{aligned}$$

according to [6]. An equivalent characterization of a Bingham distributed random vector is describing it as a zero-mean Gaussian random vector conditioned on unit length. Thus, $\mathrm{Cov}(\underline{y}) = \mathrm{Cov}(\underline{x} | \, ||\underline{x}|| = 1)$ with $\underline{x} \sim \mathcal{N}_d(\underline{0}, -0.5(\mathbf{M}(\mathbf{Z} + c\mathbf{I})\mathbf{M}^\top)^{-1})$, where $c \in \mathbb{R}$ can be chosen arbitrarily as long as $\mathbf{M}(\mathbf{Z} + c\mathbf{I})\mathbf{M}^\top$ is negative definite.

For estimating parameters of a Bingham distribution based on moment matching, we consider a given covariance matrix $\mathbf{C}$ from a spherical distribution with mean $\underline{0}$. Let $\mathbf{M} \cdot \mathrm{diag}(\omega_1, \ldots \omega_d) \cdot \mathbf{M}^\top$ be the eigendecomposition of $\mathbf{C}$. Then, the columns of $\mathbf{M}$ consist of orthogonal eigenvectors. Solving

$$\frac{\frac{\partial}{\partial z_i} N(\mathbf{Z})}{N(\mathbf{Z})} = \omega_i \ , \qquad i = 1, \ldots, d \qquad (2)$$

yields $\mathbf{Z} = \mathrm{diag}(z_1, \ldots, z_d)$. From (1), one of the parameters $z_d$ can be chosen arbitrarily (for computational reasons, the matrix $\mathbf{Z}$ is usually chosen so that $z_d = 0$). The resulting $\mathrm{Bingham}(\mathbf{M}, \mathbf{Z})$ distribution has a covariance matrix $\mathbf{C}$. In order to obtain a unique estimate, the original covariance matrix must not be degenerate, that is, none of its eigenvalues must be 0 (otherwise one or more entries of $\mathbf{Z}$ diverge to $-\infty$). This parameter estimation procedure is discussed in more depth in [1], [6].

Typically, orientations can be represented by unit quaternions [22]. A clockwise rotation of $\theta$ degrees around the unit length axis $\underline{a} = (a_x, a_y, a_z)^\top$ is represented by the quaternion

$$q = \cos\left(\frac{\theta}{2}\right) + (a_x i + a_y j + a_z k) \sin\left(\frac{\theta}{2}\right) \ .$$

This representation can also be used to define the power of a unit quaternion raised to $c \in \mathbb{R}$ by

$$q^c = \cos\left(c \cdot \frac{\theta}{2}\right) + (a_x i + a_y j + a_z k) \sin\left(c \cdot \frac{\theta}{2}\right) \ .$$

The orientations represented by the quaternions $q$ and $-q$ are identical. Thus, a BD on $S^3$ is suitable for representing uncertainty of the quaternion $q = q_1 + q_2 i + q_3 j + q_4 k$, which will be represented by the Bingham distributed random vector $\underline{y} = (q_2, q_3, q_4, q_1)^\top$. Putting the real part of the quaternion $q_1$ at the last position of our vector is motivated by the fact that the BD has its maxima at $\pm\mathbf{M}^\top[0,0,0,1]^\top$. The advantage of this notation is the fact that choosing $\mathbf{M} = \mathbf{I}$ corresponds to an orientation equivalent of zero-mean.

Another advantage of the quaternion representation is the fact that a composition of two rotations can be represented by a quaternion multiplication (Hamilton product). We denote this operation by $\oplus$ which, for two Bingham random vectors $\underline{x}$ and $\underline{y}$, is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \oplus \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} := \begin{pmatrix} x_4 y_1 + x_1 y_4 + x_2 y_3 - x_3 y_2 \\ x_4 y_2 - x_1 y_3 + x_2 y_4 + x_3 y_1 \\ x_4 y_3 + x_1 y_2 - x_2 y_1 + x_3 y_4 \\ x_4 y_4 - x_1 y_1 - x_2 y_2 - x_3 y_3 \end{pmatrix} .$$

Unfortunately, this product of two Bingham distributed random vectors is not itself a Bingham distributed random vector. A proof that the Bingham distribution is not closed under this composition is given in [16]. Thus, moment matching is used to approximate $\underline{x} \oplus \underline{y}$ by a Bingham distribution, i.e., we obtain a new Bingham distributed random vector $\hat{z}$ with $\mathrm{Cov}(\hat{z}) = \mathrm{Cov}(\underline{x} \oplus \underline{y})$. The covariance matrix of the quaternion product can be computed directly as in [13], [16] and the parameters for the BD of $\hat{z}$ are found by computing an eigendecomposition of this covariance matrix and then solving (2). Here, precomputation can also be used for speed-up.

## III. Deterministic Sampling of Uncertain Orientation

Our use of the Bingham distribution is motivated by its ability to characterize uncertainty over orientations parametrized by unit quaternions. Only simple transformations of such uncertain quaternions preserve the Bingham distribution, for example changing of the current orientation in a predefined direction without introducing any further noise.

Computing the transformation $g(\underline{y})$ of a $\mathrm{Bingham}(\mathbf{M}, \mathbf{Z})$ distributed random variable $\underline{y}$ is not possible in closed form for arbitrary functions $g(\cdot)$. Thus, we propose a technique to approximate a Bingham distribution on $S^{d-1}$ by $4d - 2$ deterministically placed samples adapting the basic idea of the Unscented Transform [2] to the manifold of orientations. With regard to quaternions, we consider the case $d = 4$. Thus, each sample can be considered as a quaternion describing an orientation. One sample is placed at the pole $(0,0,0,1)^\top$, which can be thought of as a mode on the sphere. Furthermore, six samples

$$\begin{pmatrix} \pm\sin(\alpha_1) \\ 0 \\ 0 \\ \cos(\alpha_1) \end{pmatrix} , \begin{pmatrix} 0 \\ \pm\sin(\alpha_2) \\ 0 \\ \cos(\alpha_2) \end{pmatrix} , \begin{pmatrix} 0 \\ 0 \\ \pm\sin(\alpha_3) \\ \cos(\alpha_3) \end{pmatrix}$$

are placed around this pole. Negation yields samples around the pole $(0,0,0,-1)$ to account for antipodal symmetry. That is, we obtain another seven samples which are mirror images of the first set. For actual computation considered in this paper, it is sufficient to consider the samples placed around one pole. However, the full sample set is necessary in this theoretical derivation for correctly matching the covariance, because otherwise the mean of the sample set would not be zero. Each pole is assigned the probability mass $p_0/2$ and each sample corresponding to the angle $\alpha_i$ is assigned the probability mass $p_i/4$. Each sample is multiplied by $\mathbf{M}$. Thus, the sample based probability distribution generated by this method has covariance $\mathbf{M}\mathbf{C}\mathbf{M}^\top$ where $\mathbf{M}$ stems from the original Bingham distribution and $\mathbf{C}$ is given by

$$\mathbf{C} = \mathrm{diag}\Bigg( p_1 \sin(\alpha_1)^2,\ p_2 \sin(\alpha_2)^2,\ p_3 \sin(\alpha_3)^2, \\ p_0 + \sum_{i=1}^{3} p_i \cos(\alpha_i)^2 \Bigg) .$$

In the next step, we use moment matching to find $\alpha_i$ and $p_i$ so that our samples have the same uncertainty as the approximated Bingham distribution. This is achieved by solving

$$\mathbf{C} = \mathrm{diag}(\omega_1, \ldots, \omega_4) , \qquad (3)$$

with $\omega_i$ as defined in (2), which gives

$$\alpha_i = \arcsin\left(\sqrt{\frac{\omega_i}{p_i}}\right) , \qquad i = 1,2,3 .$$

Thus, we require $\omega_i \le p_i$ for $i = 1,2,3$. Choosing

$$p_0 = \lambda \omega_4 ,$$
$$p_i = \omega_i + (1-\lambda)\frac{\omega_4}{3} , \qquad i = 1,2,3$$

gives a feasible solution to (3) for every $\lambda \in [0,1)$. Finally, our deterministically sampled distribution has covariance $\mathbf{M} \cdot \mathrm{diag}(\omega_1, \ldots, \omega_4) \cdot \mathbf{M}^\top$, which corresponds to the covariance of our Bingham distribution and completes the approximation.

## IV. Quaternion Based Filtering of Orientation

The recursive filter is separated into a prediction step and a measurement update step. The prediction step makes use of the proposed deterministic sampling scheme. Thus, it can be seen as an orientation counterpart to the prediction step in the UKF. The measurement update step is implemented as in [13], [16] and thus considers noisy but direct measurements of the true system state.

### A. Prediction Step

We consider system models given by

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

where $\underline{w}_t \sim \mathrm{Bingham}(\mathbf{M}_t^w, \mathbf{Z}_t^w)$ and $g : S^3 \mapsto S^3$ satisfies $g(\underline{x}) = -g(-\underline{x})$ in order to respect antipodal symmetry. We use deterministic sampling as described in the preceding section to approximate our current system state $\underline{x}_t \sim \mathrm{Bingham}(\mathbf{M}_t^e, \mathbf{Z}_t^e)$:

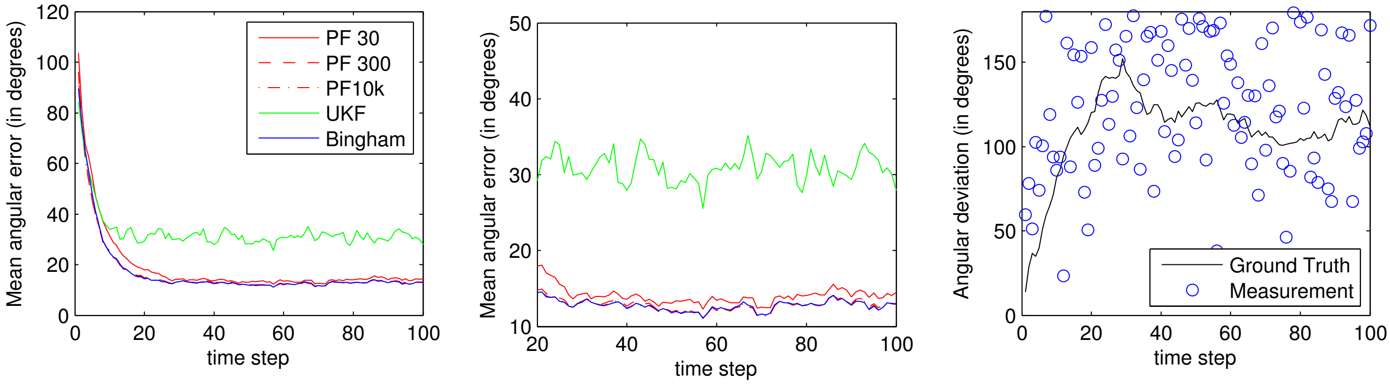
1) Approximate $\mathrm{Bingham}(\mathbf{M}_t^e, \mathbf{Z}_t^e)$ using deterministic sampling.
2) Propagate each sample through the system function $g(\cdot)$.
3) Compute approximation of $\mathrm{Cov}(g(\underline{x}_t))$ by computing the sample covariance after propagation.
4) Use this approximated covariance and $\mathrm{Cov}(\underline{w}_t)$ to compute $\mathrm{Cov}(g(\underline{x}_t) \oplus \underline{w}_t)$. This is possible in closed form according to the formula given in Appendix A.9 of [16].
5) Obtain $\mathbf{M}_{t+1}^p$ and $\mathbf{Z}_{t+1}^p$ from $\mathrm{Cov}(g(\underline{x}_t) \oplus \underline{w}_t)$ by moment matching.

The true distribution of $g(\underline{x}_t) \oplus \underline{w}_t$ is also antipodally symmetric, as $g(\cdot)$ and $\oplus$ preserve this property. This is one of the motivating reasons for the approximation made in step 4). For more general system models, such as $\underline{x}_{t+1} = g(\underline{x}_t, \underline{w}_t)$ one would replace steps 3) and 4) by deterministic sampling of $\underline{x}_t$ and $\underline{w}_t$, and approximating $\mathrm{Cov}(g(\underline{x}_t, \underline{w}_t))$ by the sample covariance. After this procedure our predicted system state is described by a $\mathrm{Bingham}(\mathbf{M}_{t+1}^p, \mathbf{Z}_{t+1}^p)$ distribution and the corresponding maximum likelihood estimate $\pm\hat{\underline{x}}_{t+1}^p$ of the orientation is described by the last column of $\pm\mathbf{M}_{t+1}^p$ (because the order of the entries in $\mathbf{Z}_{t+1}^p$ was chosen to be increasing and $z_d = 0$).

### B. Measurement Update Step

The measurement equation is assumed to be given by

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




(a) Mean angular error.

(b) Mean angular error after initialization.

(c) A typical run.

Fig. 2: Error evaluation after 100 Monte Carlo runs in the case involving high measurement noise. The typical run is given as the Rodrigues angle between the orientation represented by the true system state and the orientation represented by the quaternion $[0, 0, 0, 1]^\top$. The error is given as the angle between the true system state and the estimates of the considered filters.

where $\underline{v}_t \sim \text{Bingham}(\mathbf{M}_t^v, \mathbf{Z}_t^v)$. Choosing the identity matrix as the first parameter of the Bingham distribution, i.e., $\mathbf{M}_t^v = \mathbf{I}$, corresponds to the concept of zero-mean noise in the Euclidean space. Furthermore, this can also be used to consider certain measurements that involve only information about one or two orientation axes (rather than all three fully describing the orientation). This is done by a suitable choice of $\mathbf{M}_t^v$ and by choosing the last two (respectively three) entries of $\mathbf{Z}_t^v$ to be zero. Then, any quaternion representing the measured axes correctly can be used as measurement in this step.

The proposed measurement model results in the Bayesian estimator

$$f(\underline{x}_t|\hat{\underline{z}}_t) = c \cdot f(\hat{\underline{z}}_t|\underline{x}_t) \cdot f(\underline{x}_t) \ ,$$

where $c$ is a normalization constant. Here, $f(\underline{x}_t)$ is the pdf of a $\text{Bingham}(\mathbf{M}_t^p, \mathbf{Z}_t^p)$ distribution. For obtaining $f(\hat{\underline{z}}_t|\underline{x}_t)$ we first observe that the considered measurement model implies $f(\hat{\underline{z}}_t|\underline{x}_t) = f(\underline{x}_t^{-1} \oplus \hat{\underline{z}}_t; \mathbf{M}_t^v, \mathbf{Z}_t^v)$. Furthermore, the quaternion inverse $\underline{a}^{-1}$ (which can also be thought of as the inverse rotation) of a Bingham distributed random quaternion $\underline{a} \sim \text{Bingham}(\mathbf{M}, \mathbf{Z})$ also follows a Bingham distribution, because inversion of a unit quaternion is simply the quaternion conjugation. That is $\underline{a}^{-1} = \text{diag}(-1, -1, -1, 1) \cdot \underline{a}$ and thus $\underline{a}^{-1} \sim \text{Bingham}(\bar{\mathbf{M}}, \mathbf{Z})$ where $\bar{\mathbf{M}} = \text{diag}(-1, -1, -1, 1) \cdot \mathbf{M}$. We also make use of the fact that quaternion multiplication $\underline{a} \oplus \underline{b}$ can be expressed as a matrix vector multiplication $\mathbf{A} \cdot \underline{b}$, where $\mathbf{A}$ depends on $\underline{a}$. Bringing all together yields

$$f(\hat{\underline{z}}_t|\underline{x}_t) = f(\underline{x}_t^{-1} \oplus \hat{\underline{z}}_t; \mathbf{M}_t^v, \mathbf{Z}_t^v) = f(\underline{x}_t; \bar{\mathbf{M}}_t^v \oplus \hat{\underline{z}}_t, \mathbf{Z}_t^v) \ .$$

As already mentioned in the preceding section, we make use of the fact that the product of two Bingham pdfs is again a rescaled Bingham pdf and thus the measurement update yields

$$f(\underline{x}_t|\hat{\underline{z}}_t) \propto \exp(\underline{x}_t^\top \underbrace{((\hat{\underline{z}}_t \oplus \bar{\mathbf{M}}_t^v)\mathbf{Z}_t^v(\hat{\underline{z}}_t \oplus \bar{\mathbf{M}}_t^v)^\top + \mathbf{M}_t^p\mathbf{Z}_t^p(\mathbf{M}_t^p)^\top)}_{\mathbf{A}:=}\underline{x}_t) \ .$$

Applying an eigendecomposition to $\mathbf{A}$ yields the parameters ($\mathbf{M}_t^e$, $\mathbf{Z}_t^e$) describing our posterior Bingham distribution. Thus, the whole measurement update is performed by computing the matrix $\mathbf{A}$ and its eigendecomposition.

## V. Evaluation

We consider a stabilization scenario of a robotic ball joint in order to evaluate the proposed filtering approach. This problem arises in servoing a sensor on a moving platform to always point in a given direction. Thus, in this scenario, the true system state is stabilized towards a predefined goal state $\underline{y}$. In the considered scenario the impact

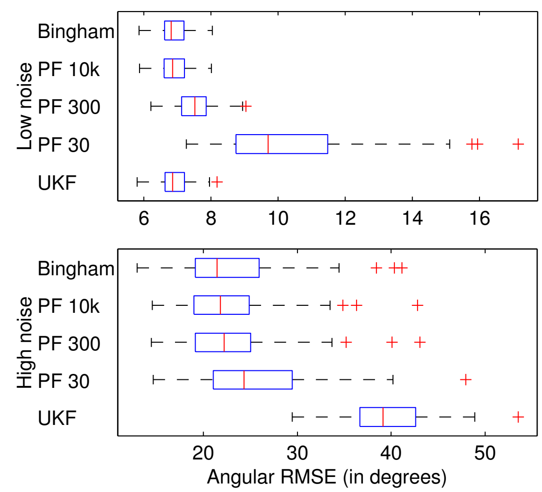



Fig. 3: The proposed filter performs better for strong noise by taking its periodic nature into account.

of the stabilization system directly depends on the deviation from the goal state. For implementing a simple function satisfying these criteria, we can now make use of quaternion multiplication and quaternion exponentiation. The proposed functionality can be represented by the following function

$$g(\underline{x}) = \underline{x} \oplus (\underline{x}^{-1} \oplus \underline{y})^a \ ,$$

for $||\underline{x} - \underline{y}|| \leq || - \underline{x} - \underline{y}||$ and $g(\underline{x}) = -g(-\underline{x})$ otherwise. Here, computation of $\underline{x}^{-1}$ and $(\underline{x}^{-1}\underline{y})^u$ is performed by interpreting the vectors as quaternions. The magnitude of stabilization feedback imposed by $g$ is controlled by $a \in (0, 1)$. The considered system function also appears in the context of spherical interpolation [23].

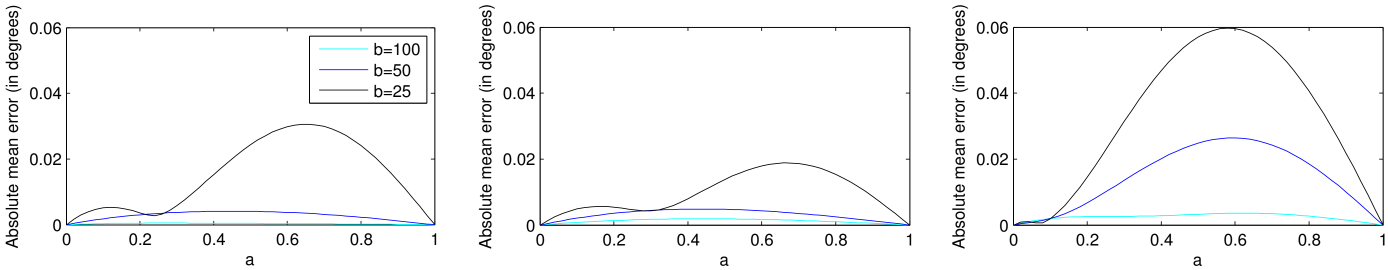
Based on this setup, we perform two different simulations. First, we apply our proposed sampling scheme in a filtering scenario. Second, we evaluate the propagation of a Bingham random variable using the proposed sampling scheme.

### A. Filtering

The considered system and measurement models are given by

$$\underline{x}_{t+1} = \underline{x}_t \oplus (\underline{x}_t^{-1} \oplus \underline{y})^a \oplus \frac{\underline{w}_t}{||\underline{w}_t||} \ ,$$

$$\underline{z}_t = \underline{x}_t \oplus \frac{\underline{v}_t}{||\underline{v}_t||} \ ,$$




(a) $\lambda = 0.1$ (b) $\lambda = 0.5$ (c) $\lambda = 0.9$

Fig. 4: Absolute mean error (in degrees) of propagation when computing $E(g(\underline{x}))$.

where $\underline{w}_t \sim \mathcal{N}(\underline{\mu}_w, \mathbf{C}_w)$, $\underline{v}_t \sim \mathcal{N}(\underline{\mu}_v, \mathbf{C}_v)$, and the parameter $a$ controls the magnitude of the system model. Here, the use of Gaussian noise is motivated by avoiding an unfair advantage of the Bingham filter. This advantage would arise in a scenario, where the Bingham filter handles Bingham distributed noise, while being compared to filters making a Gaussian noise assumption.

The proposed filter was compared with a modified UKF and a modified particle filter with importance resampling after each update step. In order to further help the UKF and the particle filter to handle antipodal symmetry, the measurement is checked and, if necessary, multiplied by $-1$ to ensure $||\underline{z}_t - \underline{x}_t^p|| < ||-\underline{z}_t - \underline{x}_t^p||$ before performing the measurement update step. For the particle filter, this check needs to be done for each particle. In order to ensure feasible results of the UKF, the estimate is projected to the unit sphere at the end of the measurement update step. Furthermore, the measurement likelihood in the particle filter is approximated by a Gaussian distribution according to

$$p(\underline{z}_t | \underline{x}_t) \approx f_v(\underline{x}_t^{-1} \oplus \underline{z}_t) \ ,$$

where $f_v(\cdot)$ denotes the Gaussian pdf of $\underline{v}_t$ and $\underline{x}_t^{-1}$ is the vector representing the inverse quaternion of $\underline{x}_t$. On the one hand, this approximation reduces the quality of the particle fiter. On the other hand, use of the true likelihood would be computationally burdensome. Three instances of the particle filter (with 30, 300, and $10^4$ particles) were used.

The proposed filter assumes $\underline{w}_t / ||\underline{w}_t||$ and $\underline{v}_t / ||\underline{v}_t||$ to be Bingham distributed (which is an approximation, because a Gaussian vector *renormalized* to unit length does not follow the Bingham distribution, i.e., the distribution of a Gaussian vector *conditioned* to unit length). Thus, the parameters $(\mathbf{Z}^w, \mathbf{M}^w)$ and $(\mathbf{Z}^v, \mathbf{M}^v)$ need to be matched to our system model. This can be done by generating random samples for each Gaussian distribution involved (we used 10 000 samples). Afterwards, these samples are normalized to length 1 and the Bingham distribution parameters are matched as described in Sec. II.

Using the classical RMSE as an error measure would be misleading in several ways. First, it does not sufficiently consider the spherical nature of the Bingham distribution. Second, it does not consider antipodal symmetry and would consider $q$ to be a wrong estimate of $-q$ even though both represent the same orientation. Both problems are tackled by introducing an angular error

$$\alpha(\underline{x}_t, \underline{x}_t^e) := 2 \cdot \min(\operatorname{acos}(\underline{x}_t^\top \cdot \underline{x}_t^e), \pi - \operatorname{acos}(\underline{x}_t^\top \cdot \underline{x}_t^e)) \ . \quad (4)$$

Here, the minimization procedure accounts for antipodal symmetry. This definition is used for computing the mean angular error and the angular RMSE. The angular error corresponds to the angle in the Rodrigues rotation formula [24] for a rotation from $\underline{x}_t$ to $\underline{x}_t^e$, which can be seen by considering two arbitrary quaternions $\mathbf{a}, \mathbf{b} \in \mathbb{H}$. We know, that the Rodrigues angle describing a transformation from orientation $\mathbf{a}$ into orientation $\mathbf{b}$ is given by $\theta := 2 \cdot \cos\left(\operatorname{Re}\left(\mathbf{b} \oplus \mathbf{a}^{-1}\right)\right)$, where $\oplus$ once again denotes the classical quaternion multiplication, i.e., the Hamilton product. Using $\underline{a}$ and $\underline{b}$ as vector representations for the quaternions $\mathbf{a}$, $\mathbf{b}$, it follows

$$\begin{aligned} \theta/2 &= \operatorname{acos}\left(\operatorname{Re}\left(\mathbf{b} \oplus \mathbf{a}^{-1}\right)\right) \\ &= \operatorname{acos}(b_4 a_4 - b_1(-a_1) - b_2(-a_2) - b_3(-a_3)) \\ &= \operatorname{acos}(\underline{b}^\top \underline{a}) \ . \end{aligned}$$

The proposed angular notion is also used to plot the true system state evolution as the angular deviation from $[0, 0, 0, 1]^\top$. This vector is chosen, because it represents the identity in the skew-field of quaternions and, thus, stands for no change in orientation.

The initial parameters used to generate the ground truth were $\underline{\mu}_0 = \underline{\mu}_w = \underline{\mu}_v = [0, 0, 0, 1]^\top$, $\mathbf{C}_0 = 0.01 \cdot \mathbf{I}$, and $\mathbf{C}_w = 0.001 \cdot \mathbf{I}$ which corresponds to an expected angular deviation of $18.2^\circ$ and $5.8^\circ$ respectively. The covariance of measurement noise was $\mathbf{C}_v = 0.003 \cdot \mathbf{I}$ for a low noise scenario and $\mathbf{C}_v = 0.3 \cdot \mathbf{I}$ for a high noise scenario (corresponding to an expected angular deviation of $10.0^\circ$ and $85.8^\circ$ respectively). Here, $\mathbf{I}$ once again denotes the identity matrix. The goal region was chosen as $\underline{y} = [0.5, 0.5, 0.5, 0.5]^\top$ and the exponent $u$ was chosen as 0.1. An initial estimate was given by $\underline{\mu}_e = [1, 0, 0, 0]^\top$ and $\mathbf{C}_e = \mathbf{I}$. The corresponding Bingham distribution parameters for the initial estimate were found in the same way as described above. Expected angular deviations have been obtained by approximating $\mathbb{E}\left(\alpha\left(\underline{\mu}, \underline{x} \cdot ||\underline{x}||^{-1}\right)\right)$ using $10^7$ random samples. For deterministic sampling in the prediction step of the proposed filter, we used $\lambda = 0.5$.

Using this setup, 100 Monte Carlo runs were performed for both measurement noise scenarios described above. In each run 100 time steps were simulated. The mean angular error in each time step and ground truth of a typical run (in the high noise scenario) are shown in Fig. 2. Evaluations of the angular RMSE are shown in Fig. 3. Depending on the noise levels, the proposed approach has either similar performance or outperforms the UKF and the particle filters.

We compared the computation time using Matlab 2014a on a system using an Intel i7-2620M processor. In our setup, the average computation time was 5.5ms for one Bingham filter step, 45ms for one particle filter step (with 300 particles), and 3.3ms for the UKF. The libbingham [21] statistics library was used for computing the Bingham normalization constant and performing moment matching based parameter estimation procedures.

### *B. Propagation*

At this point, it might be argued that the superiority of the Bingham filter based on the proposed deterministic sampling scheme might result from using a wrong Gaussian assumption or a likelihood approximation in the particle filter. Thus, it is of interest to evaluate the proposed sampling scheme in comparison to the ground truth in a pure propagation scenario. This is performed by computing the mode of $g(\underline{x})$, where $\underline{x}$ is a Bingham-distributed random vector and $g(\cdot)$ is defined as in (4). Ground truth is obtained by propagating $10^7$ random samples (which are ensured to have a zero-mean in the spherical sense) and then computing the spherical mean of the resulting propagated samples. For obtaining the result of our proposed method, we perform one propagation step using the deterministically obtained samples and the considered function $g(\cdot)$ and then reapproximate the result by a Bingham distribution.

Similarly to the simulations above, the angular mean error is used as an error measure. We choose $\underline{y} = [0.5, 0.5, 0.5, 0.5]^\top$ as the goal region in this simulation. For our Bingham distribution we have chosen $\mathbf{M} \in \mathbb{R}^{4\times 4}$ to be the identity matrix and $\mathbf{Z} = -\operatorname{diag}(b, b, b, 0)$ with $b = 25$, $b = 50$, and $b = 100$. Furthermore, different values for the deterministic sampling parameter $\lambda$ and the magnitude parameter $a$ were used. The results are shown in Fig. 4. According to our experiments, $\lambda = 0.5$ is a good choice in many cases.

## VI. Discussion and Conclusions

In orientation estimation, use of directional statistics makes a correct consideration of the underlying domain possible. Particularly, applications involving strong system and measurement noise benefit from this development. In this work, we contribute to this line of research by proposing a novel deterministic sampling scheme for the Bingham distribution. This gives rise to a recursive filter for orientation estimation based on uncertain quaternions represented by a Bingham distributed random vector.

Use of the Bingham distribution has two important theoretical advantages. First, the Bayesian measurement update can be performed in closed form in a computationally efficient way. Second, the underlying domain is considered correctly in this distributional assumption. Particularly, for scenarios with strong noise, the proposed approach outperforms classical filtering techniques based on the Gaussian distribution. These scenarios are of particular interest when using bad sensors such as magnetometers, handling weak features in robotic perception applications, or dealing with combinations of high and low uncertainties of different axes.

On the other hand, there are also two drawbacks. First, the considered composition (i.e., the quaternion multiplication) of two Bingham distributed random vectors is not itself a Bingham distributed random vector. Thus, even in simple scenarios, an approximation is used based on moment matching. However, to the best of our knowledge there is currently no filter avoiding this kind of approximation when using a correct distribution assumption defined on the underlying nonlinear domain. The second drawback is the computational burden involved in evaluating the Bingham normalization constant. For most applications, this does not present a problem because precomputed lookup tables and approximate computation techniques can be applied without harming the quality of the Bingham filter.

Finally, it is important to note that the proposed deterministic sampling scheme can be easily adapted to other antipodally symmetric distributions on the hypersphere. Furthermore, in future work we may consider more efficient computational techniques, measurement update algorithms for more complex measurement functions, and using deterministic sampling for a combination of directional and non-directional quantities including platform position or inertial sensor measurement biases.

## 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", as well as the European Commission within the BEAMING project.

## References

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

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

[3] E. Kraft, "A Quaternion-based Unscented Kalman Filter for Orientation Tracking," in *Proceedings of the Sixth International Conference of Information Fusion*, vol. 1, 2003, pp. 47–54.

[4] J. LaViola, "A Comparison of Unscented and Extended Kalman Filtering for Estimating Quaternion Motion," in *Proceedings of the 2003 American Control Conference*, vol. 3, 2003, pp. 2435–2440.

[5] S. J. Julier and J. J. LaViola, Jr., "On Kalman Filtering With Nonlinear Equality Constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2774–2784, Jun. 2007.

[6] C. Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, Nov. 1974.

[7] N. I. Fisher, *Statistical Analysis of Spherical Data*. Cambridge University Press, 1993.

[8] G. S. Watson, "Equatorial Distributions on a Sphere," *Biometrika*, vol. 52, no. 1/2, pp. 193–201, 1965.

[9] E. Dimroth, "Fortschritte der Gefügestatistik," *Neues Jahrbuch für Mineralogie: Monatshefte*, 1963.

[10] D. E. Tyler, "Statistical Analysis for the Angular Central Gaussian Distribution on the Sphere," *Biometrika*, vol. 74, no. 3, pp. 579–589, 1987.

[11] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A Recursive Fusion Filter for Angular Data," in *International Conference on Robotics and Biometrics (ROBIO)*, Dec. 2009, pp. 882–887.

[12] 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)*, Washington D. C., USA, Jun. 2013, pp. 5439–5445.

[13] G. Kurz, I. Gilitschenski, S. J. Julier, and U. D. Hanebeck, "Recursive Estimation of Orientation Based on the Bingham Distribution," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013, pp. 1487–1494.

[14] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun, "A Multi-temporal Multi-sensor Circular Fusion Filter," *Information Fusion*, vol. 18, pp. 86–100, Jul. 2014.

[15] 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.

[16] J. Glover and L. P. Kaelbling, "Tracking 3-D Rotations with the Quaternion Bingham Filter," MIT CSAIL, Tech. Rep., Mar. 2013.

[17] C. S. Herz, "Bessel Functions of Matrix Argument," *Annals of Mathematics*, vol. 61, no. 3, pp. 474–523, May 1955.

[18] P. Koev and A. Edelman, "The Efficient Evaluation of the Hypergeometric Function of a Matrix Argument," *Mathematics of Computation*, vol. 75, pp. 833–846, 2005.

[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] T. Koyama, H. Nakayama, K. Nishiyama, and N. Takayama, "Holonomic Gradient Descent for the Fisher-Bingham Distribution on the d-dimensional Sphere," *arXiv:1201.3239*, Jan. 2012.

[21] J. Glover, "libbingham Bingham Statistics Library," 2013. [Online]. Available: http://code.google.com/p/bingham/

[22] J. B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality*. Princeton University Press, 2002.

[23] K. Shoemake, "Animating Rotation with Quaternion Curves," *ACM SIGGRAPH Computer Graphics*, vol. 19, no. 3, pp. 245–254, Jul. 1985.

[24] B. K. P. Horn, "Closed-form Solution of Absolute Orientation Using Unit Quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, Apr. 1987.