Matrix Fisher–Gaussian Distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$ and Bayesian Attitude Estimation

Weixin Wang, and Taeyoung Lee

***Abstract*—In this paper, a new probability distribution, referred to as the matrix Fisher–Gaussian distribution, is proposed on the product manifold of three-dimensional special orthogonal group and Euclidean space. It is constructed by conditioning a multivariate Gaussian distribution from the ambient Euclidean space into the manifold, while imposing a certain geometric constraint on the correlation term to avoid over-parameterization. The unique feature is that it may represent large uncertainties in attitudes, linear variables of an arbitrary dimension, and angular–linear correlations between them in a global fashion without singularities. Various stochastic properties and an approximate maximum likelihood estimator are developed. Furthermore, two methods are developed to propagate uncertainties through a stochastic differential equation representing attitude kinematics. Based on these, a Bayesian estimator is proposed to estimate the attitude and time-varying gyro bias concurrently. Numerical studies indicate that the proposed estimator provides more accurate estimates against the multiplicative extended Kalman filter and unscented Kalman filter for challenging cases.**

***Index Terms*—Matrix Fisher–Gaussian distribution, Matrix Fisher distribution, attitude estimation, special orthogonal group, correlation**

I. Introduction

Attitude estimation using a gyroscope with time varying biases has been studied since 1960s [1]. The unique challenge in attitude estimation is that the attitude evolves on the nonlinear manifold, referred to as the three-dimensional special orthogonal group $\mathrm{SO}(3)$, which cannot be globally identified with the Euclidean space of the same dimension. Any three-dimensional parameterization of $\mathrm{SO}(3)$ leads to singularities [2].

One of the key milestones in attitude estimation is the development of multiplicative extended Kalman filter (MEKF) [3]–[5], where the mean attitude is represented by a quaternion, and the uncertainty distribution about the mean attitude is described by a three-dimensional attitude parameter following a Gaussian distribution. MEKF typically relies on the assumption that the uncertainty distribution is highly concentrated about the mean attitude, and it adopts the framework of extended Kalman filter to estimate the attitude and bias simultaneously through the linearized attitude kinematics equation. MEKF has been a de facto algorithm for attitude estimation in various disciplines, including aerospace engineering [1], robotics [6], [7], etc. Besides MEKF, deterministic attitude observers on $\mathrm{SO}(3)$ have been presented in [8], [9].

The approach of formulating uncertainties around the mean has also been utilized in estimation problems on other manifolds. For example, in [10], [11] uncertainty propagation and measurement update have been developed on the special Euclidean group by formulating a covariance in the tangent space at the mean to quantify uncertainties. The same approach has also been generalized to matrix Lie groups in [12], [13], and to Riemannian manifolds in [14], [15].

Despite its success, MEKF is limited by the fundamental issues inherited from local parameterizations and linearization. First, only the first order term of the attitude error is considered in the linearized kinematics equation. Therefore the error attitude has to be small enough for the linearization to be valid. Second, the Gaussian distribution of local coordinates does not properly represent attitude uncertainties on the compact manifold. For example, suppose that the exponential coordinate is employed. In this case, the same attitude corresponds to infinitely many different rotation vectors with the difference in length by multiples of 2π. Thus the probability density function describing the uncertainty of the rotation vector should be wrapped, i.e., the densities at rotation vectors representing the same attitude need to be added up. This is ignored in MEKF, which is problematic especially when the attitude uncertainty is widely dispersed. In short, MEKF is not suitable for attitude estimation with large uncertainties.

In order to overcome the shortcomings of MEKF, there have been efforts to construct attitude estimators with probability distributions defined directly on $\mathrm{SO}(3)$. In [16], a probability density function on $\mathrm{SO}(3) \times \mathbb{R}^n$ is inspired by harmonic analysis, and the Fokker-Planck equation describing the evolution of the density is solved in the ambient Euclidean space to construct an attitude filter on $\mathrm{SO}(3)$.

On the other hand, rotational random variables and matrices in compact manifolds have been studied in directional statistics [17], which has provided various models for probability distributions. In fact, the attitude part of the density function in [16] is exactly the matrix Fisher distribution on $\mathrm{SO}(3)$ [18], [19]. And the matrix Fisher distribution has been shown to be equivalent to the Bingham distribution defined on the unit-sphere for quaternions with antipodal points identified [20], [21]. Utilizing these, attitude estimators have been developed by using the Bingham distribution in [22], [23], and by using the matrix Fisher distribution in [24]. However, these are based on probability distributions on $\mathrm{SO}(3)$, so the gyro bias cannot be estimated concurrently with the attitude.

W. Wang is a doctoral candidate at the Department of Mechanical and Aerospace Engineering, The George Washington University, Washington DC, USA. wwang442@gwu.edu

T. Lee is a professor at the Department of Mechanical and Aerospace Engineering, The George Washington University, Washington DC, USA. tylee@gwu.edu

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

To address this, a probability density function on the product space of $\mathrm{SO}(3) \times \mathbb{R}^n$ should be utilized such that the angular-linear correlation between attitudes and gyro biases can be modeled. There have been several efforts to formulate angular-linear correlations. For example, the distribution on $\mathbb{S}^1 \times \mathbb{R}^1$, where a circular variable is coupled with a linear variable, has been studied in [25]–[28]. In [29], the Bingham distribution on the unit-sphere $\mathbb{S}^r$ is extended to the Gauss-Bingham distribution on $\mathbb{S}^r \times \mathbb{R}^n$, which can be interpreted as a distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$. Nevertheless, the angular-linear correlation is introduced by making the orientation parameter of the Bingham distribution dependent on the linear variable, which lacks a geometric interpretation. One potential issue of this approach is that the orientation parameter lies in $\mathrm{SO}(r+1)$ which needs to be *parameterized* again, and this brings back the issue of local parameterizations. More importantly, it does not have a closed form solution for maximum likelihood estimation (MLE), which is essential in Bayesian estimators assuming that uncertainties are modeled by the Gauss-Bingham distribution always. This may cause difficulties for real-time implementations in practice.

In [30], we have introduced a new distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$, referred to as the matrix Fisher–Gaussian (MFG) distribution. It is constructed by conditioning a (9+n)-variate Gaussian distribution from $\mathbb{R}^{9+n}$ into $\mathrm{SO}(3) \times \mathbb{R}^n$, which leads to the matrix Fisher distribution for the attitude. The correlation term between the three-dimensional attitude and n-dimensional linear variable is inherited from the correlation between $\mathbb{R}^9$ and $\mathbb{R}^n$ before conditioning, which is constrained to be non-zero only in the tangent space of the mean attitude. The desirable feature is that MFG provides an intuitive geometric interpretation for the angular-linear correlation, which is represented by $3n$ parameters, thereby avoiding a potential over-parameterization in [16] relying on $9n$ parameters. Moreover, although the MLE of MFG cannot be solved analytically, there is a closed form approximation by first solving the marginal MLE for the attitude part. This reduces the computational load significantly compared with [29] that requires numerical optimizations. Furthermore, MFG can be approximated by a (3+n)-variate Gaussian distribution if its attitude part is concentrated. Therefore, it can be interpreted as a generalization of the common approach relying on the Gaussian distribution of three-dimensional attitude parameters.

Based on the preliminary result in [30], in this paper we propose a new form of MFG that is more suitable for attitude and gyro bias estimation. The difference stems from how the angular-linear correlation is formulated: in [30], the correlation is interpreted with the axis of rotation resolved in the inertial frame as the linear variable is varied; and in this paper, the axis of rotation is resolved in the body-fixed frame. Therefore the distribution presented in [30] is denoted by MFGI, and the new distribution proposed in this paper is denoted by MFGB. It is shown that uncertainties characterized by MFGI are distinct from MFGB when the attitude distribution is non-isotropic. In particular, it is illustrated that MFGB is more appropriate for attitude and gyro bias estimation, as the gyro bias is resolved in the body-fixed frame.

Based on the proposed MFG distributions, we design an intrinsic Bayesian estimator to estimate attitude and gyro bias concurrently. The Bayesian estimator is composed of two parts, namely uncertainty propagation and measurement update. In [30], we proposed an unscented transform of MFG to propagate uncertainties with selected sigma points. In this paper, we improve the corresponding computational efficiency by eliminating the need to sample sigma points from the bias random walk noise. More importantly, we propose another method for uncertainty propagation, where selected moments of MFG required for MLE are calculated with analytical expressions. Next, the measurement update is accomplished by matching the posterior distribution obtained using Bayes' rule to a posterior MFG using MLE. Finally, numerical simulations demonstrate the advantage of the proposed estimator over the conventional MEKF and unscented Kalman filter (UKF) in convergence and accuracy.

II. Matrix Fisher Distribution

A. Notations and Facts

The three-dimensional special orthogonal group $\mathrm{SO}(3)$ is

$$\mathrm{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \big| \; RR^T = I_{3\times 3}, \; \det(R) = 1\},$$

which is commonly used to represent the attitude of a rigid body in the right-handed frame. Its Lie algebra, denoted by $\mathfrak{so}(3)$, is the tangent space of $\mathrm{SO}(3)$ at $I_{3\times 3}$, given by

$$\mathfrak{so}(3) = \{A \in \mathbb{R}^{3\times 3} \; \big| \; A = -A^T\}.$$

The Lie algebra $\mathfrak{so}(3)$ can be identified with $\mathbb{R}^3$ by the *hat* $\wedge$ map and the *vee* $\vee$ map defined as follows.

$$\mathfrak{so}(3) \ni \begin{bmatrix} 0 & -\Omega_z & \Omega_y \\ \Omega_z & 0 & -\Omega_x \\ -\Omega_y & \Omega_x & 0 \end{bmatrix} \begin{array}{c} \xrightarrow{\text{vee } \vee} \\ \xleftarrow{\text{hat } \wedge} \end{array} \begin{bmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{bmatrix} \in \mathbb{R}^3.$$

In this paper, e_i is used to represent the i-th standard basis vector of $\mathbb{R}^n$, i.e., the i-th column of $I_{n\times n}$. Any 3-by-3 diagonal matrix with the diagonal entries of 1 or -1 is denoted by $\mathcal{D} \in \mathbb{R}^{3\times 3}$. In particular, when it is augmented with subscripts, the subscripts correspond to the diagonal indices for 1. For example, $\mathcal{D}_1 = \mathrm{diag}(1,-1,-1)$. Let the set of circular shifts of $\{1,2,3\}$ be $\mathcal{I} = \{(1,2,3),(2,3,1),(3,1,2)\}$. The n-dimensional unit sphere is denoted by $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \big| \; x^T x = 1\}$. The trace of a matrix is denoted by $\mathrm{tr}(\cdot)$, and the function $\exp(\mathrm{tr}(\cdot))$ is abbreviated as $\mathrm{etr}(\cdot)$. Also, the operator $\mathrm{vec}(\cdot)$ is used to concatenate the columns of a matrix into a vector, and the Kronecker product is denoted by $\otimes$.

Throughout this paper, the following identities regarding the trace of a matrix and the hat map will be repeatedly used. For any $A, B \in \mathbb{R}^{3\times 3}$,

$$\mathrm{tr}(AB) = \mathrm{tr}(BA) = \mathrm{tr}\big(B^T A^T\big). \tag{1}$$

For any $R \in \mathrm{SO}(3)$, $A \in \mathbb{R}^{3\times 3}$ and $x \in \mathbb{R}^3$,

$$\widehat{Rx} = R\hat{x}R^T, \tag{2}$$

$$\hat{x}^2 = xx^T - x^T x I_{3\times 3}, \tag{3}$$

$$(\hat{x}A + A^T\hat{x})^\vee = (\mathrm{tr}(A)\, I_{3\times 3} - A)x, \tag{4}$$

$$\mathrm{tr}(\hat{x}A) = x^T(A^T - A)^\vee. \tag{5}$$

B. Matrix Fisher Distribution on $\mathrm{SO}(3)$

The matrix Fisher distribution of a random matrix $R \in \mathrm{SO}(3)$ is defined by the following density function

$$p(R;F) = \frac{1}{c(F)}\mathrm{etr}\big(F^T R\big)\,, \tag{6}$$

with respect to the uniform distribution on $\mathrm{SO}(3)$. The matrix $F \in \mathbb{R}^{3\times 3}$ is the parameter describing the shape of the distribution, and $c(F) \in \mathbb{R}$ is the normalizing constant. This is denoted by $R \sim \mathcal{M}(F)$. From (6), it is straightforward to show the matrix Fisher distribution is closed under rotations: if $R \sim \mathcal{M}(F)$ then $RA \sim \mathcal{M}(FA)$ and $AR \sim \mathcal{M}(AF)$ for any fixed $A \in \mathrm{SO}(3)$.

Various properties of a matrix Fisher distribution can be accessed through the *proper* singular value decomposition (SVD) of F [19], [24].

Definition 1. *Let the singular value decomposition of F be given by $F = U'S'V'^T$, where $S' \in \mathbb{R}^{3\times 3}$ is a diagonal matrix composed of the singular values $s'_1 \geq s'_2 \geq s'_3 \geq 0$ of F, and $U', V' \in \mathrm{O}(3)$ are orthogonal matrices. The proper singular value decomposition of F is*

$$F = USV^T, \tag{7}$$

where the rotation matrices $U, V \in \mathrm{SO}(3)$, and the diagonal matrix $S \in \mathbb{R}^{3\times 3}$ are defined as

$$\begin{aligned} U &= U'\mathrm{diag}(1,1,\det[U']),\\ S &= \mathrm{diag}(s_1,s_2,s_3) = \mathrm{diag}(s'_1,s'_2,\det[U'V']s'_3),\\ V &= V'\mathrm{diag}(1,1,\det[V']). \end{aligned} \tag{8}$$

The motivation of the above proper SVD is to ensure $U, V \in \mathrm{SO}(3)$, while allowing s_3 to be negative. It should be noted that the usual convention $F = U'S'V'$ with a positive semidefinite S' and $U', V' \in \mathrm{O}(3)$ can also be used to define MFG [31].

The normalizing constant $c(F)$ depends only on the proper singular values of F, i.e.,

$$c(F) = c(S) = \int_{Q\in\mathrm{SO}(3)} \mathrm{etr}\big(SQ^T\big)\,\mathrm{d}Q, \tag{9}$$

where $\mathrm{d}Q$ is the bi-invariant Haar measure for $\mathrm{SO}(3)$, normalized such that $\int_{Q\in\mathrm{SO}(3)}\mathrm{d}Q = 1$. The first order moment of R is given by

$$\mathrm{E}[R] = UDV^T = U\mathrm{diag}(d_1,d_2,d_3)\,V^T, \tag{10}$$

where $d_i \in \mathbb{R}$ for $i \in \{1,2,3\}$ is

$$d_i = \frac{1}{c(S)}\frac{\partial c(S)}{\partial s_i}. \tag{11}$$

However, $\mathrm{E}[R] \in \mathbb{R}^{3\times 3}$ does not belong to $\mathrm{SO}(3)$ in general. Instead, the *mean attitude* of R is usually interpreted as $UV^T \triangleq M \in \mathrm{SO}(3)$, which maximizes the density (6), and also minimizes the Frobenius mean squared error.

Similar to the Gaussian distribution, the matrix Fisher distribution has three principal axes, given by the columns of U resolved in the standard coordinates of $\mathbb{R}^3$, or equivalently the columns of V when resolved in the coordinates specified by the columns of M. The attitude rotated from the mean attitude M about the i-th principal axis for an angle θ, i.e., $R(\theta) = \exp(\theta\widehat{Ue_i})M = M\exp(\theta\widehat{Ve_i})$, has the density

$$p(R(\theta);F) = \frac{e^{s_i}}{c(S)}\exp((s_j+s_k)\cos\theta), \tag{12}$$

where $(i,j,k) \in \mathcal{I}$. This corresponds to the von Mises distribution [17] for θ defined on the unit circle $\mathbb{S}^1$, and its concentration around the mean angle $\theta = 0$ is specified by $s_j + s_k$. An interesting property is that when $s_j + s_k$ is sufficiently large, it is approximated by the Gaussian density with the zero mean and variance $1/(s_j+s_k)$. This implies that the distribution of R can be approximated by a three-dimensional Gaussian distribution when R is concentrated around its mean attitude M [32].

III. Matrix Fisher–Gaussian Distribution

In this section, we first present a new form of matrix Fisher–Gaussian distribution (MFGB) on $\mathrm{SO}(3)\times\mathbb{R}^n$. In contrast to the matrix Fisher–Gaussian distribution (MFGI) in [30], [31], the angular–linear correlation is formulated such that varying the linear random variable causes the attitude distribution to be rotated by an axis resolved in the body-fixed frame. As such, it is more suitable for attitude estimation where the angular velocity is measured by a gyroscope fixed to the body. Several stochastic properties of MFGB are presented, and the difference between MFGB and MFGI is discussed.

Definition 2 (MFGB)**.** *The random variables $(R,x) \in \mathrm{SO}(3)\times\mathbb{R}^n$ follow the matrix Fisher–Gaussian distribution with parameters $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n\times n}$, $U, V \in \mathrm{SO}(3)$, $S = \mathrm{diag}(s_1,s_2,s_3) \in \mathbb{R}^{3\times 3}$ with $s_1 \geq s_2 \geq |s_3| \geq 0$ and $P \in \mathbb{R}^{n\times 3}$, if it has the following density function:*

$$\begin{aligned} p(R,x;\mu,\Sigma,V,S,U,P) = \frac{1}{c(S)\sqrt{(2\pi)^n\det(\Sigma_c)}}\times\\ \exp\big\{-\tfrac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)\big\}\,\mathrm{etr}\big\{FR^T\big\}\,, \end{aligned} \tag{13}$$

where $\mu_c \in \mathbb{R}^n$ is given by

$$\mu_c = \mu + P\nu_R, \tag{14}$$

with

$$\nu_R = (SQ - Q^TS)^\vee, \tag{15}$$

for $Q = U^TRV$, and $0 \prec \Sigma_c \in \mathbb{R}^{n\times n}$ is defined as

$$\Sigma_c = \Sigma - P(\mathrm{tr}(S)\,I_{3\times 3} - S)P^T, \tag{16}$$

Also, $F = USV^T \in \mathbb{R}^{3\times 3}$, and $c(S) \in \mathbb{R}$ is the normalizing constant of the corresponding matrix Fisher distribution. This distribution is denoted by $\mathcal{MG}(\mu,\Sigma,P,U,S,V)$.

The probability density function of MFG given by (13) 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. From its definition, it is straightforward to see that the marginal distribution of R is a matrix Fisher distribution with parameter F, and the distribution of x conditioned by R is Gaussian with $x|R \sim \mathcal{N}(\mu_c(R),\Sigma_c)$.

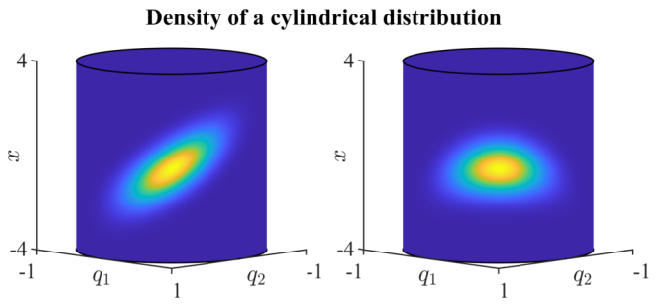

Fig. 1. The illustration of the cylindrical distribution with the mean angle $\pi/4$ [25]. In the left figure, the correlation between the linear variable $x \in \mathbb{R}^1$ and the angular variable $q = (q_1, q_2) \in \mathbb{S}^1$ is only nonzero along the tangent direction of $\mathbb{S}^1$ at $\pi/4$, and the distribution of q conditioned by x varies as x is altered, and vice versa. However in the right figure, the correlation is specified along the radial direction of $\mathbb{S}^1$, and there is no clear correlation between x and q. These illustrate the correlation along the tangent direction captures what we expect as a *linear* correlation between two random variables.

When $P = 0$, x and R are independent, and they further satisfy $x \sim \mathcal{N}(\mu, \Sigma)$ and $R \sim \mathcal{M}(F)$. As such, the attitude-linear correlation is specified by the $3n$ elements of P. More specifically, the correlation between x and R is caused by the fact that the conditional mean $\mu_c(R)$ of x is dependent on R. When conditioned by R, it is shifted by $P\nu_R$ in (14), where ν_R in (15) indicates how R deviates from the mean attitude UV^T. For example, when $R = UV^T$, or equivalently when $Q = I_{3\times3}$, we have $\nu_R = 0$ and $\mu_c = \mu$.

Remark 1. *Definition 2 (MFGB) is the same as Definition 1 (MFGI) in [30] except the expression of ν_R in* (15)*. In [30], it is defined as $\nu_R = (QS - SQ^T)^\vee$. As discussed later in Section III-F, this difference leads to distinct interpretations for the angular–linear correlation. Due to the similarity of MFGB and MFGI, they share several stochastic properties. Throughout this paper, theorems and propositions exclusively applied to MFGB are labeled with '(MFGB)', and those shared by both MFGB and MFGI with respect to their own definition of ν_R are labeled with '(MFG)'. The counterpart of this section that is devoted to MFGI is available in [31].*

A. Geometric Construction and Interpretation

The construction of MFG is motivated by [25], where a distribution on the cylinder is constructed by conditioning a three-dimensional Gaussian distribution from $\mathbb{R}^3$ into $\mathbb{S}^1 \times \mathbb{R}^1$. However, in [25] two parameters are used to quantify the angular-linear correlation, despite both of the angular part $\mathbb{S}^1$ and the linear part $\mathbb{R}^1$ are one-dimensional. This is because the correlation is formulated between the ambient space of $\mathbb{S}^1$, namely $\mathbb{R}^2$, and $\mathbb{R}^1$.

In general, the correlation between two random variables describes how one random variable is linearly varied from its mean, when the other random variable is perturbed from its own mean. As such, we wish that the correlation between $\mathbb{S}^1$ and $\mathbb{R}^1$ is described by a single parameter. The key observation is that the unit-vector in $\mathbb{S}^1$ cannot be varied from the mean exclusively along the radial direction due to the unit-length constraint. Further, when it is rotated, the variation along the radial direction is constrained by the variation along the tangential direction. Thus, the correlation between $\mathbb{S}^1$ and $\mathbb{R}^1$ can be described by the correlation between the linear variable and the *tangential* direction of the angular variable at the mean angle. This is illustrated in Fig. 1.

Similarly, if the correlation between $R \in \mathrm{SO}(3)$ and $x \in \mathbb{R}^n$ is defined as in [25], we would need $9n$ parameters. Instead, the ambient space of $\mathrm{SO}(3)$, namely $\mathbb{R}^{3\times3}$, is decomposed into two parts, namely the tangent space at the mean attitude and its orthogonal complement. The correlation with linear variables along the orthogonal complement is set to zero, thereby reducing the number of free parameters into $3n$. More specifically, a Gaussian distribution is formulated on $\mathbb{R}^9 \times \mathbb{R}^n$ such that the correlation along the orthogonal complement to the tangent space at the mean attitude is annihilated before conditioning onto $\mathrm{SO}(3) \times \mathbb{R}^n$.

Depending on how $\mathrm{SO}(3)$ is embedded in $\mathbb{R}^9$, two forms of MFG can be constructed. In [30], $\mathrm{SO}(3)$ is embedded by concatenating row vectors of the rotation matrix, whereas in this paper, it is completed by concatenating column vectors. The geometric construction of MFGB is summarized below.

Theorem 1 (MFGB)**.** *Consider the parameters $(\mu, \Sigma, V, S, U, P)$ defining MFGB as introduced in Definition 2. Let $M = UV^T \in \mathrm{SO}(3)$, $K = USU^T \in \mathbb{R}^{3\times3}$, $\mu_R = \mathrm{vec}(M) \in \mathbb{R}^9$, and $\Sigma_R^{-1} = I_{3\times3} \otimes K \in \mathbb{R}^{9\times9}$.*

Also let $t_i = \mathrm{vec}[M\widehat{Ve_i}]$ for $i \in \{1, 2, 3\}$ be the basis for the tangent space of $\mathrm{SO}(3)$ at M embedded in $\mathbb{R}^9$. And let $\{t_4, \dots, t_9\}$ be the orthogonal complement of $\{t_1, t_2, t_3\}$ in $\mathbb{R}^9$. Define $T = [t_1, \dots, t_9]^T \in \mathbb{R}^{9\times9}$, and $P_R = [P, 0_{n\times6}]T \in \mathbb{R}^{n\times9}$.

Suppose $(x_R, x) \in \mathbb{R}^{9+n}$ follows the Gaussian distribution

$$\begin{bmatrix} x_R \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_R \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_R & P_R^T \\ P_R & \Sigma \end{bmatrix} \right). \tag{17}$$

Then for $R = \mathrm{vec}^{-1}(x_R) \in \mathbb{R}^{3\times3}$, the distribution of (R, x) conditioned on $R^TR = I_{3\times3}$ and $\det(R) = 1$ is $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$.

Proof. The joint density of (x_R, x) can be written in the form of a conditional-marginal density as

$$\begin{aligned} p(x_R, x) = {} & \tfrac{1}{c}\exp\left\{-\tfrac{1}{2}(x - \mu_c)^T\Sigma_c^{-1}(x - \mu_c)\right\} \\ & \times\exp\left\{-\tfrac{1}{2}(x_R - \mu_R)^T\Sigma_R^{-1}(x_R - \mu_R)\right\}, \end{aligned} \tag{18}$$

where c is the normalizing constant, $\mu_c = \mu + P_R\Sigma_R^{-1}(x_R - \mu_R) \in \mathbb{R}^n$, and $\Sigma_c = \Sigma - P_R\Sigma_R^{-1}P_R^T \in \mathbb{R}^{n\times n}$.

The exponent of the last term of (18) can be written as

$$\begin{aligned} & -\tfrac{1}{2}(x_R - \mu_R)^T\Sigma_R^{-1}(x_R - \mu_R) \\ = {} & -\tfrac{1}{2}\mathrm{tr}\big(K(\mathrm{vec}^{-1}(x_R) - M)(\mathrm{vec}^{-1}(x_R) - M)^T\big) \\ = {} & \mathrm{tr}\big(KM\mathrm{vec}^{-1}(x_R)^T\big) + C = \mathrm{tr}\big(FR^T\big) + C, \end{aligned}$$

where C is a constant independent of x_R or x, because $\mathrm{vec}^{-1}(x_R)\mathrm{vec}^{-1}(x_R)^T = I_{3\times3}$ when conditioned on $R^TR = I_{3\times3}$. So the second term on the right hand side of (18) reduces to the matrix Fisher density after conditioning.

Next, for the first term on the right hand side of (18), the second part of Σ_c is

$$P_R\Sigma_R^{-1}P_R^T = P[t_1, t_2, t_3]^T\Sigma_R^{-1}[t_1, t_2, t_3]P^T \triangleq P\tilde{\Sigma}_R^{-1}P^T,$$

for some $\tilde{\Sigma}_R^{-1} \in \mathbb{R}^{3\times3}$. Let $t_i \in \mathbb{R}^9$ be equally split into three vectors $\{t_{i1}, t_{i2}, t_{i3}\}$, then the i,j-th entry of $\tilde{\Sigma}_R^{-1}$ can be written as

$$(\tilde{\Sigma}_R^{-1})_{ij} = \sum_{k=1}^{3} t_{ik}^T K t_{jk} = \mathrm{tr}\big([t_{i1}, t_{i2}, t_{i3}]^T K [t_{j1}, t_{j2}, t_{j3}]\big)$$
$$= \mathrm{tr}(\widehat{Ve_i}^T M^T K M \widehat{Ve_j}) = \mathrm{tr}(S\hat{e}_j\hat{e}_i^T),$$

which implies $\tilde{\Sigma}_R^{-1} = \mathrm{tr}(S)\, I_{3\times3} - S$. Thus Σ_c has the same expression as in (16). Besides, the second part of μ_c is

$$P_R \Sigma_R^{-1}(x_R - \mu_R) = P[t_1, t_2, t_3]^T \Sigma_R^{-1} \mathrm{vec}(R - M),$$

and

$$t_i^T \Sigma_R^{-1} \mathrm{vec}(R - M) = \mathrm{tr}(\widehat{Ve_i}^T M^T K (R - M))$$
$$= \mathrm{tr}(S\hat{e}_i - SQ\hat{e}_i) = e_i^T (SQ - Q^T S)^\vee,$$

where $Q = U^T R V$. This shows μ_c also has the same expression as in (14). In conclusion, the density function (18) is the same as (13) after conditioning on $RR^T = I_{3\times3}$ and $\det(R) = 1$. □

In (17), the covariance between $x_R = \mathrm{vec}(R)$ and x is $P_R = [P, 0_{n\times6}]T$, i.e., it is non-zero only along $\{t_1, t_2, t_3\}$, which is a basis for the tangent space of $\mathrm{SO}(3)$ at UV^T. The basis is chosen as the principal axes of the matrix Fisher part, so that the correlation P is expressed with respect to the principal axes, which simplifies the corresponding mathematical analysis and provides geometric interpretations. Also, there are only $3n$ parameters required to quantify the linear correlation between the three-dimensional attitude and the n-dimensional linear variable, instead of $9n$.

This construction of MFG leads to a clear geometric interpretation of the correlation parameter P: if $P_{ij} > 0$, as x_i becomes increased, the distribution of R conditioned on x rotates about the j-th principal axis of the matrix Fisher part, and vice versa. See Fig. 2 for a simple example.

B. Equivalent Distributions

As presented in Theorem 1 and (15), the definition of MFG relies on the proper SVD of F, since the correlation is defined along the principal axes given by U and V. There are two uniqueness issues associated with the definition of SVD. The trivial one is that when U and V undergo simultaneous sign changes of two columns, they are still the left and right proper singular vectors of F. The other issue is more interesting: when F has repeated singular values, the corresponding proper singular vectors in U and V are only unique up to a rotation. This means the same MFG can be potentially parameterized with different proper SVDs. However, as the next proposition shows, MFG can be uniquely parameterized by the intermediate parameters F, μ_c and Σ_c.

Proposition 1 (MFG)**.** *Two matrix Fisher–Gaussian distributions, namely $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$ and $\mathcal{MG}(\tilde{\mu}, \tilde{\Sigma}, \tilde{P}, \tilde{U}, \tilde{S}, \tilde{V})$ are equivalent if and only if $F = \tilde{F}$, $\mu_c = \tilde{\mu}_c$ for all $R \in \mathrm{SO}(3)$, and $\Sigma_c = \tilde{\Sigma}_c$.*

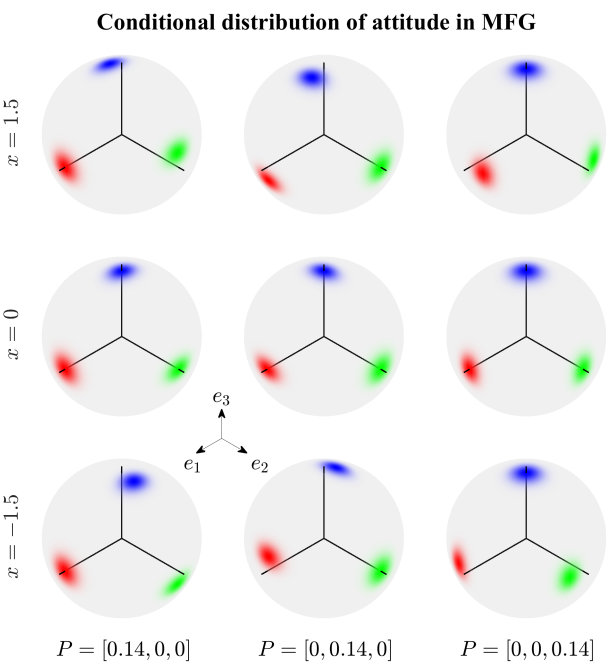

Fig. 2. Visualization of the attitude-linear correlation for $(R, x) \in \mathrm{SO}(3) \times \mathbb{R}^1$: the density for R conditioned on x is illustrated on the unit-sphere for three correlation matrices P. More specifically, the marginal distribution for each column of R is shown on the unit sphere as red, green and blue shades. The parameters are $n = 1$, $\mu = 0$, $\Sigma = 1$, $U = V = I_{3\times3}$, $S = 10 I_{3\times3}$. For each P, three conditioning values of x are considered. The first column is for $P = [0.14, 0, 0]$. When x is increased from -1.5 to 1.5, the conditional distribution for R is rotated about the e_1 axis. Similarly, in the second column for $P = [0, 0.14, 0]$, the variation of x is correlated with rotating the distribution of R along the e_2 axis. The third column shows rotations about e_3 due to the correlation.

The proof of this proposition requires the following lemma, which states that a matrix Fisher distribution is uniquely parameterized by its parameter F.

Lemma 1. *Two matrix Fisher distributions, namely $\mathcal{M}(F)$ and $\mathcal{M}(\tilde{F})$ are equivalent if and only if $F = \tilde{F}$.*

Proof. It is trivial to show $F = \tilde{F}$ implies $\mathcal{M}(F) = \mathcal{M}(\tilde{F})$. Next, suppose $\mathcal{M}(F) = \mathcal{M}(\tilde{F})$, i.e., $\frac{\mathrm{etr}(FR^T)}{c(F)} = \frac{\mathrm{etr}(\tilde{F}R^T)}{c(\tilde{F})}$ for all $R \in \mathrm{SO}(3)$. Let $\Delta F = F - \tilde{F}$ and $\Delta c = \log(c(F)/c(\tilde{F}))$, the above equation is equivalent to

$$\mathrm{tr}\big(\Delta F R^T\big) - \Delta c = 0. \tag{19}$$

Substitute $R = I_{3\times3}$, $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ into (19) to obtain

$$\begin{aligned}
\Delta F_{11} + \Delta F_{22} + \Delta F_{33} - \Delta c = 0,\\
\Delta F_{11} - \Delta F_{22} - \Delta F_{33} - \Delta c = 0,\\
-\Delta F_{11} + \Delta F_{22} - \Delta F_{33} - \Delta c = 0,\\
-\Delta F_{11} - \Delta F_{22} + \Delta F_{33} - \Delta c = 0,
\end{aligned}$$

which shows $\Delta F_{11} = \Delta F_{22} = \Delta F_{33} = \Delta c = 0$. Next, substituting $R = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ and $\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ into (19) yields $\Delta F_{12} = 0$. Similarly, other entries of ΔF can be shown to be zeros. Therefore, $F = \tilde{F}$. □

Next, we present the proof for Proposition 1.

Proof for Proposition 1. The sufficiency directly follows from (13) since the density function is determined by F, μ_c and Σ_c. Next we show the necessity. Let p and $\tilde{p}$ be the density functions for the two sets of parameters respectively. Suppose $p(x,R) = \tilde{p}(x,R)$ for all $x \in \mathbb{R}^n$ and $R \in \mathrm{SO}(3)$. Since $\mathrm{SO}(3)$ is compact, and μ_c is continuous in R as seen from (14), $\|\mu_c\|$ has an upper bound. Therefore,

$$\lim_{\|x\|\to\infty} p(x,R) = \frac{\mathrm{etr}\big(FR^T\big)}{c(F)\sqrt{(2\pi)^n \det(\Sigma_c)}}.$$

Because $p(x,R) = \tilde{p}(x,R)$ for all $x \in \mathbb{R}^n$, the above equation implies $\frac{\mathrm{etr}\big(FR^T\big)}{c(F)\sqrt{(2\pi)^n \det(\Sigma_c)}} = \frac{\mathrm{etr}\big(\tilde{F}R^T\big)}{c(\tilde{F})\sqrt{(2\pi)^n \det(\tilde{\Sigma}_c)}}$, and by following the same argument in Lemma 1, $F = \tilde{F}$.

Next, since $F = \tilde{F}$ and $p(x,R) = \tilde{p}(x,R)$, we have $(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c) = (x-\tilde{\mu}_c)^T\tilde{\Sigma}_c^{-1}(x-\tilde{\mu}_c)$ for all $x \in \mathbb{R}^n$ and $R \in \mathrm{SO}(3)$. Substituting $x = \mu_c$ yields $\mu_c = \tilde{\mu}_c$ for all $R \in \mathrm{SO}(3)$, since $\tilde{\Sigma}_c$ is positive-definite. Also, substituting $x = \mu_c + e_i$ shows $(\Sigma_c^{-1})_{ii} = (\tilde{\Sigma}_c^{-1})_{ii}$, and similarly, substituting $x = \mu_c + e_i + e_j$ for $i \neq j$ yields $(\Sigma_c^{-1})_{ij} = (\tilde{\Sigma}_c^{-1})_{ij}$ since Σ_c^{-1} and $\tilde{\Sigma}_c^{-1}$ are symmetric. Therefore $\Sigma_c = \tilde{\Sigma}_c$. □

To further break down the equivalent conditions for MFG into the parameters U, S and V, a detailed analysis on the multiplicity of singular values S is required. In Appendix A, we provide a complete characterization of the equivalent parameters of MFGB. It shows that MFGB can be parameterized differently by rotating U, V, and P in a consistent way, if S has repeated values.

C. Gaussian Approximation

It has been shown that a matrix Fisher distribution can be approximated by a three-dimensional Gaussian distribution when it is highly concentrated, i.e., when $s_3 \gg 0$ [32]. The same property holds for MFG as stated in the following theorem.

Theorem 2 (MFG)**.** *Suppose $(x,R) \sim \mathcal{MG}(\mu,\Sigma,P,U,S,V)$. Let $R = U\exp(\hat{\eta})V^T$. If $s_3 \gg 0$, then (x,η) approximately follows a $(3+n)$-dimensional Gaussian distribution with*

$$\begin{bmatrix} x \\ \eta \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma & P \\ P^T & (\mathrm{tr}(S)\, I_{3\times 3} - S)^{-1} \end{bmatrix}\right). \tag{20}$$

Proof. For the matrix Fisher density part in (13), we have $\mathrm{etr}\big(F^TR\big) = \mathrm{etr}(S\exp(\hat{\eta}))$. Let $\Sigma' = (\mathrm{tr}(S)\, I_{3\times 3} - S)^{-1} = \mathrm{diag}\Big(\frac{1}{s_2+s_3}, \frac{1}{s_1+s_3}, \frac{1}{s_1+s_2}\Big) \in \mathbb{R}^{3\times 3}$, and let $\xi \in \mathbb{R}^3$ be defined as $\sqrt{\Sigma'}\xi = \eta$. Then,

$$\begin{aligned}
&\mathrm{etr}\big(F^TR\big) = \mathrm{etr}\Big(S\exp\Big((\sqrt{\Sigma'}\xi)^\wedge\Big)\Big) \\
&= \mathrm{etr}\Big(S\Big(I_{3\times 3} + (\sqrt{\Sigma'}\xi)^\wedge + \tfrac{1}{2}\big((\sqrt{\Sigma'}\xi)^\wedge\big)^2 + o(\Sigma')\Big)\Big) \\
&\propto \mathrm{etr}\Big(\tfrac{1}{2}S\big((\sqrt{\Sigma'}\xi)^\wedge\big)^2 + o(\Sigma')\Big) \\
&\approx \exp\left(-\tfrac{1}{2}\eta^T(\Sigma')^{-1}\eta\right).
\end{aligned}$$

Also, for the ν_R term in the conditional mean (14), we have

$$\begin{aligned}
\hat{\nu}_R =& S\exp\big((\sqrt{\Sigma'}\xi)^\wedge\big) - \exp\big((\sqrt{\Sigma'}\xi)^\wedge\big)^T S \\
=& S\big(I_{3\times 3} + (\sqrt{\Sigma'}\xi)^\wedge\big) - \big(I_{3\times 3} + (\sqrt{\Sigma'}\xi)^\wedge\big)^T S + o(\sqrt{\Sigma'}) \\
\approx& \big((\Sigma')^{-1}\eta\big)^\wedge,
\end{aligned}$$

which yields $\mu_c \approx \mu + P(\Sigma')^{-1}\eta$. Furthermore $\Sigma_c = \Sigma - P(\Sigma')^{-1}P^T$. Therefore, (13) is approximated by a $(3+n)$-dimensional Gaussian density written in the conditional-marginal form, which is identical to (20). □

D. Moments

Next, we present selected moments of MFG, which are used in the approximate MLE of the next subsection.

Theorem 3 (MFG)**.** *Suppose $(R,x) \sim \mathcal{MG}(\mu,\Sigma,P,U,S,V)$. Then,*

$$\mathrm{E}[R] = UDV^T, \tag{21}$$

where $D = \mathrm{diag}(d_1,d_2,d_3)$ is given in (11)*. Also,*

$$\mathrm{E}[x] = \mu, \tag{22}$$
$$\mathrm{E}[\nu_R] = 0, \tag{23}$$
$$\mathrm{E}[xx^T] = \Sigma_c + \mu\mu^T + P\mathrm{E}[\nu_R\nu_R^T]P^T, \tag{24}$$
$$\mathrm{E}[x\nu_R^T] = P\mathrm{E}[\nu_R\nu_R^T], \tag{25}$$

where $\mathrm{E}[\nu_R\nu_R^T] \in \mathbb{R}^{3\times 3}$ is a diagonal matrix with the i-th diagonal element given by

$$(\mathrm{E}[\nu_R\nu_R^T])_{ii} = (s_j^2 + s_k^2)\mathrm{E}[Q_{jk}^2] - 2s_js_k\mathrm{E}[Q_{jk}Q_{kj}], \tag{26}$$

for $(i,j,k) \in \mathcal{I}$. The explicit expressions for $\mathrm{E}[Q_{jk}^2]$ and $\mathrm{E}[Q_{jk}Q_{kj}]$ can be found in [33].

Proof. Equation (21) follows immediately from the fact that the marginal distribution of R is a matrix Fisher distribution with parameter F. Next, for (23), we have:

$$\mathrm{E}[\nu_R] = \mathrm{E}[SQ - Q^TS]^\vee = (SD - D^TS)^\vee = 0. \tag{27}$$

Also, for (24), we can integrate xx^T directly and get

$$\begin{aligned}
\mathrm{E}[xx^T] &= \int_{\mathrm{SO}(3)}\int_{\mathbb{R}^n} xx^T p(R,x)\mathrm{d}x\mathrm{d}R \\
&= \frac{1}{c(F)}\int_{\mathrm{SO}(3)} [\Sigma_c + (\mu + P\nu_R)(\mu + P\nu_R)^T]\mathrm{etr}(FR^T)\mathrm{d}R \\
&= \Sigma_c + \mu\mu^T + P\mathrm{E}[\nu_R\nu_R^T]P^T.
\end{aligned} \tag{28}$$

The remaining (22), (25) and (26) can be derived similarly. □

E. Maximum Likelihood Estimation

Here we consider the maximum likelihood estimation (MLE) problem to construct an MFG from its samples. Given a set of samples $(R_i,x_i)_{i=1}^{N_s}$, the log-likelihood function of the parameters, after omitting some constants, is given by

$$\begin{aligned}
l =& -\log(c(S)) + \mathrm{tr}(F\bar{\mathrm{E}}[R]^T) - \tfrac{1}{2}\log(\det(\Sigma_c)) \\
&- \tfrac{1}{2}\bar{\mathrm{E}}[(x-\mu-P\nu_R)^T\Sigma_c^{-1}(x-\mu-P\nu_R)],
\end{aligned} \tag{29}$$

where $\bar{\mathrm{E}}[\cdot]$ represents the sample mean of a random variable. For example, $\bar{\mathrm{E}}[R] = \frac{1}{N_s}\sum_{i=1}^{N_s} R_i$.

As the log-likelihood function should be maximized jointly for the matrix Fisher part and the Gaussian part, it is challenging to obtain a closed form solution for (μ,Σ,V,S,U,P).

From the construction of MFG, this is comparable to the MLE of a $(9+n)$-variate Gaussian distribution with prescribed linear constraints on the covariance matrix, which is known as a challenging problem [34].

Instead of jointly maximizing the likelihood, we exploit the fact that the marginal distribution for R is a matrix Fisher distribution, and the conditional distribution for $x|R$ is Gaussian. More specifically, the log-likelihood for the marginal distribution corresponds to the first two terms on the right hand side of (29), and the marginal MLE for parameters U, S, V is solved by the MLE of the matrix Fisher distribution.

Theorem 4 (MFG [19], [24])**.** *The marginal maximum likelihood estimates for U, V are given by the proper singular value decomposition $\bar{\mathrm{E}}[R] = UDV^T$, and the marginal MLE for S is given by solving* (11) *for S using D.*

After obtaining U, S, V, they are used in the conditional log-likelihood for $x|R$ corresponding to the last two terms on the right hand side of (29), and the resulting conditional MLE is addressed as follows.

Theorem 5 (MFG)**.** *Let $U, V \in \mathrm{SO}(3)$ and $S \in \mathbb{R}^{3\times 3}$ be the solution of the marginal MLE for R. Define $Q_i = U^T R_i V$, and $\nu_{R_i} = (SQ_i - Q_i^T S)^\vee$ for $i = 1, \ldots, N_s$. Also define the following sample covariance matrices:*

$$\overline{\mathrm{cov}}(x, x) = \bar{\mathrm{E}}(xx^T) - \bar{\mathrm{E}}[x]\bar{\mathrm{E}}[x]^T, \tag{30}$$

$$\overline{\mathrm{cov}}(x, \nu_R) = \bar{\mathrm{E}}[x\nu_R^T] - \bar{\mathrm{E}}[x]\bar{\mathrm{E}}[\nu_R]^T, \tag{31}$$

$$\overline{\mathrm{cov}}(\nu_R, \nu_R) = \bar{\mathrm{E}}[\nu_R\nu_R^T] - \bar{\mathrm{E}}[\nu_R]\bar{\mathrm{E}}[\nu_R]^T. \tag{32}$$

Then the solution of the conditional MLE for P, μ, and Σ is given by

$$P = \overline{\mathrm{cov}}(x, \nu_R)\overline{\mathrm{cov}}(\nu_R, \nu_R)^{-1}, \tag{33}$$

$$\mu = \bar{\mathrm{E}}[x] - P\bar{\mathrm{E}}[\nu_R], \tag{34}$$

$$\Sigma = \overline{\mathrm{cov}}(x, x) - P\overline{\mathrm{cov}}(x, \nu_R)^T + P(\mathrm{tr}(S)\, I_{3\times 3} - S)P^T. \tag{35}$$

Proof. Take the derivatives of (29) with respect to μ and P to obtain

$$\frac{\partial l}{\partial \mu} = \Sigma_c^{-1}\bar{\mathrm{E}}\left[x - \mu - P\nu_R\right],$$
$$\frac{\partial l}{\partial P} = \Sigma_c^{-1}\bar{\mathrm{E}}\left[(x-\mu)\nu_R^T - P\nu_R\nu_R^T\right].$$

By setting the derivatives zero, the MLE of μ and P can be obtained as in (34) and (33). Next, take the derivative of (29) with respect to Σ_c^{-1} to have

$$\frac{\partial l}{\partial \Sigma_c^{-1}} = \frac{1}{2}\Sigma_c - \frac{1}{2}\bar{\mathrm{E}}\left[(x - \mu - P\nu_R)(x - \mu - P\nu_R)^T\right].$$

Setting the derivative to zero and substituting (33) and (34), we obtain (35). □

The given marginal-conditional MLE is an approximation to the joint MLE, because the information of U, S and V encoded in $\{x_i\}$ is discarded over marginalization. Intuitively, the correlation between x and $\mathrm{vec}(R)$ indicated by the samples is not necessarily constrained in the tangent space at UV^T calculated from $\bar{\mathrm{E}}[R]$, as required by MFG in Theorem 1.

To understand how well the marginal-conditional MLE approximates the joint MLE, we perform the following analysis to compare the information that R carries about the unknown parameter S, with that of $x|R$. We focus on the specific case when the dimension of the linear part is one, i.e., $n = 1$. Define a metric $\lambda_{s_i} \in \mathbb{R}^1$ as

$$\lambda_{s_i} = \frac{g_{s_i s_i}(R)}{g_{s_i s_i}(x|R)}, \tag{36}$$

where $g_{s_i s_i}(R)$ is the diagonal element of the Fisher information matrix for the marginal distribution $p(R)$ with respect to s_i. Similarly, $g_{s_i s_i}(x|R)$ is for the conditional distribution $p(x|R)$ [35, Chapter 9.8]. The quantity λ_{s_i} indicates the ratio of the information of the concentration parameter s_i contained in R, to that in x, due to the fact that $g_{s_i s_i}(R, x) = g_{s_i s_i}(R) + g_{s_i s_i}(x|R)$. The higher λ_{s_i} is, the less information is discarded in the marginal MLE for parameter s_i.

Proposition 2 (MFG)**.** *Suppose $(R, x) \in \mathrm{SO}(3) \times \mathbb{R}^1$ follows $\mathcal{MG}(\mu, \sigma^2, P, U, sI_{3\times 3}, V)$, where $P = \frac{\rho\sigma}{\sqrt{2s}}\begin{bmatrix}1 & 1 & 1\end{bmatrix}$ for $\rho \in \mathbb{R}^1$. Then,*

$$\lambda_{s_i} = \frac{1 - 3\rho^2}{\rho^2}\frac{2s\left(c(S)\partial_{11}c(S) - (\partial_1 c(S))^2\right)}{c(s)\left(c(S) - \partial_{11}c(S)\right)}, \tag{37}$$

where $\partial_i c(S) = \frac{\partial c(S)}{\partial s_i}\big|_{S=sI}$ and $\partial_{ij} c(S) = \frac{\partial^2 c(S)}{\partial s_i \partial s_j}\big|_{S=sI}$.

Proof. By Chapter 2.6 in [35], we have

$$g_{s_i s_i}(R, x) = -\mathrm{E}\left[\frac{\partial^2 \log p(R, x)}{\partial s_i^2}\right]$$
$$= \frac{\partial_{ii} c(S)}{c(S)} - \frac{(\partial_i c(s))^2}{c(S)^2} + \mathrm{E}\left[\frac{\partial \nu_R^T}{\partial s_i} P^T \Sigma_c^{-1} P \frac{\partial \nu_R}{\partial s_i}\right],$$

where the first two terms are the marginal information $g_{s_i s_i}(R)$, and the last expectation is the conditional information $g_{s_i s_i}(x|R)$. Substitute $n = 1$, $\Sigma = \sigma^2$, $S = sI_{3\times 3}$, and $P = \frac{\rho\sigma}{\sqrt{2s}}\begin{bmatrix}1 & 1 & 1\end{bmatrix}$ into the conditional information to obtain

$$g_{s_i s_i}(x|R) = \frac{\rho^2}{1 - 3\rho^2}\frac{\mathrm{E}\left[Q_{ij}^2\right] + \mathrm{E}\left[Q_{ik}^2\right]}{2s}$$
$$= \frac{\rho^2}{1 - 3\rho^2}\frac{1 - \mathrm{E}\left[Q_{ii}^2\right]}{2s} = \frac{\rho^2}{1 - 3\rho^2}\frac{c(S) - \partial_{ii} c(S)}{2sc(S)}.$$

And (37) follows from the above two equations. □

In Proposition 2, ρ can be interpreted as the *correlation coefficient* between R and x. It is clearly seen from (37) when ρ is close to zero, i.e., when the correlation between R and x is weak, the information of s_i is mainly contained in R. Therefore, the marginal-conditional MLE is close to the joint MLE. Next, we examine the effect of concentration level of the attitude. Let $r(s)$ be the second fraction term on the right hand side of (37), whose value is illustrated in Fig. 3 for varying s. This indicates when the marginal attitude distribution is close to uniform, i.e., $s \to 0$, relatively more information of s_i is carried by x. On the other hand, when the attitude is more concentrated, say $s > 6$, the fraction $r(s)$ does not vary much, and λ_{s_i} is mainly determined by the level of correlation at the first part on the right hand side of (37).

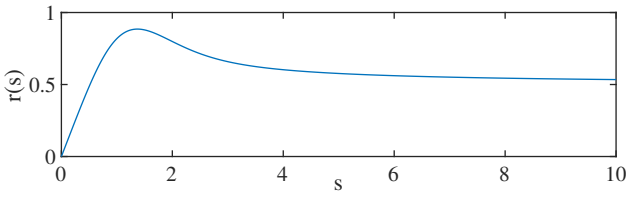

Fig. 3. The graph of $r(s) = \frac{2s\big(c(S)\partial_{11}c(S) - (\partial_1 c(S))^2\big)}{c(s)\big(c(S) - \partial_{11}c(S)\big)}$ against s.

The proposed closed form solution of MLE is essential for designing and implementing an attitude estimator based on MFG, as it is inevitably used in each step of uncertainty propagation and measurement update, which typically run at more than 100 Hz. It is not feasible to solve the joint MLE numerically at every step for real-time implementations.

F. Distinction Between MFGB and MFGI

At first glance, it appears that the new definition of MFGB presented in this paper is very close to MFGI in [30], as the difference is caused by the single term ν_R. Here, we discuss the implication of the two definitions of ν_R and its role in formulating the angular–linear correlation.

More specifically, we highlight the difference by examining $R|x$, the attitude distribution of R conditioned by the linear random variable x, when (R, x) is close to the mean $(M = UV^T, \mu)$. For simplicity, assume the dimension of x is $n = 1$, and let $(R, x) \sim \mathcal{MG}(\mu, \sigma^2, P, U, S, V)$. For both definitions, we have

$$p(R|x) \propto \operatorname{etr}\big(SQ^T\big) \exp\left\{\frac{x-\mu}{\sigma_c^2} P\nu_R - \frac{(P\nu_R)^2}{2\sigma_c^2}\right\},$$

where $Q = U^T R V$, and $\sigma_c^2 = \sigma^2 - P(\operatorname{tr}(S)\, I_{3\times 3} - S)P^T \in \mathbb{R}$. As $R \to M$, $Q \to I_{3\times 3}$ and $\nu_R \to 0$. Therefore, when the attitude is close to the mean attitude, the second order term $(P\nu_R)^2$ may be omitted.

For MFGB, $P\nu_R = \operatorname{tr}(\widehat{P^T} Q^T S)$ from (5) and (15). Thus,

$$\begin{aligned} p(R|x) &\propto \operatorname{etr}\big(SQ^T\big) \operatorname{etr}\left(\frac{x-\mu}{\sigma_c^2} \widehat{P^T} Q^T S\right) \\ &= \operatorname{etr}\big((I_{3\times 3} + \hat{\upsilon}(x))\, Q^T S\big)\,, \end{aligned}$$

where $\upsilon(x) \triangleq (x-\mu)/\sigma_c^2 P^T \in \mathbb{R}^{3\times 1}$. Also, we have $\exp(\hat{\upsilon}(x)) = I_{3\times 3} + \hat{\upsilon}(x) + O(\|\upsilon(x)\|^2)$. Therefore, when x is also close to μ, it can be rewritten as

$$p(R|x) \approx \operatorname{etr}\Big(USV^T \exp(\widehat{V\upsilon}(x)) R^T\Big)\,.$$

In short, when (R, x) follows MFGB, the conditional distribution $R|x$ follows the matrix Fisher distribution with

$$(R|x)\big|_{\text{MFGB}} \approx \mathcal{M}(USV^T \exp(\widehat{V\upsilon}(x))) \tag{38}$$

near the mean value (M, μ). Similarly, $R|x$ of MFGI can be approximated by

$$(R|x)\big|_{\text{MFGI}} \approx \mathcal{M}(\exp(\widehat{U\upsilon}(x)) USV^T). \tag{39}$$

Now we compare (38) and (39). When $x = \mu$, they are identical, and $R|x \sim \mathcal{M}(USV^T)$. Further when $x \neq \mu$, they have the same mean attitude given by $U\exp(\hat{\upsilon}(x))V^T$.

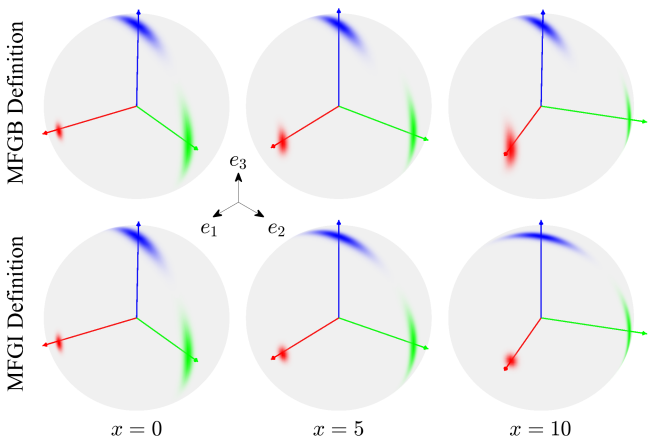

Fig. 4. Difference between MFGB in this paper (top row) and MFGI in [30] (bottom row). The conditional density of $R|x$ is illustrated on the unit sphere for varying x, with the red, green, and blue arrows representing the first, second, and third columns of the mean attitude, respectively. Parameters are chosen as $n = 1$, $\mu = 0$, $\Sigma = 1$, $P = [0, 0, 0.069]$, $U = V = I_{3\times 3}$, $S = \operatorname{diag}(100, 3, 3)$. For both of MFGB and MFBI, the mean attitude of $R|x$ rotates identically about the third inertial axis. For MFGI (bottom row), the attitude distribution rotates in the same way as the mean attitude. But for MFGB (top row), it is as if each sample attitude rotates about its own third body-fixed axis. As a result, the distribution of the third axis (blue shade) remains unchanged in the inertial frame.

The difference between (38) and (39) is caused by how the principal axes of rotations are changed as x is varied. For (38), as U is fixed, the principal axes remain unchanged with respect to x when perceived from the inertial frame. However, they rotate about $V\upsilon(x)$ when observed in the body-fixed frame, as V is changed into $\exp(-\widehat{V\upsilon}(x))V$. Also, (38) indicates $R|x$ has the same density as $R\exp(\widehat{V\upsilon}(x))$, i.e., the rotation is applied on the right. Therefore, the correlation term $\upsilon(x)$ causes the attitude distribution to be rotated about the axis $V\upsilon(x)$ resolved in the body-fixed frame. For example, suppose that the distribution is represented by a set of sample attitudes $\{R_i\}_{i=1}^N$. Then, each sample attitude R_i has a distinct axis of rotation given by $R_i V\upsilon(x)$ in the inertial frame.

On the other hand, in (39), the principal axes represented in the inertial frame are rotated, as U is changed into $\exp(\widehat{U\upsilon}(x))U$. But, when observed from the body-fixed frame, they remain unchanged. Also, (39) implies $R|x$ has the same density as $\exp(\widehat{U\upsilon}(x))R$, i.e., the rotation is applied on the left. This means the correlation term $\upsilon(x)$ causes the attitude distribution to be rotated about the axis $U\upsilon(x)$ resolved in the inertial frame, which is identical for all sample attitudes.

This distinction motivates the naming convention for the two definitions: for MFGB, the correlation causes the attitude distribution to be rotated about the body-fixed frame; for MFGI, it causes the distribution to be rotated about the inertial frame. These are illustrated in Fig. 4. For estimation of attitude and gyro bias, MFGB presented in this paper is more suitable as the gyro bias is resolved in the body-fixed frame.

Finally, when the attitude uncertainty is isotropic, i.e., when $S = sI_{3\times 3}$ for $s \geq 0$, MFGB is identical to MFGI. As such, the difference between MFGB and MFGI becomes more significant as the attitude distribution is more non-isotropic.

IV. Bayesian Estimation for Attitude and Gyroscope Bias

In this section, we apply the proposed MFG to the attitude estimation with a time-varying gyro bias. We design a Bayesian estimator composed of uncertainty propagation and measurement update. The uncertainty propagation step is to propagate MFG describing the attitude and gyroscope bias along the stochastic differential equation of the attitude kinematics. This is accomplished by calculating the moments required for MLE in Theorem 4 and Theorem 5 using two methods: (i) by approximated analytical expressions and (ii) by the sampling based unscented transform. In the measurement update step, we consider two types of measurements, namely attitude and vector measurements. It is shown that both types lead to the same form of posterior density, which is then matched to a new MFG.

When propagating uncertainties of attitude and bias, the following kinematics model [16], [24] is considered

$$R^T \mathrm{d}R = (\hat{x} + \hat{\Omega})\mathrm{d}t + (H_u dW_u)^\wedge, \tag{40}$$

$$\mathrm{d}x = H_v dW_v, \tag{41}$$

where $R \in \mathrm{SO}(3)$ is the attitude of a rigid body and $x \in \mathbb{R}^3$ is the bias of the onboard gyroscope. The vector $\Omega \in \mathbb{R}^3$ is the angular velocity measured by the gyroscope that is resolved in the body-fixed frame. Next, W_u and $W_v \in \mathbb{R}^3$ are two independent three-dimensional Wiener processes, and $H_u, H_v \in \mathbb{R}^{3\times 3}$ are two matrices describing the strengths of noises. The angular velocity measurement has two sources of noises: the bias term x and the Gaussian white noise contributed by $H_u dW_u$. The bias is slowly varying while being driven by another white noise $H_v dW_v$.

The stochastic differential equation (40) is interpreted in the Stratonovich sense to guarantee the process does not leave $\mathrm{SO}(3)$ [16], [36]. Let the time be discretized by a sequence $\{t_0, t_1, \ldots\}$. For convenience, it is assumed that the time step $h \in \mathbb{R}^1$ is fixed, i.e., $h = t_{k+1} - t_k$ for any k. According to [36, Eqn. 14], the kinematics model can be discretized into

$$R_{k+1} = R_k \exp\left\{h(\hat{\Omega}_k + \hat{x}_k) + (H_u \Delta W_u)^\wedge\right\}, \tag{42}$$

$$x_{k+1} = x_k + H_v \Delta W_v, \tag{43}$$

where $\Delta W_u, \Delta W_v \in \mathbb{R}^3$ are the stochastic increments of the Wiener processes over a time step, which are Gaussian with

$$H_u \Delta W_u \sim \mathcal{N}(0, hG_u), \quad H_v \Delta W_v \sim \mathcal{N}(0, hG_v), \tag{44}$$

where $G_u = H_u H_u^T$ and $G_v = H_v H_v^T \in \mathbb{R}^{3\times 3}$.

The initial attitude and bias $(R(t_0), x(t_0))$ at t_0 are assumed to follow MFG with $n = 3$ and $(\mu_0, \Sigma_0, P_0, U_0, S_0, V_0)$ of appropriate dimensions. We wish to propagate it through the discretized equations (42) and (43) to construct the propagated MFG. The evolution of the probability density over a stochastic differential equation is governed by the Fokker-Planck equation, and in general, the propagated density is not necessarily MFG. This is addressed by calculating the moments of the propagated density, and constructing the corresponding MFG from the solution of MLE presented in Section III-E. Depending on how the moments of the propagated density are calculated, we present two methods: an analytical method and an unscented method.

A. Analytical Uncertainty Propagation

Suppose $(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$. In this subsection, we present an approach to construct a new MFG corresponding to the propagated density of (R_{k+1}, x_{k+1}) by calculating its moments analytically.

First, the exponent in (42) can be decomposed into

$$\{h(\Omega_k + \mu_k)\} + \{h(x_k - \mu_k) + H_u \Delta W_u\},$$

after taking the hat map off. The first part is deterministic, and the second part is a random vector with zero mean. This leads to the following approximation to (42).

Lemma 2. *Equation* (42) *is almost surely equivalent to*

$$R_{k+1} = R_k e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h)} e^{h(\hat{\Omega}_k + \hat{\mu}_k)}. \tag{45}$$

Proof. Equation (42) is rewritten into

$$R_{k+1} = R_k \left[e^{h(\hat{\Omega}_k + \hat{x}_k) + (H_u \Delta W_u)^\wedge} e^{-h(\hat{\Omega}_k + \hat{\mu}_k)}\right] e^{h(\hat{\Omega}_k + \hat{\mu}_k)}.$$

The Baker-Campbell-Hausdorff (BCH) formula [37] provides the solution of Z to the equation $e^X e^Y = e^Z$ for given X, Y. Applying this to the expression in the square brackets,

$$R_{k+1} = R_k e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + A} e^{h(\hat{\Omega}_k + \hat{\mu}_k)}, \tag{46}$$

where the additional term $A \in \mathfrak{so}(3)$ is composed of at least twice iterated Lie brackets, and it is of the order of h^2 and $h\Delta W_u$. Since $\lim_{h\to 0} \Delta W_u = 0$ almost surely, $A \sim o(h)$. □

This lemma is helpful in making use of the closed form expression of the exponential map $\exp : \mathfrak{so}(3) \to \mathrm{SO}(3)$ (the Rodrigues rotation formula) for the deterministic components of (42). The uncertainty of R_{k+1} contributed by the noises is quantified by the centered stochastic component $h(x_k - \mu_k) + H_u \Delta W_u$ with zero mean. Next, we present an expression for $\mathrm{E}[R_{k+1}]$ to solve the marginal MLE.

Theorem 6 (MFG)**.** *The expectation of the propagated attitude R_{k+1} is given by*

$$\mathrm{E}[R_{k+1}] = \Big\{\mathrm{E}[R_k]\left(I_{3\times 3} + \tfrac{h}{2}(G_u - \mathrm{tr}(G_u)\, I_{3\times 3})\right) + hU_k \mathrm{E}\Big[Q_k V_k^T \widehat{P_k \nu_{R_k}}\Big]\Big\} e^{h(\hat{\Omega}_k + \hat{\mu}_k)} + O(h^2), \tag{47}$$

where $Q_k = U_k^T R_k V_k$, $\nu_{R_k} = (S_k Q_k - Q_k^T S_k)^\vee$.

Proof. First, expand the first exponential term in (45) into an infinite sum as

$$e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h)} = \textstyle\sum_{i=0}^{\infty} \frac{1}{i!} \left\{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h)\right\}^i. \tag{48}$$

Note that (i) ΔW_u is a zero mean Gaussian vector with covariance matrix $hI_{3\times 3}$, so its odd order moments are zero, and $\mathrm{E}\Big[((H_u \Delta W_u)^\wedge)^{2n}\Big] \sim O(h^n)$; (ii) $o(h)$ in the above equation only has terms of order at least h^2 or $h\Delta W_u$ as

shown in (46). Combining these two observations, we have the first order approximation of $\mathrm{E}[R_{k+1}]$ as

$$\begin{aligned}\mathrm{E}[R_{k+1}] = \big\{&\mathrm{E}[R_k] + h\mathrm{E}[R_k(\hat{x}_k - \hat{\mu}_k)] \\ &+ \tfrac{1}{2}\mathrm{E}\big[R_k((H_u\Delta W_u)^\wedge)^2\big]\big\}e^{h(\hat{\Omega}_k+\hat{\mu}_k)} + O(h^2). \end{aligned} \tag{49}$$

Then, since (R_k, x_k) follows MFG,

$$\mathrm{E}[R_k(\hat{x}_k - \hat{\mu}_k)] = \mathrm{E}\Big[R_k\widehat{P\nu_{R_k}}\Big] = U_k\mathrm{E}\Big[Q_kV_k^T\widehat{P_k\nu_{R_k}}\Big].$$

In addition, due to the independence of R_k and ΔW_u,

$$\mathrm{E}\big[R_k((H_u\Delta W_u)^\wedge)^2\big] = h\mathrm{E}[R_k]\,(G_u - \mathrm{tr}(G_u)\,I_{3\times 3}).$$

Substituting the above two equations into (49) yields (47). □

With the given $\mathrm{E}[R_{k+1}]$, the marginal MLE for the attitude part of MFG can be solved as discussed in Theorem 4, which yields the estimates of U_{k+1}, S_{k+1} and V_{k+1}. Define $Q_{k+1} = U_{k+1}^TR_{k+1}V_{k+1}$ and $\nu_{R_{k+1}} = (S_{k+1}Q_{k+1} - Q_{k+1}^TS_{k+1})^\vee$ as the intermediate parameters for the MFG at time t_{k+1}. Then the conditional MLE for the rest of parameters is solved as in Theorem 5 with the moments given as follows.

Theorem 7 (MFGB)**.** *Let $\tilde{U}, \tilde{V} \in \mathrm{SO}(3)$ and $\tilde{S}, \tilde{\tilde{V}}, \Gamma_Q \in \mathbb{R}^{3\times 3}$ be*

$$\begin{aligned}&\tilde{U} = U_{k+1}^TU_k, \quad \tilde{V} = V_{k+1}^Te^{-h(\hat{\Omega}_k+\hat{\mu}_k)}V_k, \quad \tilde{S} = \tilde{U}^TS_{k+1}\tilde{V}, \\ &\tilde{\tilde{V}} = V_{k+1}^Te^{-h(\hat{\Omega}_k+\hat{\mu})}G_u^TV_k, \quad \Gamma_Q = \mathrm{tr}\Big(Q_k^T\tilde{S}\Big)\,I_{3\times 3} - Q_k^T\tilde{S}.\end{aligned}$$

Also, let $\tilde{\nu}_R, \tilde{\tilde{\nu}}_R \in \mathbb{R}^3$ be

$$\tilde{\nu}_R = (\tilde{S}^TQ_k - Q_k^T\tilde{S})^\vee, \tag{50}$$

$$\tilde{\tilde{\nu}}_R = (S_{k+1}\tilde{U}Q_k\tilde{\tilde{V}}^T - \tilde{\tilde{V}}Q_k^T\tilde{U}^TS_{k+1})^\vee. \tag{51}$$

Then, the moments of x_{k+1} and $\nu_{R_{k+1}}$ required for the conditional MLE are given by

$$\mathrm{E}[x_{k+1}] = \mu_k, \tag{52}$$

$$\mathrm{E}\big[\nu_{R_{k+1}}\big] = 0, \tag{53}$$

$$\mathrm{E}\big[x_{k+1}x_{k+1}^T\big] = \mathrm{E}\big[x_kx_k^T\big] + hG_v, \tag{54}$$

$$\begin{aligned}\mathrm{E}\Big[x_{k+1}\nu_{R_{k+1}}^T\Big] = &\Big[P_k\Big(\mathrm{E}\big[\nu_{R_k}\tilde{\nu}_R^T\big] - \tfrac{h\mathrm{tr}(G_u)}{2}\mathrm{E}\big[\nu_{R_k}\tilde{\nu}_R^T\big] \\ &+ h\mathrm{E}\big[\nu_{R_k}\nu_{R_k}^TP_k^TV_k\Gamma_Q^T\big]\Big) + \mu_k\Big(\mathrm{E}\big[\tilde{\nu}_R^T\big] - \tfrac{h\mathrm{tr}(G_u)}{2}\mathrm{E}\big[\tilde{\nu}_R^T\big] \\ &+ h\mathrm{E}\big[\nu_{R_k}^TP_k^TV_k\Gamma_Q^T\big]\Big) + h\Sigma_{c_k}V_k\mathrm{E}\big[\Gamma_Q^T\big]\Big]\tilde{V}^T \\ &+ \tfrac{h}{2}\Big(\mu_k\mathrm{E}\big[\tilde{\tilde{\nu}}_R^T\big] + P_k\mathrm{E}\big[\nu_{R_k}\tilde{\tilde{\nu}}_R^T\big]\Big) + O(h^2),\end{aligned} \tag{55}$$

$$\begin{aligned}\mathrm{E}\Big[\nu_{R_{k+1}}\nu_{R_{k+1}}^T\Big] = &\tilde{V}\Big(\mathrm{E}\big[\tilde{\nu}_R\tilde{\nu}_R^T\big] + h\mathrm{E}\big[\Gamma_QV_k^TP_k\nu_{R_k}\tilde{\nu}_R^T\big] \\ &+ h\mathrm{E}\big[\tilde{\nu}_R\nu_{R_k}^TP_k^TV_k\Gamma_Q^T\big] + h\mathrm{E}\big[\Gamma_QV_k^TG_uV_k\Gamma_Q^T\big] \\ &- h\mathrm{tr}(G_u)\,\mathrm{E}\big[\tilde{\nu}_R\tilde{\nu}_R^T\big]\Big)\tilde{V}^T + \tfrac{h}{2}\big(\tilde{V}\mathrm{E}\big[\tilde{\nu}_R\tilde{\tilde{\nu}}_R^T\big] \\ &+ \mathrm{E}\big[\tilde{\tilde{\nu}}_R\tilde{\nu}_R^T\big]\,\tilde{V}^T\big) + O(h^2).\end{aligned} \tag{56}$$

Proof. The proof of this theorem is a straightforward but tedious extension of Theorem 6, and it is omitted. The counterpart of this theorem for MFGI is available in [31] along with the detailed proof, which is readily adapted to this theorem by replacing the expression for ν_R. □

TABLE I
ANALYTICAL UNCERTAINTY PROPAGATION

1:	**procedure** $\mathcal{MG}(t_{k+1}) =$ ANALYTICAL PROPAGATION$(\mathcal{MG}(t_k), \Omega_k)$
2:	Calculate $\mathrm{E}[R_{k+1}]$ using (47).
3:	Obtain $U_{k+1}, S_{k+1}, V_{k+1}$ according to Theorem 4 using $\mathrm{E}[R_{k+1}]$.
4:	Calculate the moments in Theorem 7.
5:	Obtain $\mu_{k+1}, \Sigma_{k+1}, P_{k+1}$ according to Theorem 5 using the moments calculated in Step 4.
6:	Set $\mathcal{MG}(t_{k+1}) = \mathcal{MG}(\mu_{k+1}, \Sigma_{k+1}, P_{k+1}, U_{k+1}, S_{k+1}, V_{k+1})$.
7:	**end procedure**

Remark 2. *Note that ν_{R_k}, $\tilde{\nu}_R$, $\tilde{\tilde{\nu}}_R$, and Γ_Q are linear in Q_k. Therefore, the expectations on the right hand side of* (47)*,* (55) *and* (56) *can be calculated using the moments of $Q \sim \mathcal{M}(S_k)$ up to the third order. More specifically, these expectations can be expressed as linear combinations of $\mathrm{E}[Q_{ij}]$, $\mathrm{E}[Q_{ij}Q_{kl}]$, $\mathrm{E}[Q_{ij}Q_{kl}Q_{mn}]$ for $i, j, k, l, m, n \in \{1, 2, 3\}$. The moments of Q can be calculated using* (11) *and the method in [33].*

With these moments, the estimates for $(\mu_{k+1}, \Sigma_{k+1}, P_{k+1})$ can be constructed through the conditional MLE given in Theorem 5.

In summary, Theorem 6 and Theorem 7 provide an analytical approach to propagate $(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$ into $(R_{k+1}, x_{k+1}) \sim \mathcal{MG}(\mu_{k+1}, \Sigma_{k+1}, P_{k+1}, U_{k+1}, S_{k+1}, V_{k+1})$, up to $O(h^2)$ in moments. This is a typical procedure in assumed density filters, and its validity is guaranteed by the fact that the maximum likelihood estimator minimizes the Kullback–Leibler divergence from the MFG family to the true density [38]. The pseudocode for this analytical uncertainty propagation scheme is shown in Table I.

B. Unscented Uncertainty Propagation

In this subsection, we present an alternative, sampling-based method to propagate the uncertainty. In contrast to the preceding analytical approach, the unscented transform selects so called sigma points from the distribution of (R_k, x_k), which are propagated through (42) and (43), and then they are matched to a new MFG using MLE. The sigma points are selected in a deterministic fashion to characterize the mean and dispersion of the distribution.

In [30, Definition 2], we have introduced an unscented transform to select sigma points from MFGI. For MFGB of this paper, the unscented transform remains the same, except that the new definition of ν_R in (15) is used. Given $(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$, we select $7 + 2n = 13$ sigma points from MFG, together with 7 sigma points from the noise $H_u\Delta W_u$ in (44) according to the common unscented transform for a Gaussian distribution in $\mathbb{R}^3$ (for example, see [39, Chapter 9]). These sigma points are propagated to t_{k+1} through the discrete kinematics model (42) and (43) without the noise term $H_v\Delta W_v$. Then a new MFG at t_{k+1} is recovered from these propagated sigma points using MLE. The effect of the noise term $H_v\Delta W_v$ driving the gyro bias in (43) is accounted by adding hG_v to the new covariance matrix Σ_{k+1} for x_{k+1}, according to the following Proposition 3. Compared with [30], this eliminates the need to take the unscented

TABLE II
UNSCENTED UNCERTAINTY PROPAGATION

procedure $\mathcal{MG}(t_{k+1}) = \text{UNSCENTED PROPAGATION}(\mathcal{MG}(t_k), \Omega_k)$
Select sigma points and weights $\{(R, x, w)_i\}_{i=1}^{13}$ from $\mathcal{MG}(t_k)$.
Select sigma points and weights $\{(H_u \Delta W_u, w)_j\}_{j=1}^{7}$ from $\mathcal{N}(0, hG_u)$ according to the common unscented transform of a Gaussian distribution [39].
Propagate the sigma points through (42) and (43) without the noise $H_v \Delta W_v$, i.e.,

$$R_{i,j} = R_i \exp(h(\hat{\Omega}_k + \hat{x}_i) + (H_u \Delta W_u)_j^{\wedge}), \qquad x_{i,j} = x_i,$$

and calculate the weights as $w_{i,j} = w_i w_j$.
Obtain $(\mu_{k+1}, \Sigma_{k+1}, P_{k+1}, U_{k+1}, S_{k+1}, V_{k+1})$ from these $13 \times 7 = 91$ sigma points $(R, x, w)_{i,j}$ using Theorem 4 and Theorem 5.
Let $\Sigma_{k+1} = \Sigma_{k+1} + hG_v$
Set $\mathcal{MG}(t_{k+1}) = \mathcal{MG}(\mu_{k+1}, \Sigma_{k+1}, P_{k+1}, U_{k+1}, S_{k+1}, V_{k+1})$.
end procedure

transform of $H_v \Delta W_v$. The pseudocode for this uncertainty propagation scheme is summarized in Table II.

Proposition 3 (MFG)**.** *Suppose* $(R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$ *and* $x' \sim \mathcal{N}(\mu', \Sigma')$*, and they are mutually independent. Then* $(R, x + x') \sim \mathcal{MG}(\mu + \mu', \Sigma + \Sigma', P, U, S, V)$.

Proof. Let $y = x + x'$, then the density function for (R, y) is

$$\begin{aligned} f_{R,y}(R, y) &= \int_{x \in \mathbb{R}^n} f_{R,x}(R, x) f_{x'}(y - x) \mathrm{d}x \\ &= \frac{1}{c} \int_{x \in \mathbb{R}^n} \operatorname{etr}\big(F R^T\big) \exp\big(-\tfrac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\big) \\ &\quad \times \exp\big(-\tfrac{1}{2}(y - x - \mu')^T (\Sigma')^{-1} (y - x - \mu')\big) \, \mathrm{d}x \\ &= \frac{1}{c'} \operatorname{etr}\big(F R^T\big) \exp\big\{-\tfrac{1}{2}(y - \mu_c - \mu')^T (\Sigma_c + \Sigma')^{-1} \\ &\quad \times (y - \mu_c - \mu')\big\}, \end{aligned}$$

where c, c' are some normalizing constants, and the last equality is from the addition of two independent Gaussian random vectors. Comparing the above equation with (13) yields the proposition. □

C. Measurement Update

Finally, we present how to update the propagated MFG when measurements are available. Here we consider two cases when the attitude is directly measured, or when vectors associated with attitude, such as the direction of magnetic field or gravity, are measured. As the measurement update is assumed to be completed instantaneously, the subscript k denoting the time step is omitted throughout this subsection. The variables relevant to the posterior distribution conditioned by measurements are denoted by the superscript $+$.

First, suppose the attitude is measured by N_a attitude sensors as $Z_i \in \mathrm{SO}(3)$, whose error is distributed by the matrix Fisher distribution. More specifically, given the true attitude $R_t \in \mathrm{SO}(3)$, the measurement error $R_t^T Z_i \in \mathrm{SO}(3)$ follows the matrix Fisher distribution with the parameter $F_{Z_i} \in \mathbb{R}^{3\times3}$ for $i = 1, \dots, N_a$, which characterizes the accuracy and bias of the i-th attitude sensor.

Next, suppose there are also N_v fixed reference vectors $a_j \in \mathbb{S}^2$ in the inertial reference frame, which are measured by direction sensors in the body-fixed frame as $z_j \in \mathbb{S}^2$ for $j = 1, \dots, N_v$. Furthermore, given the true attitude R_t, the noisy measurement z_j is assumed to follow the von Mises Fisher distribution [17] with mean direction $R_t^T B_j a_j \in \mathbb{S}^2$ and concentration parameter $\kappa_j > 0$. The parameter $B_j \in \mathrm{SO}(3)$ specifies the constant bias of the direction sensor, and κ_j specifies the concentration of its random noise.

Suppose the prior distribution of (R, x) before measurement update follows MFG with parameters $(\mu, \Sigma, P, U, S, V)$. By Bayes' rule and Theorem 3.2 in [24], the posterior density conditioned on all of the available measurements $\mathcal{Z} = \{Z_1, \dots, Z_{N_a}, z_1, \dots, z_{N_v}\}$ is

$$\begin{aligned} p(R, x | \mathcal{Z}) \propto \operatorname{etr}\Bigg(\Bigg(F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T\Bigg) R^T\Bigg) \\ \times \exp\big(-\tfrac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\big), \end{aligned} \tag{57}$$

where F, μ_c and Σ_c are defined as in Definition 2 with respect to $(\mu, \Sigma, P, U, S, V)$. The above posterior density of $(R, x)|\mathcal{Z}$ is no longer MFG, as the tangent space at the mean attitude of the updated matrix Fisher part is altered, i.e., the correlation does not satisfy the constraint described in Theorem 1. Similar to the previous two subsections, we match a new MFG with parameters $(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ to this density through MLE after calculating the required moments.

Theorem 8 (MFGB)**.** *Define* $F^+ \in \mathbb{R}^{3\times3}$ *as*

$$F^+ = F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T, \tag{58}$$

and let its proper singular value decomposition be $F^+ = U^+ S^+ (V^+)^T$*. Also, let*

$$\nu_R^+ = (S^+ Q^+ - (Q^+)^T S^+)^{\vee}, \tag{59}$$

for $Q^+ = (U^+)^T R V^+ \in \mathrm{SO}(3)$*. Then the moments of the posterior density* (57)*, namely* $\mathrm{E}[R|\mathcal{Z}]$*,* $\mathrm{E}\big[\nu_R^+ | \mathcal{Z}\big]$ *and* $\mathrm{E}\big[\nu_R^+ (\nu_R^+)^T | \mathcal{Z}\big]$ *are identical to their counterparts in Theorem 3 after replacing* U, S, V *with* U^+, S^+, V^+*, and*

$$\mathrm{E}[x|\mathcal{Z}] = \mu + P \mathrm{E}[\nu_R | \mathcal{Z}], \tag{60}$$

$$\begin{aligned} \mathrm{E}\big[x x^T | \mathcal{Z}\big] &= \mu \mu^T + \mu \mathrm{E}[\nu_R | \mathcal{Z}]^T P^T + P \mathrm{E}[\nu_R | \mathcal{Z}] \mu^T \\ &\quad + P \mathrm{E}\big[\nu_R \nu_R^T | \mathcal{Z}\big] P^T + \Sigma_c, \end{aligned} \tag{61}$$

$$\mathrm{E}\big[x (\nu_R^+)^T | \mathcal{Z}\big] = P \mathrm{E}\big[\nu_R (\nu_R^+)^T | \mathcal{Z}\big], \tag{62}$$

where

$$\mathrm{E}[\nu_R | \mathcal{Z}] = \tilde{V} (\tilde{S}^T \mathrm{E}\big[Q^+ | \mathcal{Z}\big] - \mathrm{E}\big[Q^+ | \mathcal{Z}\big]^T \tilde{S})^{\vee}, \tag{63}$$

$$\mathrm{E}\big[\nu_R \nu_R^T | \mathcal{Z}\big] = \tilde{V} \mathrm{E}\big[\tilde{\nu}_R^+ (\tilde{\nu}_R^+)^T | \mathcal{Z}\big] \tilde{V}^T, \tag{64}$$

$$\mathrm{E}\big[\nu_R (\nu_R^+)^T | \mathcal{Z}\big] = \tilde{V} \mathrm{E}\big[\tilde{\nu}_R^+ (\tilde{\nu}_R^+)^T | \mathcal{Z}\big], \tag{65}$$

with $\tilde{U} = U^T U^+$*,* $\tilde{V} = V^T V^+ \in \mathrm{SO}(3)$*,* $\tilde{S} = \tilde{U}^T S \tilde{V} \in \mathbb{R}^{3\times3}$*, and* $\tilde{\nu}_R^+ \in \mathbb{R}^3$ *is*

$$\tilde{\nu}_R^+ = (\tilde{S}^T Q^+ - (Q^+)^T \tilde{S})^{\vee}.$$

Proof. The expressions for $\mathrm{E}[R|\mathcal{Z}]$, $\mathrm{E}\big[\nu_R^+ | \mathcal{Z}\big]$, $\mathrm{E}\big[\nu_R^+ (\nu_R^+)^T | \mathcal{Z}\big]$, and (60)-(62) can be obtained by integrating

TABLE III
BAYESIAN ESTIMATION FOR ATTITUDE AND GYROSCOPE BIAS

procedure ESTIMATION($\mathcal{MG}(t_0), \Omega(t)$)
  Let $k = 0$.
  repeat
    Either $\mathcal{MG}(t_{k+1})$ = Analytical Propagation($\mathcal{MG}(t_k)$,$\Omega(t_k)$) or $\mathcal{MG}(t_{k+1})$ = Unscented Propagation($\mathcal{MG}(t_k)$,$\Omega(t_k)$).
    $k = k + 1$.
  until $Z(t_{k+1})$ or $z(t_{k+1})$ ia available
  $\mathcal{MG}(t_{k+1})$ = Measurement Update($\mathcal{MG}(t_{k+1})$,$Z(t_{k+1})$,$z(t_{k+1})$).
  Obtain the estimates as $R(t_{k+1}) = U_{k+1}V_{k+1}^T$, $x(t_{k+1}) = \mu_{k+1}$ for $\mathcal{MG}(t_{k+1})$.
  go to step 3.
end procedure

procedure $\mathcal{MG}^+$ = MEASUREMENT UPDATE($\mathcal{MG}^-$,Z,z)
  Compute F^+ from (58), and calculate its proper SVD as $U^+S^+(V^+)^T = F^+$.
  Calculate the moments of the posterior density in Theorem 8.
  Obtain μ^+, Σ^+, P^+ according to Theorem 5.
  Set $\mathcal{MG}^+ = \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$
end procedure

these variables with respect to the density (57). Since $Q = U^TRV = \tilde{U}Q^+\tilde{V}^T$, we have

$$\nu_R = (S\tilde{U}Q^+\tilde{V}^T - \tilde{V}(Q^+)^T\tilde{U}^TS)^\vee = \tilde{V}(\tilde{S}^TQ^+ - (Q^+)^T\tilde{S})^\vee,$$

from which (63)-(65) follow. □

Remark 3. *$\mathrm{E}\big[\tilde{\nu}_R^+(\tilde{\nu}_R^+)^T|\mathcal{Z}\big]$ and $\mathrm{E}\big[\tilde{\nu}_R^+(\nu_R^+)^T|\mathcal{Z}\big]$ can be expressed as linear combinations of $\mathrm{E}\big[Q_{ij}^+Q_{kl}^+|\mathcal{Z}\big]$ for $i, j, k, l \in \{1, 2, 3\}$, therefore they can be calculated using the second order moments of the matrix Fisher distribution $Q^+|\mathcal{Z} \sim \mathcal{M}(S^+)$.*

Since the attitude part of (57) is already a matrix Fisher density, $U^+S^+(V^+)^T = F^+$ is the solution to the marginal MLE for the matrix Fisher part. The conditional MLE is solved by Theorem 5 with the moments calculated above, which yields μ^+, Σ^+ and P^+. These provide the measurement update to represent the posterior distribution conditioned by the measurement as MFG.

The proposed uncertainty propagation and measurement update steps constitute a Bayesian attitude and gyro bias estimator. The current belief represented by MFG can be propagated until an additional measurement is available, based on which the propagated belief is updated. The estimates for the attitude and gyro bias are given by UV^T and μ, respectively. The pseudocode for the proposed Bayesian estimator is presented in Table III. A set of MATLAB codes for the proposed MFG and estimators are available at [40].

V. NUMERICAL SIMULATIONS

In this section, we compare the proposed Bayesian estimators based on MFG with the well-established MEKF and UKF [41] through numerical studies. We consider a rotational motion of a rigid body where three Euler angles (body-fixed 3-2-1) follow sinusoidal waves with the frequency at $0.35\,\mathrm{Hz}$, and the amplitudes of π, $\pi/2$, and π, respectively. The corresponding average angular speed is $6.17\,\mathrm{rad/s}$. The measured angular velocity is obtained from its true value by adding a bias and a white noise. The gyroscope bias is modeled as a Wiener process (bias-instability noise) starting at zero with the isotropic strength $\sigma_v = 500\,\mathrm{deg/h/\sqrt{s}}$, i.e., $H_v = \sigma_v I_{3\times3}$ in (41). The white noise of angular velocity (angle random walk) has the isotropic strength $\sigma_u = 10\,\mathrm{deg/\sqrt{s}}$, i.e., $H_u = \sigma_u I_{3\times3}$ in (40). These two values are greater than those of typical gyroscopes, but they are selected to generate large uncertainties.

We consider two vector measurements, given by

$$z_i = R_t^T a_i + v_i \tag{66}$$

for $i \in \{1, 2\}$, and they are normalized to have unit lengths. In the above equation, the reference directions fixed in the inertial frame are chosen as $a_1 = e_2$ and $a_2 = e_1$. The true attitude is denoted by R_t, and v_1, v_2 are Gaussian noises distributed by $v_1 \sim \mathcal{N}(0, 0.01I_{3\times3})$, $v_2 \sim \mathcal{N}(0, \sigma_2^2 I_{3\times3})$. These follow the common MEKF and UKF frameworks in attitude estimation. To simulate the proposed MFG filters, $a_i + v_i$ for $i = 1, 2$ are matched to von Mises Fisher distributions by Monte Carlo sampling to obtain κ_1, κ_2 in (58) with $B_1 = B_2 = I_{3\times3}$.

The initial attitude R_0 is set as the true attitude rotated about its first body-fixed axis by $180°$, and the initial bias is set as $x_0 = [0.2, 0.2, 0.2]^T\ \mathrm{rad/s}$. The initial attitude uncertainty is set as $\delta R \sim \mathcal{N}(0, 10^{10}I_{3\times3})$ for MEKF and UKF, which is very close to the uniform distribution, and it is matched to a matrix Fisher distribution through Monte Carlo sampling for MFG filters. The initial bias uncertainty is $0.1^2 I_{3\times3}$, and the correlation between attitude and bias is zero. The gyroscope measurement frequency f_{gyro} varies from $150\,\mathrm{Hz}$ to $10\,\mathrm{Hz}$, and vector measurements are sampled at every five gyroscope samples. The simulation period is $300\,\mathrm{s}$.

Six estimation schemes are compared, namely MEKF, UKF, two estimators with MFGB (one with the analytical propagation and the other with the unscented propagation, denoted respectively by MFGBA and MFGBU), and their counterparts with MFGI (denoted by MFGIA and MFGIU respectively). We perform simulations with respect to the varying second vector measurement accuracy represented by σ_2, and the varying gyroscope measurement frequency f_{gyro}. For each case, one hundred Monte Carlo simulations (with respect to the random noise) are carried out. Then, the attitude and bias errors are averaged across all time steps in one simulation, and further averaged across all simulations. Paired t-tests ($N = 100$, $\alpha = 0.001$) are performed between MEKF, UKF and MFG filters, and between MFGB and MFGI to indicate any statistically significant difference.

A. Effects of the Second Vector Measurement

In this subsection, the effect of the accuracy of the second vector measurement on estimation errors is investigated. More specifically, its error variance is varied as $\sigma_2^2 \in \{0.01,\ 0.1,\ 1,\ 5,\ 10,\ 50,\ 200,\ \infty\}$, with the fixed $f_{\text{gyro}} = 150\,\mathrm{Hz}$. The first case of $\sigma_2^2 = 0.01$ corresponds to when the second vector measurement is as accurate as the first one. As σ_2 is increased, the measurement becomes gradually less accurate, until it is no longer used when $\sigma_2^2 = \infty$.

The *full* attitude error is defined as the angle between the true attitude and the estimated attitude. As the second

TABLE IV
ATTITUDE (deg) AND BIAS (deg/s) ERRORS (S.D.) FOR DIFFERENT SECOND VECTOR MEASUREMENT ACCURACIES

σ_2^2		MEKF	UKF	MFGIA	MFGIU	MFGBA	MFGBU
0.01	attitude error (full)	5.05(0.08)	5.02(0.08)	4.82(0.04)^{a,b}	4.82(0.04)^{a,b}	4.82(0.04)^{a,b}	4.82(0.04)^{a,b}
	bias error	3.6(0.8)	2.5(0.5)	2.6(0.5)^{a}	2.6(0.5)^{a}	2.6(0.5)^{a}	2.6(0.5)^{a}
0.1	attitude error (full)	7.06(0.25)	6.95(0.16)	6.67(0.11)^{a,b}	6.67(0.11)^{a,b}	6.67(0.11)^{a,b}	6.67(0.11)^{a,b}
	bias error	4.0(0.9)	2.6(0.5)	2.8(0.5)^{a}	2.8(0.5)^{a}	2.8(0.5)^{a}	2.8(0.5)^{a}
1	attitude error (full)	13.8(1.4)	13.3(1.3)	10.1(0.4)^{a,b}	10.1(0.4)^{a,b}	10.1(0.4)^{a,b}	10.1(0.4)^{a,b}
	bias error	4.1(0.9)	2.6(0.5)	2.9(0.5)^{a,d}	2.9(0.5)^{a,d}	2.9(0.5)^{a,d}	2.9(0.5)^{a,d}
5	attitude error (full)	51.0(17.4)	44.4(11.7)	14.4(0.9)^{a,b}	14.4(0.9)^{a,b}	14.4(0.9)^{a,b}	14.4(0.9)^{a,b}
	bias error	5.6(2.3)	3.1(0.8)	3.0(0.5)^{a}	3.0(0.5)^{a}	2.9(0.5)^{a}	2.9(0.5)^{a}
10	attitude error (full)	75.0(17.1)	72.2(16.9)	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}
	attitude error (partial)	4.19(0.12)	4.11(0.05)	4.00(0.04)^{a,b}	4.00(0.04)^{a,b}	4.00(0.04)^{a,b,e}	4.00(0.04)^{a,b,e}
	bias error	6.7(4.1)	3.4(0.9)	3.0(0.5)^{a,b}	3.0(0.5)^{a,b}	3.0(0.5)^{a,b,e}	3.0(0.5)^{a,b,e}
50	attitude error (full)	85.0(17.1)	86.9(16.0)	25.6(4.3)^{a,b}	25.5(4.4)^{a,b}	25.3(4.4)^{a,b,e}	25.4(4.5)^{a,b}
	attitude error (partial)	4.19(0.06)	4.11(0.05)	4.01(0.04)^{a,b}	4.01(0.04)^{a,b}	4.00(0.04)^{a,b,e}	4.00(0.04)^{a,b,e}
	bias error	6.5(2.7)	3.5(1.0)	3.3(0.6)^{a}	3.3(0.6)^{a}	3.1(0.6)^{a,b,e}	3.1(0.6)^{a,b,e}
200	attitude error (partial)	4.19(0.06)	4.11(0.05)	4.02(0.04)^{a,b}	4.01(0.04)^{a,b}	4.00(0.04)^{a,b,e}	4.00(0.04)^{a,b,e}
	bias error	6.5(2.6)	3.5(1.0)	5.0(1.0)^{a,d}	4.3(1.0)^{a,d}	3.4(0.7)^{a,e}	3.4(0.8)^{a,e}
∞	attitude error (partial)	4.24(0.11)	4.10(0.06)	4.20(0.05)^{a,d}	4.16(0.05)^{a,d}	3.99(0.04)^{a,b,e}	3.99(0.04)^{a,b,e}
	bias error	6.5(2.4)	3.7(1.1)	20.1(1.4)^{c,d}	18.7(1.6)^{c,d}	3.6(0.8)^{a,e}	3.7(0.8)^{a,e}

$^{a(c)}$ MFG filter is significantly better (worse) than MEKF; $^{b(d)}$ MFG filter is significantly better (worse) than UKF;
^{e} MFGB filter presented in this paper is significantly better than MFGI filter of [30].

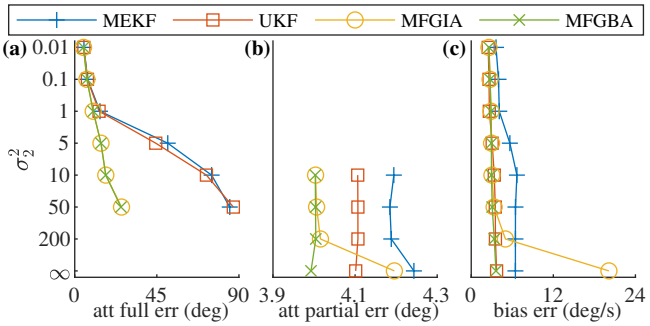

Fig. 5. Estimation errors for varying accuracies of the second vector measurement: (a) full attitude error; (b) partial attitude error; (c) bias error for MEKF, UKF, MFGIA and MFGBA with varying σ_2^2. The errors of the unscented MFG filters are similar with the analytical MFG filters, and they are omitted for readability.

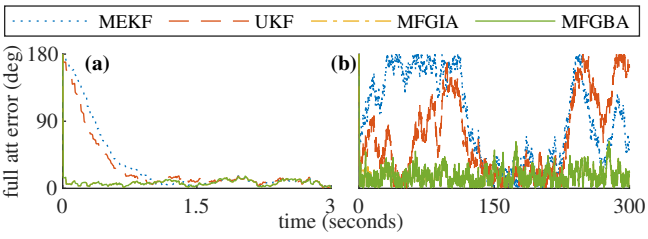

Fig. 6. (a) Full attitude error of MEKF, UKF, MFGIA and MFGBA for a sample simulation ($\sigma_2^2 = 0.1$, $f_{\text{gyro}} = 150$Hz) during the first three seconds. (b) Full attitude error (0 s to 300 s) for a sample simulation ($\sigma_2^2 = 10$, $f_{\text{gyro}} = 150$Hz). The error of the unscented MFG filters is similar to the analytical MFG filters.

measurement becomes more inaccurate, the full attitude cannot be completely estimated because the rotation around the first reference vector becomes unobservable [42]. Thus, the *partial* attitude error is defined as the angle between $R_t^T a_1$ and $R^T a_1$, where R_t and R are the true attitude and the estimated attitude, respectively. The partial attitude error only captures

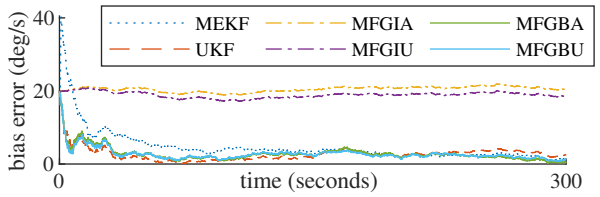

Fig. 7. Bias error (0 s to 300 s) in an example simulation ($\sigma_2^2 = \infty$, f_{gyro} =150 Hz).

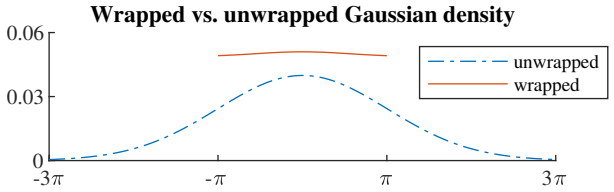

Fig. 8. Gaussian density vs. wrapped Gaussian density with $\mu = 0$ and $\sigma^2 = 10$. The wrapped Gaussian distribution is defined in the circular space $\mathbb{R}/2\pi$ by identifying θ with $\theta + 2k\pi$ for $k \in \mathbb{Z}$, so the densities beyond $[-\pi, \pi]$ are wrapped into $[-\pi, \pi]$.

the accuracy along the first reference vector. The simulation results are summarized in Table IV and Fig. 5.

When $\sigma_2^2 \in \{0.01,\ 0.1,\ 1\}$, the attitude error of the MFG filters is slightly lower than MEKF and UKF, and this advantage is mainly contributed by the faster initial convergence of MFG filters as illustrated in Fig. 6(a). This is because for MEKF, the linearization of the measurement function is accurate only if the attitude error is small; and for UKF, the sigma points from a very large variance (10^{10} in the initial step) suffer from the wrapping error, so they are unable to capture the large initial uncertainty. The bias error of the MFG filters is also lower then MEKF, and is comparable to UKF (except when $\sigma_2^2 = 1$, UKF is statistically lower).

When $\sigma_2^2 \in \{5,\ 10,\ 50\}$, the full attitude error of the MFG

TABLE V
ATTITUDE (deg) AND BIAS (deg/s) ERRORS (S.D.) FOR DIFFERENT GYROSCOPE MEASUREMENT FREQUENCIES

f_{gyro} (Hz)		MEKF	UKF	MFGIA	MFGIU	MFGBA	MFGBU
150	attitude error (full)	75.0(17.1)	72.2(16.9)	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}	17.0(1.3)^{a,b}
	attitude error (partial)	4.19(0.12)	4.11(0.05)	4.00(0.04)^{a,b}	4.00(0.04)^{a,b}	4.00(0.04)^{a,b,e}	4.00(0.04)^{a,b,e}
	bias error	6.7(4.1)	3.4(0.9)	3.0(0.5)^{a,b}	3.0(0.5)^{a,b}	3.0(0.5)^{a,b,e}	3.0(0.5)^{a,b,e}
100	attitude error (full)	80.9(16.5)	77.3(17.8)	19.0(1.6)^{a,b}	19.0(1.6)^{a,b}	19.0(1.7)^{a,b}	19.0(1.7)^{a,b}
	attitude error (partial)	4.66(0.08)	4.54(0.07)	4.41(0.05)^{a,b}	4.41(0.05)^{a,b}	4.41(0.05)^{a,b,e}	4.41(0.05)^{a,b,e}
	bias error	6.9(2.5)	3.7(1.1)	3.3(0.6)^{a,b}	3.2(0.6)^{a,b}	3.2(0.6)^{a,b,e}	3.1(0.6)^{a,b,e}
50	attitude error (full)	85.7(16.0)	84.2(15.3)	23.2(2.2)^{a,b}	23.2(2.3)^{a,b}	23.2(2.2)^{a,b}	23.2(2.3)^{a,b}
	attitude error (partial)	5.61(0.12)	5.41(0.08)	5.23(0.07)^{a,b}	5.23(0.07)^{a,b}	5.22(0.07)^{a,b,e}	5.22(0.07)^{a,b,e}
	bias error	6.9(2.2)	4.2(1.4)	3.6(0.6)^{a,b}	3.6(0.6)^{a,b}	3.4(0.7)^{a,b,e}	3.4(0.7)^{a,b,e}
25	attitude error (full)	89.6(11.9)	90.2(10.6)	29.2(3.6)^{a,b}	29.0(3.6)^{a,b}	29.1(3.6)^{a,b}	29.0(3.7)^{a,b}
	attitude error (partial)	6.82(0.23)	6.55(0.14)	6.23(0.10)^{a,b}	6.22(0.09)^{a,b}	6.21(0.09)^{a,b,e}	6.21(0.09)^{a,b,e}
	bias error	8.5(3.0)	5.8(2.1)	4.1(0.8)^{a,b}	4.0(0.8)^{a,b}	3.7(0.8)^{a,b,e}	3.7(0.8)^{a,b,e}
10	attitude error (full)	90.1(10.0)	91.2(2.9)	44.0(7.4)^{a,b}	43.6(9.5)^{a,b}	43.7(8.5)^{a,b}	43.1(9.2)^{a,b}
	attitude error (partial)	10.5(0.4)	10.4(0.3)	9.4(0.15)^{a,b}	9.3(0.15)^{a,b}	9.3(0.15)^{a,b,e}	9.3(0.15)^{a,b,e}
	bias error	14.0(5.5)	17.8(6.6)	6.3(1.4)^{a,b}	5.7(2.0)^{a,b}	5.1(1.6)^{a,b,e}	5.0(1.7)^{a,b,e}

$^{a(c)}$ MFG filter is significantly better (worse) than MEKF; $^{b(d)}$ MFG filter is significantly better (worse) than UKF;
^{e} MFGB filter presented in this paper is significantly better than MFGI filter of [30].

filters is much lower than MEKF and UKF (Fig 6(b)). This is because when the second vector measurement is inaccurate, the attitude uncertainty becomes highly dispersed along the rotation about the first reference vector, and the Gaussian distribution used by MEKF and UKF is incapable of modeling such large dispersion due to the wrapping error. This is illustrated in Fig. 8 on the circular space, where the Gaussian densities beyond $[-\pi, \pi]$ are wrapped into $[-\pi, \pi]$ since they represent the same angle, and the resulting wrapped Gaussian distribution is much more dispersed if the original Gaussian has large variance. Also, the partial attitude error and bias error of the MFG filters are slightly lower than MEKF and UKF. Next, comparing MFGB with MFGI, MFGB begins to exhibit some statistical advantages in partial attitude and bias errors, although their relative difference is still very small. This advantage can be attributed to that the attitude uncertainty becomes more non-isotropic as the second vector measurement becomes less accurate, so the distinction between the two definitions discussed in Section III-F begins to emerge.

When $\sigma_2^2 \in \{200, \infty\}$, MFGBA and MFGBU are still slightly more accurate than MEKF and UKF in partial attitude estimates, and more accurate than MEKF in bias estimation. On the other hand, the performance of MFGIA and MFGIU in bias estimation degrades greatly, which affects the attitude error especially when $\sigma_2^2 = \infty$. In Fig. 7, the bias error of a sample simulation ($\sigma_2^2 = \infty$) is shown, where MFGIA and MFGIU make little corrections to the bias. It turns out that there is little correlation built between the bias and attitude for MFGI during the uncertainty propagation, indicating that the attitude-bias correlation with non-isotropic attitude measurements cannot be modeled properly with MFGI.

B. Effects of the Measurement Frequency

Next, the effect of the gyroscope measurement frequency is investigated with $f_{\text{gyro}} \in \{150,\ 100,\ 50,\ 25,\ 10\}$ Hz. The variance for the second direction measurement is fixed with $\sigma_2^2 = 10$. The corresponding simulation results are summarized in Table V and Fig. 9.

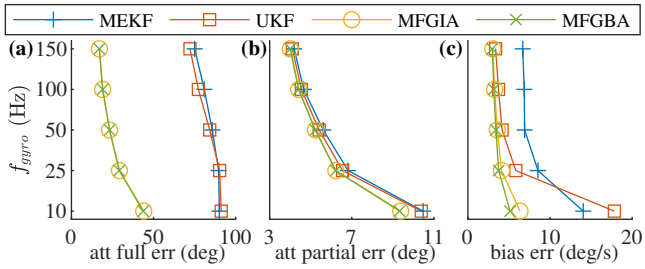

Fig. 9. Estimation errors for different f_{gyro}: (a) full attitude error; (b) partial attitude error; (c) bias error for MEKF, UKF, MFGIA and MFGBA. The error of the unscented MFG filters is similar to the analytical MFG filters.

In contrast to MFG filters, MEKF and UKF are unable to estimate the full attitude for all measurement frequencies (the estimation errors are greater than 72°), although the full attitude error for MFG filters gradually becomes large as f_{gyro} is decreased. The partial attitude error of MFG filters is slightly lower than MEKF and UKF for all f_{gyro}, due to their faster initial convergence. The bias estimation accuracy of MFG filters is also better than MEKF and UKF, and this advantage becomes larger as f_{gyro} is decreased. At $f_{\text{gyro}} = 10$ Hz, the bias error of MEKF and UKF is unable to converge, whereas it still remains relatively low for MFG filters. Comparing MFGB with MFGI, the advantage of MFGB in the bias estimation becomes greater as f_{gyro} is decreased. This is because the advantage of MFGB in the uncertainty propagation step gets intensified by lowering the measurement frequency.

C. Computation Efficiency

Although the proposed MFG filters exhibit better accuracy and faster convergence rate, they require substantially more computation time than MEKF and UKF. Table VI lists the computation time for one discrete time step ($\sigma_2^2 = 10$, $f_{\text{gyro}} = 150$ Hz) of all filters measured in MATLAB R2020b with an AMD Ryzen 3900X CPU. It shows the MFG filters take more than 100 times of computation time than MEKF,

TABLE VI
COMPUTATION TIME (S.D.) FOR ONE DISCRETE TIME STEP

	MEKF	UKF	MFGIA	MFGIU	MFGBA	MFGBU
step time (ms)	0.075(0.009)	1.6(0.2)	14.3(1.6)	20.6(2.4)	14.4(1.6)	20.5(2.2)

and they take about 10 times more than UKF. For analytical MFG filters, solving (11) takes 14.1% of the computation time, and calculating the moments in Theorem 6, Theorem 7 and Theorem 8 takes 85.5%. For unscented MFG filters, solving (11) takes 56.0% of the computation time, selecting and propagating sigma points take 37.6%, and calculating the moments in Theorem 5 and Theorem 8 takes 6.0%.

VI. CONCLUSIONS

In this paper, a new definition of the matrix Fisher–Gaussian distribution is proposed to describe the angular-linear correlation between $\mathrm{SO}(3)$ and $\mathbb{R}^n$ in a global fashion. It is constructed by conditioning a $(9+n)$-variate Gaussian distribution from $\mathbb{R}^{9+n}$ into $\mathrm{SO}(3)\times\mathbb{R}^n$, while constraining the correlation between $\mathbb{R}^9$ and $\mathbb{R}^n$ to be non-zero only along the tangent space of the mean attitude, thereby avoiding over-parameterization. Compared with the prior definition of the matrix Fisher–Gaussian distribution in [30], the proposed distribution formulates the correlation with the axis of rotation resolved in the body-fixed frame. This is more suitable in attitude estimation as the bias of the gyroscope is resolved in the body-fixed frame, and it is particularly advantageous when the attitude uncertainty is non-isotropic.

Next, based on the proposed MFG, two Bayesian estimators are formulated depending on how the uncertainty is propagated. Numerical simulations demonstrate the advantage of the proposed MFG filters over the well-established multiplicative extended Kalman filter and unscented Kalman filter in accuracy and convergence, particularly when the measurement noise is large and the measurement frequency is low.

While this paper focuses on the attitude and gyro bias estimation, the proposed MFG can characterize the angular-linear correlation between the attitude of a rigid body and any linear random variable of an arbitrary dimension in a global fashion, which is the fundamental contribution of this paper. There is a great potential that the proposed MFG becomes utilized in any statistical analysis involving the rotation motion coupled with linear dynamics in the board area of science and engineering.

APPENDIX A
EQUIVALENT CONDITIONS FOR MFGB

In this appendix, we present the equivalent conditions for MFGB. The characterization for MFGI is given in [31]. First, to deal with the uniqueness problem for repeated singular values of F, we augment Definition 1 by the following uniqueness condition: *the first nonzero element of each column of U' is positive* [19]. This condition ensures that the columns of U and V cannot undergo simultaneous sign changes.

Theorem 9 (MFGB)**.** *Suppose $F=USV$, $\tilde{F}=\tilde{U}\tilde{S}\tilde{V}^T$ are the proper SVD of F and $\tilde{F}$ with the augmented uniqueness condition. Then $\mathcal{MG}(\mu,\Sigma,P,U,S,V)$ and $\mathcal{MG}(\tilde{\mu},\tilde{\Sigma},\tilde{P},\tilde{U},\tilde{S},\tilde{V})$ are equivalent if and only if $\mu=\tilde{\mu}$, $S=\tilde{S}$, and one of the following conditions is satisfied:*

1) *if $s_1=s_2=s_3=0$, then $\Sigma=\tilde{\Sigma}$.*
2) *if $s_1\neq s_2=s_3=0$, then $\exists\theta_1,\theta_2\in\mathbb{R}$ such that $\tilde{U}=UT_1$, $\tilde{V}=VT_2$ where $T_1=\exp(\theta_1\hat{e}_1)$ and $T_2=\exp(\theta_2\hat{e}_1)$, $[\tilde{P}_{:,2},\tilde{P}_{:,3}]=[P_{:,2},P_{:,3}]\begin{bmatrix}\cos\theta_2 & -\sin\theta_2\\ \sin\theta_2 & \cos\theta_2\end{bmatrix}$ where $P_{:,i}$ is the i-th column of P, and $\Sigma=\tilde{\Sigma}$.*
3) *if $s_1=s_2=s_3\neq 0$, then $\exists T\in\mathrm{SO}(3)$ such that $\tilde{U}=UT$, $\tilde{V}=VT$, $\tilde{P}=PT$ and $\Sigma=\tilde{\Sigma}$.*
4) *if $s_1\neq s_2=s_3\neq 0$, then $\exists\theta\in\mathbb{R}$ such that $\tilde{U}=UT$, $\tilde{V}=VT$, $\tilde{P}=PT$, where $T=\exp(\theta\hat{e}_1)$, and $\Sigma=\tilde{\Sigma}$.*
5) *if $s_1=s_2\neq|s_3|$, then $\exists\theta\in\mathbb{R}$ such that $\tilde{U}=UT$, $\tilde{V}=VT$, $\tilde{P}=PT$, where $T=\exp(\theta\hat{e}_3)$, and $\Sigma=\tilde{\Sigma}$.*
6) *if $s_1\neq s_2\neq|s_3|$, then $U=\tilde{U}$, $V=\tilde{V}$, $P=\tilde{P}$, and $\Sigma=\tilde{\Sigma}$.*
7) *if $s_1=s_2=-s_3\neq 0$, then $\exists T\in\mathrm{SO}(3)$ such that $\tilde{U}=U\mathcal{D}_{12}T\mathcal{D}_{12}$, $\tilde{V}=VT$, $\tilde{P}=PT$, and $\tilde{\Sigma}=\Sigma+P\left[T(\mathrm{tr}(S)\,I_{3\times3}-S)T^T-(\mathrm{tr}(S)\,I_{3\times3}-S)\right]P^T$.*
8) *if $s_1\neq s_2=-s_3\neq 0$, then $\exists\theta\in\mathbb{R}$ such that $\tilde{U}=UT^T$, $\tilde{V}=VT$, $\tilde{P}=PT$, and $\tilde{\Sigma}=\Sigma+P\left[T(\mathrm{tr}(S)\,I_{3\times3}-S)T^T-(\mathrm{tr}(S)\,I_{3\times3}-S)\right]P^T$, where $T=\exp(\theta\hat{e}_1)$.*

Proof. The proof is very similar to Theorem 12 for MFGI in [31], which can be adapted to this theorem by replacing the expression for ν_R without difficulties. The detailed proof is therefore omitted. □

In other words, Theorem 9 states that if F has repeated singular values and $s_3\geq 0$, then MFG can be parameterized differently by rotating U, V, and P in a consistent way. When $s_3<0$, it is noticeable that Σ and $\tilde{\Sigma}$ are also different as indicated in case 7) and 8). The reason for this is when $s_2=-s_3$, $T(\mathrm{tr}(S)\,I_{3\times3}-S)T^T\neq\mathrm{tr}(S)\,I_{3\times3}-S$.

The uniqueness problem caused by simultaneous sign changes of any two columns of U and V is addressed in the following proposition.

Proposition 4 (MFG)**.** *Let $i\in\{1,2,3\}$, then $\mathcal{MG}(\mu,\Sigma,P,U,S,V)$ and $\mathcal{MG}(\mu,\Sigma,P\mathcal{D}_i,U\mathcal{D}_i,S,V\mathcal{D}_i)$ are equivalent.*

Proof. Let F, μ_c, Σ_c and $\tilde{F}$, $\tilde{\mu}_c$, $\tilde{\Sigma}_c$ be the intermediate parameters of the two MFGs respectively, as stated in Proposition 1. Then it is clear that

$$\tilde{F}=U\mathcal{D}_iS\mathcal{D}_i^TV^T=USV^T=F,$$
$$\tilde{\Sigma}_c=\Sigma-P\mathcal{D}_i(\mathrm{tr}(S)\,I_{3\times3}-S)\mathcal{D}_i^TP^T=\Sigma_c,$$

and for all $R \in \mathrm{SO}(3)$,

$$\begin{aligned}\tilde{\mu}_c &= \mu + P\mathcal{D}_i(S\mathcal{D}_i^T U^T RV\mathcal{D}_i - \mathcal{D}_i^T V^T R^T U D_i S)^\vee \\ &= \mu + P(\mathcal{D}_i S\mathcal{D}_i^T U^T RV - V^T R^T U\mathcal{D}_i S\mathcal{D}_i^T)^\vee = \mu_c.\end{aligned}$$

Therefore by Proposition 1, the two MFGs are equivalent. □

REFERENCES

[1] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of nonlinear attitude estimation methods," *Journal of guidance, control, and dynamics*, vol. 30, no. 1, pp. 12–28, 2007.

[2] J. Stuelpnagel, "On the parametrization of the three-dimensional rotation group," *SIAM review*, vol. 6, no. 4, pp. 422–430, 1964.

[3] N. Toda, J. Heiss, and F. Schlee, "Spars: The system, algorithms, and test results," in *Symposium on Spacecraft Attitude Determination, Aerospace Corp. Rept. TR-0066 (6306)-12*, vol. 1, 1969, pp. 361–370.

[4] 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.

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

[6] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3565–3572.

[7] A. R. Jiménez, F. Seco, J. C. Prieto, and J. Guevara, "Indoor pedestrian navigation using an INS/EKF framework for yaw drift reduction and a foot-mounted imu," in *Workshop on Positioning, Navigation and Communication*. IEEE, 2010, pp. 135–143.

[8] R. Mahony, T. Hamel, and J. M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on automatic control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[9] H. F. Grip, T. I. Fossen, T. A. Johansen, and A. Saberi, "Attitude estimation using biased gyro and vector measurements with time-varying reference vectors," *IEEE Transactions on automatic control*, vol. 57, no. 5, pp. 1332–1338, 2011.

[10] Y. Wang and G. S. Chirikjian, "Nonparametric second-order theory of error propagation on motion groups," *The International journal of robotics research*, vol. 27, no. 11-12, pp. 1258–1273, 2008.

[11] G. Chirikjian and M. Kobilarov, "Gaussian approximation of non-linear measurement models on Lie groups," in *IEEE Conference on Decision and Control*. IEEE, 2014, pp. 6401–6406.

[12] A. Barrau and S. Bonnabel, "Intrinsic filtering on Lie groups with applications to attitude estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 2, pp. 436–449, 2014.

[13] ——, "The invariant extended Kalman filter as a stable observer," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2016.

[14] H. M. Menegaz, J. Y. Ishihara, and H. T. Kussaba, "Unscented Kalman filters for Riemannian state-space systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1487–1502, 2018.

[15] R. Darling, "Geometrically intrinsic nonlinear recursive filters I: algorithms," University of California, Berkeley, Tech. Rep. 494, 1998.

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

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

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

[21] 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.

[22] 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*. IEEE, 2014, pp. 4133–4140.

[23] 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.

[24] 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, 10 2018.

[25] 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.

[26] R. A. Johnson and T. E. Wehrly, "Some angular-linear distributions and related regression models," *Journal of the American Statistical Association*, vol. 73, no. 363, pp. 602–606, 1978.

[27] S. Kato and K. Shimizu, "Dependent models for observations which include angular ones," *Journal of Statistical Planning and Inference*, vol. 138, no. 11, pp. 3538 – 3549, 2008.

[28] T. Abe and C. Ley, "A tractable, parsimonious and flexible model for cylindrical data, with applications," *Econometrics and Statistics*, vol. 4, pp. 91 – 104, 2017.

[29] 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.

[30] 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*. IEEE, 2020, pp. 4429–4434.

[31] ——, "Matrix Fisher-Gaussian distribution on $\mathrm{SO}(3) \times \mathbb{R}^n$ and bayesian attitude estimation," *arXiv preprint arXiv:2003.02180*, 2020.

[32] T. Lee, "Bayesian attitude estimation with approximate matrix Fisher distributions on SO(3)," in *IEEE Conference on Decision and Control*. IEEE, 2018, pp. 5319–5325.

[33] W. Wang and T. Lee, "Higher-order central moments of matrix Fisher distribution on SO(3)," *Statistics & Probability Letters*, vol. 169, p. 108983, 2021.

[34] P. Zwiernik, C. Uhler, and D. Richards, "Maximum likelihood estimation for linear Gaussian covariance models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 4, pp. 1269–1292, 2017.

[35] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[36] A. Barrau and S. Bonnabel, "Stochastic observers on Lie groups: a tutorial," in *IEEE Conference on Decision and Control*. IEEE, 2018, pp. 1264–1269.

[37] B. Hall, *Lie groups, Lie algebras, and representations: an elementary introduction*. Springer, 2015, vol. 222.

[38] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society*, pp. 1–25, 1982.

[39] A. J. Haug, *Bayesian estimation and tracking: a practical guide*. John Wiley & Sons, 2012.

[40] W. Wang and T. Lee, "Matrix Fisher-Gaussian distribution," [Online]. Available: https://github.com/fdcl-gwu/Matrix-Fisher-Gaussian-Code, 2019.

[41] J. L. Crassidis and F. L. Markley, "Unscented filtering for spacecraft attitude estimation," *Journal of guidance, control, and dynamics*, vol. 26, no. 4, pp. 536–542, 2003.

[42] W. Wang, K. Gamagedara, and T. Lee, "On the observability of attitude with single direction measurements," *arXiv preprint arXiv:2008.13067*, 2020.

Weixin Wang received his M.S. degree at University of Wisconsin-Madison, WI, USA, in 2018, and the B.E. degree at Tsinghua University, Beijing, China, in 2016, both in mechanical engineering. He is currently pursuing his Ph.D. degree at George Washington University. His research interests include nonlinear estimation theory, inertial sensors, and human movement tracking.

Taeyoung Lee is a professor of the Department of Mechanical and Aerospace Engineering at the George Washington University. He received his doctoral degree in Aerospace Engineering and his master's degree in Mathematics at the University of Michigan in 2008. His research interests include geometric mechanics and control with applications to complex aerospace systems.