Bayesian Attitude Estimation with Approximate Matrix Fisher Distributions on $\mathsf{SO}(3)$

Taeyoung Lee

***Abstract*— A matrix Fisher distribution on the special orthogonal group is a compact, global form of attitude uncertainty distribution that has been successfully utilized for Bayesian attitude estimations in an intrinsic fashion. This paper addresses two computational issues in implementing matrix Fisher distributions, namely numerical stability and computational efficiency. More precisely, an exponentially scaled normalizing constant of the matrix Fisher distribution and its mathematical properties are introduced for robust numerical implementations. Next, two approximate matrix Fisher distributions are formulated for a highly concentrated case and an almost uniformly distributed case respectively. These approximate forms yield an explicit form of attitude estimation schemes for the considered cases, and it also illustrate the similarity between the Gaussian distribution and the matrix Fisher distribution in the highly concentrated cases.**

I. Introduction

Quaternion-based extended Kalman filters and their variations have been the workhorse of attitude estimation. While these approaches have been verified successfully in a variety of applications, it is still desirable to formulate attitude estimation directly on the special orthogonal group.

In [1], an attitude estimation scheme is proposed by representing a probability density function on the special orthogonal group with noncommutative harmonic analysis [2]. This utilizes the property that irreducible matrix representations of a compact Lie group constitute an orthonormal basis for complex-valued square integrable functions on the group. This idea has been applied for uncertainty propagation and attitude estimation in robotics and aerospace engineering [3], [4], [5]. While an arbitrary shape of distributions can be constructed via irreducible matrix representations, harmonic analysis on a Lie group may become impractical for real-time implementation especially if higher order terms in the Fourier coefficients are considered.

Another notable approach is representing highly concentrated probability distributions on a Lie group via normal distribution in the tangent space [6], [7]. These approaches take the advantage that the exponential map is a local diffeomorphism between the lie algebra and the group, and the resulting estimation scheme resembles the familiar Kalman filers in the Euclidean space. However, the issue of singularities remains and it is not generally applied to the cases of large uncertainties.

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

This research been supported in part by NSF under the grants CMMI-1335008, CMMI-1760928, CNS-1837382, and CNS-1337722.

Recently, a specific form of exponential density of random matrices, namely the matrix Fisher distribution [8], [9] has been utilized in attitude estimation. Various stochastic properties of the matrix Fisher distribution are presented on the special orthogonal group, and two types of Bayesian attitude estimation schemes are proposed [10], [11]. As they are constructed on the special orthogonal group, the issue of singularities in quaternion-based filters is naturally avoided, and there is no restriction on the degree of concentration as opposed to the aforementioned approaches based on exponential coordinates. It has been illustrated that the Bayesian attitude estimator based on the matrix Fisher distribution performs successfully for challenging cases of large uncertainties when the initial estimate is the uniform distribution on the special orthogonal group.

This paper addresses computational issues to improve robustness and efficiency in implementing the matrix Fisher distributions. First, the normalizing constant of the matrix Fisher distribution may become exceedingly large especially when the distributions becomes more concentrated, thereby causing numerical overflow. Here we introduce an exponentially scaled form of the normalizing constant and reformulated the Bayesian attitude filter so as to avoid such numerical instabilities. Next, we present two approximate forms of the matrix Fisher distribution for two extreme cases of high concentration and wide dispersion, and we present an explicit expression for the normalizing constant. This avoids numerical iterations required for Bayesian attitude filtering constructed in [10], and results in an explicit attitude estimator. Additionally, it is shown that the matrix Fisher distribution in highly concentrated cases is well approximated by a normal distribution of Euler angles.

In short, this paper proposes an alternative formulation of the matrix Fisher distributions for numerical robustness, and approximate matrix Fisher distributions to improve computational efficiency. These are illustrated by several numerical examples.

II. Bayesian Attitude Filtering with Matrix Fisher Distribution on $\mathsf{SO}(3)$

Here we summarize the properties of the matrix Fisher distribution on the special orthogonal group, and a Bayesian estimation scheme proposed in [10].

The configuration manifold for the attitude dynamics of a rigid body is the three-dimensional special orthogonal group,

$$\mathsf{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \,|\, R^T R = I_{3\times 3},\, \det[R] = 1\},$$

where each rotation matrix corresponds the linear transformation of the representation of a vector from the body-fixed frame to the inertial frame. The lie algebra $\mathfrak{so}(3)$ is the set of 3×3 skew-symmetric matrices, i.e., $\mathfrak{so}(3)=\{S\in\mathbb{R}^{3\times 3}\,|\,S=-S^T\}$. The *hat* map: $\wedge:\mathbb{R}^3\rightarrow\mathfrak{so}(3)$ is defined such that $\hat{x}=-(\hat{x})^T$, and $\hat{x}y=x\times y$ for any $x,y\in\mathbb{R}^3$. The inverse of the hat map is denoted by the *vee* map: $\vee:\mathfrak{so}(3)\rightarrow\mathbb{R}^3$. The two-sphere is the set of unit-vectors in $\mathbb{R}^3$, i.e., $\mathsf{S}^2=\{q\in\mathbb{R}^3\,|\,\|q\|=1\}$, and the i-th standard basis of $\mathbb{R}^3$ is denoted by $e_i\in\mathsf{S}^2$ for $i\in\{1,2,3\}$. The set of circular shifts of $(1,2,3)$ is defined as $\mathcal{I}=\{(1,2,3),(2,3,1),(3,1,2)\}$. The Frobenius norm of a matrix $A\in\mathbb{R}^{3\times 3}$ is defined as $\|A\|_F=\sqrt{\mathrm{tr}[A^TA]}$.

We utilize the modified Bessel function of the first kind [12]. For any $x\in\mathbb{R}$, the zeroth order function, and the first order function are given by

$$I_0(x)=\frac{1}{\pi}\int_0^\pi \exp(x\cos\theta)\,d\theta=\sum_{n=0}^{\infty}\frac{(\frac{1}{2}x)^{2n}}{(n!)^2}, \tag{1}$$

$$I_1(x)=\frac{1}{\pi}\int_0^\pi \cos\theta\exp(x\cos\theta)\,d\theta=\frac{d}{dx}I_0(x). \tag{2}$$

A. Matrix Fisher Distribution on $\mathsf{SO}(3)$

Definition 1 *A random rotation matrix $R\in\mathsf{SO}(3)$ is distributed according to a matrix Fisher distribution, if its probability density function is defined relative to the uniform distribution on $\mathsf{SO}(3)$ as*

$$p(R)=\frac{1}{c(F)}\exp(\mathrm{tr}[F^TR]), \tag{3}$$

where $F\in\mathbb{R}^{3\times 3}$, and $c(F)\in\mathbb{R}$ is a normalizing constant defined such that $\int_{\mathsf{SO}(3)}p(R)\,dR=1$. This is denoted by $R\sim\mathcal{M}(F)$.

Throughout this paper, the measure dR is scaled such that $\int_{\mathsf{SO}(3)}dR=1$ [2]. As such, the uniform distribution is given by $p(R)=1$. For the subsequent study, it is convenient to decompose the matrix parameter F as follows.

Definition 2 *([13]) For a given $F\in\mathbb{R}^{3\times 3}$, let the singular value decomposition 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 $U',V'\in\mathbb{R}^{3\times 3}$ are orthonormal matrices. The 'proper' singular value decomposition of F is defined as*

$$F=USV^T, \tag{4}$$

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

$$U=U'\mathrm{diag}[1,\,1,\,\det[U']], \tag{5}$$

$$S=\mathrm{diag}[s_1,\,s_2,\,s_3]=\mathrm{diag}[s_1',\,s_2',\,\det[U'V']s_3'], \tag{6}$$

$$V=V'\mathrm{diag}[1,\,1,\,\det[V']]. \tag{7}$$

The main motivation of the *proper* singular value decomposition is to ensure $\det[U]=\det[V]=+1$ thereby guaranteeing $U,V\in\mathsf{SO}(3)$. Various stochastic properties of (3) have been presented in [10]. The following theorem lists selected properties that are directly relevant to this paper.

Theorem 1 ([10]) *Consider a matrix Fisher distribution given by (3), where the matrix parameter F is decomposed as (4). Suppose $R\sim\mathcal{M}(F)$, and let $Q=U^TRV\in\mathsf{SO}(3)$.*

(i) *$Q\sim\mathcal{M}(S)$.*

(ii) *The normalizing constant satisfies $c(F)=c(S)$, which is given by*

$$c(S)=\int_{-1}^{1}\frac{1}{2}I_0\left[\frac{1}{2}(s_i-s_j)(1-u)\right]\times I_0\left[\frac{1}{2}(s_i+s_j)(1+u)\right]\exp(s_ku)\,du, \tag{8}$$

for any $(i,j,k)\in\mathcal{I}$.

(iii) *The first moment of Q is given by*

$$\mathrm{E}[Q_{ij}]=\begin{cases}\dfrac{1}{c(S)}\dfrac{\partial c(S)}{\partial s_i}=\dfrac{\partial\log c(S)}{\partial s_i} & \text{if } i=j,\\ 0 & \text{otherwise},\end{cases} \tag{9}$$

where $Q_{ij}\in\mathbb{R}$ denotes (i,j)-th element of Q for $i,j\in\{1,2,3\}$.

(iv) *The first moment of R is given by*

$$\mathrm{E}[R]=U\mathrm{E}[Q]V^T. \tag{10}$$

(v) *The max mean attitude and the minimum mean squared error estimate are identical and they are given by*

$$\mathrm{M}_{\max}[R]=\mathrm{M}_{\mathrm{mse}}[R]=UV^T\in\mathsf{SO}(3). \tag{11}$$

Proof: See [10]. ■

The property (ii) implies that the normalizing constant depends only on the proper singular values of F, and it is given by an one-dimensional integral. The property (iv) shows that the first moment, or the arithmetic mean of R is constructed by that of Q, which is given by the normalizing constant and its derivatives. Such arithmetic mean is not necessarily a rotation matrix. Instead the *mean attitude* is formulated as the attitude that maximizes the density value or the attitude that minimizes mean squared error. The property (v) implies both are equivalent, and are given by UV^T. Geometric interpretation of U,S,V in determining the shape of the distribution is presented in [10].

B. First Order Attitude Filter

Consider a stochastic differential equation on $\mathsf{SO}(3)$,

$$(R^TdR)^\vee=\Omega dt+HdW, \tag{12}$$

where $\Omega\in\mathbb{R}^3$ is the angular velocity in the body-fixed frame, $H\in\mathbb{R}^{3\times 3}$ is a diagonal matrix, and $W\in\mathbb{R}^3$ denotes an array of independent, identically distributed Wiener processes. The time variable t is discretized with a fixed step size $h>0$, and let the value of a variable at the k-th time step be denoted by the subscript k. Assuming $R_0\sim\mathcal{M}(F_0)$ for $F_0\in\mathbb{R}^3$, we wish to determine $F_k\in\mathbb{R}^{3\times 3}$ so that $R_k\sim\mathcal{M}(F_k)$.

TABLE I

FIRST-ORDER ATTITUDE ESTIMATION

1: **procedure** FIRST-ORDER ATTITUDE ESTIMATION
2: $R_0 \sim \mathcal{M}(F_0)$, $k = 0$
3: **repeat**
4: $F_{k+1} = $PROPAGATION$(F_k, \Omega_k, H_k)$
5: $k = k + 1$
6: **until** Z_{k+1} or $z_{i_{k+1}}$ is available
7: $F_{k+1} = $CORRECTION$(F_{k+1}, Z_{k+1}, z_{k+1})$
8: **go to** Step 3
9: **end procedure**
10: **procedure** F_{k+1}=PROPAGATION(F_k, Ω_k, H_k)
11: Compute $\mathrm{E}[R_k]$ with F_k from (10)
12: Compute $\mathrm{E}[R_{k+1}]$ with Ω_k and H_k from (13)
13: Perform the singular value decomposition of $\mathrm{E}[R_{k+1}]$ to obtain $\mathrm{E}[R_{k+1}] = U_{k+1}\mathrm{E}[Q_{k+1}]V_{k+1}^T$ as (10)
14: Solve (9) for S_{k+1}
15: Compute F_{k+1} with $U_{k+1}, S_{k+1}, V_{k+1}$ from (4)
16: **end procedure**
17: **procedure** F^+=CORRECTION(F^-, Z)
18: Compute F^+ from (15) with Z
19: **end procedure**

Two types of attitude estimators are introduced in [10]. Here, the first order filtering method is summarized for a single attitude measurements for simplicity. For the prediction step of Bayesian estimation, it is shown that the first moment is propagated as

$$\mathrm{E}[R_{k+1}] = \mathrm{E}[R_k]\left\{I_{3\times 3} + \frac{h}{2}(-\mathrm{tr}[G_k]I_{3\times 3} + G_k)\right\} \times \exp(h\hat{\Omega}_k) + \mathcal{O}(h^{1.5}), \tag{13}$$

where $G_k = H_k H_k^T \in \mathbb{R}^{3\times 3}$. Next, for the correction step, suppose an attitude measurement $Z \in \mathsf{SO(3)}$ is available, which is distributed according to

$$p(Z|R) = \frac{1}{c(F_Z)}\exp(\mathrm{tr}[F_Z^T R^T Z]), \tag{14}$$

with $F_Z \in \mathbb{R}^{3\times 3}$ that specifies the accuracy and the bias of the attitude sensor. It has been shown that for a priori distribution $R_k^- \sim \mathcal{M}(F_k^-)$, the a posteriori distribution conditioned by the measurement is also another matrix Fisher distribution, given by

$$R|Z \sim \mathcal{M}(F + ZF_Z^T). \tag{15}$$

The corresponding filtering scheme integrating the prediction step and the correction step is summarized in Table I.

III. ROBUST NUMERICAL IMPLEMENTATION

When implementing the preceding attitude estimation scheme, one may encounter numerical overflow especially if the proper singular values are relatively large. It is because the modified Bessel function $I_0(x)$, that appears in the calculation of the normalizing constant $c(S)$ in (8), exponentially increases with respect to x as illustrated by the following expansion for large x,

$$I_0(x) = \frac{e^x}{\sqrt{2\pi x}}\left[1 + \frac{1}{8x}\left(1 + \frac{9}{16x}(1 + \cdots)\right)\right].$$

This may cause numerical instability in attitude estimation when the distribution becomes concentrated. In this section, we present an alternative form of the normalizing constant to avoid such issues, and the preceding attitude estimation scheme is reformulated accordingly.

A. Exponentially Scaled Normalizing Constant

In several numerical libraries to compute the modified Bessel function of the first kind, there is an option to compute an exponentially scaled value. More specifically, the exponentially scaled modified Bessel functions of the first kind are defined as

$$\bar{I}_0(x) = \exp(-|x|)I_0(x), \tag{16}$$

$$\bar{I}_1(x) = \exp(-|x|)I_1(x). \tag{17}$$

For example, in Matlab, $I_0(x)$ is computed by the command `besseli(0,x)`, and the scaled value $\bar{I}_0(x)$ can be obtained by `besseli(0,x,1)`. The function $\bar{I}_0(x)$ is differentiable when $x \neq 0$, and from (2)

$$\frac{d\bar{I}_0(x)}{dx} = \bar{I}_1(x) - \mathrm{sgn}[x]\bar{I}_0(x). \tag{18}$$

It is desirable to utilize the scaled modified Bessel functions as $\lim_{x\to\infty}\bar{I}_0(\pm x) = \lim_{x\to\infty}\bar{I}_1(\pm x) = 0$.

Motivated by these, we define the *exponentially scaled* normalizing constant as follows.

Definition 3 *For a matrix Fisher distribution $\mathcal{M}(F)$, let the proper singular value decomposition is given by (4). Its exponentially scaled normalizing constant is defined as*

$$\bar{c}(S) = \exp(-\mathrm{tr}[S])c(S). \tag{19}$$

We show that the exponentially scaled normalizing constant and its derivatives are written in terms of the scaled modified Bessel functions as summarized below.

Theorem 2 *The exponentially scaled normalizing constant for the matrix Fisher distribution (19) satisfies the following properties for any $(i, j, k) \in \mathcal{I}$.*

(i) *$\bar{c}(S)$ is evaluated by*

$$\bar{c}(S) = \int_{-1}^{1}\frac{1}{2}\bar{I}_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \times \exp((\min\{s_i, s_j\} + s_k)(u-1))\,du. \tag{20}$$

(ii) *The first order derivatives of $\bar{c}(S)$ are given by*

$$\frac{\partial\bar{c}(S)}{\partial s_k} = \int_{-1}^{1}\frac{1}{2}\bar{I}_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \times \exp((\min\{s_i, s_j\} + s_k)(u-1))(u-1)\,du. \tag{21}$$

(iii) *The second order derivatives of $\bar{c}(S)$ with respect to s_k are given by*

$$\frac{\partial^2 \bar{c}(S)}{\partial s_k^2} = \int_{-1}^{1} \frac{1}{2}\bar{I}_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \times \exp((\min\{s_i, s_j\} + s_k)(u-1))(u-1)\,du. \quad (22)$$

Also, the second order mixed derivatives are

$$\begin{aligned}\frac{\partial^2 \bar{c}(S)}{\partial s_i \partial s_j} = &\int_{-1}^{1} \frac{1}{4}\bar{I}_1\left[\frac{1}{2}(s_j - s_k)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_j + s_k)(1+u)\right] \\ &\times u(1-u)\exp((s_i + \min\{s_j, s_k\})(u-1)) \\ &+ \frac{1}{4}\bar{I}_0\left[\frac{1}{2}(s_j - s_k)(1-u)\right]\bar{I}_1\left[\frac{1}{2}(s_j + s_k)(1+u)\right] \\ &\times u(1+u)\exp((s_i + \min\{s_j, s_k\})(u-1))\,du \\ &- \frac{\partial \bar{c}(S)}{\partial s_i} - \frac{\partial \bar{c}(S)}{\partial s_j} - \bar{c}(S). \end{aligned} \quad (23)$$

(iv) *The derivatives of $c(S)$ can be rediscovered by*

$$\frac{\partial c(S)}{\partial s_i} = e^{\mathrm{tr}[S]}\left(\bar{c}(S) + \frac{\partial \bar{c}(S)}{\partial s_i}\right), \quad (24)$$

$$\frac{\partial^2 c(S)}{\partial s_i \partial s_j} = e^{\mathrm{tr}[S]}\left(\bar{c}(S) + \frac{\partial \bar{c}(S)}{\partial s_i} + \frac{\partial \bar{c}(S)}{\partial s_j} + \frac{\partial^2 \bar{c}(S)}{\partial s_i \partial s_j}\right). \quad (25)$$

Proof: Substitute (16) into (8), and rearrange. When $s_i \geq s_j$, it reduces to

$$c(S) = e^{s_i}\int_{-1}^{1} \frac{1}{2}\bar{I}_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \times \exp((s_j + s_k)u)\,du,$$

or when $s_j \geq s_i$,

$$c(S) = e^{s_j}\int_{-1}^{1} \frac{1}{2}\bar{I}_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \times \exp((s_i + s_k)u)\,du,$$

for any $(i, j, k) \in \mathcal{I}$. From (19), these show (20). Taking the derivatives of (20) with respect to s_k, it is straightforward to show (21) and (22).

Next, taking the derivatives of (21) with respect to s_i or s_j is cumbersome as $\bar{I}_0(x)$ is not differentiable at $x = 0$. Instead, in [10], the mixed second order derivatives of $c(S)$ has been given by

$$\begin{aligned}\frac{\partial^2 c(S)}{\partial s_i \partial s_j} = &\int_{-1}^{1} \frac{1}{4} I_1\left[\frac{1}{2}(s_j - s_k)(1-u)\right] \\ &\times I_0\left[\frac{1}{2}(s_j + s_k)(1+u)\right] u(1-u)\exp(s_i u) \\ &+ \frac{1}{4} I_0\left[\frac{1}{2}(s_j - s_k)(1-u)\right] \\ &\times I_1\left[\frac{1}{2}(s_j + s_k)(1+u)\right] u(1+u)\exp(s_i u)\,du, \end{aligned} \quad (26)$$

Substitute (16) and (17) to (26) to obtain

$$\begin{aligned}\frac{\partial^2 c(S)}{\partial s_i \partial s_j} = e^{\mathrm{tr}[S]} &\int_{-1}^{1} \frac{1}{4}\bar{I}_1\left[\frac{1}{2}(s_j - s_k)(1-u)\right]\bar{I}_0\left[\frac{1}{2}(s_j + s_k)(1+u)\right] \\ &\times u(1-u)\exp((s_i + \min\{s_j, s_k\})(u-1)) \\ &+ \frac{1}{4}\bar{I}_0\left[\frac{1}{2}(s_j - s_k)(1-u)\right]\bar{I}_1\left[\frac{1}{2}(s_j + s_k)(1+u)\right] \\ &\times u(1+u)\exp((s_i + \min\{s_j, s_k\})(u-1))\,du. \end{aligned} \quad (27)$$

It is straightforward to show (24) and (25) from (19). Substituting (20), (21), and (27) to (25) yields (23). ■

B. Alternative Formulation of Bayesian Attitude Estimation

We can rewrite all of the stochastic properties of the matrix Fisher distribution and the attitude estimation schemes proposed in [10] in terms of the exponentially scaled normalizing constant and its derivatives. These yield more numerically robust implementation that avoid numerical overflows that possibly appear particularly when the proper singular values are large. The following theorem shows two selected properties reformulated in terms of the scaled normalizing constant.

Theorem 3 *Consider a matrix Fisher distribution given by (3), where the matrix parameter F is decomposed as (4). Suppose $R \sim \mathcal{M}(F)$, and let $Q = U^T R V \in \mathsf{SO}(3)$.*

(i) *The probability density can be rewritten as*

$$p(R) = \frac{1}{\bar{c}(S)}\exp(\mathrm{tr}[F^T R] - \mathrm{tr}[S]). \quad (28)$$

(ii) *The first moment of Q is given by*

$$\mathrm{E}[Q_{ij}] = \begin{cases} 1 + \dfrac{1}{\bar{c}(S)}\dfrac{\partial \bar{c}(S)}{\partial s_i} = 1 + \dfrac{\partial \log \bar{c}(S)}{\partial s_i} & \textit{if } i = j, \\ 0 & \textit{otherwise}, \end{cases} \quad (29)$$

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

Using this, (29) replaces (9), and this is useful for the step 14 of Table I. More specifically, the following set of equations should be solved for $s = (s_1, s_2, s_3) \in \mathbb{R}^3$ with given $(\mathrm{E}[Q_{11}], \mathrm{E}[Q_{22}], \mathrm{E}[Q_{33}])$,

$$f(s) = \frac{1}{\bar{c}(S)}\frac{\partial \bar{c}(S)}{\partial s} - \begin{bmatrix} \mathrm{E}[Q_{11}] - 1 \\ \mathrm{E}[Q_{22}] - 1 \\ \mathrm{E}[Q_{33}] - 1 \end{bmatrix} = 0. \quad (30)$$

This can be solved via the following Newton's iteration

$$s^{(q+1)} = s^{(q)} - \left(\left.\frac{\partial f(s)}{\partial s}\right|_{s=s^{(q)}}\right)^{-1} f(s^{(q)}), \quad (31)$$

where the superscript (q) denotes the number of iterations, and the gradient can be computed by

$$\frac{\partial f(s)}{\partial s} = \frac{1}{\bar{c}(S)}\frac{\partial^2 \bar{c}(S)}{\partial s^2} - \frac{1}{\bar{c}(S)^2}\frac{\partial \bar{c}(S)}{\partial s}\left(\frac{\partial \bar{c}(S)}{\partial s}\right)^T. \quad (32)$$

These alternative formulation based on the proposed scaled normalizing constant would avoid numerical overflow for robust implementation.

IV. Approximate Matrix Fisher Distribution on SO(3)

In this section, we present two types of an approximate matrix Fisher distribution when the proper singular values are close to zero, or when they are sufficiently large. The former corresponds to the case of almost uniformly distributed, and the latter represents highly concentrated distributions.

There are two desirable properties. First, the evaluation of the normalizing constant requires a one-dimensional integration as shown by (8) or (19). We formulate an explicit form of the normalizing constant for approximate distributions, and we also provide an explicit solution of (30) for computationally efficient implementation of the filtering schemes. Second, we aim to identify the similarity between the matrix Fisher distribution and other well-known distributions for those special cases.

A. Almost Uniformly Distributed Cases

Suppose $|s_3| \leq s_2 \leq s_1 \ll 1$. When $S = 0_{3\times 3}$, the expression of the probability density (3) reduces to $p(R) = 1$, which represents the uniform distribution on $\mathsf{SO}(3)$. As such, this case corresponds to when the attitude uncertainty distribution is almost uniform. In this case, the normalizing constant is approximated as follows.

Theorem 4 *When $|s_3| \leq s_2 \leq s_1 \ll 1$, the normalizing constant of $\mathcal{M}(S)$ and its derivatives are approximated by*

$$c(S) = 1 + \frac{1}{6}(s_1^2 + s_2^2 + s_3^2) + \frac{1}{6}s_1 s_2 s_3 + \mathcal{O}(3), \quad (33)$$

$$\frac{\partial c(S)}{\partial s} = \frac{1}{3}s + \frac{1}{6}\begin{bmatrix} s_2 s_3 \\ s_3 s_1 \\ s_1 s_2 \end{bmatrix} + \mathcal{O}(2). \quad (34)$$

Proof: We have

$$I_0(x) = \sum_{r=0}^{\infty}(\frac{1}{2}x)^{2r}/(r!)^2 = 1 + \frac{1}{4}x^2 + \mathcal{O}(x^4)$$

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \mathcal{O}(x^3).$$

Substituting these into (8),

$$c(S) \approx \int_{-1}^{1}\frac{1}{2}\left\{1 + \frac{1}{16}(s_1 - s_2)^2(1-u)^2\right\} \times \left\{1 + \frac{1}{16}(s_1 + s_2)^2(1+u)^2\right\} \times \left\{1 + s_3 u + \frac{1}{2}s_3^2 u^2\right\} + \mathcal{O}(s^3)\, du,$$

which follows (33) and (34). ∎

Ignoring the second or higher order terms, the implicit equation (9) can be solved for s_i as

$$s_i = 3\mathrm{E}[Q_{ii}], \quad (35)$$

for $i \in \{1, 2, 3\}$.

B. Highly Concentrated Cases

Next, suppose $1 \ll s_2 + s_3 \leq s_3 + s_1 \leq s_1 + s_2$. Let $Q = U^T R V$ be parameterized as

$$Q(\eta) = \exp(\hat{\eta}) = I_{3\times 3} + \frac{\sin\|\eta\|}{\|\eta\|}\hat{\eta} + \frac{1 - \cos\|\eta\|}{\|\eta\|^2}\hat{\eta}^2$$

for $\eta \in \mathbb{R}^3$ with $\|\eta\| \leq \pi$. The Haar measure is given by

$$dQ = \frac{1 - \cos\|\eta\|}{4\pi^2\|\eta\|^2}d\eta.$$

When $\|\eta\| \ll 1$, it is expanded about $\eta = 0$ as

$$\begin{aligned} Q(\eta) &= I_{3\times 3} + \hat{\eta} + \frac{1}{2}\hat{\eta}^2 + \mathcal{O}(3) \\ &= \begin{bmatrix} 1 - \frac{1}{2}(\eta_2^2 + \eta_3^2) & \frac{1}{2}\eta_1\eta_2 - \eta_3 & \frac{1}{2}\eta_1\eta_3 + \eta_2 \\ \frac{1}{2}\eta_1\eta_2 + \eta_3 & 1 - \frac{1}{2}(\eta_3^2 + \eta_1^2) & \frac{1}{2}\eta_2\eta_3 - \eta_1 \\ \frac{1}{2}\eta_1\eta_3 - \eta_2 & \frac{1}{2}\eta_2\eta_3 + \eta_3 & 1 - \frac{1}{2}(\eta_1^2 + \eta_2^2) \end{bmatrix} \\ &\quad + \mathcal{O}(3). \end{aligned}$$

Therefore, using $dR = d(UQV^T) = dQ \approx \frac{1}{8\pi^2}d\eta$,

$$\begin{aligned} p(R) &\propto \exp(\mathrm{tr}[F^T R])\, dR \\ &= \exp(\mathrm{tr}[SQ(\theta)])\, dQ \\ &= \frac{1}{8\pi^2}\exp(\mathrm{tr}[S])\exp\left\{\prod_{(i,j,k)\in\mathcal{I}} -\frac{1}{2}(s_j + s_k)\eta_i^2\right\} d\eta \\ &\quad + \mathcal{O}(2). \end{aligned} \quad (36)$$

This shows that highly concentrated matrix Fisher distribution $\mathcal{M}(F)$ can be approximated by mutually independent Gaussian distributions for η with the zero mean and the covariance,

$$\mathrm{E}[\eta\eta^T] = \begin{bmatrix} \frac{1}{s_2+s_3} & 0 & 0 \\ 0 & \frac{1}{s_3+s_1} & 0 \\ 0 & 0 & \frac{1}{s_1+s_2} \end{bmatrix}$$

where $\eta \in \mathbb{R}^3$ is defined such hat

$$R = U\exp(\hat{\eta})V^T = UV^T\exp(\widehat{V\eta}) = \exp(\widehat{U\eta})UV^T.$$

As such, η_i corresponds to the angle of rotation of UV^T about the axis whose representation is given by Ue_i in the inertial frame, or equivalently Ve_i in the body-fixed frame.

Theorem 5 *When $1 \ll s_2 + s_3 \leq s_3 + s_1 \leq s_1 + s_2$, the normalizing constant of $\mathcal{M}(S)$ and its derivatives are approximated by*

$$c(S) \approx \frac{\exp(\mathrm{tr}[S])}{\sqrt{8\pi(s_1 + s_2)(s_2 + s_3)(s_3 + s_1)}}, \quad (37)$$

$$\frac{1}{c(S)}\frac{\partial c(S)}{\partial s_i} \approx 1 - \frac{1}{2}\left(\frac{1}{s_i + s_j} + \frac{1}{s_k + s_i}\right). \quad (38)$$

Proof: From (36),

$$c(S) \approx \frac{\exp(\mathrm{tr}[S])}{8\pi^2} \int_{\|\eta\|\leq\pi} \exp\left\{-\frac{1}{2} \sum_{(i,j,k)\in\mathcal{I}} (s_i+s_j)\theta_k^2\right\} d\eta,$$

which can be further approximated with $1 \ll s_i + s_j$ as

$$c(S) \approx \frac{\exp(\mathrm{tr}[S])}{8\pi^2} \int_{\mathbb{R}^3} \exp\left\{-\frac{1}{2} \sum_{(i,j,k)\in\mathcal{I}} (s_i+s_j)\theta_k^2\right\} d\eta.$$

In multivariate Gaussian distributions, it is well known that the above integral is evaluated as

$$\frac{\sqrt{8\pi^3}}{\sqrt{(s_1+s_2)(s_2+s_3)(s_3+s_1)}},$$

which yields (37) and (38). ∎

Similar with the almost uniform distributions, rearranging with (38), the implicit equation (9) can be solved for s_i explicitly as

$$\begin{aligned} s_i = \frac{1}{2}\Bigg\{ &-\frac{1}{1+\mathrm{E}[Q_{ii}]-\mathrm{E}[Q_{jj}]-\mathrm{E}[Q_{kk}]} \\ &+\frac{1}{1-\mathrm{E}[Q_{ii}]+\mathrm{E}[Q_{jj}]-\mathrm{E}[Q_{kk}]} \\ &+\frac{1}{1-\mathrm{E}[Q_{ii}]-\mathrm{E}[Q_{jj}]+\mathrm{E}[Q_{kk}]}\Bigg\}, \end{aligned} \tag{39}$$

for any $(i,j,k) \in \mathcal{I}$.

V. Numerical Examples

A. Normalizing Constant

First, the probability density given by (3) is numerically compared with the alternative expression (28) based on the scaled normalizing constant. For a single variable s, define

$$S(s) = \mathrm{diag}[\frac{1}{2}s,\ \frac{1}{3}s,\ \frac{1}{6}s], \tag{40}$$

with $\mathrm{tr}[S(s)] = s$. The probability density for $\mathcal{M}(S(s))$ at $R = I_{3\times3}$ is numerically evaluated for varying s. When simulated in Matlab with the largest possible floating number of 1.7977×10^{308}, numerical overflow appears when $s > 700$ approximately for (3). However, when using the exponentially scaled expression (28), the overflow happens when $s > 10^8$. This illustrates robustness of the scaled expressions in numerical implementation.

Next, we compare the approximate normalizing constants and their derivatives presented in Section IV with the true values. Figure 1 illustrates the computed values for $S(s)$ in (40) with varying s. It is shown that the approximation for the almost uniform cases is reasonable when $\mathrm{tr}[S] \leq 1$, and the approximation for highly concentrated cases is appropriate when $s_i + s_j \geq 10$.

B. Attitude Estimation

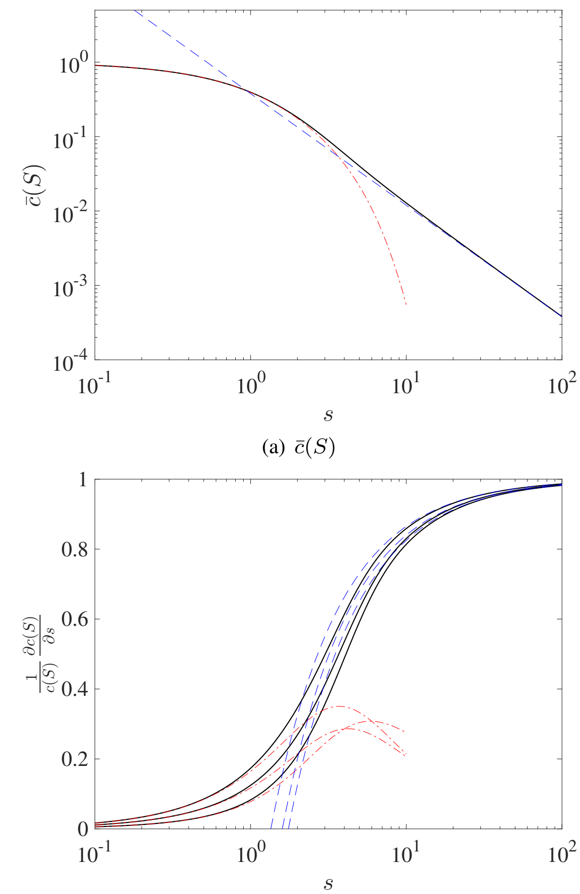
We consider one of the numerical examples of attitude estimation presented in [10]. The reference attitude trajectories is constructed by a complex rotational maneuver of a 3D pendulum [14]. The angular velocity is measured at the rate of 50 Hz with $H = \mathrm{diag}[1.8,\ 1.6,\ 2.4]$, and the attitude is measured at 10 Hz following (14) with $F_Z = \mathrm{diag}[40, 50, 35]$. The mean angular velocity measurement error is $0.45\,\mathrm{rad/s}$, and the mean attitude measurement error is $10.45°$. The initial distribution is given by $R \sim \mathcal{M}(F())$ with

$$F(0) = 0_{3\times3},$$

which corresponds to the uniform distribution on $\mathsf{SO}(3)$, i.e., the initial attitude is completely unknown.

Two cases are considered, depending on how the implicit equation (9) is solved (at the step 14 of Table I): in the first case, it is solved via Newton iteration as (31), and in the second case, the approximate solutions are utilized according to the following logic:

- If $\max\{\mathrm{E}[Q_{ii}]\} < 0.1$, use the almost uniform approximation (35),
- Else if $\max\{\mathrm{E}[Q_{ii}]\} > 0.9$, use the highly concentrated approximation (39),
- Otherwise, perform the Newton iteration with (31).



(b) $\frac{1}{c(S)}\frac{\partial c(S)}{\partial s}$

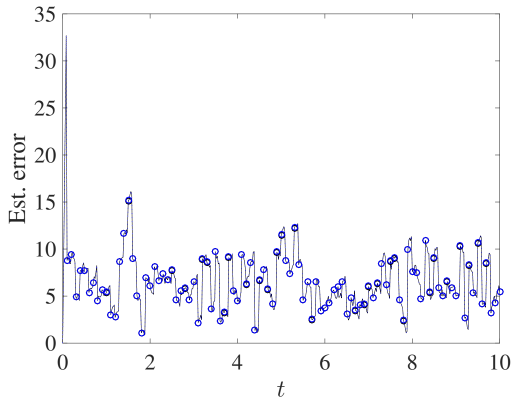
Fig. 1. Comparison between the normalizing constants and approximate ones (solid/black: true, dotted/red:(33),(34), dashed/blue:(37),(38))

The corresponding simulation results are illustrated in Table II and Figure 2. The estimation based on the approximation solution yields slightly greater attitude estimation error, but the difference is negligible. Whereas the computation time becomes drastically reduced.

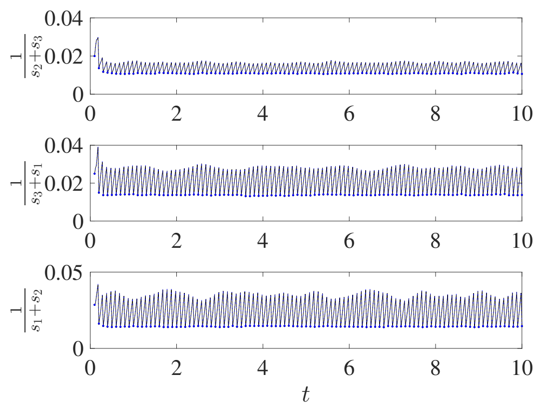
TABLE II

SIMULATION RESULTS

	estimation error	CPU time
Newton iteration	6.48°	95.75 sec
Approx. solution	6.51°	3.70 sec



(a) Attitude estimation error (deg)



(b) Uncertainty measured by $\frac{1}{s_i+s_j}$

Fig. 2. Simulation results: Newton-iteration (black), approximate solution (blue)

VI. CONCLUSIONS

The matrix Fisher distribution on the special orthogonal group is a compact form of attitude probability densities that resemble the Gaussian distribution on $\mathbb{R}^3$, and it has been successfully utilized in Bayesian attitude estimation. This paper has presented a robust numerical implementation based on an exponentially scaled normalizing constant, and two approximate distributions for almost uniformly distributed cases and highly concentrated cases. It turns out that a highly concentrated matrix Fisher distribution can be approximated by a Gaussian distribution for the rotation angles from the mean attitude. Also, it is shown that the approximated distributions yield an explicit form of Bayesian attitude estimation that reduces the computational load substantially, while causing minimal performance degradation. Future works include generalizing the matrix Fisher distribution to model the coupling between the uncertainties in the rotational dynamics and the translational dynamics of a rigid body.

REFERENCES

[1] J. T. H. Lo and L. R. Eshleman, “Exponential Fourier densities on SO(3) and optimal estimation and detection for rotational processes,” *SIAM Journal on Applied Mathematics*, vol. 36, no. 1, pp. 73–82, 1979.

[2] G. Chirikjian and A. Kyatkin, *Engineering applications of noncommutative harmonic analysis*. Boca Raton, FL: CRC Press, 2001.

[3] F. Markley, “Attitude filtering on SO(3),” in *Proceedings of the AAS Malcolm D. Shuster Astronautics Symposium*, 2005, AAS 05-460.

[4] W. Park, J. S. Kim, Y. Zhou, N. J. Cowan, A. M. Okamura, and G. S. Chirikjian, “Diffusion-based motion planning for a nonholonomic flexible needle model,” in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 2005.

[5] T. Lee, M. Leok, and N. McClamroch, “Global symplectic uncertainty propagation on SO(3),” in *Proceedings of the IEEE Conference on Decision and Control*, Cancun, Mexico, Dec. 2008, pp. 61–66.

[6] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. Chirikjian, “Kinematic state estimation and motion planning for stochastic nonholonomic system using the exponential map,” *Robotica*, vol. 26, pp. 419–434, 2008.

[7] G. Chirikjian and M. Kobilarov, “Gaussian approximation of nonlinear measurement models on Lie groups,” in *Proceedings of the IEEE Conference on Decision and Control*, 2014, pp. 6401–6406.

[8] T. Downs, “Orientation statistics,” *Biometrika*, vol. 59, pp. 665–676, 1972.

[9] C. Khatri and K. Mardia, “The von Mises-Fisher matrix distribution in orientation statistics,” *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 95–106, 1977.

[10] T. Lee, “Bayesian attitude estimation with the matrix Fisher distribution on SO(3),” *IEEE Transactions on Automatic Control*, 2018, accepted.

[11] ——, “Global unscented attitude estimation via the matrix Fisher distributions on SO(3),” in *Proceedings of the American Control Conference*, Boston, MA, Jul. 2016, pp. 4942–4947.

[12] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. Dover, 1965.

[13] F. Markley, “Attitude determination using vector observations and the singular value decomposition,” *The Journal of the Astronautical Sciences*, vol. 36, no. 3, pp. 245–259, 1988.

[14] T. Lee, N. Chaturvedi, A. Sanyal, M. Leok, and N. McClamroch, “Propagation of uncertainty in rigid body attitude flows,” in *Proceedings of the IEEE Conference on Decision and Control*, New Orleans, LA, Dec. 2007, pp. 2689–2694.