Bingham–Gaussian Distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ for Unscented Attitude Estimation

Weixin Wang and Taeyoung Lee*

Abstract— **In this paper, we propose the Bingham–Gaussian (BG) distribution defined on the Cartesian product of the three-sphere and the Euclidean space of an arbitrary dimension. BG is constructed by transforming the matrix Fisher–Gaussian distribution using the homeomorphism between unit quaternions and the three dimensional rotation group. BG can be used to model the correlation between the attitude and the Euclidean space in a global fashion without singularities, and it properly deals with the cyclic nature of the space representing attitude, which is particularly advantageous when the uncertainty is large. An unscented attitude filter is designed based on BG which estimates the attitude and gyroscope bias concurrently. The proposed filter is compared against the conventional extended Kalman filter using numerical simulations.**

I. INTRODUCTION

Estimating the attitude of a rigid body is one of the most fundamental problems in aerospace engineering and robotics. In practice, the attitude is often coupled with other Euclidean quantities that need to be estimated concurrently, such as biases of the onboard gyroscope, and positions of the rigid body, etc. The conventional method for this estimation problem is the multiplicative extended Kalman filter (MEKF) [1], [2]. In MEKF, the uncertainty of attitude is modeled by assuming that the three dimensional parameterization of attitude error follows a Gaussian distribution, so that the estimation problem can be handled by an extended Kalman filter. However, MEKF has two inherent limitations: (i) The linearizations of attitude kinematics and measurement functions introduce inaccuracies of the uncertainty, and (ii) The Gaussian distribution fails to represent the uncertainty faithfully when the dispersion of attitude error is large, due to the wrapping error [3].

To address the limitations of MEKF, probability distributions defined on the manifold representing attitude have been used to replace the Gaussian distribution in filter designs. For example, the Bingham distribution on the three sphere $\mathbb{S}^3$ for unit quaternions is used in the quaternion Binghan filter [4], [5] for attitude estimation. As an alternative, the matrix Fisher distribution on the three dimensional special orthogonal group $\mathrm{SO}(3)$ has been shown equivalent to the Bingham distribution on $\mathbb{S}^3$ [6], which is also used to design an attitude filter [7]. However, these methods are unable to quantify the correlation between attitude and other Euclidean dimensions, and hence are unable to estimate them concurrently.

W. Wang and T. Lee are with the Department of Mechanical and Aerospace Engineering, The George Washington University, Washington DC, USA. {wwang442,tylee}@gwu.edu

*This research has been supported in part by NSF under the grant CNS-1837382, and AFOSR under the grant FA9550-18-1-0288.

Several approaches have been presented to address this concurrent estimation problem. In [8], [9], a new probability distribution is proposed in the space of dual quaternions which is able to model the correlation between attitude and position. However, this correlation must respect the structure of $\mathrm{SE}(3)$ when it is propagated, i.e., the Euclidean part is restricted to be the three dimensional position. In [10], a more general probability distribution is proposed on $\mathrm{SO}(3) \times \mathbb{R}^n$, and it is propagated by solving the Fokker-Planck equation of attitude kinematics and updated using the Bayes rule in the application of attitude estimation with sensor biases. Similarly, in [11] the Gauss-Bingham distribution is proposed on $\mathbb{S}^r \times \mathbb{R}^n$ which is able to model any correlation between the attitude and Euclidean dimensions. However, the maximum likelihood estimation (MLE) of the Gauss-Bingham distribution must be optimized numerically, which limits its usage in real time implementations.

In [12], [13], we proposed the matrix Fisher–Gaussian (MFG) distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$. It has several appealing features compared with previous attempts: (i) It has a geometric construction which has been successfully used to construct the distribution on the cylinder $\mathbb{S}^1 \times \mathbb{R}^1$ [14]; (ii) It uses $3n$ parameters to quantify *linear* correlations only; and (iii) It has a closed form approximate solution to the MLE problem, so the attitude filter based on MFG has the potential to be implemented in real time.

Considering that it is a common practice to use unit quaternions in attitude filters, and that the matrix Fisher distribution is equivalent to the Bingham distribution, in this paper we introduce the Bingham–Gaussian (BG) distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ which is equivalent to MFG. Then, we develop an unscented estimator for the attitude and gyroscope bias using BG. The unscented transform of BG is partly adapted from [15], and it is different from that of MFG adapted from [7]. Simulations indicate that the unscented BG filter is more accurate than the conventional MEKF, and it is very similar to the unscented MFG filter.

II. MATHEMATICAL PRELIMINARIES

A. Unit Quaternions

Unit quaternions are vectors on the three sphere $\mathbb{S}^3 \subset \mathbb{R}^4$. In this paper, the Hamilton convention is used [16]:

$$q = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix}^T = \begin{bmatrix} q_r & q_v^T \end{bmatrix}^T \in \mathbb{S}^3, \qquad (1)$$

where $q_r = q_0 \in \mathbb{R}$ is the scalar part, and $q_v = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}^T \in \mathbb{R}^3$ is the vector part. A multiplication operation $\otimes$ can be defined on $\mathbb{S}^3$:

$$p \otimes q = \begin{bmatrix} p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3 \\ p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2 \\ p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1 \\ p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0 \end{bmatrix}, \tag{2}$$

which makes $\mathbb{S}^3$ a Lie group isomorphic to $\mathrm{SU}(2)$, and hence a double cover of $\mathrm{SO}(3)$. Quaternion multiplication (2) can also be written in matrix form as

$$p \otimes q = [p]_L q = [q]_R p, \tag{3}$$

where

$$[p]_L = \begin{bmatrix} p_0 & -p_1 & -p_2 & -p_3 \\ p_1 & p_0 & -p_3 & p_2 \\ p_2 & p_3 & p_0 & -p_1 \\ p_3 & -p_2 & p_1 & p_0 \end{bmatrix}, \tag{4}$$

and

$$[q]_R = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix}. \tag{5}$$

The maps $[\cdot]_L$ and $[\cdot]_R : \mathbb{S}^3 \to \mathrm{GL}(4)$ define two group homomorphisms. The ranges are denoted by $\mathbb{S}^3_L$ and $\mathbb{S}^3_R$, and are named left- and right-isoclinic rotation groups respectively. $\mathbb{S}^3_L$ and $\mathbb{S}^3_R$ commute with each other, i.e., for any $p, q \in \mathbb{S}^3$, we have $[p]_L[q]_R = [q]_R[p]_L$. It is straightforward to check $\mathbb{S}^3_L$, $\mathbb{S}^3_R \subset \mathrm{SO}(4)$, and in fact, they are two normal subgroups of $\mathrm{SO}(4)$. Furthermore, their direct product $\mathbb{S}^3_L \times \mathbb{S}^3_R$ is a double cover of $\mathrm{SO}(4)$. Specifically, any $M \in \mathrm{SO}(4)$ can be uniquely decomposed into

$$M = [p]_L[q]_R \tag{6}$$

for some $p, q \in \mathbb{S}^3$ up to the signs of p and q.

The homomorphism between $\mathbb{S}^3$ and $\mathrm{SO}(3)$ is denoted by $\varphi : \mathbb{S}^3 \to \mathrm{SO}(3)$,

$$\varphi(q) = \begin{bmatrix} 1-2(q_2^2+q_3^2) & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & 1-2(q_1^2+q_3^2) & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & 1-2(q_1^2+q_2^2) \end{bmatrix}. \tag{7}$$

Note that $\varphi(q) = \varphi(-q)$ for any $q \in \mathbb{S}^3$. The map $\wedge$ is used to denote the vector space isomorphism from $\mathbb{R}^3$ to $\mathfrak{so}(3)$ or $\mathrm{Im}(\mathbb{H})$, i.e., the Lie algebra of $\mathrm{SO}(3)$ or $\mathbb{S}^3$, depending on the context. More specifically, for $v = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T \in \mathbb{R}^3$,

$$\hat{v} = (v)^\wedge = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \in \mathfrak{so}(3), \tag{8}$$

or

$$\hat{v} = (v)^\wedge = \begin{bmatrix} 0 & v_1 & v_2 & v_3 \end{bmatrix}^T \in \mathrm{Im}(\mathbb{H}). \tag{9}$$

The inverse map of $\wedge$ is denoted by $\vee$. The homomorphism φ satisfies the following identity [6]: for any q, $p \in \mathbb{S}^3$,

$$\mathrm{tr}\left(\varphi(p)\varphi(q)^T\right) + 1 = 4(p^Tq)^2. \tag{10}$$

B. Bingham Distribution

A random vector $q \in \mathbb{S}^r$ follows a Bingham distribution [17] with parameter $A = A^T \in \mathbb{R}^{(r+1)\times(r+1)}$ if it has the following density function:

$$f(q; A) = \frac{1}{c(A)} \exp\left(q^T A q\right), \tag{11}$$

where $c(A) = {}_1F_1(\frac{1}{2}, \frac{r+1}{2}, A)$ is the normalizing constant, and ${}_1F_1$ denotes a confluent hypergeometric function of matrix argument [17]. This distribution is denoted by $q \sim \mathcal{B}(A)$. The Bingham distribution is antipodally symmetric, i.e., $f(q) = f(-q)$ for any $q \in \mathbb{S}^r$.

The properties of a Bingham distribution are determined by the eigenvalue decomposition of the parameter $A = MZM^T$, where $M = \begin{bmatrix} m_0 & \dots & m_r \end{bmatrix}$ are the eigenvectors of A, and $Z = \mathrm{diag}\left(\begin{bmatrix} z_0 & \cdots & z_r \end{bmatrix}\right)$ are the corresponding eigenvalues. For uniqueness, we assume the eigenvalues are in descending order. Due to the unit length constraint, it is straight forward to show $\mathcal{B}\left(MZM^T\right) = \mathcal{B}\left(M(Z + zI_{r+1})M^T\right)$ for any $z \in \mathbb{R}$. Thus, in this paper we make the following assumption:

$$0 = z_0 \geq \cdots \geq z_r. \tag{12}$$

The normalizing constant depends only on the eigenvalues, i.e., $c(A) = c(Z)$. Since the Bingham distribution is antipodally symmetric, its first order moment $\mathrm{E}[q] = 0$. Its second order moment, or the moment of inertia is given by

$$\mathrm{E}\left[qq^T\right] = MDM^T \triangleq M\mathrm{diag}(d_0, \dots, d_r)M^T, \tag{13}$$

where

$$d_i = \frac{1}{c(Z)}\frac{\partial c(Z)}{\partial z_i}, \qquad i = 0, \dots, r. \tag{14}$$

Since $q^Tq = 1$, the matrix D satisfies $\sum_{i=0}^r d_i = 1$. The mode of Bingham distribution is $q = m_0$, which maximizes the density (11).

A random rotation matrix $R \in \mathrm{SO}(3)$ follows the matrix Fisher distribution [18], [19] with parameter $F \in \mathbb{R}^{3\times3}$ if it has the following density function:

$$f(R; F) = \frac{1}{c(F)} \mathrm{etr}\left(FR^T\right), \tag{15}$$

where $c(F)$ is the normalizing constant, and $\mathrm{etr}(\cdot)$ is an abbreviation for $\exp(\mathrm{tr}(\cdot))$. This distribution is denoted by $\mathcal{M}(F)$. When $r = 3$, the Bingham distribution is equivalent to the matrix Fisher distribution using the homomorphism φ.

Lemma 1 ([6]): Let $M \in \mathrm{SO}(4)$, and $Z = \mathrm{diag}\left(\begin{bmatrix} 0 & z_1 & z_2 & z_3 \end{bmatrix}\right)$ with $0 \geq z_1 \geq z_2 \geq z_3$. Define $S = \mathrm{diag}\left(\begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}\right)$ as $s_1 = \frac{1}{4}(z_1 - z_2 - z_3)$, $s_2 = \frac{1}{4}(z_2 - z_1 - z_3)$, and $s_3 = \frac{1}{4}(z_3 - z_1 - z_2)$. Let $M = [u]_L[v^{-1}]_R$ be its isoclinic decomposition. Define $U = \varphi(u)$ and $V = \varphi(v)$. Then $q \sim \mathcal{B}(MZM^T)$ if and only if $\varphi(q) \sim \mathcal{M}(USV^T)$.

Proof: It suffices to check for all $q \in \mathbb{S}^3$, $c \cdot \exp\left(q^T MZM^T q\right) = \mathrm{etr}\left(USV^T\varphi(q)^T\right)$ for some constant $c \in \mathbb{R}$. Let $M = \begin{bmatrix} m_0 & m_1 & m_2 & m_3 \end{bmatrix}$, $I_{4\times4} =$

$\begin{bmatrix} e_0 & e_1 & e_2 & e_3 \end{bmatrix}$, and $I_{3\times 3} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 \end{bmatrix}$, then

$$\varphi(m_0) = \varphi([u]_L[v^{-1}]_R e_0) = UV^T,$$
$$\varphi(m_i) = \varphi([u]_L[v^{-1}]_R e_i) = U\exp(\pi\hat{\varepsilon}_i)V^T.$$

Therefore, we have

$$USV^T = \frac{1}{4}\sum_{i=1}^{3} z_i \varphi(m_i).$$

Then, using (10) we have

$$\operatorname{tr}\left(USV^T\varphi(q)^T\right) = \frac{1}{4}\sum_{i=1}^{3} z_i \cdot \left(4(q^Tm_i)^2 - 1\right)$$
$$= \sum_{i=1}^{3} z_i(q^Tm_i)^2 - \frac{1}{4}\sum_{i=1}^{3} z_i = q^TMZM^Tq + c,$$

which shows the equivalence. ■

C. Matrix Fisher–Gaussian Distribution

The matrix Fisher–Gaussian (MFG) distribution [12], [13] is defined on $\mathrm{SO}(3)\times\mathbb{R}^n$. Random elements $(R, x) \in \mathrm{SO}(3)\times\mathbb{R}^n$ follow a matrix Fisher–Gaussian distribution with parameters $U, V \in \mathrm{SO}(3)$, $S = \operatorname{diag}\left(\begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}\right) \in \mathbb{R}^{3\times 3}$ where $s_1 \geq s_2 \geq |s_3| \geq 0$, $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n\times n}$, and $P \in \mathbb{R}^{n\times 3}$, if they have the density function:

$$f(R, x; U, S, V, \mu, \Sigma, P) = \frac{1}{c(S)\sqrt{(2\pi)^n \det(\Sigma_c)}} \times \exp\left(-\frac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)\right)\operatorname{etr}\left(USV^TR^T\right), \tag{16}$$

where $c(S)$ is the normalizing constant for the corresponding matrix Fisher distribution, and $\Sigma_c \in \mathbb{R}^{n\times n}$ satisfying $\Sigma_c = \Sigma_c^T \succ 0$ is given by

$$\Sigma_c = \Sigma - P(\operatorname{tr}(S)\, I_{3\times 3} - S)P^T. \tag{17}$$

Next, there are two formulations of MFG, referred to as MFGI and MFGB, depending on the expression of $\mu_c \in \mathbb{R}^n$ [13]. For MFGI, it is given by

$$\mu_c = \mu + P(QS - SQ^T)^\vee \triangleq \mu + P\nu_R^I, \tag{18}$$

or, for MFGB,

$$\mu_c = \mu + P(SQ - Q^TS)^\vee \triangleq \mu + P\nu_R^B, \tag{19}$$

where $Q = U^TRV$. This distribution is denoted by $(R, x) \sim \mathcal{MG}(U, S, V, \mu, \Sigma, P)$.

The probability density function (16) is composed of three terms: the first one is for normalization; the second term is for x and it has the form as $\mathcal{N}(\mu_c, \Sigma_c)$; the last term is for R and it is identical to the matrix Fisher Distribution. It is straightforward to see that the marginal distribution of R is a matrix Fisher distribution with parameter USV^T, and the distribution of x conditioned by R is Gaussian with $x|R \sim \mathcal{N}(\mu_c(R), \Sigma_c)$. The correlation between R and x is encoded in the $(3n)$-element matrix P. The vector $\nu_R \in \mathbb{R}^3$ indicates the deviation of R from the mean attitude UV^T, and it has two forms. If (18) is used, x is correlated with rotations specified by P resolved in the inertial frame (MFGI); if instead (19) is used, x is correlated with those rotations resolved in the body-fixed frame (MFGB) of each $R \in \mathrm{SO}(3)$.

III. Bingham–Gaussian Distribution

In this section we introduce the Bingham–Gaussian (BG) distribution on $\mathbb{S}^3\times\mathbb{R}^n$ which is equivalent to the matrix Fisher–Gaussian distribution on $\mathrm{SO}(3)\times\mathbb{R}^n$.

A. Definition

Definition 1: Random elements $(q, x) \in \mathbb{S}^3\times\mathbb{R}^n$ follow Bingham-Gaussian distribution with parameters $M \in \mathrm{SO}(4)$, $Z = \operatorname{diag}(0, z_1, z_2, z_3)$ where $0 \geq z_1 \geq z_3 \geq z_3$, $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n\times n}$, and $P \in \mathbb{R}^{n\times 3}$, if they have the following density function:

$$f(q, x; M, Z, \mu, \Sigma, P) = \frac{1}{c(Z)\sqrt{(2\pi)^n\det(\Sigma_c)}} \times \exp\left(-\frac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)\right)\exp\left(q^TMZM^Tq\right), \tag{20}$$

where $c(Z)$ is the normalizing constant for the corresponding Bingham distribution, and $\Sigma_c \in \mathbb{R}^{n\times n}$ satisfying $\Sigma_c = \Sigma_c^T \succ 0$ is given by

$$\Sigma_c = \Sigma + \frac{1}{2}P\operatorname{diag}\left(\begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix}\right)P^T. \tag{21}$$

Depending on the choice of $\mu_c \in \mathbb{R}^n$, there are two formulations of BG. For BGI, $\mu_c \in \mathbb{R}^n$ is given by

$$\mu_c = \mu + P\begin{bmatrix} (z_2 - z_3)p_2p_3 - z_1p_0p_1 \\ (z_3 - z_1)p_1p_3 - z_2p_0p_2 \\ (z_1 - z_2)p_1p_2 - z_3p_0p_3 \end{bmatrix} \triangleq \mu + P\nu_q^I, \tag{22}$$

or, for BGB,

$$\mu_c = \mu + P\begin{bmatrix} (z_3 - z_2)p_2p_3 - z_1p_0p_1 \\ (z_1 - z_3)p_1p_3 - z_2p_0p_2 \\ (z_2 - z_1)p_1p_2 - z_3p_0p_3 \end{bmatrix} \triangleq \mu + P\nu_q^B, \tag{23}$$

with $p = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 \end{bmatrix}^T \triangleq M^Tq$. This distribution is denoted by $\mathcal{BG}(M, Z, \mu, \Sigma, P)$.

BG is antipodally symmetric in $\mathbb{S}^3$, i.e., for any $(q, x) \in \mathbb{S}^3\times\mathbb{R}^n$, $f(q, x) = f(-q, x)$. Similar with MFG, the density (20) has three terms: the first one is for normalization; the second term is for x and it has the form as $\mathcal{N}(\mu_c, \Sigma_c)$; the last term is for q and it is identical to the Bingham distribution. The marginal distribution of q is a Bingham distribution with parameter MZM^T, and the distribution of x conditioned by q is Gaussian with $x|q \sim \mathcal{N}(\mu_c(q), \Sigma_c)$. Also, the correlation between q and x is represented by the $(3n)$-element matrix P. The vector $\nu_q \in \mathbb{R}^3$ indicates the deviation of q from the mode m_0, i.e., if $q = m_0$, then $p = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T$ and hence $\nu_q = 0$. The two expressions for ν_q again have the same interpretations as MFG: If (22) is used, x is correlated with rotations specified by P resolved in the inertial frame (BGI); if (23) is used, the rotations are resolved in the body-fixed frame (BGB) of each $q \in \mathbb{S}^3$.

B. Construction

The Bingham-Gaussian distribution is constructed by transforming the matrix Fisher-Gaussian distribution using the homomorphism $\varphi : \mathbb{S}^3 \to \mathrm{SO}(3)$.

Theorem 1: Let (M, Z, μ, Σ, P) be the parameters of a Bingham-Gaussian distribution. Define $S = \mathrm{diag}\left(\begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}\right)$, and $U, V \in \mathrm{SO}(3)$ in the same way as in Lemma 1. Then $(q, x) \sim \mathcal{BG}(M, Z, \mu, \Sigma, P)$ if and only if $(\varphi(q), x) \sim \mathcal{MG}(U, S, V, \mu, \Sigma, P)$.

Proof: In Lemma 1, we have proven $\exp(q^T MZM^T q) \propto \mathrm{etr}\left(USV^T \varphi(q)^T\right)$. Also, it is straightforward to show $\Sigma_{c,\mathcal{MG}} = \Sigma_{c,\mathcal{BG}}$. Therefore, it remains to check for all $q \in \mathbb{S}^3$, $\mu_{c,\mathcal{BG}}(q) = \mu_{c,\mathcal{MG}}(\varphi(q))$. Note that $\nu_q^I = (\varphi(p)S - S\varphi(p)^T)^\vee = \nu^I_{\varphi(q)}$, and $\nu_q^B = (S\varphi(p) - \varphi(p)^T S)^\vee = \nu^B_{\varphi(q)}$. So we only need to show $\varphi(M^T q) = U^T \varphi(q) V$, which is straightforward

$$\varphi(M^T q) = \varphi([u^{-1}]_L [v]_R q) = U^T \varphi(q) V.$$

This concludes the proof. ∎

Because BG and MFG are equivalent, all of the properties of MFG given in [12], [13] also apply to BG. In this paper we list a few of them which is necessary for the subsequent development of the unscented attitude filter based on BG.

It should be noted that in [13] we give a geometric construction of MFG by conditioning a $(9+n)$-variate Gaussian distribution form the ambient space $\mathbb{R}^9 \times \mathbb{R}^n$ to $\mathrm{SO}(3) \times \mathbb{R}^n$. On the other hand, $\mathbb{S}^3$ with antipodal points identified (namely the real projective space $\mathbb{RP}^3$) does not embed into $\mathbb{R}^4$. Therefore we cannot give a similar geometric construction of BG by conditioning, which may be pursued in future works.

C. Moments

Next, we present the first and the second order moments of BG.

Theorem 2: Let $(q, x) \sim \mathcal{BG}(M, Z, \mu, \Sigma, P)$. Then $\mathrm{E}[q] = 0$, and $\mathrm{E}\left[qq^T\right]$ is given in (13) and (14). Other moments are

$$\mathrm{E}[x] = \mu, \tag{24}$$

$$\mathrm{E}[\nu_q] = 0, \tag{25}$$

$$\mathrm{E}\left[xx^T\right] = \Sigma_c + \mu\mu^T + P\mathrm{E}\left[\nu_q \nu_q^T\right] P^T, \tag{26}$$

$$\mathrm{E}\left[x\nu_q^T\right] = P\mathrm{E}\left[\nu_q \nu_q^T\right], \tag{27}$$

where $\mathrm{E}\left[\nu_q \nu_q^T\right]$ is a diagonal matrix with the i-th diagonal element given by

$$\frac{z_j - z_k}{2c(Z)} \left(\frac{\partial c(Z)}{\partial z_j} - \frac{\partial c(Z)}{\partial z_k} \right) + \frac{z_i}{2c(Z)} \left(\frac{\partial c(Z)}{\partial z_i} - \frac{\partial c(Z)}{\partial z_0} \right), \tag{28}$$

for $\{i, j, k\} = \{1, 2, 3\}$.

Proof: Equations (24) to (27) are straightforward to show by integrating the density (20) directly and using (13). To prove (28), first note that $\mathrm{E}\left[p_0^{n_0} p_1^{n_1} p_2^{n_2} p_3^{n_3}\right] = 0$ if any of $\{n_0, n_1, n_2, n_3\}$ is odd, because the Bingham distribution is antipodally symmetric. Thus, $\mathrm{E}\left[\nu_q \nu_q^T\right]$ is diagonal with the i-th diagonal term given by

$$\mathrm{E}\left[\nu_q \nu_q^T\right]_{ii} = (z_j - z_k)^2 \mathrm{E}\left[p_j^2 p_k^2\right] + z_i^2 \mathrm{E}\left[p_0^2 p_i^2\right]. \tag{29}$$

Equation (28) is then shown by using the following two identities [17]:

$$\mathrm{E}\left[p_j^2 p_k^2\right] = \frac{1}{c(Z)} \frac{\partial^2 c(Z)}{\partial z_j \partial z_k}, \tag{30}$$

$$2(z_j - z_k) \frac{\partial^2 c(Z)}{\partial z_j \partial z_k} = \frac{\partial c(Z)}{\partial z_j} - \frac{\partial c(Z)}{\partial z_k}, \tag{31}$$

for $j, k \in \{0, 1, 2, 3\}$, and that $\mathrm{E}\left[\nu_q \nu_q^T\right]$ is continuous in Z. ∎

The normalizing constant $c(Z)$ and its first order derivatives can be calculated using the one dimensional integral formula used in [7], or the saddle point approximation used in [20]. The second order derivatives of $c(Z)$ can be calculated by solving a linear system involving the first order moments, derived from (30), (31) and $q^T q = 1$.

D. Maximum Likelihood Estimation

Suppose we have a set of weighted samples $\{(x_i, q_i, w_i)\}_{i=1}^{N_s}$ from a BG. The log-likelihood function for the parameters of BG, after omitting some constants, is given by

$$\begin{aligned} l = &-\log c(Z) + \bar{\mathrm{E}}\left[q^T MZMq\right] \\ &- \log\det \Sigma_c - \frac{1}{2}\bar{\mathrm{E}}\left[(x - \mu_c)^T \Sigma_c^{-1} (z - \mu_c)\right], \end{aligned} \tag{32}$$

where $\bar{\mathrm{E}}[\cdot]$ represents the sample mean of a random variable. As in the MLE for MFG, maximizing the log-likelihood simultaneously for all parameters requires numerical iterations. Instead, the log-likelihood function is split into two parts:

$$l = l_q + l_{x|q}, \tag{33}$$

where l_q corresponds to the first two terms on the right hand side of (32), which is the marginal likelihood for the Bingham distribution; and $l_{x|q}$ corresponds to the last two terms on the right hand side of (32), which is the conditional likelihood for the Gaussian distribution.

The marginal likelihood l_q is first maximized using the MLE for the Bingham distribution.

Theorem 3 ([17]): The marginal maximum likelihood estimate for M is given by the eigen-decomposition $\bar{\mathrm{E}}\left[qq^T\right] = MDM^T$, and that for Z is given by solving (14) from D.

After obtaining the estimates for M and Z, let $p_i = M^T q_i$, and ν_{q_i} be given in (22) or (23) using p_i. Also, denote $\overline{\mathrm{cov}}(a, b) = \bar{\mathrm{E}}\left[ab^T\right] - \bar{\mathrm{E}}[a]\, \bar{\mathrm{E}}[b]^T$ as the sample covariance between $a \in \mathbb{R}^m$, $b \in \mathbb{R}^n$. Then the estimates for μ, Σ, and P are addressed by maximizing the conditional likelihood.

Theorem 4: The conditional maximum likelihood estimates for P, μ, and Σ are

$$P = \overline{\mathrm{cov}}(x, \nu_R)\, \overline{\mathrm{cov}}(\nu_R, \nu_R)^{-1}, \tag{34}$$

$$\mu = \bar{\mathrm{E}}[x] - P\bar{\mathrm{E}}[\nu_R], \tag{35}$$

$$\Sigma = \overline{\mathrm{cov}}(x, x) - P\overline{\mathrm{cov}}(x, \nu_q)^T - \tfrac{1}{2} P \mathrm{diag}([z_1, z_2, z_3]) P^T. \tag{36}$$

Proof: The proof is the same as Theorem 5 in [13]. ■

The estimates obtained by maximizing the marginal and conditional likelihood separately are only approximations to maximizing the full likelihood. However, this is a necessary compromise for efficient implementations of attitude filters using BG, where the MLE must be solved typically at more than 100 Hz in the case of a gyroscope. The accuracy of this approximation is analyzed from an information theoretic point of view in [13].

IV. Unscented Attitude Filter

In this section, we apply the proposed BG in the classical problem of estimating the attitude from a gyroscope with time-varying biases, and other attitude or direction sensors.

The discrete time gyroscope kinematics is assumed to be

$$q_{k+1} = q_k \otimes \exp\left\{h(\hat{\Omega}_k + \hat{x}_k) + (H_u \Delta W_u)^\wedge\right\}, \quad (37)$$

$$x_{k+1} = x_k + H_v \Delta W_v, \quad (38)$$

where the subscripts denote discrete time steps. The vector $\Omega \in \mathbb{R}^3$ is the measured angular velocity, and $x \in \mathbb{R}^3$ is the gyroscope bias. Next, $\Delta W_u, \Delta W_v \in \mathbb{R}^3$ are independent white noises, and $H_u, H_v \in \mathbb{R}^{3\times 3}$ describe the strengths of noises. Finally, $h = t_{k+1} - t_k$ is the constant sampling period.

The attitude q is also measured by N_a attitude sensors as $\tilde{q}_i$, $i = 1, \ldots, N_a$, with the noises distributed by $q^{-1} \otimes \tilde{q}_i \sim \mathcal{B}(A_i)$. Also, N_v reference directions $v_{r_i} \in \mathbb{S}^2$ fixed in the inertial frame are measured in the gyroscope body-fixed frame as $\tilde{v}_i \in \mathbb{S}^2$, $i = 1, \ldots, N_v$, with the noise distributed by $\tilde{v}_i \sim \mathcal{VM}(\varphi(q^{-1})v_{r_i}, \kappa_i)$, where $\mathcal{VM}(\mu, \kappa)$ denotes the von Mises Fisher distribution on $\mathbb{S}^2$ [3]. The set of all measurements at time t_k is denoted by $\mathcal{Z}_k$. Next, we introduce an unscented attitude filter using BG.

A. Unscented Transform of BG

To perform sigma point selection, we first introduce the canonical form of BG.

Theorem 5: Let $(q, x) \sim \mathcal{BG}(M, Z, \mu, \Sigma, P)$. Define $p = M^T q \in \mathbb{S}^3$, and $y = \Sigma_c^{-1/2}(x - \mu - \nu_q) \in \mathbb{R}^n$. Then $(p, y) \sim \mathcal{BG}(0, I, 0, I, Z)$.

Proof: Substitute $q = Mp$, and $x = \Sigma_c^{1/2} y + \mu + \nu_q$ into (20), then the desired result is proven. ■

The unscented transform of BG is split into selecting sigma points for the decoupled Bingham [15] and Gaussian parts of the canonical BG, which are then transformed back to the original BG according to Theorem 5.

Definition 2: Let $(q, x) \sim \mathcal{BG}(M, Z, \mu, \Sigma, P)$. Choose w_B, w_G, $w_I > 0$ as the weights for sigma points of the Bingham, Gaussian, and identity parts respectively, such that $w_B + w_G + w_I = 1$. Select $(7 + 2n)$ sigma points for the canonical BG as

$$(p, y)_{1,2} = ([\cos\theta_1, \pm\sin\theta_1, 0, 0]^T, [0, \ldots, 0]^T),$$
$$(p, y)_{3,4} = ([\cos\theta_2, 0, \pm\sin\theta_2, 0]^T, [0, \ldots, 0]^T),$$
$$(p, y)_{5,6} = ([\cos\theta_3, 0, 0, \pm\sin\theta_3]^T, [0, \ldots, 0]^T),$$
$$(p, y)_{7,8} = \left([1, 0, 0, 0]^T, \left[\pm\sqrt{\tfrac{n}{w_G}}, 0, \ldots, 0\right]^T\right),$$
$$\vdots$$
$$(p, y)_{5+2n,6+2n} = \left([1, 0, 0, 0]^T, \left[0, \ldots, 0, \pm\sqrt{\tfrac{n}{w_G}}\right]^T\right),$$
$$(p, y)_{7+2n} = ([1, 0, 0, 0]^T, [0, \ldots, 0]^T), \quad (39)$$

where $\theta_i,\ i = 1, 2, 3$ are given by

$$\theta_i = \arcsin\left(\sqrt{\frac{d_i}{2w_i}}\right), \qquad d_i = \frac{1}{c(Z)}\frac{\partial c(Z)}{\partial z_i}, \quad (40)$$

and w_i are the weights for the first three pairs of sigma points, given by

$$2w_i = d_i + (1 - \alpha)\tfrac{d_0}{3}, \quad (41)$$

where α is chosen such that $2(w_1 + w_2 + w_3) = w_B$. Also, the weights for the next n pairs sigma points are $\frac{w_G}{2n}$, and the weight for the last sigma point is w_I. Let $q_i = Mp_i$, and $x_i = \Sigma_c^{1/2} y_i + \mu + P\nu_{q_i}$ for $i = 1, \ldots, 2n + 7$, then the sigma points for $\mathcal{BG}(\mu, \Sigma, P, M, Z)$ are defined as $\{(q, x, w)_i\}_{i=1}^{2n+7}$.

After obtaining these sigma points, they can be propagated through the gyroscope kinematics equations (37)-(38), and a new BG can be constructed from these propagated sigma points using the MLE introduced in Section III-D. The next theorem validates the proposed unscented transform of BG.

Theorem 6: The marginal-conditional MLE for the sigma points in Definition 2 is exactly $\mathcal{BG}(M, Z, \mu, \Sigma, P)$.

Proof: The marginal MLE is proven in [15]. The conditional MLE can be proven trivially by Theorem 4. ■

While BG is equivalent to MFG, the presented unscented transform of BG is not equivalent to that of MFG presented in [12]. More specifically, the sigma points from the Bingham part of BG on $\mathbb{S}^3$ [15] do not necessarily correspond to the sigma points from the matrix Fisher part of MFG on SO(3) [7] under φ. The comparison of these two sigma point selection schemes will be pursued in future works.

B. Measurement Update

The likelihood functions of the attitude q are

$$f(\tilde{q}_i|q) \propto \exp\left((q^{-1} \otimes \tilde{q}_i)^T A_i (q^{-1} \otimes \tilde{q}_i)\right), \quad (42)$$

$$f(\tilde{v}_i|q) \propto \exp\left(\kappa_i \tilde{v}_i^T \varphi(q^{-1}) v_{r_i}\right), \quad (43)$$

for the attitude and direction measurements respectively. Suppose before update, $(q, x) \sim \mathcal{BG}(M, Z, \mu, \Sigma, P)$. According to the Bayes' formula, the posterior density for (q, x) is the density (20) multiplied by the likelihood functions (42), (43), which is calculated as follows.

Theorem 7: Let $D = \mathrm{diag}\left(\begin{bmatrix}1 & -1 & -1 & -1\end{bmatrix}\right)$. Then the posterior density is

$$p(q, x|\mathcal{Z}) = \exp\left(-\tfrac{1}{2}(x - \mu_c)^T \Sigma_c^{-1}(x - \mu_c)\right) \exp\left(q^T A^+ q\right) \quad (44)$$

where μ_c is defined with respect to M, Z, and

$$A^+ = MZM^T + \sum_{i=1}^{N_a} [\tilde{q}_i]_L D A_i D [\tilde{q}_i]_L^T + \sum_{i=1}^{N_v} \kappa_i [\hat{\tilde{v}}_i]_R^T [\hat{v}_{r_i}]_L. \quad (45)$$

TABLE I

UNSCENTED ATTITUDE FILTER BASED ON BG

1: **procedure** ESTIMATION($\mathcal{BG}(t_0), \Omega(t), \mathcal{Z}(t)$)
2: Let $k = 0$.
3: **repeat**
4: Select sigma points $\{q_i, x_i, w_i\}_{i=1}^{13}$ from $\mathcal{BG}(t_k)$.
5: Select sigma points $\{W_j, w_j\}_{j=1}^{7}$ from $\mathcal{N}(0, hH_uH_u^T)$.
6: Propagate the sigma points according to

$$q_{i,j} = q_i \otimes \exp(h(\hat{\Omega}(t_k) + \hat{x}_i) + \hat{W}_j),$$
$$x_{i,j} = x_i, \qquad w_{i,j} = w_i w_j.$$

7: Estimate $\{M, Z, \mu, \Sigma, P\}_{k+1}$ from $\{q, x, w\}_{i,j}$ using MLE.
8: Let $\Sigma_{k+1} = \Sigma_{k+1} + hH_vH_v^T$.
9: Set $\mathcal{BG}(t_{k+1}) = \mathcal{BG}(\{M, Z, \mu, \Sigma, P\}_{k+1})$
10: $k = k + 1$.
11: **until** $\mathcal{Z}(t_{k+1})$ ia available
12: Calculate A^+ using (45) from $\mathcal{BG}(t_{k+1})$ and $\mathcal{Z}(t_{k+1})$.
13: Let $M^+Z^+(M^+)^T = A^+$ be the eigen-decomposition of A^+.
14: Calculate the moments in (48)-(50).
15: Obtain $\{\mu, \Sigma, P\}^+$ using Theorem 4.
16: Let $\mathcal{BG}(t_{k+1}) = \mathcal{BG}(\{M, Z, \mu, \Sigma, P\}^+)$.
17: Obtain the estimates of attitude and bias: $q = m_0^+$, and $x = \mu^+$.
18: **go to** step 3.
19: **end procedure**

Proof: We only need to rearrange (42) and (43) into the form of Bingham densities.

$$\begin{aligned} f(\tilde{q}_i|q) &\propto \exp\left((\tilde{q}_i^{-1} \otimes q)^T DA_iD(\tilde{q}_i^{-1} \otimes q)\right) \\ &= \exp\left(q^T \left([\tilde{q}_i]_L DA_iD[\tilde{q}_i]_L^T\right) q\right). \end{aligned} \tag{46}$$

Also, we have

$$\begin{aligned} f(\tilde{v}_i|q) &\propto \exp\left(\kappa_i \hat{\tilde{v}}_i^T (q^{-1} \otimes \hat{v}_{r_i} \otimes q)\right) \\ &= \exp\left(\kappa_i \hat{\tilde{v}}_i^T [q]_L^T [q]_R \hat{v}_{r_i}\right) \\ &= \exp\left(q^T (\kappa_i [\hat{\tilde{v}}_i]_R^T [\hat{v}_{r_i}]_L) q\right). \end{aligned} \tag{47}$$

And the desired result is proven. ■

The posterior density (44) is no longer a BG density, since μ_c is defined with respect to $MZM^T = A$, not A^+. Therefore, it needs to be matched to a BG using the MLE. The marginal MLE is given trivially by the eigen-decomposition $A^+ = M^+Z^+(M^+)^T$. And the conditional MLE are calculated using the following expectations.

$$\mathrm{E}[x] = \mu + P\mathrm{E}[\nu_q], \tag{48}$$

$$\begin{aligned} \mathrm{E}\left[xx^T\right] &= \mu\mu^T + \mu\mathrm{E}[\nu_q]^T P^T \\ &\quad + P\mathrm{E}[\nu_q]\mu^T + P\mathrm{E}\left[\nu_q\nu_q^T\right] P^T + \Sigma_c, \end{aligned} \tag{49}$$

$$\mathrm{E}\left[x(\nu_q^+)^T\right] = P\mathrm{E}\left[\nu_q(\nu_q^+)^T\right], \tag{50}$$

where ν_q^+ is defined with respect to M^+ and Z^+. Let $\tilde{M} = M^TM^+$, then $\mathrm{E}[\nu_q] = \tilde{M}D^+\tilde{M}^T$, where D^+ is the moment of inertia for $\mathcal{B}(Z^+)$. Also, $\mathrm{E}\left[\nu_q\nu_q^+\right]$, and $\mathrm{E}[\nu_q\nu_q]$ can be expressed as linear combinations of $\mathrm{E}\left[(p_i^+)^2(p_j^+)^2\right]$, $i, j \in \{0, 1, 2, 3\}$, which can be calculated using (30) and (31).

The unscented propagation and measurement update of BG constitute an unscented attitude filter which estimates the attitude and gyroscope bias concurrently. The pseudo-code is summarized in Table I.

V. SIMULATIONS

The proposed unscented attitude filter based on BG is compared against the conventional MEKF through simulations. The gyroscope noise parameters are chosen as $H_u = \sigma_u I_{3\times3}$ where $\sigma_u = 10\,\mathrm{deg}/\sqrt{\mathrm{s}}$, and $H_v = \sigma_v I_{3\times3}$ where $\sigma_v = 500\,\mathrm{deg}/\mathrm{h}/\sqrt{\mathrm{s}}$. Two reference vectors $v_{r_1} = \begin{bmatrix}0 & 1 & 0\end{bmatrix}^T$ and $v_{r_2} = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^T$ fixed in the inertial frame are assumed to be measured in the body-fixed frame. The direction sensor noises are assumed to be isotropic zero-mean Gaussian and addictive, with the variances given by $\sigma_1^2 = 0.01$, and $\sigma_2^2 \in \{0.01, 0.1, 1, 10\}$ respectively. The measurement noises are then matched to von Mises Fisher distributions as required in Section IV through moment matching for the unscented BG filter. The sampling frequency for gyroscope is $150\,\mathrm{Hz}$, and for the direction sensors is $30\,\mathrm{Hz}$. The simulation lasts $5\,\mathrm{min}$. One hundred Monte Carlo simulations with respect to random noises are performed for each filter under each noise level σ_2^2.

The initial attitude is set as the true attitude rotated about its first body-fixed axis by 180°, and the initial bias is chosen as $\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}^T\ \mathrm{rad\,s^{-1}}$. The inertial attitude uncertainty is $\delta q \sim \mathcal{N}(0, 10^{10}I_{3\times3})$ for MEKF, and then it is matched to a Bingham distribution through moment matching for the unscented BG filter. The initial bias uncertainty is $0.1^2 I_{3\times3}$, and the correlation between attitude and bias is zero. Five filters are compared: MEKF, the unscented BG filters using the BGI/BGB definitions (BGIU/BGBU), and the unscented MFG filters using the MFGI/MFGB definitions (MFGIU/MFGBU) [13].

Three types of error are calculated: (i) the full attitude error (FAE) is the angle between the true and estimated attitude; (ii) the partial attitude error (PAE) is the angle between the first reference vector resolved in the true and estimated body-fixed frames; and (iii) the bias error (BE). PAE neglects the error of rotation about the first reference vector, so it remains low when the second direction measurement becomes very inaccurate. The errors are averaged across all time steps in one simulation, and then averaged across one hundred simulations.

The simulation results are summarized in Table II, and the error trajectory of a single simulation with $\sigma_2^2 = 10$ is presented in Fig. 1. It can be seen that when $\sigma_2^2 = 10$, i.e., when the second direction sensor is very inaccurate, the unscented BG and MFG filters are much more accurate than MEKF in full attitude estimation. This is because the uncertainty of the rotation about the first reference vector is very large when σ_2^2 is large, so the wrapping error of the Gaussian distribution assumed in MEKF becomes significant. Also, the unscented BG and MFG filters have lower partial attitude and bias errors, which is mainly contributed by their faster convergence rate. Because the attitude estimate is initialized completely incorrectly, the linearization of MEKF induces large errors, which makes it converge very slowly. Comparing the unscented BG filters with the MFG filters, they exhibit very similar estimation errors, with the minor exception that the bias error for the BG filters is slightly

TABLE II

ATTITUDE (deg) AND BIAS (deg /s) ERRORS ($\pm$S.D.) FOR DIFFERENT FILTERS

σ_2^2		MEKF	BGIU	BGBU	MFGIU	MFGBU
0.01	FAE	5.05±0.08	4.82±0.04	4.82±0.04	4.82±0.04	4.82±0.04
	PAE	3.91±0.08	3.68±0.03	3.68±0.03	3.68±0.03	3.68±0.03
	BE	3.6±0.8	2.6±0.5	2.6±0.6	2.6±0.5	2.6±0.6
0.1	FAE	7.06±0.25	6.67±0.11	6.67±0.11	6.67±0.11	6.67±0.11
	PAE	4.15±0.06	3.96±0.04	3.96±0.04	3.96±0.04	3.96±0.04
	BE	4.0±0.9	2.8±0.5	2.8±0.5	2.8±0.5	2.8±0.5
1	FAE	13.8±1.4	10.1±0.4	10.1±0.4	10.1±0.4	10.1±0.4
	PAE	4.17±0.06	4.00±0.04	4.00±0.04	4.00±0.04	4.00±0.04
	BE	4.1±0.9	3.0±0.6	3.0±0.6	2.9±0.5	2.9±0.5
10	FAE	74.9±17.0	17.1±1.3	17.1±1.3	17.0±1.3	17.0±1.3
	PAE	4.18±0.06	4.00±0.04	4.00±0.04	4.00±0.04	4.00±0.04
	BE	6.4±2.7	3.1±0.6	3.1±0.6	3.0±0.5	3.0±0.5

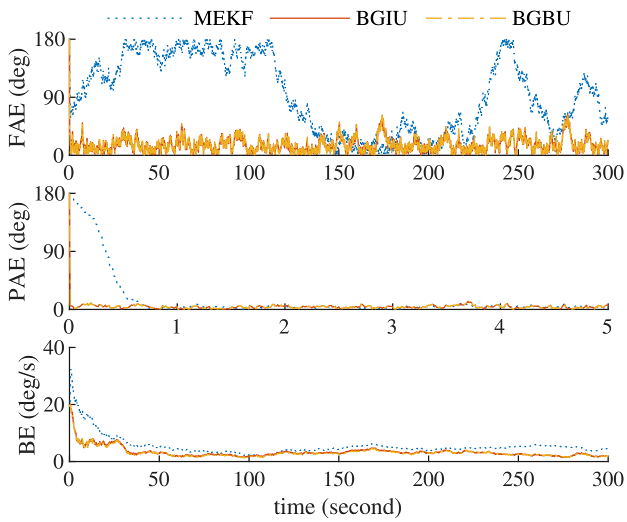

Fig. 1. Full attitude, partial attitude, and bias errors for MEKF, BGIU, and BGBU in a single simulation with $\sigma_2^2 = 10$. For partial attitude errors, only the first five seconds are shown to emphasize the initial transient responses of the filters. The MFGIU, MFGBU filters have similar error trajectories as the BG filters, and is thus omitted for better readability.

greater than MFG when σ_2^2 is relatively large.

VI. Conclusions

In this paper, we introduced the Bingham–Gaussian distribution defined on $\mathbb{S}^3 \times \mathbb{R}^n$, which is able to model the correlation between three dimensional attitude and Euclidean random variables of an arbitrary dimension. BG is equivalent to the matrix Fisher–Gaussian distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$, and thus they share the same properties. An unscented attitude filter is proposed using BG, and simulations demonstrate that it exhibits improved accuracy over the conventional MEKF especially when the uncertainties are large.

References

[1] E. J. Lefferts, F. L. Markley, and M. D. Shuster, "Kalman filtering for spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, vol. 5, no. 5, pp. 417–429, 1982.

[2] F. L. Markley, "Attitude error representations for Kalman filtering," *Journal of guidance, control, and dynamics*, vol. 26, no. 2, pp. 311–317, 2003.

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

[4] J. Glover and L. P. Kaelbling, "Tracking the spin on a ping pong ball with the quaternion Bingham filter," in *IEEE international conference on robotics and automation*, 2014, pp. 4133–4140.

[5] G. Kurz, I. Gilitschenski, S. Julier, and U. D. Hanebeck, "Recursive Bingham filter for directional estimation involving 180 degree symmetry," *Journal of Advances in Information Fusion*, vol. 9, no. 2, pp. 90–105, 2014.

[6] M. J. Prentice, "Orientation statistics without parametric assumptions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 48, no. 2, pp. 214–222, 1986.

[7] T. Lee, "Bayesian attitude estimation with the matrix Fisher distribution on SO(3)," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3377–3392, 2018.

[8] K. Li, F. Pfaff, and U. D. Hanebeck, "Geometry-driven stochastic modeling of SE(3) states based on dual quaternion representation," in *IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, 2019, pp. 254–260.

[9] S. Bultmann, K. Li, and U. D. Hanebeck, "Stereo visual SLAM based on unscented dual quaternion filtering," in *International Conference on Information Fusion (FUSION)*, 2019, pp. 1–8.

[10] F. L. Markley, "Attitude filtering on SO(3)," *The Journal of the Astronautical Sciences*, vol. 54, no. 3-4, pp. 391–413, 2006.

[11] J. E. Darling and K. J. DeMars, "Uncertainty propagation of correlated quaternion and Euclidean states using the Gauss-Bingham density," *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 1–20, 2016.

[12] W. Wang and T. Lee, "Matrix Fisher–Gaussian distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$ for attitude estimation with a gyro bias," in *American Control Conference*, 2020, pp. 4429–4434.

[13] ——, "Matrix Fisher-Gaussian distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$ and Bayesian attitude estimation," *IEEE Transactions on Automatic Control*, pp. 1–1, 2021.

[14] K. Mardia and T. Sutton, "A model for cylindrical variables with applications," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 40, no. 2, pp. 229–233, 1978.

[15] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Unscented orientation estimation based on the Bingham distribution," *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, 2015.

[16] J. Sola, "Quaternion kinematics for the error-state Kalman filter," *arXiv preprint arXiv:1711.02508*, 2017.

[17] C. Bingham, "An antipodally symmetric distribution on the sphere," *The Annals of Statistics*, pp. 1201–1225, 1974.

[18] T. D. Downs, "Orientation statistics," *Biometrika*, vol. 59, no. 3, pp. 665–676, 1972.

[19] C. Khatri and K. V. Mardia, "The von Mises–Fisher matrix distribution in orientation statistics," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 95–106, 1977.

[20] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, "Efficient Bingham filtering based on saddlepoint approximations," in *International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, 2014, pp. 1–7.