

**Geometric Formulation of Uncertainties and Estimation
for Three-Dimensional Rotations**

by Weixin Wang

B.S. in Mechanical Engineering, May 2016, Tsinghua University
M.S. in Mechanical Engineering, December 2018, University of Wisconsin-Madison

A Dissertation submitted to

The Faculty of
The School of Engineering and Applied Science
of the George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

July 21, 2022

Dissertation directed by

Taeyoung Lee
Professor of Mechanical and Aerospace Engineering

© Copyright 2022 by Weixin Wang
All rights reserved

Acknowledgments

Here you can acknowledge all of those people who have helped you to reach this point. It's rare that any work is done in a vacuum and your research is no exception. Feel free to be grateful for all those who've aided you along your way.

Abstract

Geometric Formulation of Uncertainties and Estimation for Three-Dimensional Rotations

Modeling the uncertainty of three dimensional attitude is critical in numerous estimation problems in aerospace engineering and robotics. However, the widely used Gaussian distribution is unable to accurately model large attitude uncertainties, due to the special geometric structure of the space for 3D rotations, namely the three dimensional special orthogonal group $\text{SO}(3)$. In this dissertation, the matrix Fisher distribution developed in directional statistics is used to model the attitude uncertainty. The matrix Fisher distribution is defined intrinsically on $\text{SO}(3)$, so it can accurately model arbitrarily large attitude uncertainties in a global fashion while fully respecting the underlying geometric structure of $\text{SO}(3)$.

Although various properties of the matrix Fisher distribution have been studied in directional statistics, calculating its normalizing constant and moments still remains challenging. In this dissertation, a recursive algorithm is proposed to calculate the central moment of matrix Fisher distribution up to an arbitrary order. Also, an approximation for the matrix Fisher distribution is developed when it is highly concentrated in two degrees of freedom. And based on this, approximate expressions for the normalizing constant and its derivatives are provided to reduce computation demand.

Next, a new probability density function on $\text{SO}(3) \times \mathbb{R}^n$, referred to as the matrix Fisher–Gaussian distribution (MFG) is proposed to model the correlation between attitude and other random variables in the Euclidean space. MFG greatly generalizes the application of matrix Fisher distribution, as in a lot of practical estimation problems, the attitude must be concurrently estimated with other quantities in the Euclidean space, and the correlation between them is crucial to transfer information from one to the other. In addition, the Bingham–Gaussian distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ which is equivalent to MFG is constructed using the Lie group homomorphism from \mathbb{S}^3 to $\text{SO}(3)$. Various properties of these two distributions are studied, and a closed form approximate maximum likelihood estimation is

presented for fast inference of their parameters required for real time implementations.

Finally, MFG is applied to three representative estimation problems, namely attitude estimation, IMU-GNSS integration, and visual-inertial odometry. Specifically, the observability of attitude with angular velocity and single direction measurements is studied, where two new observable conditions are discovered. Then, MFG is used to design a recursive Bayesian filter to estimate the attitude and gyroscope bias concurrently with attitude or direction measurements. The uncertainty propagation step in this filter is further generalized into a full inertial navigation setting, where the estimated attitude is used to transform the accelerometer readings into the inertial frame which is integrated twice into position. Based on this, a filter for IMU-GNSS navigation is developed with position measurements. Furthermore, the MFG is also used to model the six dimensional pose that aligns two point sets. This is combined with the uncertainty propagation scheme to design a visual-inertial odometry/navigation algorithm. The proposed filters based on MFG are compared with the Gaussian distribution based multiplicative extended Kalman filter, where it is validated through simulation studies that the MFG filters have better accuracy in very challenging cases when the attitude has large uncertainty.

Table of Contents

Acknowledgments	iv
Abstract	v
List of Figures	x
List of Tables	xiii
List of Abbreviations	xv
List of Symbols	xvi
Glossary of Terms	xviii
1 Introduction	1
1.1 Motivation and Goal	2
1.2 Literature Review	4
1.2.1 Directional Statistics	5
1.2.2 State Estimation on SO(3)	8
1.2.3 Visual-Inertial Navigation and Odometry	13
1.3 Summary of Contributions	15
1.3.1 Publications	17
1.4 Outline of Dissertation	18
2 Matrix Fisher Distribution on SO(3)	19
2.1 Properties of SO(3) as a Lie Group	19
2.2 Matrix Fisher Distribution	22
2.2.1 Definition and Properties	23
2.2.2 Construction	26
2.3 Bingham Distribution	27
2.3.1 Definition and Properties	28
2.3.2 Equivalence With the Matrix Fisher Distribution	29
2.3.3 Normalizing Constant	32
2.4 Central Moments of Matrix Fisher Distribution	34
2.4.1 The Central Moments	34
2.4.2 Calculating $\partial_{ms'_\alpha}$	37
2.4.3 Calculating $\partial^k c(S+T)/(\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k})$	41
2.4.4 Numerical Verification	42
2.5 Highly Concentrated Approximations of Matrix Fisher Distribution	43
2.5.1 Highly Concentrated in Three Degrees of Freedom	44
2.5.2 Highly Concentrated in Two Degrees of Freedom	47
3 Matrix Fisher-Gaussian Distribution	53
3.1 Definition and Construction	54

3.1.1	Definition	54
3.1.2	Construction	59
3.1.3	Distinction Between MFGI and MFGB	62
3.2	Bingham–Gaussian Distribution	66
3.3	Properties	68
4	Bayesian Attitude Estimation with Matrix Fisher-Gaussian Distribution	76
4.1	Attitude Observability with Single Direction Measurement	77
4.1.1	Problem Formulation	78
4.1.2	Uncertainty Propagation	80
4.1.3	Measurement Update	83
4.1.4	Stochastic Observability Criterion	86
4.1.5	Attitude Observability Results	92
4.1.6	Simulations and Experiments	99
4.2	Attitude Estimation	107
4.2.1	Problem Formulation	108
4.2.2	Uncertainty Propagation	110
4.2.3	Measurement Update	118
4.2.4	Numerical Simulations	122
5	Bayesian Attitude and Position Estimation with Matrix Fisher-Gaussian Dis- tribution	135
5.1	Loosely Coupled IMU-GNSS Integration	135
5.1.1	Problem Formulation	136
5.1.2	Uncertainty Propagation	138
5.1.3	Measurement Update	140
5.1.4	Numerical Simulations	146
5.2	Visual-Inertial Odometry	152
5.2.1	Pose Estimation Without Map Uncertainty	153
5.2.2	Numerical Simulations Without Map Uncertainty	169
5.2.3	Pose Estimation With Map Uncertainty	175
5.2.4	Numerical Simulations With Map Uncertainty	185
5.2.5	Applications to the VCU-RVI Dataset	193
6	Conclusions	202
6.1	Summary of Contributions	202
6.2	Future Directions	202
Bibliography	204	
A A Special Case in Theorem 2.4	214	
B The Second and Third Order Moment of Matrix Fisher Distribution	218	
B.1	The Second Order Moments	218
B.2	The Third Order Moments	220
C Non-uniqueness Parameters of MFG	222	

D Deterministic Attitude Observability	227
E Proof for Theorem 4.6	229
F Propagating Moments for IMU Kinematics	233
G Posterior Density With Map Uncertainty	239

List of Figures

1.1	Gaussian density vs. wrapped Gaussian density.	3
2.1	Maximum difference between Monte Carlo method and the proposed recursive calculation when s_2 is close to s_3 for moments of the matrix Fisher distribution.	43
2.2	Comparison of the normalizing constant $c(S)$ calculated using different methods.	46
2.3	Comparison of the first order moment $1 - d_1$ calculated using different methods.	47
2.4	Comparison of the normalizing constant $c(S)$ calculated using different methods.	51
2.5	Comparison of the first order moment $1 - d_1$ calculated using different methods.	51
2.6	Comparison of the first order moment $1 - d_2$ calculated using different methods.	52
3.1	The illustration of the cylindrical distribution with the mean angle $\pi/4$ [81]. .	59
3.2	Visualization of the attitude-linear correlation (MFGB) for $(R, x) \in \text{SO}(3) \times \mathbb{R}^1$.	63
3.3	Difference between MFGB and MFGI.	65
3.4	The graph of $r(s)$ against s	75
4.1	Illustration of propagated uncertainties by the right-trivialized (4.5) and the left-trivialized (4.6).	84
4.2	Posterior distribution with single direction measurements.	85
4.3	Estimation with the right trivialized (4.5) and the inertial direction measurement (4.3).	95
4.4	Estimation with the left trivialized (4.6) and the body-fixed direction measurement (4.4).	96
4.5	Attitude errors for the matrix Fisher (MF) estimator and MEKF in four combinations of angular velocity and reference vector measurements.	102
4.6	Attitude standard deviations for the matrix Fisher (MF) estimator and MEKF.	103
4.7	Hardware platform for experiments.	104

4.8	Attitude errors for the matrix Fisher (MF) estimator and MEKF in four combinations of measurements with experimental data.	105
4.9	Attitude uncertainty for the matrix Fisher estimator (MF) and MEKF in four combinations of measurements with experimental data.	106
4.10	Estimation errors for varying accuracies of the second vector measurement. .	126
4.11	Full attitude error for example simulations.	127
4.12	Bias error (0 s to 300 s) in an example simulation ($\sigma_2^2 = \infty$).	128
4.13	Magnetic field on the orbit with different inclinations.	130
4.14	Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 0^\circ$	132
4.15	Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 40^\circ$	133
4.16	Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 90^\circ$	133
5.1	Estimation errors with large initial attitude error.	150
5.2	Estimation standard deviations with large initial error.	151
5.3	KL divergence of the MFG and LM-optimization methods.	172
5.4	Trajectory and landmarks for visual-inertial navigation.	174
5.5	Error trajectory of a particular simulation for visual-inertial navigation without map noises.	176
5.6	Estimation error of visual-inertial navigation with map noises.	190
5.7	Trajectory and landmarks of visual-inertial navigation with map noises. . . .	191
5.8	Estimation error of visual-inertial odometry with map noises.	193
5.9	Trajectory and features of visual-inertial odometry with map noises.	194
5.10	Estimation errors for the “lab_bumper1” trial.	197
5.11	Estimation errors for the “lab_bumper2” trial.	198
5.12	Estimation errors for the “lab_bumper3” trial.	199

5.13 Estimation errors for the “lab_bumper4” trial.	200
5.14 Estimation errors for the “lab_bumper5” trial.	201

List of Tables

2.1	Maximum difference of moments between Monte Carlo method and the proposed recursive algorithm	43
4.1	Attitude observability with single direction measurements	78
4.2	Estimation error (deg) for the four combinations of measurements	101
4.3	Analytical uncertainty propagation for gyroscope kinematics	114
4.4	Unscented uncertainty propagation for gyroscope kinematics	118
4.5	Bayesian estimation for attitude and gyroscope bias	121
4.6	Attitude (deg) and bias (deg/s) errors \pm S.D. with attitude measurements . .	124
4.7	Attitude (deg) and bias (deg/s) errors (S.D.) for different σ_2	125
4.8	Estimation errors with magnetometer measurements	131
5.1	Uncertainty propagation for IMU kinematics	140
5.2	Attitude progressive update	145
5.3	Bayesian estimation for IMU-GNSS navigation	147
5.4	Estimation errors with small initial attitude error \pm S.D. (p -value)	149
5.5	Estimation errors with large initial attitude error \pm S.D. (p -value)	149
5.6	Attitude Estimation From Marginal Likelihood	161
5.7	Pose estimation without map noises	165
5.8	Measurement update without map noises	168
5.9	Estimation error of the pose that aligns two point sets	170
5.10	Estimation error \pm S.D. (p -value) of the pose for a degenerative configuration	173
5.11	Estimation error for visual-inertial navigation without map noises	175
5.12	Pose and map coordinates estimation with map noises	183
5.13	Measurement update with map noises	186

5.14	Pose and map coordinates errors with map uncertainty	187
5.15	Estimation error for visual-inertial navigation with map noises	189
5.16	Estimation error for visual-inertial odometry with map noises	192
5.17	Estimation errors for the VCU_RVI dataset	196

List of Abbreviations

BCH Baker–Campbell–Hausdorff Formula

BG Bingham–Gaussian Distribution

EKF Extended Kalman Filter

GNSS Global Navigation Satellite System

IEKF Invariant Extended Kalman Filter

IMU Inertial Measurement Unit

MEKF Multiplicative Extended Kalman Filter

MFG Matrix Fisher–Gaussian Distribution

MLE Maximum Likelihood Estimation

MMSE Minimum Mean Square Estimate

MSCKF Multi-State Constraint Kalman Filter

pSVD proper Singular Value Decomposition

SLAM Simultaneous Localization and Mapping

UKF Unscented Kalman Filter

List of Symbols

$\mathcal{BG}(\mu, \Sigma, P, M, Z)$ The Bingham–Gaussian distribution with parameters (μ, Σ, P, M, Z) .

$\mathcal{B}(A)$ The Bingham distribution with parameter A .

$[A, B]$ The Lie bracket for $A, B \in \mathfrak{so}(3)$.

$\mathcal{C}(n, k)$ The set of all k -partitions of the set $\{1, 2, \dots, n\}$. For example $\mathcal{C}(3, 2) = \left\{ \{\{1\}, \{2, 3\}\}, \{\{2\}, \{1, 3\}\}, \{\{3\}, \{1, 2\}\} \right\}$

$c(S)$ The normalizing constant of the matrix Fisher distribution with parameter S .

$\text{diag}(x_1, \dots, x_n)$ The (block) diagonal matrix with the diagonal terms given by x_1, \dots, x_n .

e_i The i -th column of $I_{n \times n}$, where n depends on the context.

$\text{etr}(A)$ An abbreviation of $\exp(\text{tr}(A))$.

$\exp(x)$ The exponential of a number $x \in \mathbb{R}$, or the exponential map of $x \in \mathfrak{so}(3)$.

$\mathbb{E}[x]$ The expectation of a random element x .

$\Gamma(x)$ Gamma function.

\hat{x} The hat map that identifies \mathbb{R}^3 to $\mathfrak{so}(3)$.

x^\vee The vee map, inverse of the hat map.

\mathcal{I} The Fisher information.

$I_\nu(x)$ Modified Bessel function of the first kind with order ν .

$\mathcal{M}(F)$ The matrix Fisher distribution with parameter F .

$\mathcal{MG}(\mu, \Sigma, P, U, S, V)$ The matrix Fisher–Gaussian distribution with parameters $(\mu, \Sigma, P, U, S, V)$.

$\mathcal{N}(\mu, \Sigma)$ Normal distribution with mean μ , and covariance matrix Σ .

\otimes Quaternion multiplication or Kronecker product depending on the context.

φ The homomorphism from \mathbb{S}^3 to $\text{SO}(3)$.

\mathbb{R}^n The n -dimensional Euclidean space.

$\text{SE}(n)$ The n -dimensional Euclidean group.

$\text{SO}(n)$ The n -dimensional special orthogonal group.

$\mathfrak{so}(n)$ The Lie algebra of $\text{SO}(n)$.

\mathbb{S}^n Unit n -sphere.

$\text{SU}(n)$ The n -dimensional special unitary group.

$\text{tr}(A)$ The trace of matrix A .

$\text{vec}(A)$ Concatenation of the columns of matrix A into a column vector.

$\mathcal{VM}(\mu, \kappa)$ The von Mises (Fisher) distribution with mean direction μ , and concentration parameter κ .

\mathbb{Z} The space of integers.

Glossary of Terms

body-fixed frame The reference frame fixed to a rigid body.

Haar measure The bi-invariant measure on $\text{SO}(3)$ for integration of functions on $\text{SO}(3)$. It is normalized such that $\text{SO}(3)$ has unit volume.

inertial frame The reference frame that is assumed to be fixed in a navigation problem.

KL-divergence A metric that measures the difference between two probability distributions.

LM algorithm An iterative gradient based algorithm to solve nonlinear minimum least square problems.

maximum likelihood estimation Inference of the parameters of a probability distribution by maximizing the likelihood function of measurements.

measurement update Calculate the posterior density function from the prior and likelihood, using the Bayes' formula.

moment The expectation of a random variable.

normalizing constant A constant for an un-normalized probability density function such that it integrates into one.

recursive Bayesian filter A stochastic estimator that estimates the state of a dynamical system by calculating the posterior distribution conditioned on the history of all measurements.

Stiefel manifold An orthogonal r frame in \mathbb{R}^n with $r \leq n$.

uncertainty propagation Calculate how the probability density function of the state of a dynamical system evolves in time.

Chapter 1: Introduction

The 3D attitude of a rigid body is a key concept in modeling our physical world. It can be defined as the directions of the three axes of an orthogonal frame fixed to the rigid body in the inertial frame, and therefore describes the orientation of the rigid body. The attitude is a state involved with rigid body mechanics, and is usually controlled in a robotic system through actuators to achieve the goal of alignment or moving to a specific position. As a feedback controller also needs the estimation of the state, a number of sensors are usually used to determine the attitude of the rigid body. These sensors can usually be classified into the following categories: (i) infrastructures that measure the attitude directly, for example an optical motion capture system; (ii) a gyroscope that measures the angular velocity; (iii) direction sensors that measure reference directions in the inertial frame, such as magnetometers, star trackers, etc; (iv) sensors that measure the surrounding environment, such as various cameras, lidars, etc.

These sensors provide either direct or indirect information of the attitude of a rigid body. But one common feature is that they are all subject to measurement noises. The noise characteristics of different types of measurement are also different. For example, the attitude integrated from angular velocity measured by a gyroscope suffers from accumulated integration of bias, and therefore it is reliable within a short period of time. It is similar for a visual odometry algorithm that estimates the attitude by capturing surrounding visual features using cameras. On the other hand, the attitude inferred from measurements of reference directions may be affected by instantaneous disturbances, but the average noise remains low for a long period of time. Thus, it is common that different sensors are combined to obtain a robust attitude estimate, by leveraging their different noise characteristics.

This motivates the need to model the uncertainty of the attitude. As when combining

the estimates from two or more different measurement sources, one must decide how much confidence should be placed on each individual source. As a rule of thumb, the measurement with more uncertainty should play a less important role in determining the final estimation. Consequently, a metric that describes the uncertainty for the estimated attitude is needed to weigh the measurements. From another point of view, uncertainty of the final estimation can play an important role in decision making for a robotic system. For example, if the uncertainty for attitude is too large, a robot may take some safety measures like slowing down, to avoid possible collisions.

This dissertation focuses on modeling the uncertainty of the rigid body attitude in a geometric and global fashion. More specifically, the uncertainty is formulated on the curved and compact manifold where the attitude naturally resides in an intrinsic fashion. Furthermore, a model for the correlation between attitude and other Euclidean quantities, for example the 3D location, is proposed. Based on these, three typical estimation problems in aerospace and robotics engineering are investigated: (i) attitude estimation with a gyroscope and direction sensors, (ii) loosely coupled IMU-GNSS integration for navigation, and (iii) visual-inertial navigation and odometry with an RGB-D camera.

1.1 Motivation and Goal

In the existing algorithms for attitude estimation from different sensors, the uncertainty of attitude is usually modeled in two different ways. In the first way, attitude parameterizations, such as Euler angles or quaternions, are treated directly as quantities in the Euclidean space, and the Gaussian distribution is used to model their uncertainty. This method entirely ignores the geometric constraints on the attitude parameterizations, such as the singularity of Euler angles, or that a quaternion representing attitude should have unit norm. As a result, its accuracy is typically low, and may have numerical issues associated with singularities, or that the covariance matrix becomes singular.

Instead of directly treating the attitude parameterizations as in Euclidean space, the more

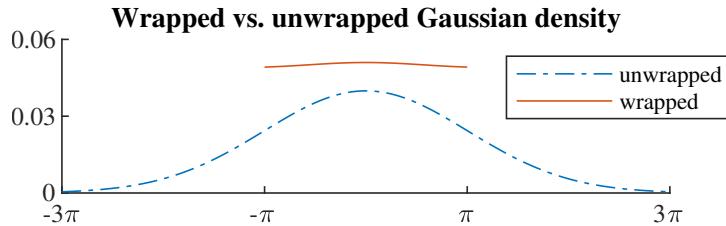


Figure 1.1: Gaussian density vs. wrapped Gaussian density with $\mu = 0$ and $\sigma^2 = 10$. The wrapped Gaussian distribution is defined in the circular space $\mathbb{R}/2\pi$ by identifying θ with $\theta + 2k\pi$ for $k \in \mathbb{Z}$, so the densities beyond $[-\pi, \pi]$ are wrapped into $[-\pi, \pi]$.

popular second approach treats the 3-dimensional parameterizations of attitude *error* as in Euclidean space. In this approach, since the attitude error is typically small, it avoids the singularity of the parameterization, which is usually at 180 degrees. Also, by using the minimal three dimensional parameterization, it avoids an equality constraint that makes the covariance matrix for a unit quaternion singular. The parameterizations frequently used include rotation vectors and modified Rodrigues parameters.

Although the second approach of modeling attitude uncertainty seems very reasonable, and usually achieves good estimation accuracy of attitude in practical applications, it still does not faithfully respect the geometric structure of the attitude space. For example, the space of rotation vectors is actually cyclic, i.e., any rotation vectors with the same direction that differ in length by $2k\pi$, $k \in \mathbb{Z}$, represent the same rotation. As a consequence, a Gaussian distribution for rotation vectors needs to be *wrapped*, or *folded* in order to faithfully represent the uncertainty. Namely, the densities representing the same rotation need to be added up. An example of a wrapped Gaussian distribution in the one dimensional rotation space is shown in Fig. 1.1. If the uncertainty is low, or the variance of the Gaussian distribution is small, the error caused by wrapping can usually be neglected. However, for large uncertainty, as indicated in Fig. 1.1, wrapping makes the uncertainty much larger than that indicated by the original Gaussian distribution.

In this dissertation, an alternative approach is sought to model large uncertainty of rigid body attitude, which cannot be accurately described by a Gaussian distribution as discussed

above. The key idea is that instead of treating parameterizations of attitude as quantities in Euclidean space, a probability density defined directly in the space of attitude is used. Classic densities include the matrix Fisher distribution defined on the three dimensional special orthogonal group $\text{SO}(3)$, and the Bingham distribution defined on the unit three sphere \mathbb{S}^3 . In this way, the geometric constraints on the parameterizations of attitude are completely avoided, and the densities are suitable for arbitrarily large uncertainty including a uniform distribution, which is made possible by the compactness of the attitude space.

To apply these densities in estimation problems encountered in engineering, the correlation between attitude and other Euclidean quantities must also be considered, since attitude is usually jointly estimated with these quantities, such as the gyroscope bias, and the position. This issue is investigated in this dissertation by introducing a new probability density defined on $\text{SO}(3) \times \mathbb{R}^n$, referred to as the matrix Fisher–Gaussian (MFG) distribution. With this new tool, classic estimation problems in engineering, including attitude determination and localization, can be tackled with probabilistic models that fully respect the underlying geometric structure of the attitude space. Therefore, in the presence of large uncertainty, for example unknown initial conditions, estimators can converge with a much faster speed than the conventional Gaussian approach, thanks to the accurate modeling of large uncertainty.

1.2 Literature Review

Probabilistic models that describe uncertainty on curved manifolds have been studied by statisticians in the branch called *directional statistics*. In this section, the developments in directional statistics that are relevant to this dissertation are reviewed. Then, classic methods for rigid body estimation using gyroscopes and direction sensors are introduced. Moreover, some attempts in applying the matrix Fisher distribution, and Bingham distribution to attitude estimation are examined. Finally, the progresses in rigid body pose estimation (attitude and position) using various cameras with an IMU (inertial measurement unit), are reviewed.

1.2.1 Directional Statistics

Although the Gaussian distribution is the dominate probabilistic model in both theoretical research and practical engineering, it falls short when the data of interest are on a curved manifold, such as the unit sphere \mathbb{S}^2 , and the special orthogonal group $\text{SO}(3)$. Statisticians have proposed new models directly defined on these manifolds, to replace the Gaussian distribution when the uncertainty of data is large. The materials in this subsection are mostly based on the book [80]. Another useful exposition of models on the Stiefel and Grassmann manifolds can be found in [18].

Models on Unit Circle and Sphere One of the most frequently encountered and simplest manifold is the one dimensional unit circle: $\mathbb{S}^1 = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$. Any element on the unit circle can be written as an angle $\theta \in \mathbb{R}$, however, θ and $\theta + 2k\pi, k \in \mathbb{Z}$ represent the same element due to the cyclic nature of \mathbb{S}^1 . As a consequence, the unit circle can also be regarded as the space of one dimensional rotation, i.e., the two dimensional special orthogonal group $\text{SO}(2) = \{R \in \mathbb{R}^{2 \times 2} \mid RR^T = I_{2 \times 2}, \det R = 1\}$, where

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Various probability densities have been developed on \mathbb{S}^1 , such as the von Mises distribution [116], wrapped normal (Gaussian) distribution, wrapped Cauchy distribution, projected normal distribution (angular Gaussian distribution), etc. The first three distributions are uni-modal and have similar shapes (see [80, Figure 3.3]), whereas the last can be bi-modal. Among these, the von Mises distribution is one of the most frequently used models, and can be regarded as an analogue of Gaussian distribution on \mathbb{S}^1 , due to their several similarities.

The von Mises distribution can be generalized into (hyper) unit spheres $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \mid \|x\| = 1\}$, for $n \geq 2$, and is usually named as von Mises–Fisher distribution [28]. If

$x \in \mathbb{S}^n$ follows a von Mises–Fisher distribution, it has the following density function

$$p(x) = \left(\frac{\kappa}{2}\right)^{(n+1)/2-1} \frac{1}{\Gamma((n+1)/2)I_{(n+1)/2-1}(\kappa)} \exp\{\kappa\mu^T x\}$$

with respect to the uniform distribution, where $\mu \in \mathbb{S}^n$, $\kappa > 0$ are the parameters of the distribution, Γ is the gamma function, and I_v is the modified Bessel function of the first kind with order v . For the unit sphere \mathbb{S}^2 , the density function reduces to

$$p(x) = \frac{\kappa}{\sinh \kappa} \exp\{\kappa\mu^T x\}. \quad (1.1)$$

This distribution is denoted by $x \sim \mathcal{VM}(\mu, \kappa)$. The parameter μ represents the mean direction, and κ represents the concentration, with larger κ meaning higher concentration. The von Mises–Fisher distribution is used to model the uncertainty of direction measurements in this dissertation.

Models on $\text{SO}(3)$ Another frequently encountered manifold in engineering and material science is the space of 3D rotations, i.e., the three dimensional special orthogonal group: $\text{SO}(3) = \{R \in \mathbb{R}^{3 \times 3} \mid RR^T = I_{3 \times 3}, \det R = 1\}$. One of the most famous models on $\text{SO}(3)$ is the matrix Fisher distribution [25, 54], which is originally defined on the Stiefel manifolds, i.e., an orthogonal r frame in \mathbb{R}^n with $r \leq n$. When $r = 2$, $n = 3$, the Stiefel manifold is $\text{SO}(3)$ [80, Chapter 13.2.1]. Similar to von Mises distribution, the matrix Fisher distribution can be regarded as an analogue of the Gaussian distribution on $\text{SO}(3)$.

Any 3D rotation can also be parameterized by a unit quaternion $q \in \mathbb{S}^3$. The space of unit quaternions \mathbb{S}^3 forms a double cover of $\text{SO}(3)$, therefore, q and $-q$ represent the same 3D rotation. The Bingham distribution [12] is formulated on \mathbb{S}^n , $n \geq 1$ with antipodal symmetry, i.e., the densities for x and $-x$ are identical. Thus, the Bingham distribution on \mathbb{S}^3 can be treated as a probability distribution on $\text{SO}(3)$. In fact, it has been shown to be equivalent to the matrix Fisher distribution on $\text{SO}(3)$ [99]. The definition, construction, and properties of

these two distributions will be introduced in detail in Chapter 2.

Besides these two popular models, a “normal” like distribution on $\text{SO}(3)$ is proposed in [94] using the central limit theorem, the Cayley distribution is introduced in [70] using the Cayley transform, and some other density functions are formulated in [17, 86]. These (including the matrix Fisher distribution) are further summarized under a generic approach of constructing probability distributions on $\text{SO}(3)$ proposed in [102] based on the exponential map, called the UARS model. Moreover, the authors in [102] introduced the wrapped tri-variate normal distribution on $\text{SO}(3)$, generalizing the wrapped normal distribution on \mathbb{S}^1 by assuming the rotation vector follows a Gaussian distribution.

Correlations with Euclidean Space In practical applications, data on manifolds are usually coupled with other data in the Euclidean space, such as the attitude and position of a rigid body. Therefore, it is also necessary to model the correlation between the manifold and Euclidean space. In [81], the von Mises distribution on \mathbb{S}^1 and the Gaussian distribution on \mathbb{R}^1 are coupled to define a density on the cylinder $\mathbb{S}^1 \times \mathbb{R}^1$, where some other models [1, 48, 52] have also been introduced to model the correlation between \mathbb{S}^1 and \mathbb{R}^1 . In addition, the von Mises distribution has also been generalized to multi-variate von Mises distribution [79] to model the correlation between pairs of several random angles in \mathbb{S}^1 .

There have also been two attempts to couple $\text{SO}(3)$ and \mathbb{R}^n . In [85], the matrix Fisher distribution is coupled with the Gaussian distribution by making the parameter of the matrix Fisher part dependent on the random variable in the Euclidean space. The same approach has also been used to couple the Bingham distribution and Gaussian distribution [24]. However, this coupling technique, although being reasonable, is not originated from a generic geometric construction principle. As a consequence, these models require very difficult computations when applied to practical problems, especially numerical optimizations are required when performing maximum likelihood estimation, i.e., recovering parameters from data, which limits their applications.

1.2.2 State Estimation on $\text{SO}(3)$

Rigid body attitude estimation is the basic algorithm for numerous robotics and aerospace applications, and has been studied extensively by researchers and practitioners. For example, in inertial navigation, the acceleration measurement must be first transformed into the inertial frame, which requires the attitude of the accelerometer to be estimated. For a bipedal robot, its inclination must be first estimated so that the controller can prevent it from falling. For a telescope operating in space, such as the Hubble telescope, its attitude must be known to direct its lens to a specific area of space.

Attitude estimation algorithms can be divided into the following three categories: (i) deterministic observers with guaranteed convergence; (ii) nonlinear Kalman filter based stochastic filters; and (iii) stochastic filters based on the matrix Fisher and Bingham distributions. But before reviewing these three categories, attitude parameterizations must be first introduced.

Attitude Parameterizations The attitude of a rigid body has three degrees of freedom. However, a rotation matrix $R \in \text{SO}(3)$ has 9 parameters, and is therefore subject to 6 equality constraints. Due to its redundancy, few stochastic filters directly operate on rotation matrices. Instead, people seek for attitude representations with fewer parameters. A comprehensive review on this topic is given in [107]. Nevertheless, it should be noted that it is topologically impossible to have a global 3-dimensional parameterization without singular points for $\text{SO}(3)$ [113]. Moreover, 5 is the minimum dimension of parameterizations to represent $\text{SO}(3)$ in a one-to-one and global fashion [113].

Perhaps the most popular 3-dimensional parameterization of attitude is Euler angles. The Euler angles have 24 different forms depending on the sequence of rotations, and in which frame the rotations are formulated (intrinsic and extrinsic). The intrinsic z-y-x Euler angles are usually used in aerospace engineering, which are also referred to as the yaw, pitch, and roll angles. The singular point for this set of Euler angles is when the pitch angle

is 90 degrees, which is famously known as the gimbal lock.

Another frequently used 3-dimensional parameterization is the rotation vector. By the Euler's rotation theorem, any 3D rotation can be described by rotating about an axis $u \in \mathbb{S}^2$ by an angle $\theta \in \mathbb{R}$. And the rotation vector is simply $v = \theta u \in \mathbb{R}^3$. The singular point for rotation vectors is when $\|v\| = \theta = \pi$, i.e., when the rotation angle is 180 degrees. Two other 3-dimensional parameterizations that are used in aerospace engineering are Rodrigues parameters and modified Rodrigues parameters, given by $\rho = \tan(\theta/2)u \in \mathbb{R}^3$, and $p = \tan(\theta/4)u \in \mathbb{R}^n$, respectively. The singular points for these two parameterizations are when $\theta = \pi$, and $\theta = 2\pi$, respectively.

As mentioned in Section 1.2.1, the attitude can also be parameterized using unit quaternions in \mathbb{S}^3 . This is a 4-dimensional singularity free parameterization, and is widely used in attitude estimation algorithms. The unit quaternion corresponding to the rotation (u, θ) can be written as

$$q = \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2)u \end{bmatrix} \in \mathbb{S}^3 \quad (1.2)$$

in the Hamilton convention. It should be noted that q and $-q$ represent the same 3D rotation.

Deterministic Observers This type of attitude estimators assume all measurements are noise free, so the convergence from a wrong initial attitude to the true attitude can be proved using a rigorous Lyapunov argument. Typically, it is assumed that the angular velocity is measured by a gyroscope, so that the attitude can be integrated forward in time. Also, various reference directions in the inertial frame are measured in the reference frame fixed to the rigid body (body-fixed frame), to correct the initial and integration errors.

The first and most famous attitude observer with guaranteed convergence is proposed in [78]. This observer has been extensively used in attitude controllers for unmanned aerial vehicles. Later, it has been generalized to deal with reference directions that are time varying

in the inertial frame [9, 37, 115]. Also, a hybrid correction scheme has been proposed in [130] to deal with the zero measure set where the initial attitude does not converge for the observer in [78]. Moreover, the rotation of the Earth has also been incorporated to enhance attitude observability when only a single reference direction is available [103].

Nonlinear Kalman Filters Different from deterministic observers, stochastic filters assume that both the gyroscope and direction measurements are subject to noises. And the filter determines how many weights are given to the integrated attitude and the direction measurements, based on their respective covariance matrices. Compared to deterministic observers, filters typically have faster convergence rate when the initial uncertainty, and sensor noise characteristics are known. On the other hand, due to the difficulty of formulating stability under stochastic settings [51], convergence properties are usually not studied for stochastic filters.

A comprehensive review of attitude estimation algorithms based on extended Kalman filters (EKF) and unscented Kalman filters (UKF) has been presented in [23]. The EKF can be naturally applied to attitude kinematics formulated in Euler angles [27, 31], Rodrigues parameters [46], and modified Rodrigues parameters [21]. However, due to the singularity of 3-dimensional attitude parameterizations, these filters must be operated in a limited range of attitude, which cannot undergo continuous spinning. Also, a less noticed problem is that the Lebesgue measure in \mathbb{R}^3 for these parameterizations does not correspond to a uniform measure on $\text{SO}(3)$, which means that the Gaussian distribution in \mathbb{R}^3 is no longer Gaussian shaped on $\text{SO}(3)$, and this causes inaccuracy in modeling the uncertainty.

The problem caused by singularity is avoided by the celebrated multiplicative EKF (MEKF) [69, 83]. Instead of directly working with 3-dimensional attitude parameterizations, MEKF treats a 3-dimensional parameterization of *attitude error* as the state vector. More specifically, a unit quaternion q is used as the estimated attitude, and an attitude error parameterized by one of the 3-dimensional parameterizations (rotation vectors, Rodrigues

and modified Rodrigues parameters) is used to model the uncertainty of the estimated q . The term “multiplicative” refers to the multiplication between the estimated quaternion and the attitude error:

$$q_t = q \otimes \delta q\{v\}, \quad (1.3)$$

where q_t is the true quaternion, $\delta q\{v\}$ is the error quaternion corresponding to the 3-dimensional parameterization $v \in \mathbb{R}^3$, and \otimes is the quaternion multiplication. Because the attitude error v is usually small, it avoids the singularity of the parameterization.

MEKF remains one of the most robust and efficient attitude filters, which has been validated by several NASA space missions [23]. The unit quaternion q in MEKF can be replaced by a rotation matrix R without changing the structure of the filter. Also, in (1.3), the attitude error v is expressed in the body-fixed frame, and it can be changed into the inertial frame by reversing the order of multiplication

$$q_t = \delta q\{v\} \otimes q. \quad (1.4)$$

This convention has some advantages over (1.3) [32, 74] which potentially injects some false information from the measurement when there is an unobserved degree of freedom. A second order MEKF has also been studied in [83], and an unscented Kalman filter parallel to MEKF has been proposed in [22].

In contrast to MEKF, the quaternion can also be directly treated as the state vector of an EKF, so that the attitude error becomes addictive

$$q_t = q + \delta q, \quad (1.5)$$

which is referred to as the addictive EKF (AEKF). The comparison of MEKF and AEKF has been studied in [84, 108, 109]. However, normalizing the quaternion to have unit norm

by brute force in AEKF makes the covariance matrix for quaternions singular, and should be avoided. On the contrary, identifying quaternions on the same line through the origin in \mathbb{R}^4 without the unit norm constraint makes AEKF equivalent to MEKF, although it is not suggested due to the increased computation [84].

The idea of MEKF has been generalized to the estimation problem of both attitude and position of a rigid body [112], which establishes a framework for Kalman filter based inertial navigation. Also, using the “error state” to represent uncertainty has been more formally studied on matrix Lie groups as the invariant extended Kalman filter (IEKF) in [5, 6, 7]. In particular, it is proved in [6] that IEKF is stable when acting as an observer on matrix Lie groups, similar to a linear Kalman filter.

Filters Based on Directional Statistics Although MEKF achieves great success in attitude estimation, it cannot deal with large uncertainty due to the underlying Gaussian assumption of the 3-dimensional parameterization of attitude error, discussed at Figure 1.1. This disadvantage can be overcome by replacing the Gaussian distribution using the probability distributions defined directly on $\text{SO}(3)$ from directional statistics, namely the matrix Fisher and Bingham distributions.

The attitude filter based on the Bingham distribution was proposed independently in [36] and [62]. Later, the attitude filter based on the matrix Fisher distribution was proposed in [64]. Given that the two distributions are equivalent, these filters are also equivalent, except some differences in numerical implementations. Furthermore, attitude filters using deterministic sampling of sigma points were developed in [35] and [64] for the two distributions respectively, similar to UKF. The Bingham distribution has also been generalized into dual quaternions to represent the uncertainty of attitude and position simultaneously in [4, 34, 71, 72]. This approach deals with the correlation between attitude in $\text{SO}(3)$ and position in \mathbb{R}^3 . However, the correlation must respect the structure of three dimensional Euclidean group $\text{SE}(3)$ when it is used, hence it cannot deal with Euclidean random variables

other than position.

In [85], an attitude filter was proposed using a new probability density on $\text{SO}(3) \times \mathbb{R}^n$ where the attitude part is a matrix Fisher distribution, so that all sensor biases can be estimated concurrently with attitude. However, the quadratic terms appeared in the uncertainty propagation and measurement update steps must be omitted to make the density function closed under the proposed functional form, which lacks theoretical soundness. In addition, an uncertainty propagation scheme based on the proposed Gauss-Bingham distribution was introduced in [24], which can also deal with the correlation between $\text{SO}(3)$ and \mathbb{R}^n . Unfortunately, the measurement update step was not considered, and the uncertainty propagation scheme requires numerical optimizations, rendering its inappropriateness for practical implementations.

In conclusion, although models from directional statistics have been used in attitude estimation algorithms to deal with large uncertainty, there is still need for a new model that (i) deals with the correlation between attitude and Euclidean random variables not restricted to position, and (ii) has a closed form maximum likelihood estimation of its parameters for efficient inference. Such a new model would greatly generalize the usage of existing distributions on $\text{SO}(3)$, and enable new estimation algorithms suitable for large uncertainty when the attitude must be estimated simultaneously with quantities in the Euclidean space, such as sensor biases and landmark locations, etc.

1.2.3 Visual-Inertial Navigation and Odometry

As the computation power increases, visual information from various forms of cameras has attracted attentions in the application of navigation problems. The image captured by cameras contains more information than an IMU or a GNSS receiver, as it not only provides information about the camera itself, but also its surrounding environments. Although cameras have been solely used [104] to locate a vehicle, and there have been several celebrated visual odometry and SLAM algorithms [30, 55, 91], the single information

source is not very robust in practical applications. For example, localization from cameras could fail when the motion is too fast, or when the illumination is very dark. Therefore, cameras are usually combined with inertial sensors to obtain more robust estimation results.

This is referred to as visual-inertial odometry, navigation, or SLAM.

Visual-inertial odometry algorithms has been reviewed in [43]. They can be roughly classified into two categories: filtering and optimization based methods. The filtering approach is usually based on the idea of MEKF (although this terminology is not frequently used in robotics community), i.e., the uncertainty of attitude is described by a 3-by-3 covariance matrix for the three dimensional parameterization of attitude error. The first practically usable filter is the MSCKF [90]. In MSCKF, the 3D coordinates for visual features are not estimated; instead, they are distilled into linear constraints on the poses in which frame the features are captured, by calculating the null space of the Jacobian of the measurement function. This method has later been improved to account for the correct observability of the state [42, 74, 75], and the Lie group IEKF has also been used to enhance consistency [16].

The optimization method can be traced back to the bundle adjustment technique [114] in computer vision. The idea is to estimate the pose and feature locations by minimizing the camera reprojection error, which is formulated into a nonlinear least square problem. The optimization can be efficiently solved using the Gauss-Newton or Levenberg–Marquardt algorithms while exploiting the sparseness of the Jacobian matrix. In [29, 77], the inertial measurement is also incorporated into this framework using the so called IMU pre-integration technique. The idea is that the IMU measurements are integrated between two keyframes with the uncertainty of the integrated pose propagated using the classic EKF approach. Then the difference between the relative pose and the integrated pose are added into the loss function of the bundle adjustment problem. IMU pre-integration has later become standard in optimization based visual inertial integration [15, 92, 101]. Also, by posing the optimization into a factor graph form [47], it can be solved incrementally to reduce

computation demand [50].

Comparing the filtering based method and optimization method, the major difference is that by using a multi-step iterated algorithm, the optimization method re-linearizes the non-linear measurement function several times, which is more accurate than the one-step linearization of the filtering method. Therefore, the optimization method is usually more accurate, but at the cost of increased computation demand [43]. Nevertheless, the iterated EKF in classical control theory has been shown to be equivalent to applying a Gauss-Newton optimization to the posterior likelihood function [10], and this has been utilized in several visual-inertial odometry algorithms [13, 14] to reduce the linearization error. Also, as discussed before, the filtering method cannot deal with large attitude uncertainty due to the inherent Gaussian assumption.

1.3 Summary of Contributions

In this dissertation, the technique of applying probability models defined on $\text{SO}(3)$ for estimation problems involving 3D rotations is investigated. Specific emphasis has been placed on using the matrix Fisher distribution to deal with large attitude uncertainty.

Matrix Fisher Distribution Although various properties of the matrix Fisher distribution on $\text{SO}(3)$ has been studied in [25, 54, 64], calculating its normalizing constant and moments remains very challenging. In this dissertation, a recursive algorithm to calculate the central moments of a matrix Fisher distribution up to arbitrary order is given, generalizing the results in [54, 64]. In particular, closed form expressions are given for the second and third order moments. Next, a new approximation of the matrix Fisher distribution is introduced when it is highly concentrated in two degrees of freedom, generalizing the results in [63]. Approximate calculations for the normalizing constant and its derivatives are given for efficient computations. These computational techniques are used in the subsequent design of Bayesian filters for fast inference of the parameters of the distribution.

Matrix Fisher–Gaussian Distribution The correlation between $\text{SO}(3)$ and Euclidean space is critical for simultaneous estimation of attitude and other Euclidean quantities. Based on the matrix Fisher distribution, a new probability density function, referred to as the matrix Fisher–Gaussian (MFG) distribution, is introduced on $\text{SO}(3) \times \mathbb{R}^n$ to model the correlation between random rotations and quantities in the Euclidean space. The MFG is constructed using a generic method with geometric intuition, and it only models the *linear* correlation with minimum number of parameters. Various properties of MFG are studied, and a closed form approximate maximum likelihood estimation of its parameters are given. This enables fast inference, with the potential to be implemented in real time estimation problems. In addition, the Bingham–Gaussian distribution which is equivalent to MFG, is defined for unit quaternions on $\mathbb{S}^3 \times \mathbb{R}^n$ using the Lie group homomorphism from \mathbb{S}^3 to $\text{SO}(3)$.

Application in Estimation Problems The proposed MFG greatly generalizes the applications of matrix Fisher distribution, enabling simultaneous estimation of attitude and arbitrary Euclidean quantities. First, the classic problem of attitude estimation with a gyroscope and direction measurements are studied. In particular, an explanation for the unobservability when only a single fixed reference direction is available in the usual IMU setting is provided, with a straightforward geometric intuition. In addition, two new observable conditions when the angular velocity and single reference direction are resolved in the same reference frame are identified. Next, the MFG is used to simultaneously estimate the attitude and time-varying gyroscope bias with direction measurements. It is demonstrated through simulation studies that the filter based on MFG has faster convergence rate from a wrong initial attitude, and better accuracy if one degree of freedom is not properly observed, than MEKF.

Then the attitude estimation is generalized into a full inertial navigation setting, where the uncertainty for attitude, linear velocity, position, and sensor biases are propagated through IMU kinematics using MFG. The loosely couple IMU-GNSS navigation is studied by proposing a filter based on MFG to deal with the position measurements. And this is later

generalized to landmark or feature measurements given by a camera in RGB-D or stereo settings, which is used in a visual-inertial navigation or odometry problem. It is shown by simulation that the MFG filter has much faster convergence rate than MEKF given a wrong initial attitude, such as a wrong heading direction when it is initially unknown. These demonstrate the advantage of MFG when dealing with large attitude uncertainty than the Gaussian distribution assumed in MEKF.

1.3.1 Publications

The results of this dissertation have been published in the following conference proceedings and journals:

- W. Wang and T. Lee, “Matrix Fisher-Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ for attitude estimation with a gyrobias,” in *American Control Conference*, 2020, pp. 4429–4434. [121]
- W. Wang and T. Lee, “Spectral uncertainty propagation for generalized stochastic hybrid systems with applications to a bouncing ball,” in *American Control Conference*, 2020, pp. 1803–1808. [123]
- W. Wang and T. Lee “Spectral Bayesian estimation for general stochastic hybrid systems,” *Automatica*, vol. 117, p. 108989, 2020. [122]
- W. Wang and T. Lee, “Higher-order central moments of matrix Fisher distribution on $\text{SO}(3)$,” *Statistics & Probability Letters*, vol. 169, p. 108983, 2021. [125]
- W. Wang and T. Lee, “Matrix Fisher–Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ and Bayesian attitude estimation,” *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2175–2191, 2021. [126]
- W. Wang and T. Lee, “Bingham-Gaussian distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ for unscented attitude estimation,” in *International Conference on Multisensor Fusion and Integration*

for Intelligent Systems, 2021, pp. 1–7. [124]

- W. Wang and T. Lee, “Spacecraft attitude and gyro-bias estimation with a single magnetometer on $\text{SO}(3) \times \mathbb{R}^n$,” in *AIAA Scitech 2022 Forum*, 2022 [127]
- W. Wang, K. Gamagedara, and T. Lee, “On the observability of attitude with single direction measurements,” *IEEE Transactions on Automatic Control*, 2022, accepted. [119]
- W. Wang and T. Lee, “Uncertainty propagation for general stochastic hybrid systems on compact Lie groups,” *SIAM Journal on Applied Dynamical Systems*, 2022, accepted. [128]

1.4 Outline of Dissertation

In Chapter 2, the matrix Fisher distribution on $\text{SO}(3)$, and the Bingham distribution on \mathbb{S}^3 are reviewed. A new algorithm to calculate the higher order moments, and a new approximation for a highly concentrated case of the matrix Fisher distribution are introduced. Chapter 3 defines the matrix Fisher–Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ and the Bingham–Gaussian distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ to model the correlation between attitude with large uncertainty and Euclidean random variables. Their various properties are studied, and a closed form approximate maximum likelihood estimation is given. In Chapter 4, attitude observability with single direction measurements is examined, and a recursive Bayesian filter for attitude estimation is designed using MFG. This filter is generalized in Chapter 5 for simultaneous attitude and position estimation, in the applications of loosely coupled IMU-GNSS navigation, and visual-inertial odometry/navigation. These filters are compared with MEKF for performance evaluation through simulation studies. Finally, future directions are discussed in Chapter 6.

Chapter 2: Matrix Fisher Distribution on SO(3)

The matrix Fisher distribution was first introduced as a density function on the Stiefel manifold to model the orientation of an object [25]. As a special case, it is defined on the 3D rotational group SO(3), and can be regarded as an analogue of the Gaussian distribution. As a tool to model random rotation matrices, it has been used in attitude estimation algorithms [64], head pose estimation [76], and regression for 3D rotations from image and point cloud inputs [89, 131].

This chapter introduces some properties of the matrix Fisher distribution, which will be useful in the subsequent construction of the matrix Fisher–Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ in Chapter 3, and the developments of estimation algorithms in Chapter 4 and Chapter 5. In particular, Chapter 2.1 reviews some basic properties of SO(3). The matrix Fisher distribution is formally introduced in Chapter 2.2, and its construction and properties are also reviewed. In Chapter 2.3, the Bingham distribution defined on \mathbb{S}^3 , which is equivalent to the matrix Fisher distribution, is reviewed. Next, some new results on the computation of higher order moments of matrix Fisher distribution are developed in Chapter 2.4. And finally, Chapter 2.5 introduces a new approximation when the matrix Fisher distribution is highly concentrated in two directions.

2.1 Properties of SO(3) as a Lie Group

The three dimensional special orthogonal group SO(3) is defined as

$$\text{SO}(3) = \{R \in \mathbb{R}^{3 \times 3} \mid RR^T = I_{3 \times 3}, \det R = 1\}. \quad (2.1)$$

An element $R \in \text{SO}(3)$ is usually called a rotation matrix, which is used to transform the coordinates of a vector after it undergoes a rotation in the 3D Euclidean space. Suppose

there are two reference frames, the inertial frame, or the world frame which is considered fixed; and the body-fixed frame, which is fixed to the rigid body. For a vector \vec{v} , let its coordinates in the inertial frame be denoted by $v^I \in \mathbb{R}^3$, and its coordinates in the body-fixed frame be denoted by $v^B \in \mathbb{R}^3$. In this dissertation, R is used to transform v^B to v^I , i.e.,

$$v^I = Rv^B, \quad (2.2)$$

if the two frames have the same origin.

The set $\text{SO}(3)$ is a Lie group, which means it is both a smooth manifold, and an algebraic group. Its tangent space at identity $I_{3 \times 3} \in \text{SO}(3)$ is the Lie algebra $\mathfrak{so}(3)$ composed of 3-by-3 skew symmetric matrices:

$$\mathfrak{so}(3) = \{A \in \mathbb{R}^{3 \times 3} \mid A = -A^T\}. \quad (2.3)$$

There is a natural identification of $\mathfrak{so}(3)$ and \mathbb{R}^3 , denoted by the “hat” and “vee” maps:

$$\mathfrak{so}(3) \ni \begin{bmatrix} 0 & -\Omega_z & \Omega_y \\ \Omega_z & 0 & -\Omega_x \\ -\Omega_y & \Omega_x & 0 \end{bmatrix} \xrightarrow{\text{vee } \vee} \begin{bmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{bmatrix} \in \mathbb{R}^3. \quad (2.4)$$

Note that for any $u, v \in \mathbb{R}^3$, $\hat{u}v = u \times v$. The tangent space at a specific $R \in \text{SO}(3)$ is $T_R \text{SO}(3) = \{RA \mid A \in \mathfrak{so}(3)\}$.

The Lie group $\text{SO}(3)$ can be locally identified with its tangent space through the exponential map. More specifically, for a given $R \in \text{SO}(3)$, there is a neighborhood of R such that the map \exp from $\mathfrak{so}(3)$ to $\text{SO}(3)$:

$$\exp : \mathfrak{so}(3) \rightarrow \text{SO}(3), \quad A \mapsto R \exp(A),$$

is one-to-one and onto in this neighborhood. The exponential map is same as the usual

matrix exponential [38]:

$$\exp(A) = \sum_{n=1}^{\infty} \frac{A^n}{n!}, \quad (2.5)$$

and it has a closed form known as the Rodrigues rotation formula for $A = \hat{\Omega}$:

$$\exp(\hat{\Omega}) = I_{3 \times 3} + \frac{\sin \|\Omega\|}{\|\Omega\|} \hat{\Omega} + \frac{1 - \cos \|\Omega\|}{\|\Omega\|^2} \hat{\Omega}^2. \quad (2.6)$$

The rotation vector introduced in Chapter 1.2.2 also satisfies (2.6) when converted to a rotation matrix. And for a rotation about the axis $u \in \mathbb{S}^2$ by an angle θ , (2.6) becomes

$$\exp(\theta \hat{u}) = I_{3 \times 3} + \sin \theta \hat{u} + (1 - \cos \theta) \hat{u}^2. \quad (2.7)$$

The local inverse of exponential map is called the logarithm map, denoted by \log as usual.

As a matrix Lie group, the multiplication of $\text{SO}(3)$ is not commutative, i.e., for $R_1, R_2 \in \text{SO}(3)$, in general $R_1 R_2 \neq R_2 R_1$. This makes $\text{SO}(3)$ much more complicated than $\text{SO}(2)$, and exhibits some interesting behaviors for density functions on $\text{SO}(3)$. The non-commutative error can be explicitly computed by the Baker–Campbell–Hausdorff (BCH) formula [38]:

$$\exp(A) \exp(B) = \exp(A + B + C), \quad (2.8)$$

where $A, B \in \mathfrak{so}(3)$, and $C \in \mathfrak{so}(3)$ is a correction term composed of iterated Lie brackets:

$$C = \frac{1}{2}[A, B] + \frac{1}{12}[A, [A, B]] - \frac{1}{12}[B, [A, B]] + \dots, \quad (2.9)$$

where the Lie bracket $[A, B] = AB - BA$ is also known as the matrix commutator. It should be noted that $\mathfrak{so}(3)$ is closed under Lie brackets.

For a curve $R(t) \in \text{SO}(3)$, its angular velocity can be expressed in two different ways

$$\frac{dR(t)}{dt} = R(t)\hat{\Omega}(t), \quad (2.10a)$$

$$\frac{dR(t)}{dt} = \hat{\omega}(t)R(t), \quad (2.10b)$$

where $\Omega(t) \in \mathbb{R}^3$ is the angular velocity expressed in the body-fixed frame, and $\omega(t) \in \mathbb{R}^3$ is the angular velocity expressed in the inertial frame. They can be transformed to each other by $\omega(t) = R(t)\Omega(t)$.

Before moving to the next section, some identities that will be used in subsequent chapters in this dissertation need to be reviewed. First, given $R^{-1} = R^T$ for any $R \in \text{SO}(3)$, it can be shown that

$$R_{kk} = R_{ii}R_{jj} - R_{ij}R_{ji}, \quad (2.11)$$

for $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$, where the subscripts denote the row and column indices of the matrix element. Also, for any $R \in \text{SO}(3)$, $A \in \mathbb{R}^{3 \times 3}$, and $x \in \mathbb{R}^3$, the following identities involving the hap and vee maps hold:

$$\widehat{Rx} = R\hat{x}R^T, \quad (2.12)$$

$$\hat{x}^2 = xx^T - x^T x I_{3 \times 3}, \quad (2.13)$$

$$(\hat{x}A + A^T \hat{x})^\vee = (\text{tr}(A)I_{3 \times 3} - A)x, \quad (2.14)$$

$$\text{tr}(\hat{x}A) = x^T (A^T - A)^\vee. \quad (2.15)$$

2.2 Matrix Fisher Distribution

In this subsection, the definition, properties and construction of the matrix Fisher distribution on $\text{SO}(3)$ are reviewed.

2.2.1 Definition and Properties

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

$$p(R; F) = \frac{1}{c(F)} \text{etr} (FR^T), \quad (2.16)$$

with respect to the uniform distribution on $\text{SO}(3)$, where $\text{etr}(\cdot)$ is the abbreviation for $\exp(\text{tr}(\cdot))$. 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 (2.16), 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 $A \in \text{SO}(3)$.

Various properties of a matrix Fisher distribution can be accessed through the *proper* singular value decomposition (pSVD) of F [54, 64, 82].

Definition 2.1. Let the singular value decomposition of F be given by $F = U'S'V'^T$, where $S' \in 3 \times 3$ is a diagonal matrix composed of the singular values $s'_1 \geq s'_2 \geq s'_3 \geq 0$ of F , and $U', V' \in \text{O}(3)$ are orthogonal matrices. The proper singular value decomposition of F is

$$F = USV^T \quad (2.17)$$

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

$$\begin{aligned} U &= U'\text{diag}(1, 1, \det[U']), \\ S &= \text{diag}(s_1, s_2, s_3) = \text{diag}(s'_1, s'_2, \det[U'V']s'_3), \\ V &= V'\text{diag}(1, 1, \det[V']). \end{aligned} \quad (2.18)$$

The motivation of the above proper SVD is to ensure $U, V \in \text{SO}(3)$, while allowing s_3 to be negative. The normalizing constant $c(F)$ depends only on the proper singular values

of F , i.e., $c(F) = c(S)$. The first order moment of R is given by

$$\mathbb{E}[R] = UDV^T = U\text{diag}(d_1, d_2, d_3)V^T, \quad (2.19)$$

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

$$d_i = \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i}. \quad (2.20)$$

However, $\mathbb{E}[R] \in 3 \times 3$ does not belong to $\text{SO}(3)$ in general. Instead, the *mean attitude* of R is usually interpreted as $UV^T \triangleq M \in \text{SO}(3)$, which maximizes the density (2.16), and also minimizes the Frobenius mean squared error [64].

Similar to the Gaussian distribution, the matrix Fisher distribution has three principal axes, given by the columns of U resolved in the inertial frame, or equivalently the columns of V when resolved in the body-fixed frame specified by the mean attitude 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

$$p(R(\theta); F) = \frac{e^{s_i}}{c(S)} \exp \left\{ (s_j + s_k) \cos \theta \right\}, \quad (2.21)$$

where $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$. This corresponds to the von Mises distribution 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, which will be discussed in detail in Chapter 2.5.

Given the relationship (2.20), it can be shown that $d_i + d_j$ and $s_i + s_j$ have the following relationship, which will be used in Chapter 4.1.

Lemma 2.1. Suppose $S = \text{diag}(s_1, s_2, s_3)$ with $s_1 \geq s_2 \geq |s_3| \geq 0$, and $D = \text{diag}(d_1, d_2, d_3)$

is given as (2.20). Then, the following properties hold:

1. $s_i + s_j = 0$ if and only if $d_i + d_j = 0$;
2. $s_i = s_j = 0$ if and only if $d_i = d_j = 0$;
3. $d_i + d_j$ is monotonically increasing with $s_i + s_j$,

for any $(i, j, k) \in \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$.

Proof. Using the one dimensional integral formula in (2.41), it can be shown that

$$\begin{aligned} \frac{\partial c(S)}{\partial s_i} + \frac{\partial c(S)}{\partial s_j} &= \int_{-1}^1 \frac{1}{2}(1+u)I_0\left[\frac{1}{2}(s_i - s_j)(1-u)\right] \\ &\quad \times I_1\left[\frac{1}{2}(s_i + s_j)(1+u)\right] \exp(s_k u) du \geq 0, \end{aligned} \quad (2.22)$$

Note that $I_0(x) \geq 1$ for any x , and $I_1(x) = 0$ if and only if $x = 0$. Also $I_1(x) > 0$ if $x > 0$.

First, if $s_i + s_j = 0$, then $I_1\left[\frac{1}{2}(s_i + s_j)(1+u)\right] = 0$, and therefore $d_i + d_j = 0$. On the other hand, if $s_i + s_j > 0$, the integrand of (2.22) is strictly positive for $u \in (-1, 1]$. Therefore, $d_i + d_j > 0$. As both of $d_i + d_j$ and $s_i + s_j$ are non-negative, it follows that $d_i + d_j = 0$ implies $s_i + s_j = 0$. Therefore, $s_i + s_j = 0$ if and only if $d_i + d_j = 0$.

Next, if $s_i = s_j = 0$, then by (15a) and (15b) in [64], $d_i = d_j = 0$. On the other hand, suppose at least one of s_i or s_j is nonzero. There are two sub-cases: (i) if $s_i + s_j \neq 0$, then $d_i + d_j \neq 0$ so d_i, d_j cannot both be zeros; (ii) if $s_i = -s_j \neq 0$, then

$$\frac{\partial c(S)}{\partial s_i} = \int_{-1}^1 \frac{1}{4}(1-u)I_1[s_i(1-u)] \exp(s_k u) du$$

where the integrand has the same sign as s_i over $u \in [-1, 1]$. So $d_i \neq 0$.

Finally to prove the monotonicity, let $\alpha = s_i + s_j$ and $\beta = s_i - s_j$, which yields $s_i = \frac{1}{2}(\alpha + \beta)$ and $s_j = \frac{1}{2}(\alpha - \beta)$. From the chain rule,

$$\frac{\partial(d_i + d_j)}{\partial \alpha} = \frac{1}{2} \left\{ \frac{\partial d_i + d_j}{\partial s_i} + \frac{\partial d_i + d_j}{\partial s_j} \right\}. \quad (2.23)$$

Let $Q \sim \mathcal{M}(S)$, then according to (2.19) and (2.20), $d_i + d_j$ is

$$d_i + d_j = \frac{1}{c(S)} \int_{Q \in \text{SO}(3)} (Q_{ii} + Q_{jj}) \text{etr}(SQ) dQ.$$

Based on the above, (2.23) becomes

$$\frac{\partial(d_i + d_j)}{\partial \alpha} = \frac{1}{2} \left(\mathbb{E}[(Q_{ii} + Q_{jj})^2] - \mathbb{E}[Q_{ii} + Q_{jj}]^2 \right)$$

which is positive by the Cauchy-Schwarz inequality, and the monotonicity follows. \square

This lemma shows that S and D have a one-to-one correspondence, so for every D a unique S can be solved from (2.20). Therefore, from (2.19), it can be seen that $\mathbb{E}[R]$ and F carry the same information. In other words, $\mathbb{E}[R]$ is a sufficient statistic for F . Given $\mathbb{E}[R]$, let its pSVD be given in (2.19), and let $S = \text{diag}(s_1, s_2, s_3)$ be solved from D using (2.20), then the maximum likelihood estimation (MLE) of F is given by $F = USV^T$ [54, 64]. This can be used to inference a matrix Fisher distribution from random samples.

2.2.2 Construction

In directional statistics, a generic way of constructing a “Gaussian” like probability distribution on a manifold is through conditioning [98]. The idea is that the manifold is first embedded into Euclidean space, where a Gaussian distribution is defined. Then the Gaussian distribution is conditioned onto the manifold, and becomes a new probability distribution on the manifold. A number of famous models can be constructed using this approach, including von Mises and von Mises–Fisher distribution on \mathbb{S}^n [80], the matrix Fisher distribution on Stiefel manifold [25], the Bingham distribution on \mathbb{S}^n [12], and the distribution on the cylinder [81], etc.

Here the construction of matrix Fisher distribution on $\text{SO}(3)$ through conditioning is given in the following theorem.

Theorem 2.1 ([25]). Suppose $R \sim \mathcal{M}(F)$, and $F = USV^T$ is its proper singular value decomposition. Let $M = UV^T$, $K_I = VSV^T$, $K_B = USU^T$. Also, let $\mu_I = \text{vec}(M^T)$, $\Sigma_I^{-1} = I_{3 \times 3} \otimes K_I$; and $\mu_B = \text{vec}(M)$, $\Sigma_B^{-1} = I_{3 \times 3} \otimes K_B$, where $\text{vec}(\cdot)$ denotes the concatenation of the columns of a matrix into a column vector, and \otimes denotes Kronecker product. Suppose $\text{vec}(R^T) \sim \mathcal{N}(\mu_I, \Sigma_I)$, or $\text{vec}(R) \sim \mathcal{N}(\mu_B, \Sigma_B)$. Then the density functions of R for these two Gaussian distributions are the density function for the matrix Fisher distribution (2.16) when restricted to $R \in \text{SO}(3)$.

Proof. For $\mathcal{N}(\mu_I, \Sigma_I)$, the density for R is

$$\begin{aligned} p_I(R) &\propto \exp \left\{ -\frac{1}{2} (\text{vec}(R^T) - \mu_I)^T \Sigma_I^{-1} (\text{vec}(R^T) - \mu_I) \right\} \\ &= \exp \left\{ -\frac{1}{2} \text{tr} (K_I(R - M)^T (R - M)) \right\} \\ &= \exp \left\{ -\frac{1}{2} \text{tr} (2K_I - K_I R^T M - K_I M^T R) \right\} \\ &\propto \exp \{FR^T\}, \end{aligned}$$

where the second equality is because $R^T R = I_{3 \times 3}$ on $\text{SO}(3)$. Similarly, for $\mathcal{N}(\mu_B, \Sigma_B)$, the density for R is

$$\begin{aligned} p_B(R) &\propto \exp \left\{ -\frac{1}{2} (\text{vec}(R) - \mu_B)^T \Sigma_B^{-1} (\text{vec}(R) - \mu_B) \right\} \\ &= \exp \left\{ -\frac{1}{2} \text{tr} (K_B(R - M)(R - M)^T) \right\} \\ &\propto \exp \{FR^T\}. \end{aligned}$$

This concludes the proof. □

2.3 Bingham Distribution

The Bingham distribution is defined on the unit sphere \mathbb{S}^n with antipodal symmetry. It is equivalent to the matrix Fisher distribution on $\text{SO}(3)$ when defined on \mathbb{S}^3 for unit quaternions. This chapter reviews its definition, properties, its equivalence with the matrix Fisher

distribution, and the computation of its normalizing constant.

2.3.1 Definition and Properties

A random vector $q \in \mathbb{S}^3$ follows a Bingham distribution with parameter $A = A^T \in \mathbb{R}^{4 \times 4}$ if it has the following density function:

$$p(q; A) = \frac{1}{c(A)} \exp \{ q^T A q \} \quad (2.24)$$

with respect to the uniform distribution on \mathbb{S}^3 , where $c(A)$ is the normalizing constant. This distribution is denoted by $q \in \mathcal{B}(A)$. The Bingham distribution is antipodally symmetric, i.e., $p(q) = p(-q)$ for any $q \in \mathbb{S}^3$.

The properties of a Bingham distribution are determined by the eigenvalue decomposition of the parameter $A = M Z M^T$, where $M = [m_0, m_1, m_2, m_3]$ are the eigenvectors of A , and $Z = \text{diag}(z_0, z_1, z_2, z_3)$ are the corresponding eigenvalues. For uniqueness, the eigenvalues are assumed to be in descending order. Due to the unit length constraint, it is straight forward to show $\mathcal{B}(M Z M^T) = \mathcal{B}(M(Z + z I_{4 \times 4})M^T)$ for any $z \in \mathbb{R}$. Thus, in this dissertation the following assumption is made:

$$0 = z_0 \geq z_1 \geq z_2 \geq z_3. \quad (2.25)$$

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

$$E[qq^T] = MDM^T \triangleq M\text{diag}(d_0, d_1, d_2, d_3)M^T, \quad (2.26)$$

where

$$d_i = \frac{1}{c(Z)} \frac{\partial c(Z)}{\partial z_i}, \quad i = 0, 1, 2, 3. \quad (2.27)$$

Since $q^T q = 1$, the matrix D satisfies $d_0 + d_1 + d_2 + d_3 = 1$.

The mode of Bingham distribution is $q = m_0$, which maximizes the density (2.24), and can be regarded as the *mean attitude*. The principal axes are given by the unit quaternions $m_1, m_2, m_3 \in \mathbb{S}^3$, and the concentration along the principal axes are described by $-z_1, -z_2, -z_3$ respectively, with larger $-z_i$ implying higher concentration. For the MLE of the parameter A , suppose the second order moment $E[qq^T]$ is given. Let its eigen decomposition be given in (2.26), and let $Z = \text{diag}(z_0, z_1, z_2, z_3)$ be solved from (2.27) using D , then $A = MZM^T$. The Bingham distribution can also be constructed by conditioning a Gaussian distribution from \mathbb{R}^4 onto \mathbb{S}^3 , the detail of this construction is available in [12].

2.3.2 Equivalence With the Matrix Fisher Distribution

The 3-sphere \mathbb{S}^3 is also a Lie group under the quaternion multiplication $\otimes : \mathbb{S}^3 \times \mathbb{S}^3 \rightarrow \mathbb{S}^3$, defined by

$$p \otimes q = \begin{bmatrix} p_0q_0 - p_1q_1 - p_2q_2 - p_3q_3 \\ p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2 \\ p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1 \\ p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0 \end{bmatrix}, \quad (2.28)$$

for any $p = [p_0, p_1, p_2, p_3]^T$ and $q = [q_0, q_1, q_2, q_3]^T \in \mathbb{S}^3$. Quaternion multiplication (2.28) can also be written in matrix form as

$$p \otimes q = [p]_L q = [q]_R p, \quad (2.29)$$

where

$$[p]_L = \begin{bmatrix} p_0 & -p_1 & -p_2 & -p_3 \\ p_1 & p_0 & -p_3 & p_2 \\ p_2 & p_3 & p_0 & -p_1 \\ p_3 & -p_2 & p_1 & p_0 \end{bmatrix}, \quad (2.30)$$

and

$$[q]_R = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix}. \quad (2.31)$$

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

$$M = [p]_L[q]_R \quad (2.32)$$

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

The Lie group \mathbb{S}^3 is isomorphic to $\text{SU}(2)$, i.e., the two dimensional special unitary group, which is a double cover of $\text{SO}(3)$, so q and $-q$ represent the same 3D rotation. The conversion from a rotation vector to a unit quaternion is given in (1.2). Together with (2.7),

this defines a Lie group homomorphism $\varphi : \mathbb{S}^3 \rightarrow \mathrm{SO}(3)$

$$q \mapsto \varphi(q) = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & 1 - 2(q_1^2 + q_3^2) & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix}. \quad (2.33)$$

The homomorphism φ satisfies the following identity [99]: for any $q, p \in \mathbb{S}^3$,

$$\mathrm{tr}(\varphi(p)\varphi(q)^T) + 1 = 4(p^T q)^2. \quad (2.34)$$

Next, the equivalence between the Bingham distribution and the matrix Fisher distribution is established in the following theorem.

Theorem 2.2 ([99]). *Let $M \in \mathrm{SO}(4)$, and $Z = \mathrm{diag}(0, z_1, z_2, z_3)$ with $0 \geq z_1 \geq z_2 \geq z_3$. Define $S = \mathrm{diag}(s_1, s_2, s_3)$ as $s_1 = \frac{1}{4}(z_1 - z_2 - z_3)$, $s_2 = \frac{1}{4}(z_2 - z_1 - z_3)$, and $s_3 = \frac{1}{4}(z_3 - z_1 - z_2)$. Let $M = [u]_L[v^{-1}]_R$ be its isoclinic decomposition. Define $U = \varphi(u)$ and $V = \varphi(v)$. Then $q \sim \mathcal{B}(MZM^T)$ if and only if $\varphi(q) \sim \mathcal{M}(USV^T)$.*

Proof. It suffices to check for all $q \in \mathbb{S}^3$, $c \cdot \exp(q^T M Z M^T q) = \mathrm{etr}(USV^T \varphi(q)^T)$ for some constant $c \in \mathbb{R}$. Let $M = [m_0, m_1, m_2, m_3]$, $I_{4 \times 4} = [e_0, e_1, e_2, e_3]$, and $I_{3 \times 3} = [\varepsilon_1, \varepsilon_2, \varepsilon_3]$, then

$$\varphi(m_0) = \varphi([u]_L[v^{-1}]_R e_0) = UV^T, \quad (2.35)$$

$$\varphi(m_i) = \varphi([u]_L[v^{-1}]_R e_i) = U \exp(\pi \hat{\varepsilon}_i) V^T, \quad (2.36)$$

for $i = 1, 2, 3$. By some straightforward calculations, it can be shown that

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

Then, using (2.34), the following equality can be proved

$$\begin{aligned}\text{tr}(USV^T \varphi(q)^T) &= \frac{1}{4} \sum_{i=1}^3 z_i \cdot (4(q^T m_i)^2 - 1) \\ &= \sum_{i=1}^3 z_i (q^T m_i)^2 - \frac{1}{4} \sum_{i=1}^3 z_i = q^T M Z M^T q + c,\end{aligned}$$

which shows the equivalence. \square

Looking at (2.35) and (2.36), the mean attitude and principal axes of the matrix Fisher and Bingham distributions are the same. Also, the concentration parameter along the i -th principal axis $2(s_j + s_k) = -z_i$ are also the same for the two distribution, for $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k$.

2.3.3 Normalizing Constant

The normalizing constant of the Bingham distribution is

$$c_B(Z) = \int_{q \in \mathbb{S}^3} \exp \{q^T Z q\} dq, \quad (2.37)$$

where dq is normalized such that $\int_{q \in \mathbb{S}^3} dq = 1$. And it can be written as a hypergeometric function of matrix argument $c_B(Z) = {}_1F_1(\frac{1}{2}, 2, Z)$ [12]. Also, the normalizing constant of the matrix Fisher distribution is

$$c_M(S) = \int_{R \in \text{SO}(3)} \text{etr}(SR^T) dR, \quad (2.38)$$

where dR is normalized such that $\int_{R \in \text{SO}(3)} dR = 1$. Similarly, it can be written as a hypergeometric function of matrix argument $c_M(S) = {}_0F_1(\frac{3}{2}; \frac{1}{4}S^2)$ [54]. Given the equivalence of matrix Fisher and Bingham distributions, their normalizing constants are also the same $c_M(S) = c_B(Z)$ [64] if S and Z satisfies the relationship in Theorem 2.2. It should be noted in some literature dq in (2.37) is the usual Lebesgue measure in \mathbb{R}^4 , and the normalizing

constant $c_B(Z)$ is scaled by the surface area of \mathbb{S}^3 which is $2\pi^2$.

The normalizing constant $c_M(S)$ or $c_B(Z)$ and their derivatives are very hard to compute. In [33], several algorithms are compared to compute $c_B(Z)$, including the series expansion [56], the saddle point approximation [58], the holonomic gradient method [106], and the new algorithm proposed in [56]. Furthermore, a one dimensional integration formula has been developed in [64, 129] as

$$c_M(S) = \int_{-1}^1 \frac{1}{2} I_0 \left[\frac{1}{2}(s_i - s_j)(1-u) \right] I_0 \left[\frac{1}{2}(s_i + s_j)(1+u) \right] \exp \{s_k u\} du \quad (2.39)$$

for $(i, j, k) \in \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$, where I_v is the modified Bessel function of the first kind with order v . This can be used with a one dimensional numerical integration scheme to compute $c_M(S)$.

As indicated in (2.20) and (2.27), evaluation of the moments $E[R]$ and $E[qq^T]$ needs the derivative of $c_M(S)$ and $c_B(Z)$. According to Theorem 2.2, the assumption $z_0 = 0$, and $d_0 + d_1 + d_2 + d_3 = 1$, $\partial c_M(S)/\partial s_i$ and $\partial c_B(Z)/\partial z_j$ for $i = 1, 2, 3$ and $j = 0, 1, 2, 3$ can be converted to each other using the following linear transform

$$\frac{\partial c_M(S)}{\partial s_1} = \frac{\partial c_B(Z)}{\partial z_0} + \frac{\partial c_B(Z)}{\partial z_1} - \frac{\partial c_B(Z)}{\partial z_2} - \frac{\partial c_B(Z)}{\partial z_3}, \quad (2.40a)$$

$$\frac{\partial c_M(S)}{\partial s_2} = \frac{\partial c_B(Z)}{\partial z_0} - \frac{\partial c_B(Z)}{\partial z_1} + \frac{\partial c_B(Z)}{\partial z_2} - \frac{\partial c_B(Z)}{\partial z_3}, \quad (2.40b)$$

$$\frac{\partial c_M(S)}{\partial s_3} = \frac{\partial c_B(Z)}{\partial z_0} - \frac{\partial c_B(Z)}{\partial z_1} - \frac{\partial c_B(Z)}{\partial z_2} + \frac{\partial c_B(Z)}{\partial z_3}. \quad (2.40c)$$

The derivative $\partial c_M(S)/\partial s_k$ can be directly obtained by taking the derivative of (2.39), as

$$\frac{\partial c_M(S)}{\partial s_k} = \int_{-1}^1 \frac{1}{2} I_0 \left[\frac{1}{2}(s_i - s_j)(1-u) \right] I_0 \left[\frac{1}{2}(s_i + s_j)(1+u) \right] u \exp \{s_k u\} du, \quad (2.41)$$

for $(i, j, k) \in \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}$. Moreover, the derivative $\partial c_B(Z)/\partial z_j$ can be evaluated using the approach proposed in [59] by computing the normalizing constant of a Bingham distribution on \mathbb{S}^5 . Higher order derivatives of $c_M(S)$ can be calculated using a

recursive algorithm in Theorem 2.5 to be introduced in Chapter 2.4. In particular, the second order derivatives of $c_M(S)$ can be evaluated by solving a simple linear system using the first order derivatives, as given in Theorem B.1.

2.4 Central Moments of Matrix Fisher Distribution

This section presents a new algorithms to compute the higher order moments of the matrix Fisher distribution on $\text{SO}(3)$. The algorithm is recursive in the order of the moments, and is able to compute moments up to an arbitrary order in theory, despite it becomes prohibitively computational expansive when the order is too high. A set of non-recursive expressions for the second and third order moments are given in Appendix B. Also, if not stated otherwise, this chapter assumes s_1, s_2 and $|s_3|$ have distinct values to avoid a singular case in the recursive formula. Nonetheless, since the moments are continuous in S , they can be approximated by making s_1, s_2 and $|s_3|$ slightly different if they have repeated values. However, this approximation becomes inaccurate when the order is very high due to the accuracy limitation of floating point arithmetic.

2.4.1 The Central Moments

For a random rotation matrix $R \sim \mathcal{M}(F)$, let $F = USV^T$ be its pSVD, and $Q = U^T RV$. This section calculates the following moments

$$\mathbb{E} [Q_{i_1 j_1} \cdots Q_{i_n j_n}] = \frac{1}{c(S)} \int_{Q \in \text{SO}(3)} Q_{i_1 j_1} \cdots Q_{i_n j_n} \text{etr}(SQ^T) dQ, \quad (2.42)$$

for the indices $i_1, j_1, i_2, j_2, \dots, i_n, j_n \in \{1, 2, 3\}$. This is referred to as the *central* moment because $Q = I_{3 \times 3}$ when $R = UV^T$. The central moments can be evaluated using the moment generating function of the matrix Fisher distribution, given in [54, 64] as

$$M(T) = \mathbb{E} [\text{etr}(T^T Q)] = \frac{c(S+T)}{c(S)}, \quad (2.43)$$

for $T \in 3 \times 3$. And the moments are calculated by differentiating $M(T)$ with respect to T evaluated at $T = 0$, namely

$$\mathbb{E}[Q_{i_1 j_1} \cdots Q_{i_n j_n}] = \frac{1}{c(S)} \left. \frac{\partial^n c(S+T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_n j_n}} \right|_{T=0}. \quad (2.44)$$

In [64, 54], the first order moment, and the second order moment of the form $\mathbb{E}[Q_{ii} Q_{jj}]$ are calculated, but no expressions for other moments are available.

Let the pSVD of $S + T$ be $S + T = U' S' V'^T$, where $S' = \text{diag}(s'_1, s'_2, s'_3) \in \mathbb{R}^{3 \times 3}$ and $U', V' \in \text{SO}(3)$. The matrices U', S', V' are considered as functions of T . For example, $S'|_{T=0} = S$. Since $c(S+T) = c(S')$, the derivative in (2.44) can be expanded using the chain rule by first taking its derivatives with respect to s'_α , and then multiplying the derivative of s'_α with respect to T_{ij} , for $\alpha = 1, 2, 3$. To do this, a notation $\mathcal{C}(n, k)$ needs to be introduced, defined as

$$\begin{aligned} \mathcal{C}(n, k) = & \left\{ \{m_1, \dots, m_k\} \mid m_1, \dots, m_k \text{ are non-empty subsets} \right. \\ & \left. \text{of } \{1, \dots, n\} \text{ and form its partition} \right\}. \end{aligned} \quad (2.45)$$

For example, $\mathcal{C}(3, 2)$ is a partition of $\{1, 2, 3\}$ with two components, i.e., $\mathcal{C}(3, 2) = \left\{ \{\{1\}, \{2, 3\}\}, \{\{2\}, \{1, 3\}\}, \{\{3\}, \{1, 2\}\} \right\}$. Then, (2.44) is expanded as follows.

Theorem 2.3. *Equation (2.44) satisfies*

$$\begin{aligned} \frac{\partial^n c(S+T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_n j_n}} = & \sum_{\alpha_1=1}^3 \frac{\partial c(S+T)}{\partial s'_{\alpha_1}} \frac{\partial^n s'_{\alpha_1}}{\partial T_{i_1 j_1} \cdots \partial T_{i_n j_n}} + \cdots \\ & + \sum_{\alpha_1=1}^3 \cdots \sum_{\alpha_k=1}^3 \frac{\partial^k c(S+T)}{\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k}} \left(\sum_{\{m_1, \dots, m_k\} \in \mathcal{C}(n, k)} \partial_{m_1} s'_{\alpha_1} \cdots \partial_{m_k} s'_{\alpha_k} \right) + \cdots \\ & + \sum_{\alpha_1=1}^3 \cdots \sum_{\alpha_n=1}^3 \frac{\partial^n c(S+T)}{\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_n}} \frac{\partial s'_{\alpha_1}}{\partial T_{i_1 j_1}} \cdots \frac{\partial s'_{\alpha_n}}{\partial T_{i_n j_n}}, \end{aligned} \quad (2.46)$$

where $\partial_{m_r} s'_{\alpha_r} = \partial^t s'_{\alpha_r} / (\partial T_{i_{m_r^1} j_{m_r^1}} \cdots \partial T_{i_{m_r^t} j_{m_r^t}})$ if $m_r = \{m_r^1, \dots, m_r^t\}$, for $r = 1, \dots, k$.

Proof. This can be shown by a straightforward but tedious algebraic manipulation with induction. Clearly (2.46) is true when $n = 1$. Suppose (2.46) holds for $n - 1$. Then for n , the k -th term is contributed by two terms from the expansion for $n - 1$: the $(k - 1)$ -th term and the k -th term for $n - 1$. Specifically, it is written as

$$\begin{aligned} \frac{\partial}{\partial T_{i_n j_n}} & \left(\frac{\partial^{n-1} c(S+T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_{n-1} j_{n-1}}} \right) = \cdots \\ & + \sum_{\alpha_1=1}^3 \cdots \sum_{\alpha_k=1}^3 \frac{\partial^k c(S+T)}{\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k}} \frac{\partial}{\partial T_{i_n j_n}} \left(\sum_{\{m_1, \dots, m_k\} \in \mathcal{C}(n-1, k)} \partial_{m_1} s'_{\alpha_1} \cdots \partial_{m_k} s'_{\alpha_k} \right) \\ & + \sum_{\alpha_1=1}^3 \cdots \sum_{\alpha_k=1}^3 \frac{\partial^k c(S+T)}{\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k}} \left(\sum_{\{m_1, \dots, m_{k-1}\} \in \mathcal{C}(n-1, k-1)} \partial_{m_1} s'_{\alpha_1} \cdots \partial_{m_{k-1}} s'_{\alpha_{k-1}} \right) \frac{\partial s'_{\alpha_k}}{\partial T_{i_n j_n}} + \cdots \end{aligned}$$

Combine these two terms with the common factor. Then, the remaining parts can be reorganized as the summation in the first parentheses above, with the summation in $\mathcal{C}'(n, k)$ given by

$$\begin{aligned} \mathcal{C}'(n, k) = & \{ \{m_1 \cup \{n\}, m_2, \dots, m_k\} \mid \{m_1, \dots, m_n\} \in \mathcal{C}(n-1, k) \} \cup \cdots \\ & \cup \{ \{m_1, \dots, m_{k-1}, m_k \cup \{n\}\} \mid \{m_1, \dots, m_n\} \in \mathcal{C}(n-1, k) \} \\ & \cup \{ \{m_1, \dots, m_{k-1}, \{n\}\} \mid \{m_1, \dots, m_{k-1}\} \in \mathcal{C}(n-1, k-1) \} = \mathcal{C}(n, k), \end{aligned}$$

which results in the k -th term of (2.46). \square

In order to evaluate (2.46), $\partial_m s'_\alpha$ and $\partial^k c(S+T)/(\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k})$ need to be calculated. In the subsequent two subsections, it is shown that both of them can be evaluated recursively. Note that $\partial_m s'_\alpha$ is the derivative of a singular value with respect to the matrix elements, so it is assumed that $s'_1 \neq s'_2 \neq |s'_3|$ to guarantee differentiability. This is legitimate because we only need to consider T varying in a small neighborhood of 0, and $s_1 \neq s_2 \neq |s_3|$ has already been assumed.

2.4.2 Calculating $\partial_m s'_\alpha$

In this subsection $\partial_m s'_\alpha$ presented in (2.46) is denoted by $\partial^t s'_\alpha / (\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t})$, i.e., the letter m in the subscripts for T is omitted for simplicity. It is first shown that some $\partial_m s'_\alpha = 0$ by just looking at the indices $\{i, j\}$. This is formulated in the following lemma:

Lemma 2.2. *Suppose $s'_1 \neq s'_2 \neq |s'_3|$, and let $n_1 \in \mathbb{N}$ be the number of occurrence of 1 in $\{i_1, j_1, \dots, i_t, j_t\}$. The integers n_2 and n_3 are defined similarly. Then $\partial_m s'_\alpha|_{T=0} = 0$ if any of n_1, n_2 , or n_3 is odd.*

Proof. Let $(\gamma_1, \gamma_2, \gamma_3) \in \{(1, -1, -1), (-1, 1, -1), (-1, -1, 1)\}$ and $\Gamma = \text{diag}(\gamma_1, \gamma_2, \gamma_3)$. Note that $\Gamma \in \text{SO}(3)$, so $S + \Gamma^T T \Gamma = \Gamma^T (S + T) \Gamma = \Gamma^T U' S' (\Gamma^T V')^T$. Because Γ^T does not change the polarity of U' or V' , s'_1, s'_2, s'_3 are also the proper singular values of $S + \Gamma^T T \Gamma$. This proves

$$\frac{\partial^t s'_\alpha(S + T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} = \frac{\partial^t s'_\alpha(S + \Gamma^T T \Gamma)}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}}.$$

On the other hand, since $(\Gamma^T T \Gamma)_{ij} = \gamma_i \gamma_j T_{ij}$, we have $\partial(\Gamma^T T \Gamma)_{i'j'}/\partial T_{ij} = \delta_i^{i'} \delta_j^{j'} \gamma_i \gamma_j$, and the higher order derivatives of $(\Gamma^T T \Gamma)_{i'j'}$ with respect to ∂T_{ij} are zero since $\delta_i^{i'} \delta_j^{j'} \gamma_i \gamma_j$ is a constant. Therefore, using a similar calculation as in (2.46), it can be shown that

$$\begin{aligned} \left. \frac{\partial^t s'_\alpha(S + \Gamma^T T \Gamma)}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \right|_{T=0} &= \left. \frac{\partial^t s'_\alpha(S + \Gamma^T T \Gamma)}{\partial(\Gamma^T T \Gamma)_{i_1 j_1} \cdots \partial(\Gamma^T T \Gamma)_{i_t j_t}} \right|_{T=0} \gamma_1 \gamma_{j_1} \cdots \gamma_t \gamma_{j_t} \\ &= \left. \frac{\partial^t s'_\alpha(S + T)}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \right|_{T=0} \gamma_1^{n_1} \gamma_2^{n_2} \gamma_3^{n_3}. \end{aligned}$$

Suppose n_1 is odd. Since $n_1 + n_2 + n_3 = 2t$ is even, either n_2 or n_3 is also odd. In case n_2 is odd and n_3 is even, substitute $(\gamma_1, \gamma_2, \gamma_3) = (1, -1, -1)$ into the above two equations to obtain $\partial_m s'_\alpha = -\partial_m s'_\alpha$, which yields $\partial_m s'_\alpha = 0$. Other cases can be shown similarly. \square

Corollary 2.1. $E[Q_{i_1 j_1} \cdots Q_{i_n j_n}] = 0$ if $\{i_1, j_1, \dots, i_n, j_n\}$ has an odd number of 1, 2, or 3.

Proof. When $s_1 \neq s_2 \neq |s_3|$, it follows directly from Lemma 2.2 and (2.46). Next, suppose $s_1, s_2, |s_3|$ have repeated values. For example, let $s_1 = s_2$. It is clearly seen from (2.38)

and (2.42) that $E[Q_{i_1 j_1} \cdots Q_{i_n j_n}]$ is a continuous function of S . Then, we can take the limit $s_1 \rightarrow s_2$ of (2.46) to show that the expectation vanishes. The other cases can be shown similarly. \square

Corollary 2.1 specifies cases for zero moments, which generalizes the similar results for the first and second order moments in [54]. Even when the condition for Corollary 2.1 is not satisfied, Lemma 2.2 simplifies (2.46) substantially. There is another trick to simplify calculations: the derivative of s'_α is unchanged after switching the subscript of T .

Lemma 2.3. *If $s'_1 \neq s'_2 \neq |s'_3|$, then*

$$\frac{\partial^n s'_\alpha}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \Big|_{T=0} = \frac{\partial^n s'_\alpha}{\partial T_{j_1 i_1} \cdots \partial T_{j_t i_t}} \Big|_{T=0}. \quad (2.47)$$

Proof. It is straightforward to show $S + T$ and $S + T^T$ have the same singular values since if $S + T = U' S' V'^T$, then $S + T^T = V' S' U'^T$. The remaining proof is the same as Lemma 2.2. \square

Together with (2.44) and (2.46), this shows the following symmetry of the moments.

Corollary 2.2.

$$E[Q_{i_1 j_1} \cdots Q_{i_n j_n}] = E[Q_{j_1 i_1} \cdots Q_{j_n i_n}]. \quad (2.48)$$

Finally, a recursive calculation for the non-zero $\partial_m s'_\alpha|_{T=0}$ is presented. In preparation for this, the derivatives of SVD [95] $S + T = U' S' V'^T$ are first introduced:

$$\frac{\partial s'_\alpha}{\partial T_{ij}} = U'_{i\alpha} V'_{j\alpha}, \quad (2.49)$$

with

$$\frac{\partial U'_{kl}}{\partial T_{ij}} = \left(U' \Omega_U^{ij} \right)_{kl} \quad \frac{\partial U'_{kl}}{\partial T_{ij}} = - \left(V' \Omega_V^{ij} \right)_{kl}, \quad (2.50)$$

where Ω_U^{ij} and Ω_V^{ij} are 3-by-3 skew-symmetric matrices with their off-diagonal elements

satisfying

$$\begin{aligned} s'_l \Omega_{U_{kl}}^{ij} + s'_k \Omega_{V_{kl}}^{ij} &= U'_{ik} V'_{jl}, \\ s'_k \Omega_{U_{kl}}^{ij} + s'_l \Omega_{V_{kl}}^{ij} &= -U'_{il} V'_{jk}. \end{aligned} \quad (2.51)$$

for $k \neq l \in \{1, 2, 3\}$. It should be noted that since $s'_1 \neq s'_2 \neq |s'_3|$, the above linear system always has a unique solution. In addition, the generalized Leibniz rule is adapted to the notation \mathcal{C} in (2.45) as below.

Lemma 2.4. Suppose $f, g \in C^t(N)$, then for any $i_1, \dots, i_t \in \{1, \dots, N\}$

$$\frac{\partial^t(fg)}{\partial x_{i_1} \cdots \partial x_{i_t}} = f \frac{\partial^t g}{\partial x_{i_1} \cdots \partial x_{i_t}} + g \frac{\partial^t f}{\partial x_{i_1} \cdots \partial x_{i_t}} + \sum_{\{m_1, m_2\} \in \mathcal{C}(t, 2)} (\partial_{m_1} f \partial_{m_2} g + \partial_{m_2} f \partial_{m_1} g). \quad (2.52)$$

The recursive calculation scheme is formulated in the following theorem.

Theorem 2.4. If $s'_1 \neq s'_2 \neq |s'_3|$, then $\partial_m s'_\alpha$ is given by

$$\frac{\partial^t s'_\alpha}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} = \frac{\partial^{t-1} (U'_{i_1 \alpha} V'_{j_1 \alpha})}{\partial T_{i_2 j_2} \cdots \partial T_{i_t j_t}}, \quad (2.53)$$

which can be evaluated recursively by the chain rule and

$$\begin{aligned} \frac{\partial^t U'_{kl}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} &= \frac{\partial^{t-1} (U'_{kp} \Omega_{U_{pl}}^{i_1 j_1})}{\partial T_{i_2 j_2} \cdots \partial T_{i_t j_t}} + \frac{\partial^{t-1} (U'_{kq} \Omega_{U_{ql}}^{i_1 j_1})}{\partial T_{i_2 j_2} \cdots \partial T_{i_t j_t}}, \\ \frac{\partial^t V'_{kl}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} &= -\frac{\partial^{t-1} (V'_{kp} \Omega_{V_{pl}}^{i_1 j_1})}{\partial T_{i_2 j_2} \cdots \partial T_{i_t j_t}} - \frac{\partial^{t-1} (V'_{kq} \Omega_{V_{ql}}^{i_1 j_1})}{\partial T_{i_2 j_2} \cdots \partial T_{i_t j_t}}, \end{aligned} \quad (2.54)$$

where $\{p, q\} = \{1, 2, 3\} \setminus \{l\}$, and

$$\begin{aligned} s'_l \frac{\partial^t \Omega_{U_{kl}}^{ij}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} + s'_k \frac{\partial^t \Omega_{V_{kl}}^{ij}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} &= \frac{\partial^t (U'_{ik} V'_{jl})}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} - \frac{\partial^t s'_l}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \Omega_{U_{kl}}^{ij} \\ &\quad - \frac{\partial^t s'_k}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \Omega_{V_{kl}}^{ij} - \sum_{\{m_1, m_2\} \in \mathcal{C}(t, 2)} \left(\partial_{m_1} s'_l \partial_{m_2} \Omega_{U_{kl}}^{ij} + \partial_{m_2} s'_l \partial_{m_1} \Omega_{U_{kl}}^{ij} \right) \\ &\quad - \sum_{\{m_1, m_2\} \in \mathcal{C}(t, 2)} \left(\partial_{m_1} s'_k \partial_{m_2} \Omega_{V_{kl}}^{ij} + \partial_{m_2} s'_k \partial_{m_1} \Omega_{V_{kl}}^{ij} \right), \end{aligned} \quad (2.55a)$$

$$\begin{aligned} s'_k \frac{\partial^t \Omega_{U_{kl}}^{ij}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} + s'_l \frac{\partial^t \Omega_{V_{kl}}^{ij}}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} &= - \frac{\partial^t (U'_{il} V'_{jk})}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} - \frac{\partial^t s'_k}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \Omega_{U_{kl}}^{ij} \\ &\quad - \frac{\partial^t s'_l}{\partial T_{i_1 j_1} \cdots \partial T_{i_t j_t}} \Omega_{V_{kl}}^{ij} - \sum_{\{m_1, m_2\} \in \mathcal{C}(t, 2)} \left(\partial_{m_1} s'_k \partial_{m_2} \Omega_{U_{kl}}^{ij} + \partial_{m_2} s'_k \partial_{m_1} \Omega_{U_{kl}}^{ij} \right) \\ &\quad - \sum_{\{m_1, m_2\} \in \mathcal{C}(t, 2)} \left(\partial_{m_1} s'_l \partial_{m_2} \Omega_{V_{kl}}^{ij} + \partial_{m_2} s'_l \partial_{m_1} \Omega_{V_{kl}}^{ij} \right). \end{aligned} \quad (2.55b)$$

Proof. Equation (2.53) is simply (2.49). Equation (2.54) is constructed by expanding the matrix multiplication of the term in the parenthesis of (2.50) with the skew-symmetry of Ω_U, Ω_V , and then, taking the derivatives with respect to the remaining $T_{i_2 j_2}, \dots, T_{i_t j_t}$. Equation (2.55) is obtained by directly differentiating (2.51). \square

For (2.53) and (2.54), the order of differentiation on the right hand side is one less than the left hand side. For (2.55), note that ∂_{m_1} and ∂_{m_2} are at most $(t-1)$ -th order for $(m_1, m_2) \in \mathcal{C}(t, 2)$. Also, the t -th order differentiation of U and V is reduced to at most the $(t-1)$ -th order of Ω_U and Ω_V as shown in (2.54). Similarly, the t -th order differentiation of s_α corresponds to the $(t-1)$ -th order of U and V , and thus the $(t-2)$ -th order of Ω_U and Ω_V . So in general, the order of differentiation on the right hand side is smaller than the left hand side. This means the recursion will finally terminate when t reaches 1, when the differentiation can be calculated explicitly using (2.49)-(2.51). The recursion stated in the above theorem is very difficult to untwist into a non-recursive expression. In Appendix A, a very special case in Theorem 2.4 is studied and shown to have a non-recursive formula, which exhibits the complexity of this recursion.

2.4.3 Calculating $\partial^k c(S+T)/(\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k})$

In this subsection, a recursive calculation scheme for $\partial^k c(S+T)/(\partial s'_{\alpha_1} \cdots \partial s'_{\alpha_k})|_{T=0} = \partial^k c(S)/(\partial s_{\alpha_1} \cdots \partial s_{\alpha_k})$ is given, such that (2.46) can be evaluated recursively together with Theorem 2.4 in the preceding subsection. First, the moments for the diagonal elements of Q are presented below.

Lemma 2.5.

$$E[Q_{i_1 i_1} \cdots Q_{i_n i_n}] = \frac{1}{c(S)} \frac{\partial^n c(S)}{\partial s_{i_1} \cdots \partial s_{i_n}}. \quad (2.56)$$

Proof. When $T = \text{diag}(T_{11}, T_{22}, T_{33})$ and the diagonal values are small enough, the singular value decomposition of $S+T$ is $I_{3 \times 3} \text{diag}(s_1 + T_{11}, s_2 + T_{22}, s_3 + T_{33}) I_{3 \times 3}$, so $\partial s'_\alpha / \partial T_{ii} = \delta_i^\alpha$. And the higher order derivatives vanish because δ_i^α is a constant function of T . Substitute these into (2.46), then (2.56) is proved when $s_1 \neq s_2 \neq |s_3|$. When $s_1, s_2, |s_3|$ have repeated values, (2.56) still holds by the continuity argument in Corollary 2.1. \square

Next, a recursive calculation for $\partial^n c(S)/(\partial s_{i_1} \cdots \partial s_{i_n})$ is presented in the following theorem.

Theorem 2.5. $\partial^n c(S)/(\partial s_{i_1} \cdots \partial s_{i_n})$ can be calculated as follows: (i) if $i_1 = i_2$,

$$\frac{\partial^n c(S)}{\partial s_{i_1}^2 \partial s_{i_3} \cdots \partial s_{i_n}} = \frac{\partial^{n-2} c(S)}{\partial s_{i_3} \cdots \partial s_{i_n}} - \left. \frac{\partial^n c(S+T)}{\partial T_{i_1 j_1}^2 \partial T_{i_3 i_3} \cdots \partial T_{i_n i_n}} \right|_{T=0} - \left. \frac{\partial^n c(S+T)}{\partial T_{i_1 k_1}^2 \partial T_{i_3 i_3} \cdots \partial T_{i_n i_n}} \right|_{T=0}, \quad (2.57)$$

where $i_1, j_1, k_1 \in \{1, 2, 3\}$, $i_1 \neq j_1 \neq k_1$, and (ii) if $i_1 \neq i_2$

$$\frac{\partial^n c(S)}{\partial s_{i_1} \partial s_{i_2} \cdots \partial s_{i_n}} = \left. \frac{\partial^{n-1} c(S+T)}{\partial T_{k_1 k_1} \partial T_{i_3 i_3} \cdots \partial T_{i_n i_n}} \right|_{T=0} + \left. \frac{\partial^n c(S+T)}{\partial T_{i_1 i_2} \partial T_{i_2 i_1} \partial T_{i_3 i_3} \cdots \partial T_{i_n i_n}} \right|_{T=0}, \quad (2.58)$$

where $i_1, i_2, k_1 \in \{1, 2, 3\}$, $i_1 \neq i_2 \neq k_1$.

Proof. Suppose $i_1 = i_2$. As $Q \in \text{SO}(3)$, $Q_{i_1 i_1}^2 = 1 - Q_{i_1 j_1}^2 - Q_{i_1 k_1}^2$ because $Q^T Q = I_{3 \times 3}$. Thus $E[Q_{i_1 i_1}^2 Q_{i_3 i_3} \cdots Q_{i_n i_n}] = E[Q_{i_3 i_3} \cdots Q_{i_n i_n}] - E[Q_{i_1 j_1}^2 Q_{i_3 i_3} \cdots Q_{i_n i_n}] - E[Q_{i_1 k_1}^2 Q_{i_3 i_3} \cdots Q_{i_n i_n}]$, and

(2.57) follows from (2.44) and Lemma 2.5. The crucial observation is that the right hand side of (2.57) only involves derivatives of $c(S)$ that have lower order than n , which is straightforward from (2.46) and $\partial s_\alpha / \partial T_{i_1 j_1} |_{T=0} = \partial s_\alpha / \partial T_{i_1 k_1} |_{T=0} = 0$ since $i_1 \neq j_1$ and $i_1 \neq k_1$. Thus, (2.57) defines a recursion which terminates in finite steps when n hits 1.

Next suppose $i_1 \neq i_2$. Note that $Q_{k_1 k_1} = Q_{i_1 i_1} Q_{i_2 i_2} - Q_{i_1 i_2} Q_{i_2 i_1}$ by (2.11), so (2.58) can be shown similarly as the case (i). \square

When $n = 1$, $\partial c(S) / \partial s_i$ can be calculated using the one dimensional integration formula in (2.41), as the terminating step for the above recursion. Also, in Appendix B, the second order derivative of $c(S)$ is obtained by solving a very simple linear system using the first order derivatives.

2.4.4 Numerical Verification

The presented algorithm is validated numerically with 32 different S , where $10 \leq s_1 \leq 50$, $10 \leq s_2 \leq s_1$, $-s_2 \leq s_3 \leq s_2$, and $|s_i| - |s_j| \geq 5$ for any $i \neq j \in \{1, 2, 3\}$. A hundred million random samples are drawn from the corresponding matrix Fisher distributions using the method in [53]. All of central moments up to the fourth order, a fifth order $E[Q_{23}Q_{32}Q_{12}Q_{23}Q_{31}]$, and a sixth order $E[Q_{11}Q_{23}Q_{32}Q_{12}Q_{23}Q_{31}]$ are calculated empirically from the random samples, and they are compared with the proposed recursive method. The differences between zero moments specified by Corollary 2.1 are at most 7.08×10^{-5} for the two methods, and the percentage differences between non-zero moments are less than 5.6%. Two specific examples are listed in Table 2.1. The computation time for the sixth order moment is about 2 minutes, and for a fourth order moment is about 0.2 seconds, using a typical desktop CPU. A MATLAB implementation of the presented algorithm can be found in [120].

When $|s_i| = |s_j|$, the linear system in (2.51) and (2.55) is singular, so this poses a challenge for floating-point arithmetic when $|s_i|$ and $|s_j|$ are very close. In Figure 2.1, the moments when $s_1 = 25$, $s_2 = 10$, and s_3 approaching to s_2 are compared between Monte

Table 2.1: Maximum difference of moments between Monte Carlo method and the proposed recursive algorithm

Zero or nonzero Order	zero ($\times 10^{-5}$)			nonzero (%)			$E[Q5]^*$	$E[Q6]^*$
	2nd	3rd	4th	2nd	3rd	4th		
$S = \text{diag}(25, 10, 5)$	3.21	3.06	2.92	0.04	0.42	0.47	0.06%	0.06%
$S = \text{diag}(25, 10, -5)$	6.69	6.33	6.00	0.03	0.22	0.37	0.002%	0.004%

* $E[Q5]$ and $E[Q6]$ denote $E[Q_{23}Q_{32}Q_{12}Q_{23}Q_{31}]$ and $E[Q_{11}Q_{23}Q_{32}Q_{12}Q_{23}Q_{31}]$ respectively.

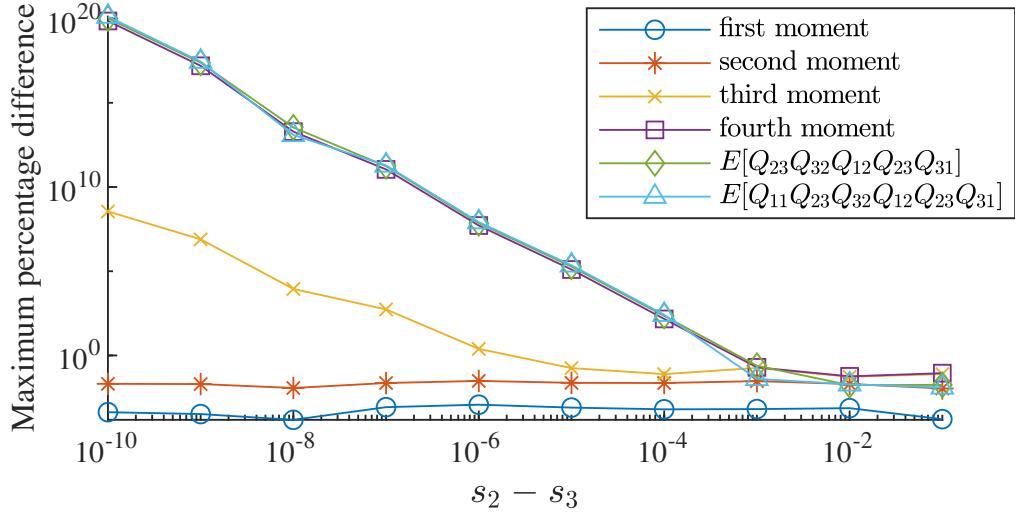


Figure 2.1: Maximum difference between Monte Carlo method and the proposed recursive calculation when s_2 is close to s_3 , with $s_1 = 25$, $s_2 = 10$. Double precision floating-point format is used in the recursive calculation.

Carlo method and the proposed recursive calculation. It is seen that the first and second order moments remain accurate for all tested S , whereas the third order moments start to deviate from the sample moments when $s_2 - s_3$ is on the level of 10^{-6} , and the fourth to sixth order moments start to deviate when $s_2 - s_3$ is on the level of 10^{-4} . Methods to eliminate this numerical inaccuracy remains to be investigated in future work.

2.5 Highly Concentrated Approximations of Matrix Fisher Distribution

The normalizing constant of the matrix Fisher distribution and its derivatives can be evaluated using the one dimensional integration formulae given in (2.39) and (2.41), or using the

saddle point approximation as discussed in [33]. However, these approaches are very computationally demanding. Also, to inference the parameters F from random samples, one must solve the partial differential equation in (2.20). Because $d_i + d_j$ monotonically increases with $s_i + s_j$ for $i \neq j \in \{1, 2, 3\}$, and $d_i + d_j$ is upper bounded by 2, the average rate of change of D with respect to S becomes very small when S is very large. This makes (2.20) extremely difficult to solve when S has very large values.

However, large S means the matrix Fisher distribution is highly concentrated. As an analog of Gaussian distribution on $\text{SO}(3)$, it can actually be approximated by a Gaussian distribution when S is large. This is similar to the von Mises and von Mises–Fisher distributions on \mathbb{S}^n [80]. Nevertheless, since $\text{SO}(3)$ has three degrees of freedom, the number of degrees of freedom in which the distribution is highly concentrated must be specified. In the subsequent two subsections, the cases when the matrix Fisher distribution is highly concentrated in all three degrees of freedom, and in two degrees of freedom, are discussed. And these approximations lead to simplifications of the normalizing constant, and its first order derivatives, which are computational costly to be evaluated. The results can be further used to expedite the MLE of the parameter F , which requires solving the partial differential equation (2.20) involving the normalizing constant. Unfortunately, we were unable to give an approximation when the matrix Fisher distribution is only highly concentrated in one degree of freedom using the same strategy.

2.5.1 Highly Concentrated in Three Degrees of Freedom

As shown in (2.21), the concentration of a matrix Fisher distribution along the i -th principal axis is controlled by $s_j + s_k$ for $i \neq j \neq k$. Thus $s_1 + s_2 \geq s_1 + s_3 \geq s_2 + s_3$, $s_2 + s_3 \gg 0$ implies the distribution is highly concentrated in all three degrees of freedom. The Gaussian approximation for this case has been studied in [63], and is reviewed in this subsection.

Theorem 2.6 ([63]). *Let $R \sim \mathcal{M}(F)$, where $F = USV^T$ is the pSVD of F . Suppose $s_2 + s_3 \gg 0$. Let $Q = U^T RV = \exp(\hat{\eta})$, then $\eta \sim \mathcal{N}(0, (\text{tr}(S)I_{3 \times 3} - S)^{-1})$, where \sim denotes*

“approximately follows”.

Proof. Let $\Sigma = (\text{tr}(S)I_{3 \times 3} - S)^{-1} = \text{diag}\left(\frac{1}{s_2+s_3}, \frac{1}{s_1+s_3}, \frac{1}{s_1+s_2}\right) \in \mathbb{R}^{3 \times 3}$, and let $\xi \in \mathbb{R}^3$ be defined as $\sqrt{\Sigma}\xi = \eta$. Then the density (2.16) for R can be written as

$$\begin{aligned} \text{etr}(FR^T) &= \text{etr}\left(S \exp\left((\sqrt{\Sigma}\xi)^\wedge\right)\right) \\ &= \text{etr}\left(S\left(I_{3 \times 3} + (\sqrt{\Sigma}\xi)^\wedge + \frac{1}{2}((\sqrt{\Sigma}\xi)^\wedge)^2 + O(\sqrt{\Sigma})\right)\right) \\ &= \text{etr}\left(\frac{1}{2}S((\sqrt{\Sigma}\xi)^\wedge)^2 + O(\sqrt{\Sigma})\right) \\ &\approx \text{etr}(S) \exp\left(-\frac{1}{2}\eta^T \Sigma^{-1} \eta\right). \end{aligned}$$

This finishes the proof. \square

Theorem 2.6 gives the following approximation of the normalizing constant.

Corollary 2.3 ([63]). *If $s_2 + s_3 \gg 0$, then*

$$c(S) \approx \frac{\text{etr}(S)}{\sqrt{8\pi(s_1+s_2)(s_1+s_3)(s_2+s_3)}} \quad (2.59)$$

$$\frac{1}{c(S)} \frac{\partial c(S)}{\partial s_i} \approx 1 - \frac{1}{2} \left(\frac{1}{s_i+s_j} + \frac{1}{s_i+s_k} \right), \quad (2.60)$$

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

Proof. According to [19, Chapter 12.1.2], (2.38) can be written as

$$c(S) = \int_{\|\eta\|<\pi} \text{etr}(S \exp(\hat{\eta})) \frac{1 - \cos \|\eta\|}{4\pi^2 \|\eta\|^2} d\eta,$$

where $d\eta$ is the Lebesgue measure on \mathbb{R}^3 . Using the calculation in Theorem 2.6, and the Taylor expansion of $1 - \cos x$, it can further be calculated as

$$\begin{aligned} c(S) &= \text{etr}(S) \int_{\|\eta\|<\pi} \exp\left\{-\frac{1}{2}\eta^T \Sigma^{-1} \eta + O(\sqrt{\Sigma})\right\} (1 + O(\Sigma)) \frac{1}{8\pi^2} d\eta \\ &\approx \int_{\mathbb{R}^3} \exp\left\{-\frac{1}{2}\eta^T \Sigma^{-1} \eta\right\} \frac{1}{8\pi^2} d\eta. \end{aligned}$$

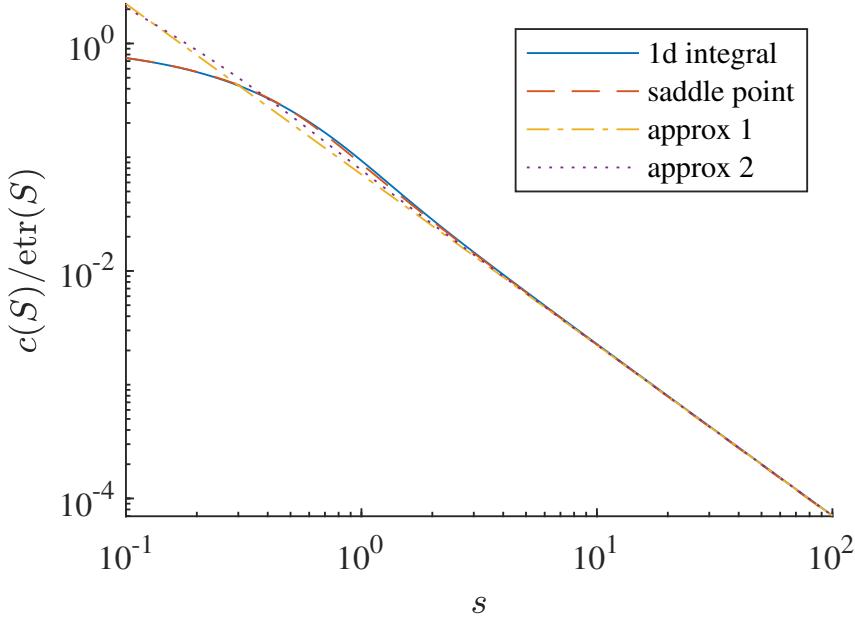


Figure 2.2: Comparison of the normalizing constant $c(S)$ calculated using the one dimensional integral formula (2.39), the saddle point approximation [58], the highly concentrated approximations (2.59) (approx 1) and (2.62) (approx 2). The parameter is $S = sI_{3 \times 3}$.

Then (2.59) can be obtained using the normalizing constant of $\mathcal{N}(0, \Sigma)$, and its derivation can also be calculated directly as (2.60). \square

Corollary 2.3 provides a very fast computation for the normalizing constant $c(S)$ and its first order derivatives, using only basic functions. Its accuracy is demonstrated in Figure 2.2 against the one dimensional integral formula (2.41), and the saddle point approximation in [58]. In addition, since the case when the matrix Fisher distribution is highly concentrated in two dimensions is a special case when it is highly concentrated in all three dimensions, the approximation in (2.62) is also compared. The tested parameter is when $S = sI_{3 \times 3}$, for s ranging from 0.1 to 100. It is shown that all calculations agree well when $s > 50$, whereas when $s < 50$, the two approximations deviate from the other two calculations, due to that the distribution is no longer highly concentrated.

The derivative of $c(S)$ is compared as $1 - d_1 = 1 - \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_1}$ in Figure 2.3. Again, all calculations are very close when $s > 10$, but the two approximations become less accurate when the distribution is not highly concentrated. It should be noted that for extremely large s ,

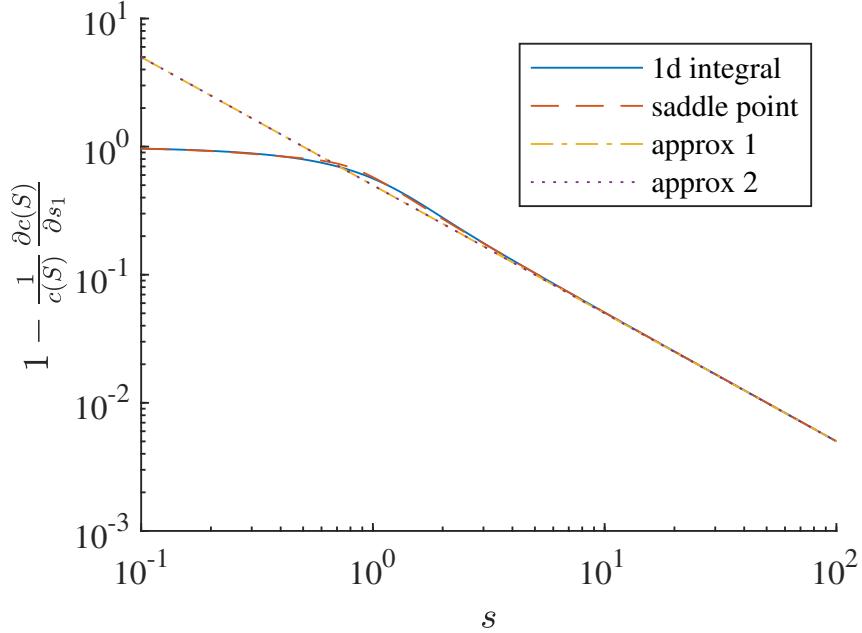


Figure 2.3: Comparison of the first order moment $1 - d_1$ calculated using the one dimensional integral formula (2.41), the saddle point approximation [58, 59], the highly concentrated approximations (2.60) (approx 1) and (2.63) (approx 2). The parameter is $S = sI_{3 \times 3}$.

for example when $s > 10^4$, the one dimensional integral formula, especially for the derivative 2.41, becomes inaccurate when implemented with MATLAB `integral` function.

2.5.2 Highly Concentrated in Two Degrees of Freedom

In this subsection, the case when $s_1 + s_3 \gg s_2 + s_3 \geq 0$ is considered, i.e., the matrix Fisher distribution is highly concentrated along the second and third principal axes, but has large dispersion along the first axis. Note that because $s_1 \geq s_2 \geq |s_3| \geq 0$, $s_1 + s_3 \gg s_2 + s_3$ is equivalent to $s_1 \gg s_2$. It is proved in this case, the matrix Fisher distribution can be approximated by a combination of a two dimensional Gaussian distribution in \mathbb{R}^2 , and a one dimensional von Mises distribution on \mathbb{S}^1 .

Theorem 2.7. *Let $R \sim \mathcal{M}(F)$, where $F = USV^T$ is the pSVD of F . Suppose $s_1 + s_3 \gg s_2 + s_3 \geq 0$. Let $Q = U^T RV = \exp(\hat{\eta}) \exp(\hat{\eta}')$, where $\eta = [0, \eta_2, \eta_3]^T$, and $\eta' = [\eta_1, 0, 0]^T$. Then $\eta_3 \sim \mathcal{VM}(0, s_2 + s_3)$, and $[\eta_2, \eta_3]^T \sim \mathcal{N}\left(0, \text{diag}\left(\frac{1}{s_1 + s_3}, \frac{1}{s_1 + s_2}\right)\right)$, and they are approximately independent.*

Proof. let $\Sigma = \text{diag} \left(0, \frac{1}{s_1+s_3}, \frac{1}{s_1+s_2} \right)$, and $\eta = \sqrt{\Sigma} \xi$. Then $\exp(\hat{\eta})$ can be written as

$$\exp(\hat{\eta}) = I_{3 \times 3} + \sqrt{\Sigma} \hat{\xi} + \frac{1}{2} \left(\sqrt{\Sigma} \hat{\xi} \right)^2 + O(s_1^{-3/2}).$$

Also, $\exp(\hat{\eta}')$ can be explicitly calculated as

$$\exp(\hat{\eta}') = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \eta_1 & -\sin \eta_1 \\ 0 & \sin \eta_1 & \cos \eta_1 \end{bmatrix}.$$

Combining the above two equations, the density function for R becomes

$$\begin{aligned} \text{etr}(FR^T) &= \text{etr}(S \exp(\hat{\eta}) \exp(\hat{\eta}')) \\ &= \exp(s_1) \cdot \exp((s_2 + s_3) \cos \eta_1) \cdot \exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\} \\ &\quad \cdot \exp \left\{ -\frac{1}{2} ((s_2 \eta_3^2 + s_3 \eta_2^2)(\cos \eta_1 - 1) + (s_3 - s_2) \eta_2 \eta_3 \sin \eta_1) + O(s_1^{-1/2}) \right\}. \end{aligned}$$

Note that the last exponential term can be simplified as

$$\begin{aligned} &(s_2 \eta_3^2 + s_3 \eta_2^2)(\cos \eta_1 - 1) + (s_3 - s_2) \eta_2 \eta_3 \sin \eta_1 \\ &= \left(\frac{s_2}{s_1+s_2} \xi_3^2 + \frac{s_3}{s_1+s_3} \xi_2^2 \right) (\cos \eta_1 - 1) + \frac{s_3 - s_2}{\sqrt{s_1+s_3} \sqrt{s_1+s_2}} \xi_2 \xi_3 \sin \eta_1 \\ &= O(s_1^{-1}). \end{aligned}$$

So the density function for R can be approximated by

$$\text{etr}(FR^T) \approx \exp(s_1) \cdot \exp((s_2 + s_3) \cos \eta_1) \cdot \exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\},$$

and the desired result is proved. \square

Before looking into how this approximation can be applied to calculating the normalizing

constant $c(S)$, a decomposition of the Haar measure dQ on $\mathrm{SO}(3)$ needs to be introduced first [26, Chapter 4]. Let $\mathbf{K} = \{\exp\{\theta\hat{e}_1\} \mid 0 \leq \theta < 2\pi\}$ be a subgroup of $\mathrm{SO}(3)$, it is clear that \mathbf{K} is isomorphic to $\mathrm{SO}(2) \cong \mathbb{S}^1$. The quotient group of \mathbf{K} is $\mathrm{SO}(3)/\mathbf{K} = \{T \in \mathrm{SO}(3) \mid T\mathbf{K}\}$, which is isomorphic to \mathbb{S}^2 under the map

$$j : \mathrm{SO}(3)/\mathbf{K} \rightarrow \mathbb{S}^2, \quad T\mathbf{K} \mapsto Te_1. \quad (2.61)$$

Therefore, for any $Q \in \mathrm{SO}(3) = \mathrm{SO}(3)/\mathbf{K} \times \mathbf{K} \cong \mathbb{S}^2 \times \mathbb{S}^1$, it can be decomposed as $Q = T(x)\exp\{\theta\hat{e}_1\}$, for $x \in \mathbb{S}^2$ and $0 \leq \theta < 2\pi$, where $T\mathbf{K} = j^{-1}(x)$. And the Haar measure dQ can be decomposed as $\frac{1}{8\pi^2}d\theta d\sigma(x)$, where $d\sigma(x)$ is the surface area measure on \mathbb{S}^2 . This decomposition can be used to obtain an approximation of the normalizing constant given in the next Corollary.

Corollary 2.4. *If $s_1 + s_3 \gg s_2 + s_3$, then*

$$c(S) \approx \frac{\exp(s_1)I_0(s_2 + s_3)}{2\sqrt{(s_1 + s_2)(s_1 + s_3)}}, \quad (2.62)$$

$$\frac{1}{c(S)} \frac{\partial c(S)}{\partial s_1} \approx 1 - \frac{1}{2} \left(\frac{1}{s_1 + s_2} + \frac{1}{s_1 + s_3} \right), \quad (2.63)$$

$$\frac{1}{c(S)} \frac{\partial c(S)}{\partial s_j} \approx \frac{I_1(s_2 + s_3)}{I_0(s_2 + s_3)} - \frac{1}{2} \frac{1}{s_1 + s_j}, \quad (2.64)$$

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

Proof. As shown in Theorem 2.7, $c(S)$ can be calculated as

$$c(S) \approx \exp(s_1) \int_{Q \in \mathrm{SO}(3)} \exp((s_2 + s_3) \cos \eta_1) \cdot \exp\left\{-\frac{1}{2}((s_1 + s_3)\eta_2^2 + (s_1 + s_2)\eta_3^2)\right\} dQ$$

Let $x = [\cos \alpha, \sin \alpha \cos \beta, \sin \alpha \sin \beta]^T \in \mathbb{S}^2$ be written in spherical coordinates, then using (2.61), it can be shown that $j^{-1}(x) = \exp\{\hat{\eta}\}\mathbf{K}$, if η is re-parameterized in polar coordinates

as $\eta_2 = -\alpha \sin \beta$, $\eta_3 = \alpha \cos \beta$. Thus, $c(S)$ can be written as

$$c(S) \approx \exp(s_1) \int_0^{2\pi} \exp((s_2 + s_3) \cos \eta_1) \frac{1}{2\pi} d\eta_1 \\ \cdot \int_0^\pi \int_0^{2\pi} \exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\} \frac{1}{4\pi} \sin \alpha d\beta d\alpha,$$

where the first integral term is the normalizing constant of the von Mises distribution [80, Chapter 3], given by $I_0(s_2 + s_3)$. Note that $\eta_2 = \frac{\xi_2}{\sqrt{s_1+s_3}}$, and $\eta_3 = \frac{\xi_3}{\sqrt{s_1+s_2}}$, so $\sin \alpha = \alpha + O(s_1^{-3/2})$. Also, the integration for $\exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\}$ becomes very small for $\alpha > \pi$ as $s_1 + s_3 \gg 0$. Therefore, the second integral can be approximated by

$$\int_0^\infty \int_0^{2\pi} \exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\} \frac{1}{4\pi} \alpha d\beta d\alpha \\ = \int_{-\infty}^\infty \int_{-\infty}^\infty \exp \left\{ -\frac{1}{2} ((s_1 + s_3) \eta_2^2 + (s_1 + s_2) \eta_3^2) \right\} \frac{1}{4\pi} d\eta_2 d\eta_3,$$

which is the normalizing constant of the Gaussian distribution. And (2.62) follows after combining the two integral terms. The derivatives (2.63) and (2.64) can be directly calculated by differentiating (2.62). \square

Although the approximations (2.62) and (2.64) involve the modified Bessel function of the first kind, then are much easier to be evaluated than $c(S)$ itself. The comparision of $c(S)$ calculated from different methods is shown in Figure 2.4, for $S = \text{diag}(s, 0.1, 0.1)$ with s rangin from 0.1 to 100. It can be seen that when $s > 50$, the approximation (2.62) is very close to the one dimensional integral and saddle point approximation. However, because $s_2 + s_3 = 0.2$ is very small, the matrix Fisher distribution has very large dispersion in the first principal axis, and the approximation (2.59) developed for high concentration in all three degrees of freedom does not approximate $c(S)$ as well as (2.62), even when $s_1 = s$ is large.

Next, the comparison of $1 - d_1 = 1 - \frac{1}{c(S)} \frac{\partial c(S)}{\partial s_1}$ is presented in Figure 2.5. Similarly, the approximation (2.63) becomes very close to the one dimensional integral and saddle

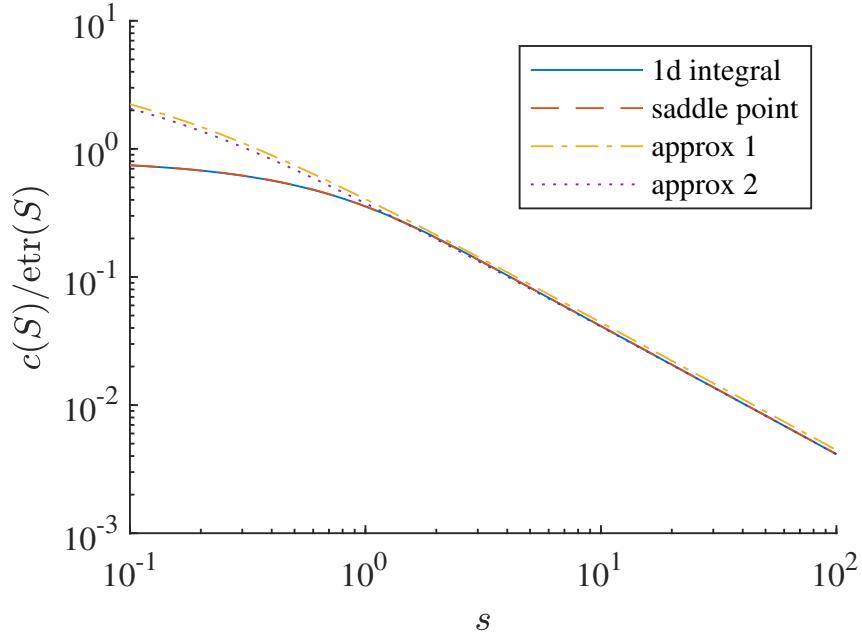


Figure 2.4: Comparison of the normalizing constant $c(S)$ calculated using the one dimensional integral formula (2.39), the saddle point approximation [58], the highly concentrated approximations (2.59) (approx 1) and (2.62) (approx 2). The parameter is $S = \text{diag}(s, 0.1, 0.1)$.

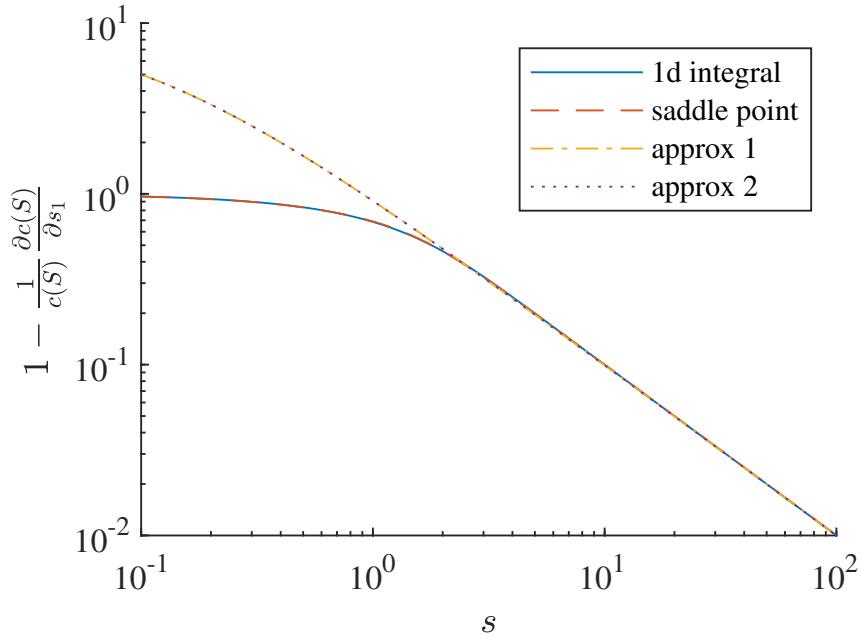


Figure 2.5: Comparison of the first order moment $1 - d_1$ calculated using the one dimensional integral formula (2.41), the saddle point approximation [58, 59], the highly concentrated approximations (2.60) (approx 1) and (2.63) (approx 2). The parameter is $S = \text{diag}(s, 0.1, 0.1)$.

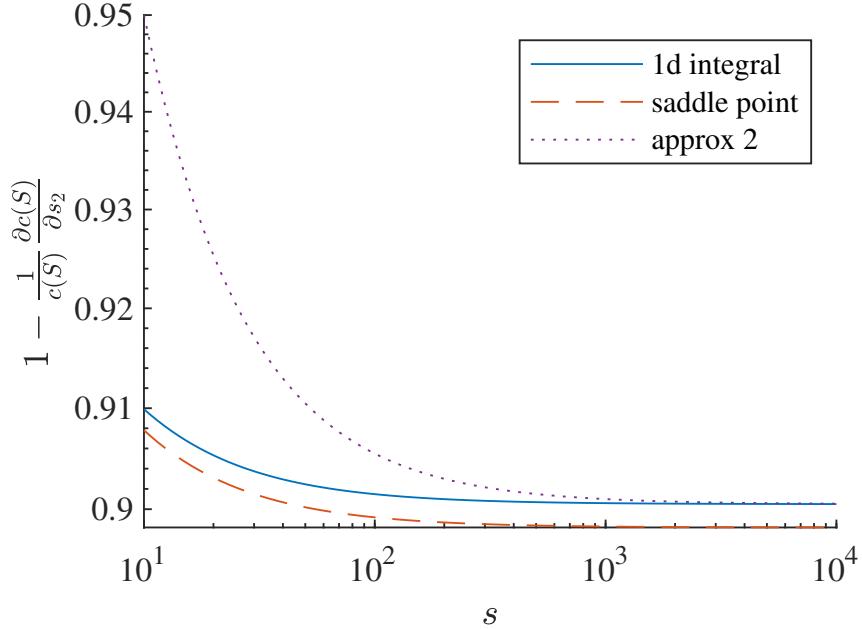


Figure 2.6: Comparison of the first order moment $1 - d_2$ calculated using the one dimensional integral formula (2.41), the saddle point approximation [58, 59], the highly concentrated approximations (2.64) (approx 2). The parameter is $S = \text{diag}(s, 0.1, 0.1)$.

point approximation. And since (2.60) is the same as (2.63), it also approximates the derivative well. The comparison of $1 - d_2 = 1 - \frac{1}{c(s)} \frac{\partial c(S)}{\partial s_2}$ is presented in Figure 2.6 for $S = \text{diag}(s, 0.1, 0.1)$ with s ranging from 10 to 10000. It can be seen that the difference between the approximation (2.64) and one dimensional integral becomes small when $s > 500$. The approximation (2.60) for high concentration in all three degrees of freedom is not shown because its error is too large. Also, it appears the that one dimensional integral is more accurate than the saddle point approximation, which is in general the case for S that are not extremely large.

In summary, this chapter reviews the matrix Fisher and Bingham distributions that are used to model large uncertainty of 3D attitude. A recursive algorithm is derived to calculate the central moments of the matrix Fisher distribution up to an arbitrary order. Also, an approximation is proposed when the matrix Fisher distribution is highly concentrated in two degrees of freedom, which is used to derive simplified expressions for the normalizing constant and its derivatives.

Chapter 3: Matrix Fisher-Gaussian Distribution

The matrix Fisher and Bingham distributions can model large uncertainty for 3D rotations, and have been used in various engineering applications. However, in practical estimation problems, the attitude is often coupled with other quantities in Euclidean space, such as sensor biases, velocity, position, and landmark locations, etc. Moreover, the correlation between attitude and these Euclidean quantities are usually very important in estimation problems. For example, in an attitude filter, the gyroscope bias cannot be directly measured, but it can be corrected because it is correlated with the attitude which can be estimated by direction sensors. Similarly, in IMU-GNSS integration, the attitude is usually not measured, but it can be corrected by its correlation with position which is directly measured. The matrix Fisher and Bingham distributions, defined on $\text{SO}(3)$ and \mathbb{S}^3 , cannot deal with the correlation between rotations and random variables in Euclidean space.

Despite the fact that there have been attempts to construct a new probability density function on $\text{SO}(3) \times \mathbb{R}^n$ [24, 85], as reviewed in Chapter 1.2.2, there are still motivations to build a new model that (i) deals with the correlation between attitude and Euclidean random variables not restricted to position, and (ii) has a closed form maximum likelihood estimation of its parameters for efficient inference. With such a new model, the matrix Fisher distribution can be generalized to be used in a lot of classic estimation problems where the attitude and other Euclidean quantities must be estimated concurrently.

This chapter introduces a new probability density function on $\text{SO}(3) \times \mathbb{R}^n$, referred to as the matrix Fisher–Gaussian (MFG) distribution. It is constructed using the conditioning approach which is a generic method of constructing new distributions in directional statistics [98]. In particular, a Gaussian distribution with a special structure is conditioned from $\mathbb{R}^9 \times \mathbb{R}^n$ onto $\text{SO}(3) \times \mathbb{R}^n$, such that only the *linear* correlation between $\text{SO}(3)$ and \mathbb{R}^n is described by $3n$ parameters. Due to this geometric construction, the correlation has

an immediate geometric interpretation. And a more desirable result is that the MLE for the parameters of MFG has a closed form approximation, which eliminates the need for a numerical optimization, hence has the potential to be implemented in real time for an estimation algorithm.

Depending on the structure of the Gaussian distribution in $\mathbb{R}^9 \times \mathbb{R}^n$ before conditioning, the MFG has two forms, referred to as the MFGI and MFGB respectively. For MFGI, the rotation in the inertial frame (multiplied on the left) is correlated with \mathbb{R}^n ; whereas for MFGB the rotation in the body-fixed frame (multiplied on the right) is correlated with \mathbb{R}^n . This highlights the non-commutative nature of $\text{SO}(3)$, and has not been seen in other probability models where the operation for the underlying space is commutative.

Based on the homomorphism between $\text{SO}(3)$ and \mathbb{S}^3 , and the equivalence between matrix Fisher and Bingham distributions, the Bingham–Gaussian (BG) distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ is also constructed. As expected, the MFG and BG are also equivalent.

From this chapter, the following notations will be used: e_i represents the i -th column of $I_{m \times m}$ for $i = 1, \dots, m$, where m will be clear from the context.

3.1 Definition and Construction

This section gives a formal definition of the matrix Fisher distribution on $\text{SO}(3) \times \mathbb{R}^n$, and its construction by conditioning a Gaussian distribution from $\mathbb{R}^9 \times \mathbb{R}^n$ is presented.

3.1.1 Definition

Definition 3.1. *The random elements $(R, x) \in \text{SO}(3) \times \mathbb{R}^n$ follow the matrix Fisher–Gaussian distribution with parameters $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n \times n}$, $U, V \in \text{SO}(3)$, $S = \text{diag}(s_1, s_2, s_3) \in \mathbb{R}^{3 \times 3}$ 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:*

$$p(R, x) = \frac{1}{c(S) \sqrt{(2\pi)^n \det(\Sigma_c)}} \exp \left\{ -\frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \right\} \text{etr} \left\{ FR^T \right\}, \quad (3.1)$$

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

$$\mu_c = \mu + Pv_R. \quad (3.2)$$

The expression for v_R has two forms, which defines the two variants of MFG, namely MFGI and MFGB:

$$(MFGI) \quad v_R^B = (QS - SQ^T)^\vee, \quad (3.3a)$$

$$(MFGB) \quad v_R^I = (SQ - Q^T S)^\vee, \quad (3.3b)$$

with $Q = U^T RV$. In addition, $0 \prec \Sigma_c \in \mathbb{R}^{n \times n}$ is defined as

$$\Sigma_c = \Sigma - P(\text{tr}(S)I_{3 \times 3} - S)P^T, \quad (3.4)$$

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 (3.1) 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 $x|R \sim \mathcal{N}(\mu_c(R), \Sigma_c)$.

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 Pv_R in (3.2), where v_R in (3.3) indicates how R deviates from the mean attitude UV^T . For example, when $R = UV^T$, or equivalently when $Q = I_{3 \times 3}$, we have $v_R = 0$ and $\mu_c = \mu$.

It should be noted that the MFG must be defined using the proper singular value decomposition $USV^T = F$, not the single parameter F . This is because the principal axes

specified by F is not unique for degenerative cases when the diagonal matrix S has repeated values, but MFG must specify the exact principal axis that is correlated with $x \in \mathbb{R}^n$. The non-uniqueness of pSVD also makes the Definition 3.1 not unique, in the sense that the same MFG can be parameterized by different parameters.

More specifically, there are two uniqueness issues associated with the definition of pSVD. 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. To study how the same MFG can be parameterized differently, it is first shown that MFG can be uniquely parameterized by the intermediate parameters F , μ_c and Σ_c .

Lemma 3.1. *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 $R \in \text{SO}(3)$, and $\Sigma_c = \tilde{\Sigma}_c$.*

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

Lemma 3.2. *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., $\frac{\text{etr}(FR^T)}{c(F)} = \frac{\text{etr}(\tilde{F}R^T)}{c(\tilde{F})}$ for all $R \in \text{SO}(3)$. Let $\Delta F = F - \tilde{F}$ and $\Delta c = \log(c(F)/c(\tilde{F}))$, the above equation is equivalent to

$$\text{tr}(\Delta F R^T) - \Delta c = 0. \quad (3.5)$$

Substitute $R = I_{3 \times 3}$, $\exp\{\pi\hat{e}_1\}$, $\exp\{\pi\hat{e}_2\}$ and $\exp\{\pi\hat{e}_3\}$ into (3.5) to obtain

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

$$\Delta F_{11} - \Delta F_{22} - \Delta F_{33} - \Delta c = 0,$$

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

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

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 (3.5) yields $\Delta F_{21} = 0$. Similarly, other entries of ΔF can be shown to be zeros. Therefore, $F = \tilde{F}$. \square

Next, the proof for Lemma 3.1 is as follows.

Proof for Proposition 3.1. The sufficiency directly follows from (3.1) 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 $R \in \text{SO}(3)$. Since $\text{SO}(3)$ is compact, and μ_c is continuous in R as seen from (3.2), $\|\mu_c\|$ has an upper bound. Therefore,

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

Because $p(x, R) = \tilde{p}(x, R)$ for all $x \in \mathbb{R}^n$, the above equation implies $\frac{\text{etr}(FR^T)}{c(F)\sqrt{(2\pi)^n \det(\Sigma_c)}} = \frac{\text{etr}(\tilde{F}R^T)}{c(\tilde{F})\sqrt{(2\pi)^n \det(\tilde{\Sigma}_c)}}$, and by following the same argument in Lemma 3.2, $F = \tilde{F}$.

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 $R \in \text{SO}(3)$. Substituting $x = \mu_c$ yields $\mu_c = \tilde{\mu}_c$ for all $R \in \text{SO}(3)$,

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 $\Sigma_c = \tilde{\Sigma}_c$. \square

Using Lemma 3.1, it can be shown that the same MFG can be parameterized by simultaneously reversing the signs of any two columns of U , V , and P . This resolves the non-uniqueness of pSVD caused by simultaneous sign changes.

Theorem 3.1. *Let $L = \exp\{\pi\hat{e}_i\}$, for $i = 1, 2$, or 3 , then $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$ and $\mathcal{MG}(\mu, \Sigma, PL, UL, S, VL)$ 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 Lemma 3.1. Then it is clear that

$$\begin{aligned}\tilde{F} &= ULSL^T V^T = USV^T = F, \\ \tilde{\Sigma}_c &= \Sigma - PL(\text{tr}(S)I_{3 \times 3} - S)L^T P^T = \Sigma_c,\end{aligned}$$

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

$$\begin{aligned}\tilde{\mu}_c &= \mu + PL(SL^T U^T RVL - L^T V^T R^T UDS)^\vee \\ &= \mu + P(LSL^T U^T RV - V^T R^T ULSL^T)^\vee = \mu_c.\end{aligned}$$

Therefore by Lemma 3.1, the two MFGs are equivalent. \square

Next, to deal with the non-uniqueness issue caused by repeated singular values, the space of eigen vectors must be analyzed in detail depending on the multiplicity of S . In Appendix C, a complete characterization of how the MFG can be parameterized differently if S has repeated values is given.

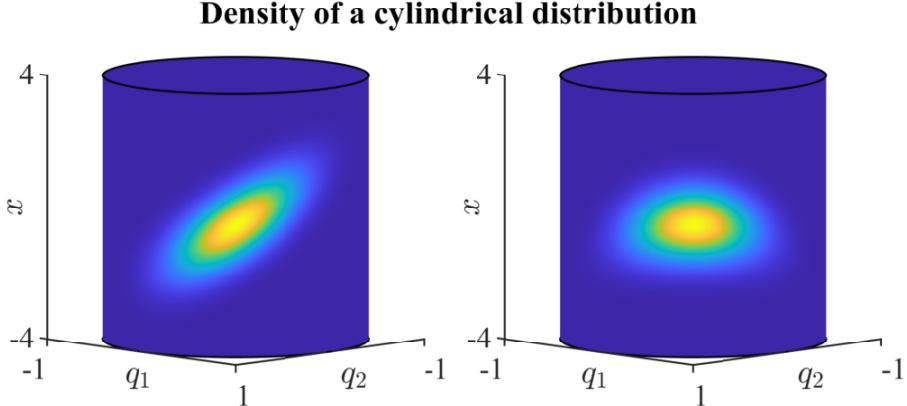


Figure 3.1: The illustration of the cylindrical distribution with the mean angle $\pi/4$ [81]. In the left figure, the correlation between the linear variable $x \in \mathbb{R}$ and the angular variable $q = (q_1, q_2) \in \mathbb{S}^1$ is only nonzero along the tangent direction of \mathbb{S}^1 at $\pi/4$, 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.

3.1.2 Construction

The construction of MFG is motivated by [81], 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 [81] 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. 3.1.

Similarly, if the correlation between $R \in \text{SO}(3)$ and $x \in \mathbb{R}^n$ is defined as in [81], it needs $9n$ parameters. Instead, the ambient space of $\text{SO}(3)$, namely $\mathbb{R}^{3 \times 3}$, 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 $\text{SO}(3) \times \mathbb{R}^n$.

The difference between MFGI and MFGB comes from how $\text{SO}(3)$ is embedded into \mathbb{R}^9 . For MFGI, the embedding is $R \mapsto \text{vec}(R^T)$, i.e., the rows of R is concatenated into a vector; whereas for MFGB, the embedding is $R \mapsto \text{vec}(R)$, i.e., the columns of R is concatenated into a vector. The construction is summarized into the following theorem.

Theorem 3.2. Consider the parameters $(\mu, \Sigma, V, S, U, P)$ satisfying the conditions in Definition 3.1. Let $M = UV^T$, $K_I = VSV^T$, $K_B = USU^T$. Also, let $\mu_I = \text{vec}(M^T)$, $\Sigma_I^{-1} = I_{3 \times 3} \otimes K_I$; and $\mu_B = \text{vec}(M)$, $\Sigma_B^{-1} = I_{3 \times 3} \otimes K_B$, where \otimes denotes Kronecker product.

Also let $t_{Ii} = \text{vec}[(\widehat{MVe}_i)^T]$ for $i \in \{1, 2, 3\}$ be the basis for the tangent space of $\text{SO}(3)$ at M embedded in \mathbb{R}^9 , and $t_{Bi} = \text{vec}[\widehat{MVe}_i]$. Define $T_I = [t_{I1}, \dots, t_{I9}] \in \mathbb{R}^{9 \times 9}$ such that $T_I \in \text{SO}(9)$, and $P_I = [P, 0_{n \times 6}]T_I$. Similarly, define $T_B = [t_{B1}, \dots, t_{B9}] \in \mathbb{R}^{9 \times 9}$ such that $T_B \in \text{SO}(9)$, and $P_B = [P, 0_{n \times 6}]T_B$.

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

$$\begin{bmatrix} \text{vec}(R^T) \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_I \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_I & P_I^T \\ P_I & \Sigma \end{bmatrix}\right), \quad (3.6)$$

then the density for (R, x) is (3.1) with v_R defined in (3.3a) when restricted to $R \in \text{SO}(3)$.

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

$$\begin{bmatrix} \text{vec}(R) \\ x \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_B \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma_B & P_B^T \\ P_B & \Sigma \end{bmatrix}\right), \quad (3.7)$$

then the density for (R, x) is (3.1) with v_R defined in (3.3b) when restricted to $R \in \text{SO}(3)$.

Proof. First consider (3.6). The joint density of $(\text{vec}(R^T), x)$ can be written in the form of a conditional-marginal density as

$$\begin{aligned} p(\text{vec}(R^T), x) &\propto \exp\left\{-\frac{1}{2}(x - \mu'_c)^T (\Sigma'_c)^{-1} (x - \mu'_c)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(\text{vec}(R^T) - \mu_I)^T \Sigma_I^{-1} (\text{vec}(R^T) - \mu_I)\right\}, \end{aligned} \quad (3.8)$$

where $\mu'_c = \mu + P_I \Sigma_I^{-1} (\text{vec}(R^T) - \mu_I) \in \mathbb{R}^n$, and $\Sigma'_c = \Sigma - P_I \Sigma_I^{-1} P_I^T \in \mathbb{R}^{n \times n}$. The second exponential term has been proved in Theorem 2.1 to be proportional to $\text{etr}(FR^T)$, where $F = USV^T$. So it remains to check for all $R \in \text{SO}(3)$, $\mu'_c = \mu_c$ in (3.2), and $\Sigma'_c = \Sigma_c$ in (3.4).

For Σ'_c , it can be shown that

$$P_I \Sigma_I^{-1} P_I^T = P[t_{I1}, t_{I2}, t_{I3}]^T \Sigma_I^{-1} [t_{I1}, t_{I2}, t_{I3}] P^T \triangleq P \tilde{\Sigma}_I^{-1} P^T,$$

for some $\tilde{\Sigma}_I^{-1} \in \mathbb{R}^{3 \times 3}$. Let $t_{Ii} \in \mathbb{R}^9$ be equally split into three vectors $\{t_{Ii1}, t_{Ii2}, t_{Ii3}\}$, then the (i, j) -th entry of $\tilde{\Sigma}_I^{-1}$ can be written as

$$\begin{aligned} (\tilde{\Sigma}_I^{-1})_{ij} &= \sum_{k=1}^3 t_{Iik}^T K_I t_{Ijk} = \text{tr}([t_{Ii1}, t_{Ii2}, t_{Ii3}]^T K_I [t_{Ij1}, t_{Ij2}, t_{Ij3}]) \\ &= \text{tr}\left(M \widehat{V e_i} K_I \widehat{V e_j}^T M^T\right) = \text{tr}(S \hat{e}_j^T \hat{e}_i), \end{aligned}$$

which implies $\tilde{\Sigma}_I^{-1} = \text{tr}(S) I_{3 \times 3} - S$, so $\Sigma'_c = \Sigma_c$ in (3.4). Next, the second part of μ'_c is

$$P_I \Sigma_I^{-1} (\text{vec}(R^T) - \mu_I) = P[t_{I1}, t_{I2}, t_{I3}]^T \Sigma_I^{-1} \text{vec}(R^T - M^T),$$

where the i -th element of $[t_{I1}, t_{I2}, t_{I3}]^T \Sigma_I^{-1} \text{vec}(R^T - M^T)$ can be calculated as

$$t_{Ii}^T \Sigma_I^{-1} \text{vec}(R^T - M^T) = \text{tr} \left(\widehat{M V e_i K} (R^T - M^T) \right) = \text{tr} (S Q^T \hat{e}_i - S \hat{e}_i) = e_i^T (Q S - S Q^T)^\vee,$$

and $Q = U^T R V$. This shows $\mu'_c = \mu_c$ in (3.2) for all $R \in \text{SO}(3)$, and finishes the proof of the case (3.6) for MFGI. The proof of (3.7) for MFGB follows exactly the same steps, and is omitted for brevity. \square

Theorem 3.2 is a generalization of Theorem 2.1, where the correlation between $\text{SO}(3)$ and \mathbb{R}^n is inherited from the correlation between \mathbb{R}^9 and \mathbb{R}^n before conditioning. Specifically, the covariance for MFGI between $\text{vec}(R^T)$ and x is $P_I = [P, 0_{n \times 6}] T_I$, i.e., it is non-zero only along $\{t_{I1}, t_{I2}, t_{I3}\}$, which is a basis for the tangent space of $\text{SO}(3)$ at the mean attitude $M = U V^T$. It is similar for MFGB where the correlation between $\text{vec}(R)$ and x is only non-zero along $\{t_{B1}, t_{B2}, t_{B3}\}$. The basis of the tangent space 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. 3.2 for a simple example exhibited for MFGB.

3.1.3 Distinction Between MFGI and MFGB

At first glance, it appears that the MFGI and MFGB are very similar, as the difference is caused by the single term v_R in (3.3). In this subsection, the implications of the two definitions of v_R and its role in formulating the angular–linear correlation are discussed.

More specifically, the difference is studied by examining $R|x$, the attitude distribution of R

Conditional distribution of attitude in MFG

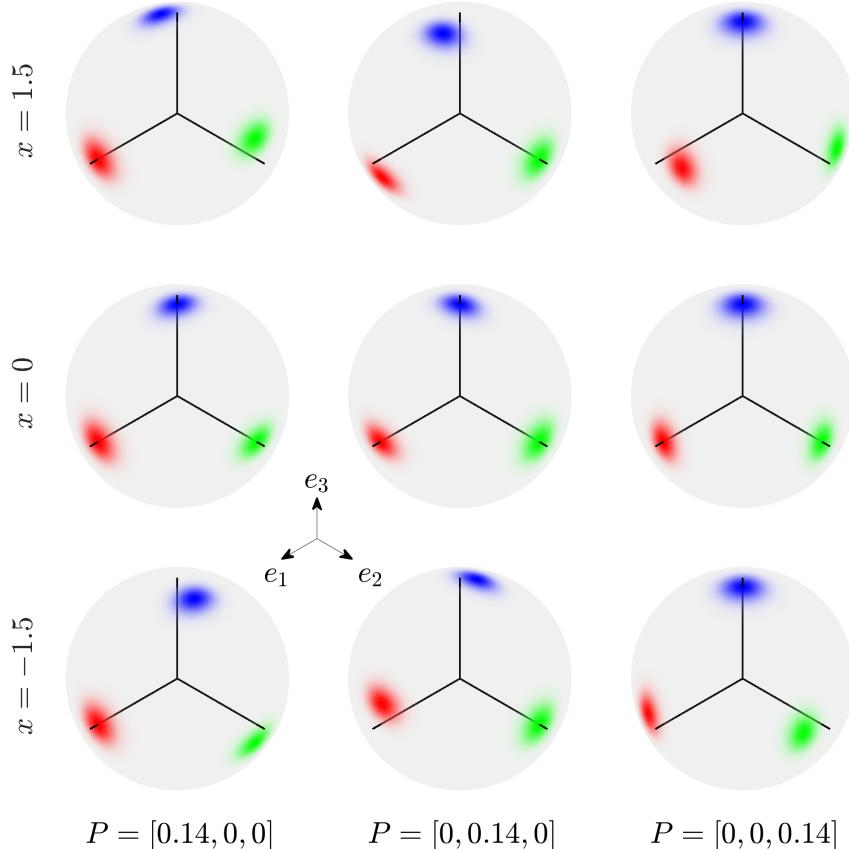


Figure 3.2: Visualization of the attitude-linear correlation (MFGB) for $(R, x) \in \mathrm{SO}(3) \times \mathbb{R}^1$: the density for R conditioned on x is illustrated on the unit-sphere for three correlation matrices P . More specifically, the marginal distribution for each column of R is shown on the unit sphere as red, green and blue shades. The parameters are $n = 1$, $\mu = 0$, $\Sigma = 1$, $U = V = I_{3 \times 3}$, $S = 10I_{3 \times 3}$. 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.

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, the density function can be written as

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

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

For MFGB, $Pv_R = \text{tr}(\widehat{P^T Q^T S})$ from (2.15) and (3.3b). Thus,

$$p(R|x) \propto \text{etr}(SQ^T) \text{etr} \left(\frac{x - \mu}{\sigma_c^2} \widehat{P^T Q^T S} \right) = \text{etr}((I_{3 \times 3} + \hat{v}(x))Q^T S),$$

where $v(x) \triangleq (x - \mu)/\sigma_c^2 P^T \in \mathbb{R}^{3 \times 1}$. Note that $\exp(\hat{v}(x)) = I_{3 \times 3} + \hat{v}(x) + O(\|v(x)\|^2)$.

Therefore, when x is also close to μ , it can be rewritten as

$$p(R|x) \approx \text{etr}(USV^T \exp(\widehat{V}v(x))R^T).$$

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

$$(R|x)|_{\text{MFGB}} \approx \mathcal{M}(USV^T \exp(\widehat{V}v(x))) \quad (3.9)$$

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

$$(R|x)|_{\text{MFGI}} \approx \mathcal{M}(\exp(\widehat{U}v(x))USV^T). \quad (3.10)$$

Now we compare (3.9) and (3.10). 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 $U \exp(\hat{v}(x))V^T$. The

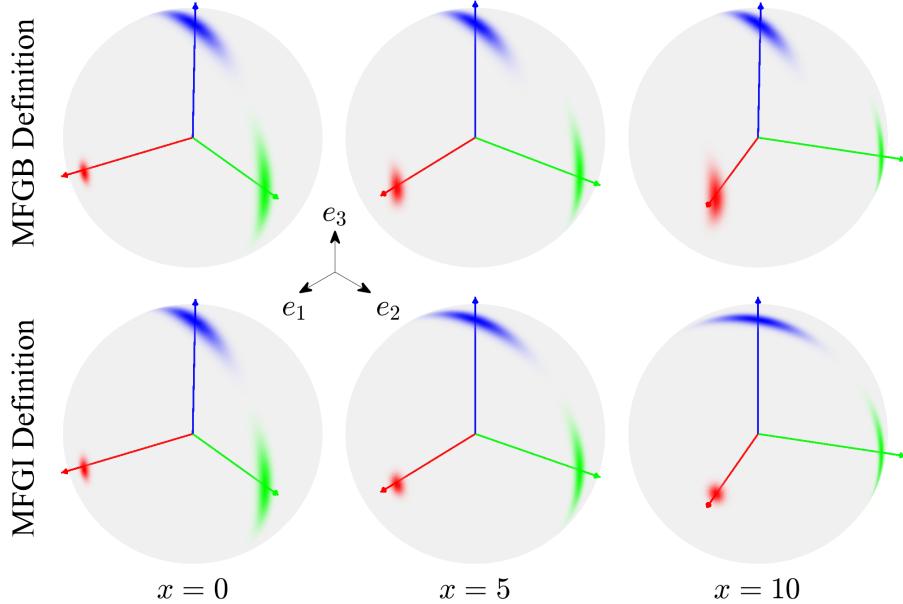


Figure 3.3: Difference between MFGB (top row) and MFGI (bottom row). The conditional density of $R|x$ is illustrated on the unit sphere for varying x , with the red, green, and blue arrows representing the first, second, and third columns of the mean attitude, respectively. Parameters are chosen as $n = 1$, $\mu = 0$, $\Sigma = 1$, $P = [0, 0, 0.069]$, $U = V = I_{3 \times 3}$, $S = \text{diag}(100, 3, 3)$. For both of MFGB and MFGI, 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.

difference between (3.9) and (3.10) is caused by how the principal axes of rotations are changed as x is varied. For (3.9), as U is fixed, the principal axes remain unchanged with respect to x when perceived from the inertial frame. However, they rotate about $Vv(x)$ when observed in the body-fixed frame, as V is changed into $\exp(-\widehat{V}v(x))V$. Also, (3.9) indicates $R|x$ has the same density as $R\exp(\widehat{V}v(x))$, i.e., the rotation is applied on the right. Therefore, the correlation term $v(x)$ causes the attitude distribution to be rotated about the axis $Vv(x)$ 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 $R_iVv(x)$ in the inertial frame.

On the other hand, in (3.10), the principal axes represented in the inertial frame are

rotated, as U is changed into $\exp(\widehat{U}\nu(x))U$. But, when observed from the body-fixed frame, they remain unchanged. Also, (3.10) implies $R|x$ has the same density as $\exp(\widehat{U}\nu(x))R$, i.e., the rotation is applied on the left. This means the correlation term $\nu(x)$ causes the attitude distribution to be rotated about the axis $U\nu(x)$ 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 (multiplied on the right); for MFGI, it causes the distribution to be rotated about the inertial frame (multiplied on the left). These are illustrated in Fig. 3.3.

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.

3.2 Bingham–Gaussian Distribution

Before diving into more properties of the MFG, this section first introduces the Bingham–Gaussian (BG) defined on $\mathbb{S}^3 \times \mathbb{R}^n$ which is equivalent to MFG. The BG is naturally constructed by transforming MFG use the Lie group homomorphism (2.33) between $\text{SO}(3)$ and \mathbb{S}^3 . Due to the prevalence of Bingham distribution in recent literature, the BG can serve as a more accessible model to those who are more familiar with Bingham distribution.

Definition 3.2. *Random elements $(q, x) \in \mathbb{S}^3 \times \mathbb{R}^n$ follow Bingham–Gaussian distribution with parameters $M \in \text{SO}(4)$, $Z = \text{diag}(0, z_1, z_2, z_3)$ where $0 \geq z_1 \geq z_3 \geq z_2$, $\mu \in \mathbb{R}^n$, $\Sigma = \Sigma^T \in \mathbb{R}^{n \times n}$, and $P \in \mathbb{R}^{n \times 3}$, if they have the following density function:*

$$f(q, x) = \frac{1}{c(Z) \sqrt{(2\pi)^n \det(\Sigma_c)}} \exp \left\{ -\frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \right\} \exp \left\{ q^T M Z M^T q \right\}, \quad (3.11)$$

where $c(Z)$ is the normalizing constant for the corresponding Bingham distribution, and

$\Sigma_c \in \mathbb{R}^{n \times n}$ satisfying $\Sigma_c = \Sigma_c^T \succ 0$ is given by

$$\Sigma_c = \Sigma + \frac{1}{2} P \text{diag}(z_1, z_2, z_3) P^T. \quad (3.12)$$

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

$$\mu_c = \mu + P \begin{bmatrix} (z_2 - z_3)p_2p_3 - z_1p_0p_1 \\ (z_3 - z_1)p_1p_3 - z_2p_0p_2 \\ (z_1 - z_2)p_1p_2 - z_3p_0p_3 \end{bmatrix} \triangleq \mu + Pv_q^I, \quad (3.13)$$

or for BGB, μ_c is

$$\mu_c = \mu + P \begin{bmatrix} (z_3 - z_2)p_2p_3 - z_1p_0p_1 \\ (z_1 - z_3)p_1p_3 - z_2p_0p_2 \\ (z_2 - z_1)p_1p_2 - z_3p_0p_3 \end{bmatrix} \triangleq \mu + Pv_q^B, \quad (3.14)$$

with $p = [p_0, p_1, p_2, p_3]^T \triangleq M^T q$. This distribution is denoted by $\mathcal{BG}(\mu, \Sigma, P, M, Z)$.

BG is antipodally symmetric in \mathbb{S}^3 , i.e., for any $(q, x) \in \mathbb{S}^3 \times \mathbb{R}^n$, $f(q, x) = f(-q, x)$. Similar with MFG, the density (3.11) has three terms: the first one is for normalization; the second term is for x and it has the form as $\mathcal{N}(\mu_c, \Sigma_c)$; the last term is for q and it is identical to the Bingham distribution. The marginal distribution of q is a Bingham distribution with parameter MZM^T , and the distribution of x conditioned by q is Gaussian with $x|q \sim \mathcal{N}(\mu_c(q), \Sigma_c)$. Also, the correlation between q and x is represented by the $(3n)$ -element matrix P . The vector $v_q \in \mathbb{R}^3$ indicates the deviation of q from the mode m_0 , i.e., if $q = m_0$, then $p = [1, 0, 0, 0]^T$ and hence $v_q = 0$. The two expressions for v_q again have the same interpretations as MFG: If (3.13) is used, x is correlated with rotations along the principal axes resolved in the inertial frame (BGI); if (3.14) is used, the rotations are resolved in the body-fixed frame (BGB) of each $q \in \mathbb{S}^3$.

The Bingham-Gaussian distribution is constructed by transforming the matrix Fisher-Gaussian distribution using the homomorphism $\varphi : \mathbb{S}^3 \rightarrow \mathrm{SO}(3)$ given in (2.33).

Theorem 3.3. *Let (μ, Σ, P, M, Z) be the parameters of a Bingham–Gaussian distribution. Define $S = \mathrm{diag}(s_1, s_2, s_3)$ as $s_1 = \frac{1}{4}(z_1 - z_2 - z_3)$, $s_2 = \frac{1}{4}(z_2 - z_1 - z_3)$, and $s_3 = \frac{1}{4}(z_3 - z_1 - z_2)$. Let $M = [u]_L[v^{-1}]_R$ be its isoclinic decomposition. Define $U = \varphi(u)$ and $V = \varphi(v)$. Then $(q, x) \sim \mathcal{BG}(\mu, \Sigma, P, M, Z)$ if and only if $(\varphi(q), x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$.*

Proof. In Theorem 2.2, it is proved that $\exp(q^T M Z M^T q) \propto \mathrm{etr}(U S V^T \varphi(q)^T)$. Also, it is straightforward to show $\Sigma_{c,\mathcal{MG}} = \Sigma_{c,\mathcal{BG}}$. Therefore, it remains to check for all $q \in \mathbb{S}^3$, $\mu_{c,\mathcal{BG}}(q) = \mu_{c,\mathcal{MG}}(\varphi(q))$. Note that $v_q^I = (\varphi(p)S - S\varphi(p)^T)^\vee = v_{\varphi(q)}^I$, and $v_q^B = (S\varphi(p) - \varphi(p)^T S)^\vee = v_{\varphi(q)}^B$. So we only need to show $\varphi(M^T q) = U^T \varphi(q)V$, which is straightforward

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

This concludes the proof. □

Because BG and MFG are equivalent, all of the properties of MFG given in Chapter 3.3 also apply to BG, although some calculations are needed using the homomorphism φ . However, the geometric construction of MFG by conditioning a $(9+n)$ -variate Gaussian distribution from the ambient space $\mathbb{R}^9 \times \mathbb{R}^n$ to $\mathrm{SO}(3) \times \mathbb{R}^n$ given in Theorem 3.2 does not apply to BG. This is because \mathbb{S}^3 with antipodal points identified (namely the real projective space \mathbb{RP}^3) does not embed into \mathbb{R}^4 . So the BG in $\mathbb{S}^3 \times \mathbb{R}^n$ does not generalize into unit sphere \mathbb{S}^r with arbitrary dimension. Constructing a new density function on $\mathbb{S}^r \times \mathbb{R}^n$ for arbitrary r may be pursued in future works.

3.3 Properties

In this section, some further properties of MFG are developed. These properties are necessary for the design of estimation algorithms using MFG in Chapter 4 and Chapter 5. First, similar

to the matrix Fisher distribution, it is shown that the MFG can also be approximated by a Gaussian distribution when the attitude part is highly concentrated.

Theorem 3.4. Suppose $(R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Let $R = U \exp(\hat{\eta}) V^T$. If $s_2 + s_3 \gg 0$, then (x, η) approximately follows a $(3+n)$ -dimensional Gaussian distribution

$$\begin{bmatrix} x \\ \eta \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma & P \\ P^T & (\text{tr}(S)I_{3 \times 3} - S)^{-1} \end{bmatrix} \right). \quad (3.15)$$

Proof. Let $\Sigma' = (\text{tr}(S)I_{3 \times 3} - S)^{-1}$. For the matrix Fisher density part in (3.1), it is proved in Theorem 2.6 that

$$\text{etr}(FR^T) \approx \text{etr}(S) \exp \left\{ -\frac{1}{2} \eta^T (\Sigma')^{-1} \eta \right\}.$$

Also, the v_R term in the conditional mean (3.3) can be calculated as

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

and

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

which yields $\mu_c \approx \mu + P(\Sigma')^{-1} \eta$. Furthermore $\Sigma_c = \Sigma - P(\Sigma')^{-1} P^T$ as in (3.4). Therefore, the density function (3.1) is approximated by a $(3+n)$ -dimensional Gaussian density written in the conditional-marginal form, which is identical to (3.15). \square

Next, selected moments of MFG are presented, which are used in the approximate MLE to recover the parameters from random samples.

Theorem 3.5. Suppose $(R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Then,

$$\mathbb{E}[R] = UDV^T, \quad (3.16)$$

where $D = \text{diag}(d_1, d_2, d_3)$ is given in (2.20). Also,

$$\mathbb{E}[x] = \mu, \quad (3.17)$$

$$\mathbb{E}[v_R] = 0, \quad (3.18)$$

$$\mathbb{E}[xx^T] = \Sigma_c + \mu\mu^T + P\mathbb{E}[v_R v_R^T]P^T, \quad (3.19)$$

$$\mathbb{E}[xv_R^T] = P\mathbb{E}[v_R v_R^T], \quad (3.20)$$

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

$$\mathbb{E}[v_R v_R^T]_{ii} = s_j d_j + s_k d_k, \quad (3.21)$$

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

Proof. Equation (3.16) follows immediately from the fact that the marginal distribution of R is a matrix Fisher distribution with parameter F . Next, for (3.18), it is shown:

$$\mathbb{E}[v_R^I] = \mathbb{E}[QS - SQ^T]^\vee = (DS - SD^T)^\vee = 0,$$

$$\mathbb{E}[v_R^B] = \mathbb{E}[SQ - Q^T S]^\vee = (SD - D^T S)^\vee = 0.$$

For (3.21), it can be shown by direct calculation that for both v_R^I and v_R^B , the off-diagonal entries of $v_R v_R^T$ only involve terms of the form $Q_{ij}Q_{ik}$ and $Q_{ij}Q_{ki}$, whose moments are zero by Corollary 2.1. Using Corollary 2.2, it can be shown that the diagonal entries can be

calculated as

$$\mathbb{E} [\mathbf{v}_R \mathbf{v}_R^T]_{ii} = (s_j^2 + s_k^2) \mathbb{E} [Q_{jk}^2] - 2s_j s_k \mathbb{E} [Q_{jk} Q_{kj}],$$

and (3.21) follows by using (B.2) to (B.5).

Next, for (3.19), xx^T is first integrated against the conditional Gaussian density, then against the marginal matrix Fisher density. The detailed procedure is

$$\begin{aligned} \mathbb{E} [xx^T] &= \int_{\text{SO}(3)} \int_{\mathbb{R}^n} xx^T p(R, x) dx dR \\ &= \frac{1}{c(F)} \int_{\text{SO}(3)} [\Sigma_c + (\mu + Pv_R)(\mu + Pv_R)^T] \text{etr}(FR^T) dR \\ &= \Sigma_c + \mu\mu^T + P\mathbb{E} [\mathbf{v}_R \mathbf{v}_R^T] P^T. \end{aligned} \quad (3.22)$$

The remaining (3.17) and (3.20) can be derived similarly. \square

Finally, the maximum likelihood estimation of the parameters for MFG is studied. The MLE is used to recover parameters from a set of sufficient statistics of the random samples. Given a set of random samples $(R_i, x_i)_{i=1}^{N_s}$ following an MFG, the log-likelihood function of the parameters, after omitting some constants, is given by

$$l = -\log(c(S)) + \text{tr} \left(F \bar{\mathbb{E}} [R]^T \right) - \frac{1}{2} \log(\det \Sigma_c) - \frac{1}{2} \bar{\mathbb{E}} [(x - \mu - Pv_R)^T \Sigma_c^{-1} (x - \mu - Pv_R)], \quad (3.23)$$

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

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 $(\mu, \Sigma, V, S, U, P)$. 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 [133]. 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 of R corresponds to the first two terms on the right hand side of (3.23), and that for the conditional distribution of $x|R$ corresponds to the last two terms.

The MLE for the marginal distribution of R can be solved using the MLE for the matrix Fisher distribution introduced in Chapter 2.2. Namely, the estimation of U, V is given by the proper singular value decomposition

$$UDV^T = \bar{\mathbb{E}}[R],$$

and the estimation of S is solved using (2.20) from D . These are used in the conditional log-likelihood for $x|R$, which is addressed in the next theorem.

Theorem 3.6. *Let $U, V \in \text{SO}(3)$ and $S \in 3 \times 3$ be the solution of the marginal MLE for R . Define $Q_i = U^T R_i V$, and $v_{R_i} = (SQ_i - Q_i^T S)^\vee$ for $i = 1, \dots, N_s$. Also define the following sample covariance matrices:*

$$\bar{\text{cov}}(x, x) = \bar{\mathbb{E}}[xx^T] - \bar{\mathbb{E}}[x]\bar{\mathbb{E}}[x]^T, \quad (3.24)$$

$$\bar{\text{cov}}(x, v_R) = \bar{\mathbb{E}}[xv_R^T] - \bar{\mathbb{E}}[x]\bar{\mathbb{E}}[v_R]^T, \quad (3.25)$$

$$\bar{\text{cov}}(v_R, v_R) = \bar{\mathbb{E}}[v_R v_R^T] - \bar{\mathbb{E}}[v_R]\bar{\mathbb{E}}[v_R]^T. \quad (3.26)$$

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

$$P = \bar{\text{cov}}(x, v_R) \bar{\text{cov}}(v_R, v_R)^{-1}, \quad (3.27)$$

$$\mu = \bar{\mathbb{E}}[x] - P\bar{\mathbb{E}}[v_R], \quad (3.28)$$

$$\Sigma = \bar{\text{cov}}(x, x) - P\bar{\text{cov}}(x, v_R)^T + P(\text{tr}(S)I_{3 \times 3} - S)P^T. \quad (3.29)$$

Proof. Taking the derivatives of (3.23) with respect to μ and P yields

$$\begin{aligned}\frac{\partial l}{\partial \mu} &= \Sigma_c^{-1} \bar{E}[x - \mu - Pv_R], \\ \frac{\partial l}{\partial P} &= \Sigma_c^{-1} \bar{E}[(x - \mu)v_R^T - Pv_R v_R^T].\end{aligned}$$

By setting the derivatives zero and noticing that Σ_c is positive definite, the MLE of μ and P can be obtained as in (3.28) and (3.27). Next, taking the derivative of (3.23) with respect to Σ_c^{-1} yields

$$\frac{\partial l}{\partial \Sigma_c^{-1}} = \frac{1}{2} \Sigma_c - \frac{1}{2} \bar{E}[(x - \mu - Pv_R)(x - \mu - Pv_R)^T].$$

Setting the derivative to zero and substituting (3.27) and (3.28), (3.29) is obtained. \square

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 R indicated by the samples is not necessarily constrained in the tangent space at UV^T calculated from $\bar{E}[R]$, as required by MFG in Theorem 3.2.

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 $x|R$. To simplify analysis, it is assumed that the dimension of the linear part is one, i.e., $n = 1$. Define a metric $\lambda_{s_i} \in \mathbb{R}$ as

$$\lambda_{s_i} = \frac{g_{s_i s_i}(R)}{g_{s_i s_i}(x|R)}, \quad (3.30)$$

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 $p(x|R)$ [57, 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 $g_{s_i s_i}(R, x) = g_{s_i s_i}(R) + g_{s_i s_i}(x|R)$. Thus, higher λ_{s_i} means less information is discarded in the marginal

MLE for parameter s_i .

Theorem 3.7. Suppose $(R, x) \in \text{SO}(3) \times \mathbb{R}^3$ follows $\mathcal{MG}(\mu, \sigma^2, P, U, sI_{3 \times 3}, V)$, where $P = \frac{\rho\sigma}{\sqrt{2s}}[1, 1, 1]$ for $\rho \in \mathbb{R}$. Then,

$$\lambda_{s_i} = \frac{1 - 3\rho^2}{\rho^2} \frac{2s(c(S)\partial_{11}c(S) - (\partial_1 c(S))^2)}{c(s)(c(S) - \partial_{11}c(S))}, \quad (3.31)$$

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

Proof. By Chapter 2.6 in [57], the Fisher information of s_i is

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

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 $P = \frac{\rho\sigma}{\sqrt{2s}}[1, 1, 1]$ into the conditional information to obtain

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

And (3.7) follows from the above two equations. \square

In Theorem 3.7, ρ can be interpreted as the *correlation coefficient* between R and x . It is clearly seen from (3.31) 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, the effect of concentration level of the attitude is examined. Let $r(s)$ be the second fraction term on the right hand side of (3.31), whose value is illustrated in Fig. 3.4 for varying s . This indicates when the marginal attitude distribution is close to uniform, i.e., $s \rightarrow 0$, 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

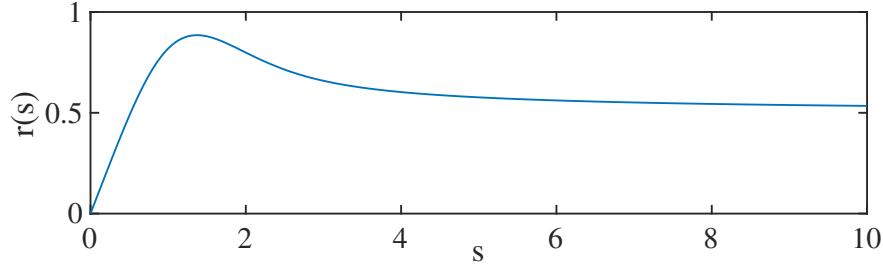


Figure 3.4: The graph of $r(s) = \frac{2s(c(s)\partial_{11}c(s) - (\partial_1 c(s))^2)}{c(s)(c(s) - \partial_{11}c(s))}$ against s .

hand side of (3.31).

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

Chapter 4: Bayesian Attitude Estimation with Matrix Fisher-Gaussian Distribution

In the previous two chapters, the matrix Fisher distribution defined on $\text{SO}(3)$ is reviewed, and a new probability model called the matrix Fisher–Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ is developed to model the correlation between $\text{SO}(3)$ and \mathbb{R}^n . The development of MFG opens a new gate to use the matrix Fisher distribution or Bingham distribution in a lot of classic estimation problems where the attitude and some quantities in Euclidean space must be estimated simultaneously, such as position, sensor biases, landmark locations, etc. Compared to MEKF and its variants [90, 112], the stochastic estimator designed using MFG has the potential to handle large attitude uncertainty, which may occur when the initial conditions are unknown, when there is a degree of freedom in attitude that is not properly observed, or when some sensors fail (for example a GPS receiver) for a long period of time. The fact that MFG models large attitude uncertainty with high fidelity could make the estimator achieve higher accuracy, or faster convergence speed from a wrong state.

In this chapter, this idea is investigated with the classic attitude estimation problem. First in Chapter 4.1, the observability of 3D attitude is studied when only angular velocity and single direction measurements are available, where the matrix Fisher distribution is utilized to analyze how the attitude uncertainty is propagated through the gyroscope kinematics, and updated by the single direction measurement. Then in Chapter 4.2, a new recursive Bayesian filter based on MFG is designed to estimate the attitude and gyroscope bias simultaneously. In particular, the linear part of MFG is used to model the gyroscope bias in \mathbb{R}^3 , and its correlation with the attitude is directly handled in the MFG.

It is demonstrated in this chapter that the estimators designed using MFG have advantages over MEKF when dealing with unknown initial conditions, and when not all degrees of freedom for attitude are adequately observed. In particular, the MFG based filter has faster

convergence speed from large initial errors, highlighting its capability to model large attitude uncertainty with high fidelity.

4.1 Attitude Observability with Single Direction Measurement

In this section, the matrix Fisher distribution is used to study the observability of attitude when the angular velocity, and a single reference direction are measured. Because the attitude has three degrees of freedom, and a single direction measurement fixes two of them, at least two reference directions are needed to fully determine the 3D attitude [82, 110]. This is widely believed to be the case even when the angular velocity is also measured [78], except in a few special cases, for example, when the single reference direction is time-varying in the inertial frame [9, 37, 68], and when the gyroscope can capture the rotation of the earth [103]. However, there is no available explanation on why introducing the information of angular velocity measurement does not improve the observability of attitude with single direction measurements.

In this section, this problem is studied by examining how the attitude uncertainty is propagated through the attitude kinematics, and how it is shaped by injecting the information from single direction measurements. In particular, it is shown that the attitude uncertainty is propagated in different ways, depending on whether the angular velocity is measured in the body-fixed frame or in the inertial frame. For the former, the direction of one-dimensional ambiguity caused by a single direction measurement remains fixed in the inertial frame, but for the latter, it is fixed in the body-fixed frame. This explains the fundamental reason why the attitude is unobservable with a single inertial reference direction and the angular velocity measured by a gyroscope: the direction of ambiguity caused by the inertial reference direction measurement remains unchanged by the angular velocity resolved in the body-fixed frame. This leads to two strategies to achieve full attitude observability with single direction measurements, namely utilizing the angular velocity resolved in the inertial frame such that the direction of ambiguity is rotated over propagation, or measuring a reference direction

Table 4.1: Attitude observability with single direction measurements

		ang. vel.	body-fixed frame	inertial frame
		ref. vec.		
body-fixed frame	body-fixed frame		observable	unobservable
	inertial frame		unobservable	observable

fixed to the body such that the next measurement can resolve the ambiguity. These results are summarized in Table 4.1.

This discovery is more rigorously studied by introducing two formulations of stochastic attitude observability. For a given probability density function on $\text{SO}(3)$, the attitude that minimizes the mean square error may not be unique [87, 96]. This gives a characterization of the attitude observability, since if multiple attitudes can minimize the mean square error, it indicates deficiency of information to distinguish them. Alternatively, the inverse of Fisher information gives a lower bound for the variance of all unbiased estimators, known as the Cramér–Rao bound, and the observability can be characterized by the positive-definiteness of the Fisher information matrix [88]. The method in [111] which generalizes the Fisher information to Riemannian manifold is used to calculate the Fisher information of the mean attitude of a matrix Fisher distribution. It is further shown that the two attitude observability criteria are consistent with each other, and they also agree with a classic deterministic observability analysis when only a single direction measurement is available.

4.1.1 Problem Formulation

In this section, a vector is denoted by a bold lower-case letter, and it is distinguished with its coordinates in some coordinate frame. The inertial frame is defined by its three orthogonal axes $\mathbf{I} = \{e_1, e_2, e_3\}$. And the coordinates of e_i in \mathbf{I} is e_i , i.e., the i -th column of $I_{3 \times 3}$, for $i = 1, 2, 3$. Similarly, the body fixed frame is defined as $\mathbf{B} = \{b_1, b_2, b_3\}$. A rotation matrix $R \in \text{SO}(3)$ is used to transform the coordinates of a vector from the body fixed frame to the inertial frame.

The following stochastic attitude kinematics is studied:

$$dR = (\boldsymbol{\omega}(t)dt + H(t)dW)^{\wedge}R, \quad (4.1)$$

$$dR = R(\boldsymbol{\Omega}(t)dt + H(t)dW)^{\wedge}, \quad (4.2)$$

where the angular velocity $\boldsymbol{\omega}(t)$ is resolved in \mathbf{I} , and $\boldsymbol{\Omega}(t)$ is resolved in \mathbf{B} , and they are assumed to be given as deterministic functions of time. The angular velocity is perturbed by the additive noise $H(t)dW$, which is a Wiener process $W \in \mathbb{R}^3$ scaled by a matrix $H(t) \in \mathbb{R}^{3 \times 3}$. The above stochastic differential equations are defined according to the Stratonovich sense so that the random matrix R evolves on $\text{SO}(3)$ [8], and they are called *right-trivialized* and *left-trivialized*, respectively.

Next, two types of direction measurements are also considered. The first type is referred to as *inertial direction measurement*, when the reference direction is known in the inertial frame, and the measurement output is resolved in the body-fixed frame. More specifically, let a be a reference vector fixed in the inertial frame, and $a \in \mathbb{S}^2$ be its coordinates in the inertial frame. The reference direction is measured with a sensor fixed to the body, which provides the measurement $x \in \mathbb{S}^2$ resolved in the body-fixed frame. It is assumed that the sensor measurement for a given attitude is distributed according to the von Mises-Fisher distribution on \mathbb{S}^2 as follows:

$$p(x|R) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa a^T Rx), \quad (4.3)$$

where the mean is $R^T a \in \mathbb{S}^2$, and $\kappa > 0$ is the concentration parameter. On the contrary, the second type of direction measurement is referred to as *body-fixed direction measurement*, when the reference direction is known in the body-fixed frame, and the measurement output is resolved in the inertial frame. In other words, a reference vector b fixed to the body with the coordinates $b \in \mathbb{S}^2$ in the body-fixed frame, is measured by a sensor as $y \in \mathbb{S}^2$ resolved

in the inertial frame. And $y|R$ is assumed to be distributed by

$$p(y|R) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa b^T R^T y). \quad (4.4)$$

It is assumed that the reference direction is repeatedly measured at discrete time instants $\{t_k\}_{k=1}^\infty$, with a fixed time step $h = t_{k+1} - t_k$. This section studies the observability of R when the continuous attitude kinematics (4.1) or (4.2) is combined with the single direction measurements (4.3) or (4.4). The approach is to investigate how the attitude uncertainty is propagated between measurements by the noisy angular velocity measurement, and how the single direction measurement injects information into the uncertainty.

4.1.2 Uncertainty Propagation

Although the continuous kinematics (4.1) and (4.2) can be directly analyzed [118], for simplicity, the discretized versions of them is studied here. As shown in [8], the solutions of

$$R_{k+1} = \exp(h\hat{\omega}_k + (H_k \Delta W)^\wedge) R_k, \quad (4.5)$$

$$R_{k+1} = R_k \exp(h\hat{\Omega}_k + (H_k \Delta W)^\wedge) \quad (4.6)$$

converge in probability to the solutions of the continuous equations (4.1) and (4.2) as $h \rightarrow 0$, respectively, where ΔW is the stochastic increment of W over the time step h . It is further assumed the noise is isotropic [8], i.e., $H_k = \gamma I_{3 \times 3}$ for $\gamma > 0$. Unfortunately, there is no explicit analytical solution to the stochastic difference equations (4.5) and (4.6). Instead, we study how the first moment of R evolves, and interpret its uncertainty utilizing the matrix Fisher distribution.

Specifically, the moments $E[R_k]$ for the right-trivialized (4.5) and left-trivialized (4.6) are propagated into $E[R_{k+1}]$ according to the following theorem.

Theorem 4.1. *Suppose R follows the right-trivialized stochastic difference equation (4.5),*

then

$$\mathbb{E}[R_{k+1}]_R = (1 - h\gamma^2) \exp(h\hat{\omega}_k) \mathbb{E}[R_k] + O(h^2). \quad (4.7)$$

If instead R follows the left-trivialized (4.6), then

$$\mathbb{E}[R_{k+1}]_L = (1 - h\gamma^2) \mathbb{E}[R_k] \exp(h\hat{\Omega}_k) + O(h^2). \quad (4.8)$$

Proof. By the Baker–Campbell–Hausdorff formula, (4.5) can be written as

$$R_{k+1} = \exp(h\hat{\omega}_k) \exp((H_k \Delta W)^\wedge + O(h\Delta W)) R_k.$$

Expanding the exponential term in the middle of the right hand side, and noting that $\mathbb{E}[\Delta W] = 0$, it can be shown that

$$\begin{aligned} \mathbb{E}[\exp((H_k \Delta W)^\wedge + O(h\Delta W))] &= I_{3 \times 3} + \frac{1}{2} \mathbb{E}\left[\left((H_k \Delta W)^\wedge + O(h\Delta W)\right)^2\right] + O(h^2) \\ &= (1 - h\gamma^2) I_{3 \times 3} + O(h^2), \end{aligned}$$

where the second equality used $\mathbb{E}[\Delta W \Delta W^T] = hI_{3 \times 3}$ and (2.13). The left-trivialized (4.8) can be derived similarly. \square

In (4.7) and (4.8), the propagation of the first moment of R from t_k to t_{k+1} is composed of two parts: the exponential term corresponds to the advection, or the rotation of the distribution, due to the deterministic angular velocity, which acts on the left or on the right of the original expectation depending on whether the right- or left- trivialized kinematics is used; and the scalar $1 - h\gamma^2$ represents the diffusion due to noise.

The expectation of R is interpreted using the matrix Fisher distribution. Suppose $R_k \sim \mathcal{M}(F_k)$ with $F_k = U_k S_k V_k^T \in \mathbb{R}^{3 \times 3}$. From (2.20), its first moment is given by $\mathbb{E}[R_k] =$

$U_k D_k V_k^T$. Therefore, (4.7) and (4.8) are rewritten as

$$\begin{aligned}\mathrm{E}[R_{k+1}]_R &= \exp(h\hat{\omega}_k) U_k \times (1 - h\gamma^2) D_k \times V_k^T, \\ \mathrm{E}[R_{k+1}]_L &= U_k \times (1 - h\gamma^2) D_k \times (\exp(-h\hat{\Omega}_k) V_k)^T\end{aligned}$$

respectively, after omitting the higher order terms of h . We assume $R_{k+1} \sim \mathcal{M}(F_{R_{k+1}})$ for (4.5), and $R_{k+1} \sim \mathcal{M}(F_{L_{k+1}})$ for (4.6), with the pSVD of the parameters given by $F_{R_{k+1}} = U_{R_{k+1}} S_{R_{k+1}} V_{R_{k+1}}^T$, and $F_{L_{k+1}} = U_{L_{k+1}} S_{L_{k+1}} V_{L_{k+1}}^T$. Then using the MLE for matrix Fisher distribution, for (4.5) it can be shown that

$$U_{R_{k+1}} = \exp(h\hat{\omega}_k) U_k, \quad V_{R_{k+1}} = V_k, \quad (4.9)$$

and for (4.6) the estimation becomes

$$U_{L_{k+1}} = U_k, \quad V_{L_{k+1}} = \exp(-h\hat{\Omega}_k) V_k. \quad (4.10)$$

Next, for both of (4.5) and (4.6),

$$S_{R_{k+1}} = S_{L_{k+1}} = \mathcal{E}^{-1}((1 - h\gamma^2) D_k) \triangleq S_{k+1}, \quad (4.11)$$

where \mathcal{E}^{-1} denotes solving S from D using (2.20).

Now, the uncertainty propagation of R is considered in three aspects: the mean attitude, the degree of dispersion, and the principal axes. First, the mean attitude is rotated from $M_k = U_k V_k^T$ into

$$M_{R_{k+1}} = U_{R_{k+1}} V_{R_{k+1}}^T = \exp(h\hat{\omega}_k) U_k V_k^T, \quad (4.12)$$

$$M_{L_{k+1}} = U_{L_{k+1}} V_{L_{k+1}}^T = U_k V_k^T \exp(h\hat{\Omega}_k), \quad (4.13)$$

for (4.5) and (4.6) respectively, which is rotated by the deterministic angular velocity as

expected. If the angular velocities are transformed to each other by the mean attitude at t_k , then the propagated mean attitudes are identical, i.e., $\Omega_k = (U_k V_k^T)^T \omega_k$ implies $M_{R_{k+1}} = M_{L_{k+1}}$. Second, the uncertainty becomes more dispersed in the same manner for (4.5) and (4.6), as S_{k+1} is reduced from S_k according to (4.11) and Lemma 2.1.

Finally, the most notable distinction between (4.5) and (4.6) is how the principal axes are rotated. For (4.5), since $U_{R_{k+1}} = \exp(h\hat{\omega}_k)U_k$, the principal axes are rotated by the rotation vector $h\hat{\omega}_k$ when perceived in the inertial frame. However, as $V_{R_{k+1}} = V_k$, the principal axes remain unchanged when observed from the body-fixed frame. For (4.6), it is exactly the opposite: since $V_{L_{k+1}} = \exp(-h\hat{\Omega}_k)V_k$, the principal axes are rotated by the rotation vector $-h\hat{\Omega}_k$ when perceived in the body-fixed frame, and they remain unchanged when observed from the inertial frame. In other words, for (4.5), the shape of uncertainty represented by the most and least uncertain directions remains fixed in the body-fixed frame, but it is rotated in the inertial frame. On the other hand, for (4.6), the most and least uncertain directions are rotated in the body-fixed frame, but they are fixed in the inertial frame. These are illustrated in Figure 4.1.

4.1.3 Measurement Update

This subsection examines how the single direction measurements (4.3) and (4.4) are used to update the uncertainty propagated from the previous time step. The additional information in the direction measurement is injected into the prior uncertainty by calculating the conditional distribution $R|x$ or $R|y$, using Bayes' formula.

Theorem 4.2. *Let $R \sim \mathcal{M}(F^-)$ be a prior distribution for a given $F^- \in \mathbb{R}^{3 \times 3}$. Then, for the inertial direction measurement, the posterior distribution of the attitude conditioned by x is also matrix Fisher with $R|x \sim \mathcal{M}(F_I^+)$, where the posterior matrix parameter $F_I^+ \in \mathbb{R}^{3 \times 3}$ is given by*

$$F_I^+ = F^- + \kappa a x^T. \quad (4.14)$$

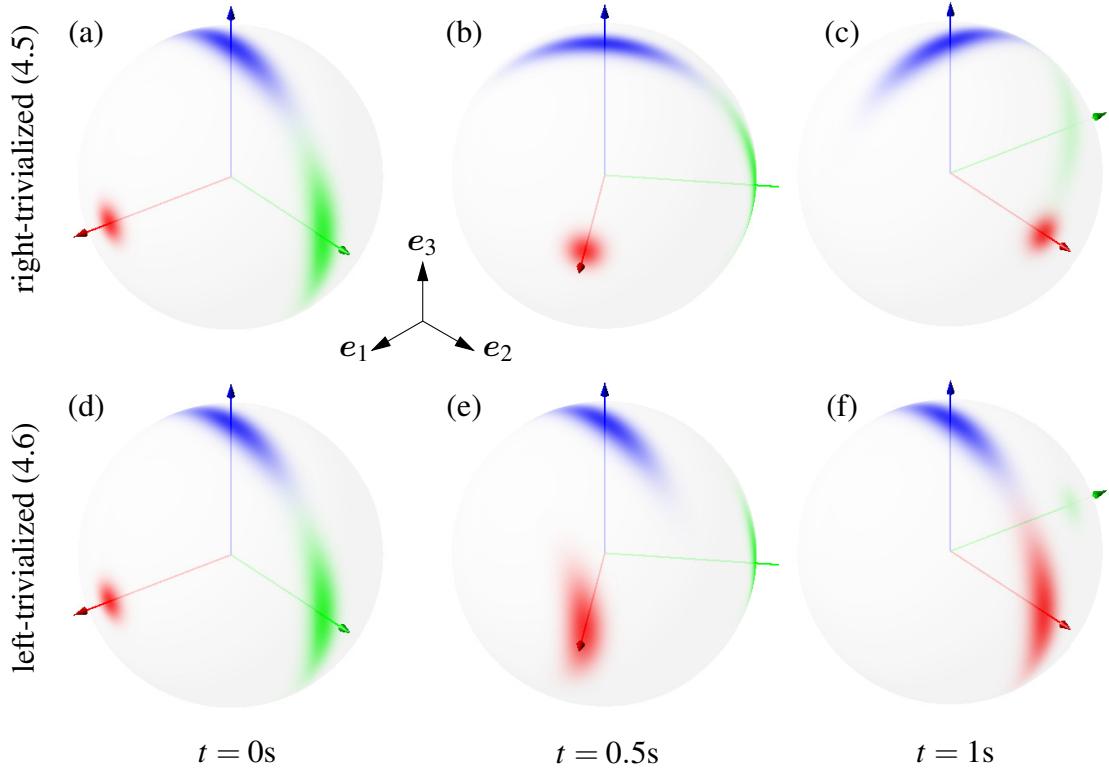


Figure 4.1: Illustration of propagated uncertainties: (a-c) by the right-trivialized (4.5); (d-f) by the left-trivialized (4.6). The initial distribution is $R_0 \sim \mathcal{M}(F_0)$ with $F_0 = \text{diag}(150, 10, 0)$. The angular velocities are $\omega = \Omega = \frac{\pi}{2}e_3 \text{rad s}^{-1}$ without any noise, which can be transformed to each other by the initial mean attitude $I_{3 \times 3}$. The red, green and blue arrows represent the mean directions of the body-fixed b_1, b_2, b_3 axes, and the corresponding shades represent the marginal distribution for each body-fixed axis. It can be observed that the mean attitude and the degree of dispersion are consistent between (4.5) and (4.6). However, for (4.5), the most uncertain direction is fixed along the body-fixed b_1 axis (red), but it is rotated in the inertial frame; and the most uncertain direction for (4.6) is rotated in the body-fixed frame (from red to green), but it remains fixed along the inertial e_1 axis, thereby causing all of the shades circular about e_1 .

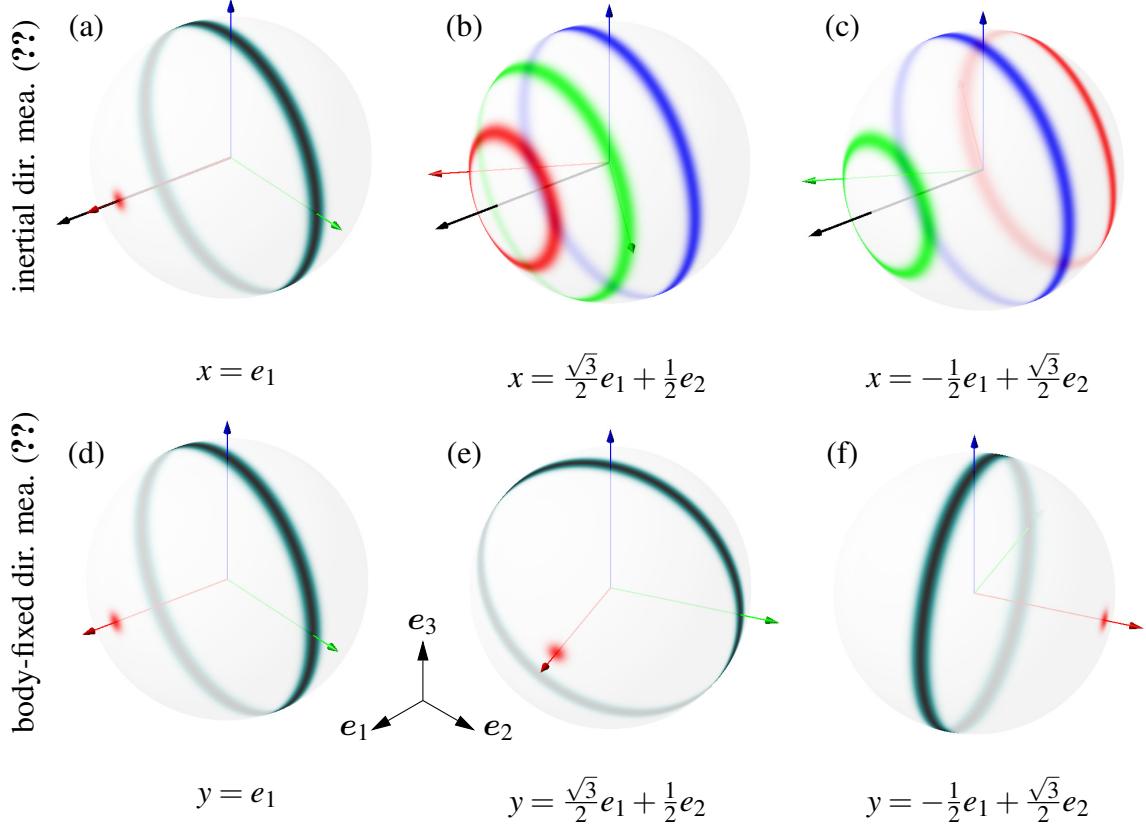


Figure 4.2: Posterior distribution with single direction measurements ($\kappa = 500$): (a-c) inertial direction measurements (4.3) with $a = e_1$; (d-f) body-fixed direction measurements (4.4) with $b = e_1$.

And for the body-fixed direction measurement, the posterior distribution is $R|y \sim \mathcal{M}(F_B^+)$, where $F_B^+ \in \mathbb{R}^{3 \times 3}$ is given by

$$F_B^+ = F^- + \kappa y b^T. \quad (4.15)$$

Proof. The proof for inertial direction measurement is given in Theorem 3.2 of [64], and the body-fixed direction measurement can be derived similarly. \square

Next, the implications of (4.14) and (4.15) are studied. Suppose that the attitude is completely unknown before the measurement, i.e., $F^- = 0_{3 \times 3}$. For the inertial direction

measurement, the matrix parameter (4.14) for the posterior distribution is decomposed into

$$F_I^+ = \begin{bmatrix} a & a' & a'' \end{bmatrix} \text{diag}(\kappa, 0, 0) \begin{bmatrix} x & x' & x'' \end{bmatrix}^T, \quad (4.16)$$

where $a', a'' \in \mathbb{S}^2$ are arbitrarily chosen such that the matrix $[a, a', a''] \in \text{SO}(3)$, and $x', x'' \in \mathbb{S}^2$ are defined similarly. Therefore, F_I^+ is written in the form of pSVD with $S = \text{diag}(\kappa, 0, 0)$. The first principal axis is a when resolved in the inertial frame, or x when resolved in the body-fixed frame. Also, the rotation about the first principal axis is completely unknown as $s_2 + s_3 = 0$. More intuitively, the marginal distribution of each body-fixed axis makes a circle normal to a (the top row of Figure 4.2), which implies the rotation about a cannot be determined. Since a is fixed in the inertial frame, the direction of this ambiguity is also fixed in the inertial frame regardless of the measurement.

Next, the matrix parameter (4.15) for the posterior distribution of a body-fixed direction measurement is decomposed into

$$F_B^+ = \begin{bmatrix} y & y' & y'' \end{bmatrix} \text{diag}[\kappa, 0, 0] \begin{bmatrix} b & b' & b'' \end{bmatrix}^T, \quad (4.17)$$

where $y', y'' \in \mathbb{S}^2$ and $b', b'' \in \mathbb{S}^2$ are chosen such that the corresponding matrices in brackets belong to $\text{SO}(3)$. The resulting first principal axis is b when resolved in the body-fixed frame, or y when resolved in the inertial frame. The marginal distribution of each body-fixed axis makes a circle normal to b (the bottom row of Figure 4.2), about which the rotation cannot be determined. Since b is fixed in the body-fixed frame, the direction of this ambiguity is also fixed to the body.

4.1.4 Stochastic Observability Criterion

To formally study the attitude observability with single direction measurements, here two stochastic attitude observability criteria are introduced. Since the posterior distribution of R conditioned by measurements contains all the information available to determine the attitude,

we study: (i) whether there is a *unique* attitude that minimizes the mean square error for the posterior distribution, and (ii) whether the Fisher information matrix for the mean attitude of R is positive-definite when R is distributed according to the matrix Fisher distribution.

One of the methods to estimate an attitude from a density function on $\text{SO}(3)$ is to solve an optimization problem that minimizes the mean square Frobenius norm.

Definition 4.1. Let $p(R)$ be the probability density function for a random $R \in \text{SO}(3)$. Its minimum mean square estimate (MMSE) is defined as

$$\mathbf{M}_{\text{MMSE}}[R] = \arg \min_{Q \in \text{SO}(3)} \{\mathbb{E} [\|R - Q\|_F^2]\}, \quad (4.18)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Finding MMSE can be addressed in terms of pSVD as follows.

Lemma 4.1. Suppose $R \in \text{SO}(3)$ is a random rotation matrix. Let the pSVD of its first moment be $\mathbb{E}[R] = UDV^T$ with $D = \text{diag}[d_1, d_2, d_3]$. Depending on D , the MMSE of R is given by

1. $d_2 + d_3 > 0$: UV^T (*unique*),
2. $d_1 \neq d_2$ and $d_2 + d_3 = 0$: $\{U \exp(\theta \hat{e}_1) V^T \mid \theta \in [-\pi, \pi]\}$ (1D),
3. $d_1 = d_2 = -d_3 > 0$: $\{U \exp(\theta \hat{a}) V^T \mid a \in \mathbb{S}^2, a_3 = 0, \theta \in [-\pi, \pi]\}$ (2D),
4. $d_1 = d_2 = d_3 = 0$: $\text{SO}(3)$ (3D),

where the number in the parentheses indicates the dimension of the set corresponding to the solution of MMSE.

Proof. The proof uses the equivalent definition [64]

$$\mathbf{M}_{\text{MMSE}}[R] = \arg \max_{R \in \text{SO}(3)} \{\text{tr}(R^T \mathbb{E}[R])\}.$$

Let $\theta \in [-\pi, \pi)$ and $a \in \mathbb{S}^2$ be defined such that $U^T RV = \exp(\theta \hat{a})$. Then, it can be shown that

$$\text{tr}(R^T E[R]) = \text{tr}(D) - (1 - \cos \theta) \sum_{i=1}^3 (d_i + d_j) a_k^2, \quad (4.19)$$

for $i \neq j \neq k$ and $i, j, k \in \{1, 2, 3\}$. Since $d_1 + d_2 \geq d_3 + d_1 \geq d_2 + d_3 \geq 0$, the above is maximized when $\theta = 0$, or equivalently $R = UV^T$.

The uniqueness is contributed by two aspects: the uniqueness of UV^T from the proper singular value decomposition and the uniqueness of the maximum at (4.19). First note that the non-uniqueness of proper singular vectors caused by any simultaneous sign change of the corresponding columns of U and V does not affect the uniqueness of UV^T , therefore only the non-uniqueness of U, V caused by repeated singular values is considered.

1. $d_2 + d_3 > 0$: If $d_3 \geq 0$, the polar decomposition of $E[R]$ is written as $E[R] = (UV^T) \cdot (VDV^T)$, where $VDV^T = (E[R]^T E[R])^{1/2}$ is uniquely determined. The rank of VDV^T is at least two, so UV^T rotates two independent columns of VDV^T to the corresponding columns of $E[R]$, and therefore it is unique.

Next, when $d_3 < 0$, consider two sub-cases: (i) if $d_1 \neq d_2$, then both U and V are unique; (ii) if $d_1 = d_2$, (U, V) can be replaced by $(U \exp(\phi \hat{e}_3), V \exp(\phi \hat{e}_3))$ for any $\phi \in [-\pi, \pi)$, but UV^T is still unique.

- 2a) $d_1 \neq d_2 = d_3 = 0$: It can be shown that

$$\text{tr}(R^T E[R]) = \text{tr}(D) - (1 - \cos \theta) d_1 (1 - a_1^2),$$

which is maximized for any R in the given set of $a = e_1$. The matrices U and V can be replaced with $U \exp(\phi_1 \hat{e}_1)$ and $V \exp(\phi_2 \hat{e}_1)$, respectively, for any $\phi_1, \phi_2 \in [-\pi, \pi)$. However, the ambiguity of U, V does not enlarge the set of R maximizing (4.19).

- 2b) $d_1 \neq d_2 = -d_3 > 0$: Similarly, (4.19) is maximized for any R in the given set. The

matrices (U, V) can be replaced with $(U \exp(\phi \hat{e}_1), V \exp(-\phi \hat{e}_1))$ for any $\phi \in [-\pi, \pi]$.

But, it does not alter the given set.

3) $d_1 = d_2 = -d_3 > 0$: It can be shown that

$$\text{tr}(R^T E[R]) = \text{tr}(D) - (1 - \cos \theta)(d_1 + d_2)a_3^2,$$

which is maximized when $a_3 = 0$. The ambiguity of U and V is written as $U \exp(\phi_1 \hat{e}_1 + \phi_2 \hat{e}_2 + \phi_3 \hat{e}_3)$ and $V \exp(-\phi_1 \hat{e}_1 - \phi_2 \hat{e}_2 + \phi_3 \hat{e}_3)$, respectively for any $\phi_1, \phi_2, \phi_3 \in [-\pi, \pi]$.

Same as above, this does not alter the set of R maximizing (4.19).

4) $d_1 = d_2 = d_3 = 0$: This is a trivial case when $\text{tr}(R^T E[R]) = 0$ for any $R \in \text{SO}(3)$.

These complete the proof. \square

For all cases, the set of MMSE contains UV^T . Nevertheless, only when $d_2 + d_3 > 0$, the MMSE is unique. Otherwise, it can only be determined up to a rotation, where the dimension of the set representing the solution of MMSE is equivalent to 3 minus the rank of $\text{tr}(D)I_{3 \times 3} - D = \text{diag}(d_2 + d_3, d_1 + d_3, d_1 + d_2)$. Therefore, it is claimed that the attitude is completely observable given a density function on $\text{SO}(3)$ if $\text{tr}(D)I_{3 \times 3} - D$ is positive-definite, i.e., when the MMSE is unique.

The above observability criterion is solely based on the first moment, and therefore, it can be applied to an unknown attitude following an arbitrary distribution. Next, it is assumed that the attitude is distributed according to a matrix Fisher distribution to present an alternative information theoretic observability criterion.

Suppose $R \sim \mathcal{M}(F)$ where the pSVD of $F \in \mathbb{R}^{3 \times 3}$ is given by $F = USV^T$. As discussed in Chapter 2.2, the MMSE of $R \sim \mathcal{M}(USV^T)$ is given by the mean attitude $R^* = UV^T$, and we want to calculate its Fisher information $\mathbb{I}(R^*)$. To do this, first a more general problem

of estimating U , S , and V from the given samples of R is studied. The log-likelihood is

$$l(R|U, S, V) = \text{tr}(USV^T R^T) - \log c(S). \quad (4.20)$$

And the corresponding Fisher information matrix is calculated as follows [111].

Lemma 4.2. *The Fisher information matrix of (4.20), namely $\mathbb{I}(U, S, V) : \text{SO}(3) \times \mathbb{R}^3 \times \text{SO}(3) \rightarrow \mathbb{R}^{9 \times 9}$ is constructed as*

$$\begin{aligned} \mathbb{I}(U, S, V) &= -\mathbb{E}[\nabla^2 l(R|U, S, V)] \\ &= \begin{bmatrix} \text{tr}(DS)I_{3 \times 3} - DS & 0 & \sum_{i=1}^3 e_i^T s \hat{e}_i D \hat{e}_i \\ 0 & \frac{\partial^2 \log c(S)}{\partial s^2} & 0 \\ \sum_{i=1}^3 e_i^T s \hat{e}_i D \hat{e}_i & 0 & \text{tr}(DS)I_{3 \times 3} - DS \end{bmatrix}, \end{aligned} \quad (4.21)$$

where $D \in \mathbb{R}^{3 \times 3}$ is the diagonal matrix composed of the proper singular values of $\mathbb{E}[R|U, S, V]$, $s = [s_1, s_2, s_3]^T$, and ∇^2 is the covariant Hessian on $\text{SO}(3) \times \mathbb{R}^3 \times \text{SO}(3)$.

Proof. Let $Q = \text{SO}(3) \times \mathbb{R}^3 \times \text{SO}(3)$, and $q = (U, S, V) \in Q$. The tangent space $T_q Q$ is identified with $T_q Q \simeq \mathbb{R}^9$ through the hat map, and the cotangent space is also identified with \mathbb{R}^9 using the dot product. More specifically, for $\xi = (u, \varsigma, v) \in \mathbb{R}^9$, the corresponding tangent vector is given by $(U\hat{u}, \varsigma, V\hat{v}) \in T_q Q$.

Since l is a real-valued function on Q , its covariant derivative $\nabla_\xi l$ along ξ is equivalent to the differential $dl(\xi)$ given by

$$\begin{aligned} \nabla_\xi l &= \frac{d}{d\varepsilon} \Big|_{\varepsilon=0} l(R|U \exp(\varepsilon\hat{u}), S + \varepsilon \text{diag}[\varsigma], V \exp(\varepsilon\hat{v})) \\ &= \text{tr}((U\hat{u}SV^T + U\text{diag}[\varsigma]V^T - US\hat{v}V^T)^T R) - \frac{\partial \log c(S)}{\partial s} \cdot \varsigma \\ &= \begin{bmatrix} (QS - SQ^T)^\vee \\ \text{diag}[Q] - \frac{1}{c(S)} \frac{\partial c(S)}{\partial s} \\ (Q^T S - SQ)^\vee \end{bmatrix} \cdot \xi, \end{aligned}$$

where $Q = U^T RV$. Because $E[Q] = D$ is diagonal, it is straightforward to show $E[dl(\xi)] = 0$ for any $\xi \in \mathbb{R}^9$.

The covariant Hessian of l along ξ_1 and ξ_2 is given by $\nabla_{\xi_1, \xi_2}^2 l = \xi_2(\xi_1 l) - (\nabla_{\xi_2} \xi_1)l = \xi_2(dl(\xi_1)) - dl(\nabla_{\xi_2} \xi_1)$, where the second term vanishes after taking expectation. The first term is bi-linear in ξ_1 and ξ_2 , thus it can be written as a matrix as in (4.21). Suppose $\xi_1 = (u_1, 0, 0)$ and $\xi_2 = (u_2, 0, 0)$, then it can be shown that

$$\xi_2(dl(\xi_1)) = (-\hat{u}_2 Q S - S Q^T \hat{u}_2)^\vee \cdot u_1 = u_1^T \left\{ \frac{1}{2}(QS + SQ^T) - \text{tr}(QS) I_{3 \times 3} \right\} u_2.$$

Taking the expectation of the expression in the braces with $E[Q] = D$ and multiplying it with -1 yield the upper-left 3-by-3 block of (4.21). The remaining blocks can be obtained similarly. \square

Next, the Fisher information $\mathbb{I}(R^*)$ can be obtained with the above information matrix. Since the variations of U and V are written as $\delta U = U\hat{u}$ and $\delta V = V\hat{v}$ for $u, v \in \mathbb{R}^3$, it can be shown that

$$\delta R^* = U\hat{u}V^T - U\hat{v}V^T.$$

Let $\eta = u - v \in \mathbb{R}^3$ so that $\delta R^* = U\hat{\eta}V^T$. Thus, the Fisher information matrix for the mean attitude R^* is constructed by left-multiplying (4.21) with the matrix $\frac{1}{2}[I_{3 \times 3}; 0_{3 \times 3}; -I_{3 \times 3}]$, and by right-multiplying (4.21) with its transpose, to obtain

$$\mathbb{I}(R^*) = \frac{1}{2} \text{diag} \begin{bmatrix} (d_2 + d_3)(s_2 + s_3) \\ (d_3 + d_1)(s_3 + s_1) \\ (d_1 + d_2)(s_1 + s_2) \end{bmatrix}. \quad (4.22)$$

According to the Cramér–Rao inequality, the inverse of Fisher information $\mathbb{I}(R^*)$ is a lower bound of the variance of all unbiased estimates, up to additional curvature terms. Therefore,

its positive-definiteness can be used to define observability [88]. Interestingly, by Lemma 2.1, the positive-definiteness of $\mathbb{I}(R^*)$ is equivalent to the uniqueness of MMSE presented in Lemma 4.1. Based on these results, we formulate stochastic attitude observability for an arbitrary density as follows.

Definition 4.2. A random rotation matrix $R \sim p(R)$ is stochastically observable if $d_2 + d_3 > 0$, or equivalently

$$\mathcal{O} = \text{tr}(D) I_{3 \times 3} - D \succ 0, \quad (4.23)$$

where $D = \text{diag}[d_1, d_2, d_3]$ is the proper singular values of $\mathbb{E}[R]$. The corresponding measure of observability is

$$\rho(R) = \det[\mathcal{O}] = (d_1 + d_2)(d_3 + d_1)(d_2 + d_3). \quad (4.24)$$

When $R \sim \mathcal{M}(USV^T)$, it is straightforward to show that (4.23) is equivalent to $\text{tr}(S) I_{3 \times 3} - S \succ 0$ with Lemma 2.1. Note that these are readily applied to the stochastic observability considered in this section, as the posterior distribution conditioned by direction measurements is assumed to be a matrix Fisher distribution according to Theorem 4.2, which is presented in the next subsection.

4.1.5 Attitude Observability Results

The attitude uncertainty propagation in Chapter 4.1.2, and the measurement update in Chapter 4.1.3 constitute a Bayesian estimator, which provides the posterior distribution of the attitude conditioned by the history of direction measurements. The posterior distribution can then be used in Definition 4.2 to determine attitude observability. As there are two cases for each of uncertainty propagation and measurement update, there are four possible combinations as summarized in Table 4.1. This subsection identifies two combinations that yield unobservability, and two other cases resulting in observability with single direction

measurements. The same results can also be derived in a deterministic sense as presented in Appendix D.

We first discuss why the common IMU cannot estimate the full attitude with single direction measurements. In a typical IMU, the angular velocity is measured in the body-fixed frame using a gyroscope, and the reference direction in the inertial frame, such as the direction of gravity, is measured in the body-fixed frame. As such it is a combination of the left-trivialized (4.6) and the inertial direction measurement (4.3). Looking at the bottom row of Figure 4.1 and the top row of Figure 4.2, it is clear why the attitude cannot be determined in this case: the direction about which the rotation cannot be determined, namely the first principal axis, remains unchanged in the inertial frame for both (4.6) and (4.3). This is formulated in the next theorem.

Theorem 4.3. *Consider the two Bayesian attitude estimators for*

- right-trivialized angular velocity in the inertial frame (4.5) and body-fixed direction measurement (4.4)
- left-trivialized angular velocity in the body-fixed frame (4.6), and inertial direction measurement (4.3)

with the initial distribution $F_0 = 0_{3 \times 3}$. For both cases, the attitude is not observable.

Proof. First consider the filter given by (4.9), (4.11), and (4.15). The propagated uncertainty before the first measurement is $F_1^- = 0_{3 \times 3}$, thus $F_1 = \kappa y_1 b^T$ after conditioning the first measurement. As shown in (4.17), $(s_2)_1 = (s_3)_1 = 0$, and $V_1 e_1 = b$. We proceed with induction. Suppose $(s_2)_k = (s_3)_k = 0$, and $V_k e_1 = b$. Then by (4.9), (4.11) and Lemma 2.1, the propagated parameters before the next measurement still satisfy $(s_2)_{k+1}^- = (s_3)_{k+1}^- = 0$ and $V_{k+1}^- e_1 = b$. Next consider the update $F_{k+1} = F_{k+1}^- + \kappa y_{k+1} b^T$, which can be written as

$$F_{k+1} = ((s_1)_{k+1}^- U_{k+1}^- e_1 + \kappa y_{k+1}) b^T \triangleq u b^T,$$

where $u \in \mathbb{R}^3$. Let $U_{k+1} = \begin{bmatrix} u \\ \|u\|, u', u'' \end{bmatrix}$, where $u', u'' \in \mathbb{S}^2$ are arbitrarily chosen such that $U_{k+1} \in \text{SO}(3)$. Also let $S_{k+1} = \text{diag}(\|u\|, 0, 0)$, and $V_{k+1} = [b, b', b'']$ as in (4.17). Then $F_{k+1} = U_{k+1} S_{k+1} V_{k+1}^T = ub^T$ is the pSVD of F_{k+1} , and it has been shown that $(s_2)_{k+1} = (s_3)_{k+1} = 0$ and $V_{k+1} e_1 = b$. Therefore, $(s_2)_k = (s_3)_k = 0$ for all $k \in \mathbb{N}$, and by Lemma 2.1, the attitude is not observable. The proof for the second case is similar. \square

Next, two examples illustrating that the other two combinations yield observability are presented. Consider the combination of right-trivialized (4.5) and inertial direction measurement (4.3) for a specific case where the true attitude evolves according to

$$R(t) = R_0 \exp(\hat{\Omega}t) = \exp(\hat{\omega}t) R_0, \quad (4.25)$$

with $R_0 = I_{3 \times 3}$ and $\Omega = \omega = -\frac{\pi}{2\sqrt{3}}[1, 1, 1] \in \mathbb{R}^3$. Initially, it is assumed that the attitude is completely unknown, i.e., $F_0 = 0_{3 \times 3}$, as illustrated in Figure 4.3(a). Then after updated by an inertial direction measurement with $a = e_1$, the rotation about the reference direction is unobservable and the resulting distribution is axially symmetric about e_1 (Figure 4.3(b)). For the right-trivialized (4.5), the distribution rotates in the inertial frame over the propagation, and therefore, the direction of ambiguity is no longer along e_1 (Figure 4.3(c)-(d)). The next inertial direction measurement is fixed in the inertial frame along e_1 (Figure 4.3(e)). Thus, it resolves the ambiguity of the propagated density to determine the attitude completely (Figure 4.3(f)).

The other combination of left-trivialized (4.6) and body-fixed direction measurement (4.4) is similar. The direction of ambiguity caused by the first body-fixed direction measurement with $b = e_1$ is fixed in the inertial frame after propagation (Figure 4.4(b)-(d)). However, the ambiguous direction of the next direction measurement is rotated in the inertial frame (Figure 4.4(e)), and this resolves the previous ambiguity (Figure 4.4(f)). The above intuition is formally presented in the next theorem.

Theorem 4.4. *Consider the two Bayesian attitude estimators for*

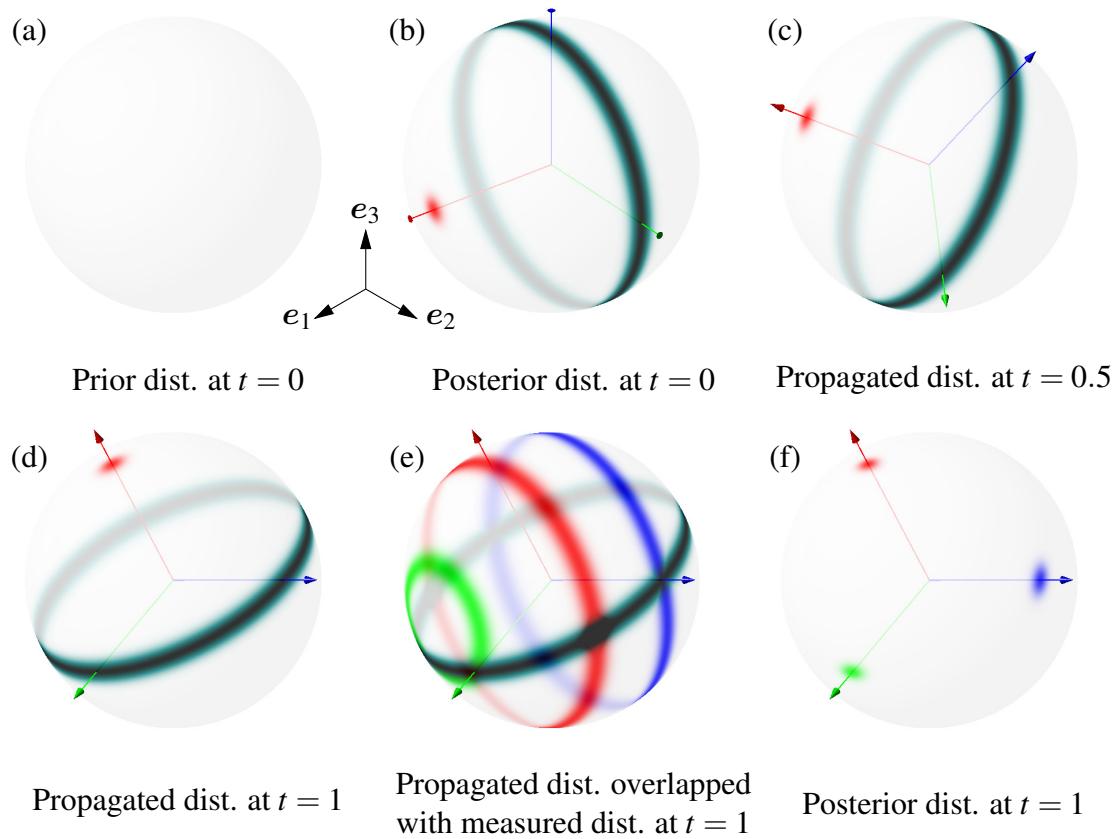


Figure 4.3: Estimation with the right trivialized (4.5) and the inertial direction measurement (4.3), where the complete attitude is estimated after incorporating two inertial single direction measurements.

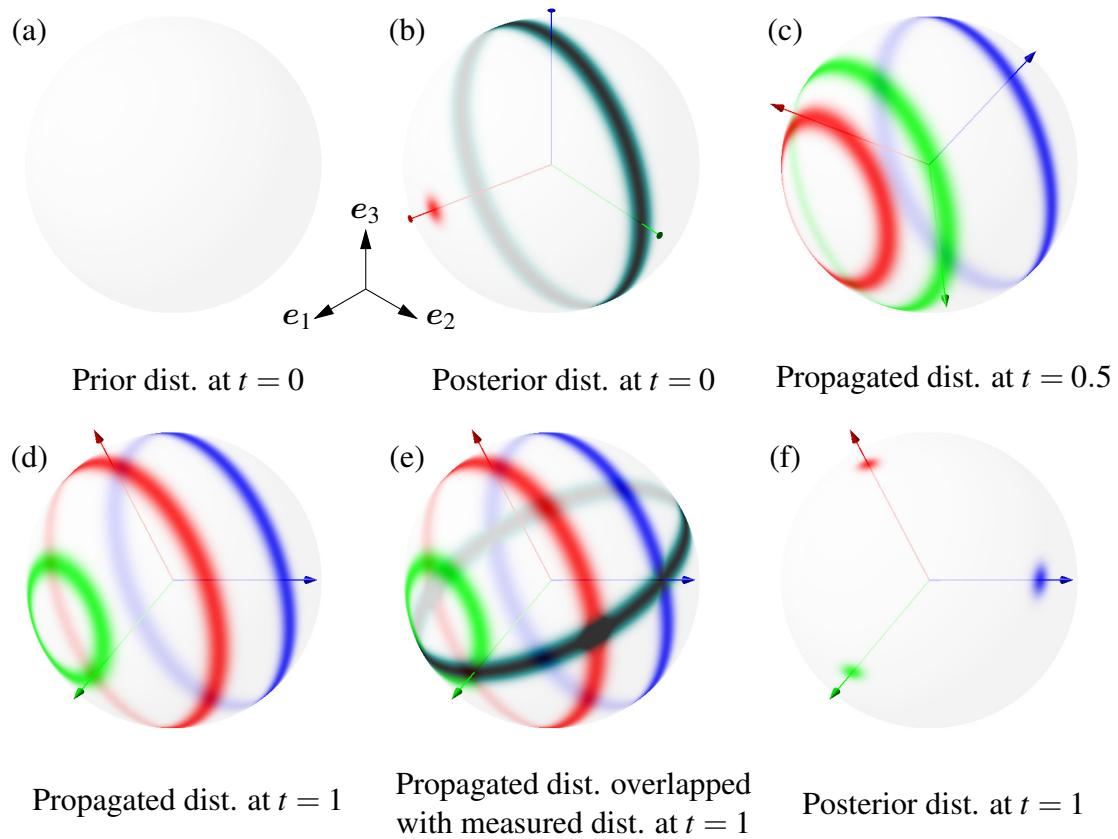


Figure 4.4: Estimation with the left trivialized (4.6) and the body-fixed direction measurement (4.4), where the complete attitude is estimated after incorporating two body-fixed single direction measurements.

- right-trivialized angular velocity in the inertial frame (4.5) and inertial direction measurement (4.3)
- left-trivialized angular velocity in the body-fixed frame (4.6), and body-fixed direction measurement (4.4)

with the initial distribution $F_0 = 0_{3 \times 3}$. Suppose there is some k_0 such that $\omega_{k_0} \times a \neq 0$ for the first case, and $\Omega_{k_0} \times b \neq 0$ for the second case. Then the attitude is observable with probability one for both cases.

Proof. Consider the first case. The posterior distribution after the first measurement is given by $F_1 = \kappa a x_1^T$. Suppose $k_0 = 1$ and denote $\exp(h\hat{\omega}_1) = \delta R$, then by (4.16), (4.9), (4.11) and Lemma 2.1, $U_2^- = \delta R[a, a', a'']$, $S_2^- = \text{diag}((s_1)_2^-, 0, 0)$, $V_2^- e_1 = x_1$, where $(s_1)_2^- > 0$ satisfies

$$\frac{1}{c(S_2^-)} \frac{\partial c(S_2^-)}{\partial (s_1)_2^-} = (1 - h\gamma^2) \frac{1}{c(S_1)} \frac{\partial c(S_1)}{\partial (s_1)_1}.$$

Thus, the posterior distribution after the second measurement is

$$F_2 = (s_1)_2^- \delta R a x_1^T + \kappa a x_2^T. \quad (4.26)$$

Let $\delta R a = \alpha a + \alpha' a' + \alpha'' a''$ for some $\alpha, \alpha', \alpha'' \in \mathbb{R}$. Since $\omega_1 \times a \neq 0$, α' and α'' cannot both be zeros. Then

$$\begin{aligned} F_2 &= (s_1)_2^- (\alpha a + \alpha' a' + \alpha'' a'') x_1^T + \kappa a x_2^T \\ &= a((s_1)_2^- \alpha x_1 + \kappa x_2)^T + (s_1)_2^- \alpha' a' x_1^T + (s_1)_2^- \alpha'' a'' x_1^T. \end{aligned}$$

Let $v = (s_1)_2^- \alpha x_1 + \kappa x_2$, and v', v'' be arbitrarily chosen such that $\left[\frac{v}{\|v\|}, v', v'' \right] \in \text{SO}(3)$. Also, let $x_1 = \beta \frac{v}{\|v\|} + \beta' v' + \beta'' v''$. Note that β_2 and β_3 cannot both be zeros almost surely.

Using these, F_2 is written as

$$F_2 = \begin{bmatrix} a & a' & a'' \end{bmatrix} \Lambda \begin{bmatrix} v \\ \|v\| & v' & v'' \end{bmatrix}^T,$$

where $\Lambda \in \mathbb{R}^{3 \times 3}$ is

$$\Lambda = \begin{bmatrix} \|v\| & 0 & 0 \\ (s_1)_2^- \alpha' \beta & (s_1)_2^- \alpha' \beta' & (s_1)_2^- \alpha' \beta'' \\ (s_1)_2^- \alpha'' \beta & (s_1)_2^- \alpha'' \beta' & (s_1)_2^- \alpha'' \beta'' \end{bmatrix}.$$

Since F_2 is obtained by multiplying rotation matrices to Λ , it is straightforward to see that F_2 and Λ share the same proper singular values. Note that $\det(\Lambda) = 0$, so there is at least one zero singular value. However, the rank of Λ is two almost surely, as at least one element of the right bottom 2-by-2 block is nonzero. Thus, Λ has only one zero singular value. By the definition of proper singular value decomposition, this concludes $\text{tr}(S_2)I_{3 \times 3} - S_2$ is positive-definite, and therefore the attitude is observable with probability one.

Next suppose $k_0 > 1$. By (4.26), $F_2 = a((s_2)_1^- x_1 + \kappa x_2)^T$ since δR does not rotate a , which means both the uncertainty propagation and update steps leave $U_k e_1 = a$, $(s_1)_k > 0$, and $(s_2)_k = (s_3)_k = 0$. Thus, the argument in the last paragraph still applies at time $t = t_{k_0}$. The proof for the other estimator is similar. \square

Finally, it should be noted that although it has been assumed that the attitude follows matrix Fisher distribution, this is not restrictive in the sense that if the initial distribution is uniform, and the angular velocity noise in (4.5) and (4.6) is zero, i.e., $\gamma = 0$, then the attitude conditioned by single direction measurements exactly follows the matrix Fisher distribution. Therefore, it is supposed that the result presented in Table 4.1 is not specific to the estimator. Instead, it is inherent to the observed stochastic dynamical system given by (4.5), (4.6) and (4.3), (4.4). Indeed, it is shown by numerical simulations and experiments that multiplicative extended Kalman filter also exhibits the same observability. Furthermore, in appendix D, it

can be proved using classic observability condition that the same observability applies to the deterministic system without noises.

4.1.6 Simulations and Experiments

In this subsection, the attitude observability results presented in Table 4.1 are validated using numerical simulations and experiments. The attitude estimation presented in Chapter 4.1.2 and Chapter 4.1.3 based on matrix Fisher distribution, and MEKF are used to estimate the attitude with angular velocity and single directions measurements.

For simulation, a rigid body is assumed to rotate about its third body-fixed axis at 6 rad s^{-1} , which simultaneously rotates about the second inertial axis at 1 rad s^{-1} . The angular velocity is measured at 150 Hz either in the inertial frame or body-fixed frame, with the isotropic random walk noise of $H = \gamma I_{3 \times 3}$ where $\gamma = 10 \text{ deg}/\sqrt{\text{s}}$. The reference vector is set as $[1, 0, 0]$, expressed either in the inertial or body-fixed frame, and measured at 30 Hz. For the matrix Fisher estimator, the measurement noise is formulated as in (4.3) or (4.4) with $\kappa = 200$, and the initial attitude distribution is uniform, i.e., $F_0 = 0_{3 \times 3}$. The noise parameters and initial conditions for MEKF are chosen similarly as the matrix Fisher estimator.

The full attitude error is defined as the angle between the estimated attitude and true attitude. For the two unobservable combinations, a partial attitude error which neglects the unobserved degree of freedom is also studied. More specifically, if the reference vector is known in the inertial frame as a , then the partial attitude error is defined as $\arccos(R_t^T a \cdot R^T a)$, where R_t is the true attitude; if the reference vector is known in the body-fixed frame as b , the partial attitude error is defined as $\arccos(R_t b \cdot Rb)$. The four combinations of angular velocity and reference vector measurements are labeled as follows:

	ang. vel.	body-fixed frame	inertial frame
ref. vec.			
body-fixed frame		AVB_RVB	AVI_RVB
inertial frame		AVB_RVI	AVI_RVI

where the boldface font indicates the cases with observability. For each case, 100 Monte Carlo simulations with the simulation period of 60 s are carried out. The attitude error is first averaged across all timestamps in one simulation, and further averaged across simulations.

The attitude estimation errors are summarized in Table 4.2. It is clearly shown that for the two observable cases (AVI_RVI and AVB_RVB), the full attitude can be estimated with the average error of about 6° . On the other hand, for the two unobservable cases (AVI_RVB and AVB_RVI), the full attitude error is around 90° , and the partial attitude error is around 3.3° . For a more straightforward comparison, the time evolution of attitude errors for a single simulation is shown in Figure 4.5, where it is illustrated that for the two observable cases, the full attitude error quickly converges from 180° to around zero; whereas for the two unobservable cases, the full attitude error never converges, but the partial attitude error remains low throughout the simulation since the initial partial error is zero.

The attitude observability is also indicated by the estimated dispersion, quantified as the standard deviation of the error rotation vector as shown in Figure 4.6. For the two observable cases, the standard deviations along all three axes remain below 10° , with the rotation about the reference vector (the first axis e_1) a little bit larger than the two other axes. For the two unobservable cases, the standard deviation along the reference vector for the matrix Fisher estimator is mostly above 10^7 deg, which indicates a uniform distribution. For MEKF, if the reference vector is known in the body-fixed frame, the standard deviation along the reference vector is similar to the matrix Fisher estimator, which is above 10^6 deg. However, if the reference vector is known in the inertial frame, the standard deviation is around 15° , which is only marginally larger than the observable cases. This discrepancy is caused by that the error rotation vector is expressed in the body-fixed frame (1.3), rather than in the inertial

Table 4.2: Estimation error (deg) for the four combinations of measurements

estimator		matrix Fisher			
case (observable , unobservable)		AVI_RVI	AVI_RVB	AVB_RVI	AVB_RVB
simulation	full error \pm s.d.	6.57 \pm 0.25	90 \pm 0	90 \pm 0	5.71 \pm 0.15
	partial error \pm s.d.	-	3.33 \pm 0.06	3.34 \pm 0.06	-
experiment	full error	7.15	90.29	90.41	1.74
	partial error	-	0.63	3.02	-

estimator		MEKF			
case (observable , unobservable)		AVI_RVI	AVI_RVB	AVB_RVI	AVB_RVB
simulation	full error \pm s.d.	7.02 \pm 0.50	135 \pm 24	89 \pm 37	6.06 \pm 0.27
	partial error \pm s.d.	-	3.33 \pm 0.06	3.33 \pm 0.06	-
experiment	full error	15.92	161.21	93.32	18.94
	partial error	-	2.95	6.43	-

frame (1.4). And if the reference vector is known in the inertial frame, MEKF has been shown to apply some slight but erroneous corrections to the rotation about the reference vector due to the linearization of the measurement function [74].

The attitude observability was also validated through experiments. We use a custom-made hardware platform (Figure 4.7), which has been developed for autonomous unmanned aerial vehicles, to collect measurements while moving it with hands. An external VICON motion capture system detects reflective markers attached to the platform to determine its attitude, which is used as the ground truth attitude. A 9-axis inertial measurement unit (VectorNav VN100) is attached to the platform, and the onboard gyroscope provides the angular velocity measurement in the body-fixed frame, which is also converted into the angular velocity measurement in the inertial frame using the ground truth attitude. For the inertial direction measurement, the direction of gravity is measured by the accelerometer in IMU. Moreover, for the body-fixed direction measurement, two additional markers are attached to the platform as a known and fixed reference vector in the body-fixed frame, which is measured by the Vicon motion system in the inertial frame. All Vicon and IMU measurements are synchronized and sampled at 100 Hz, using an onboard Nvidia Jetson TX2 computing module. The platform was rotated around its roll, pitch, and yaw axes

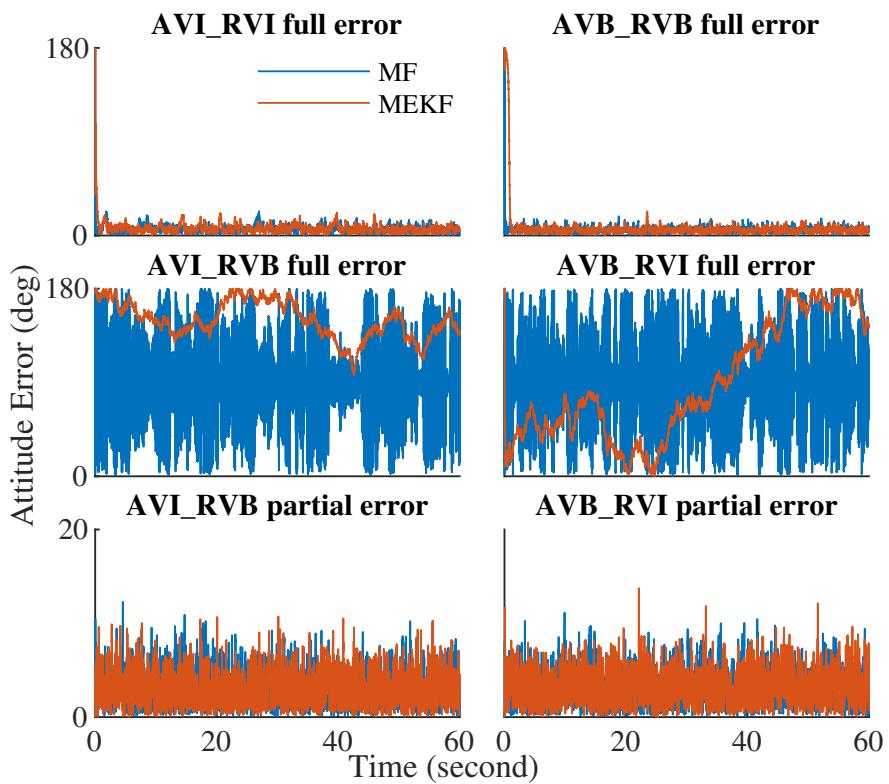


Figure 4.5: Attitude errors for the matrix Fisher (MF) estimator and MEKF in four combinations of angular velocity and reference vector measurements.

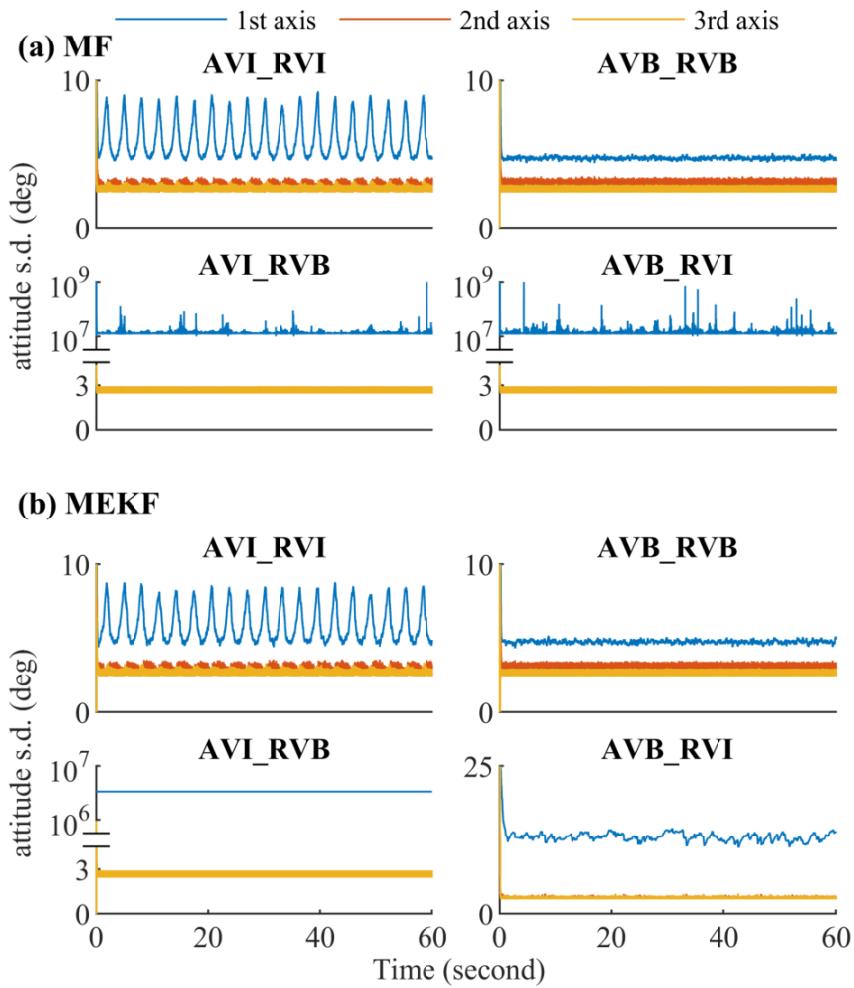
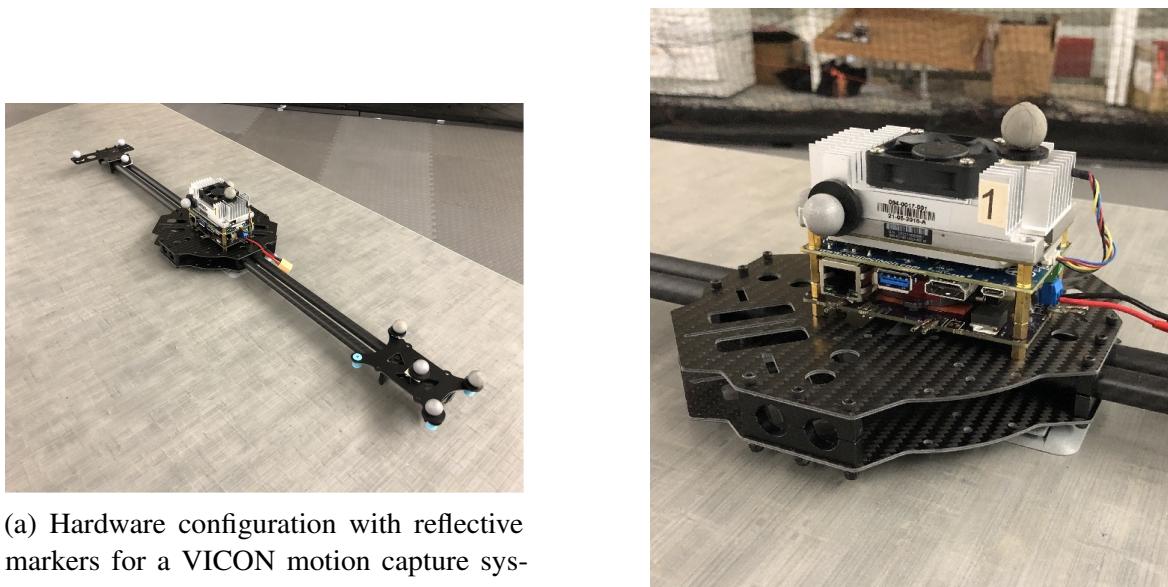


Figure 4.6: Attitude standard deviations for the matrix Fisher (MF) estimator and MEKF. In the “AVI” and “AVB” cases, the attitude covariance matrix is expressed in the inertial and body-fixed frames respectively. For the MF filter, $(\text{tr}(S)I_{3 \times 3} - S)^{-1}$ is used as the attitude covariance matrix in the principal axes frame.



(a) Hardware configuration with reflective markers for a VICON motion capture system

(b) Onboard computing module connected to IMU through a custom-made printed circuit board

Figure 4.7: Hardware platform for experiments.

during the data collection. The matrix Fisher estimator and MEKF are run off-board using the collected experimental data, with the single vector measurement update applied at 20 Hz. The noise parameters and initial conditions are set the same as in the simulation. Note that these are not carefully tuned for the specific hardware, since the main objective is to verify the observability.

The attitude errors for the four combinations of measurements are presented in Figure 4.2 and Figure 4.8. Similar with the simulation results, for the two observable cases (AVI_RVI and AVB_RVB) the full attitude error converges to around zero within 10 s; whereas for the two unobservable cases (AVI_RVB and AVB_RVI) only the partial error converges. The standard deviation of the rotation vector is demonstrated in Figure 4.9. Again the uncertainties for the two observable cases are small in all three axes, but those for the two unobservable cases are only small in two axes, which is very similar to the simulation results.

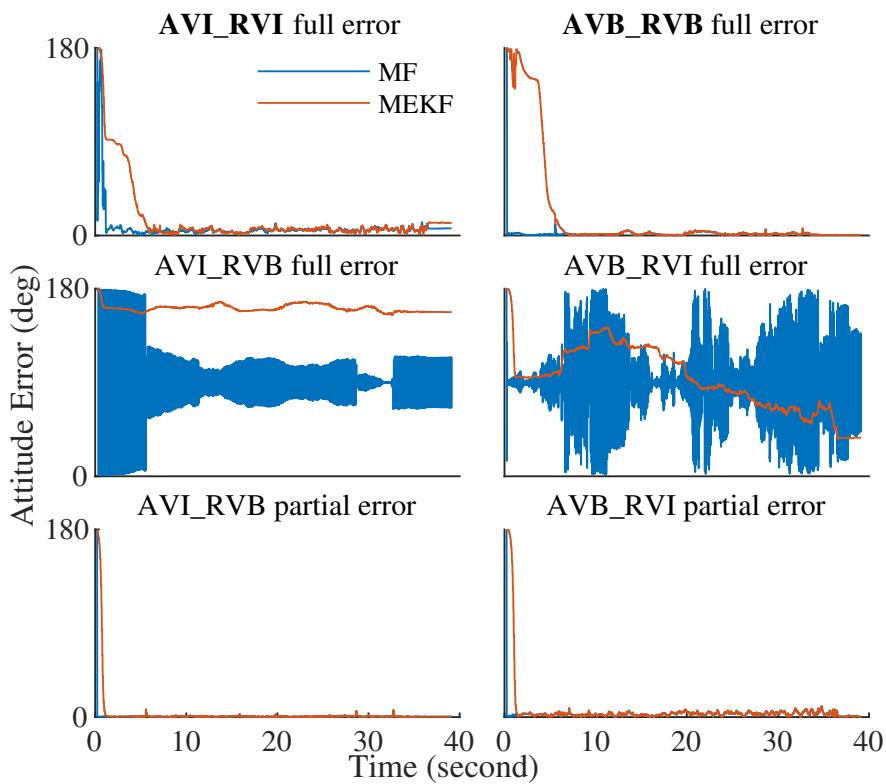


Figure 4.8: Attitude errors for the matrix Fisher (MF) estimator and MEKF in four combinations of measurements with experimental data.

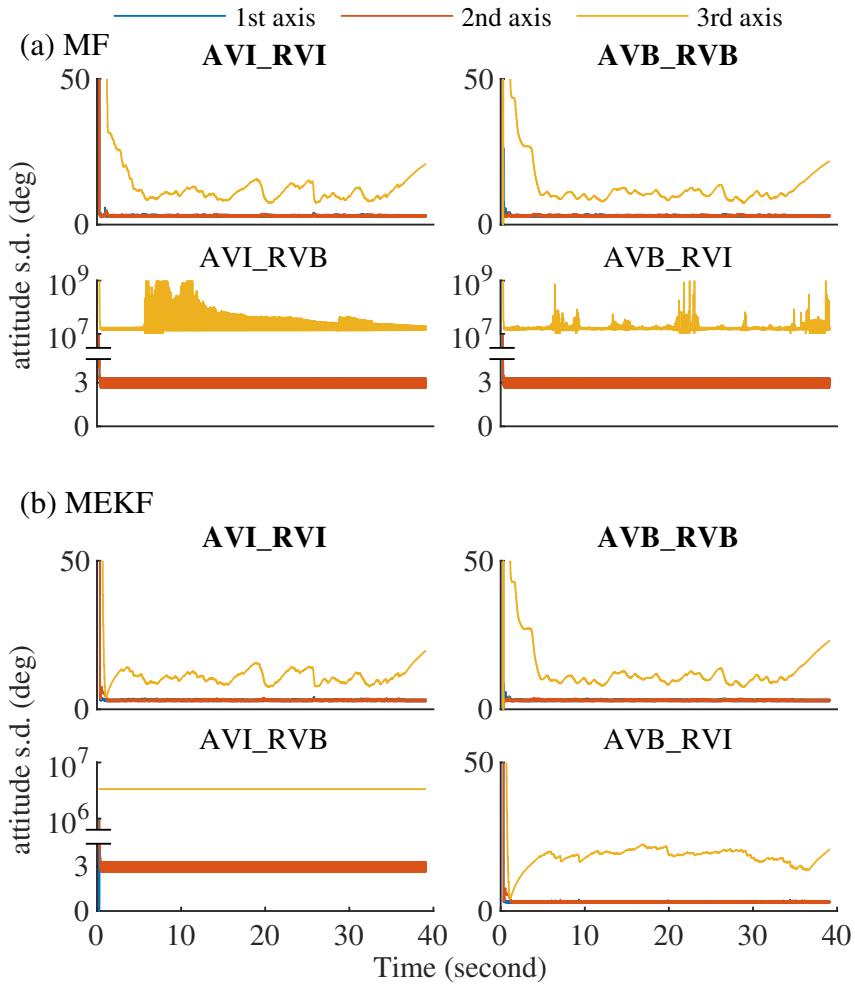


Figure 4.9: Attitude uncertainty for the matrix Fisher estimator (MF) and MEKF in four combinations of measurements with experimental data. In the “RVI” cases, the attitude covariance matrix is expressed in the inertial frame; and in the “RVB” cases, it is expressed in the $b'-b''-b$ frame, where b is the reference vector when resolved in the body-fixed frame, and b', b'' are perpendicular to b . For the MF filter, $(\text{tr}(S)I_{3 \times 3} - S)^{-1}$ is used as the attitude covariance matrix in the principal axes frame.

4.2 Attitude Estimation

In the previous section, a Bayesian attitude estimator is proposed to estimate the attitude with a gyroscope and repeated measurements of a single reference direction, in order to study attitude observability. In (4.2), it is assumed that the angular velocity measured from a gyroscope is affected by a white noise. However, usually the gyroscope noise also contains a random walk term, referred to as the bias [23]. As a random walk process, the gyroscope bias is time varying, and can easily deteriorate the integrated attitude in short time if it is not properly compensated for. Therefore, an attitude estimator also needs to estimate the bias, simultaneously with the attitude.

Concurrent estimation of attitude and gyroscope bias has been studied in numerous works, as reviewed in Chapter 1.2.2. Nevertheless, most existing algorithms rely on variations of Kalman filter, which uses Gaussian distribution to model attitude dispersion, thus they cannot handle large attitude uncertainty as shown in Figure 1.1. Although the matrix Fisher and Bingham distributions have been applied to deal with large uncertainty [36, 62, 64], they cannot estimate the gyroscope bias as the two distributions cannot model the correlation between attitude and gyroscope bias.

The introduction of matrix Fisher–Gaussian distribution in Chapter 3 was primarily motivated to use matrix Fisher distribution to deal with large attitude uncertainty, while estimating the gyroscope bias. The correlation between attitude and bias is directly handled by the MFG in a global and singularity free fashion, allowing the attitude uncertainty to be very large. In this section, a new recursive Bayesian attitude estimator is proposed based on MFG to estimate the attitude and gyroscope bias concurrently.

4.2.1 Problem Formulation

In this section, the following gyroscope kinematics is considered:

$$R^T dR = (\hat{x} + \hat{\Omega}) dt + (H_u dW_u)^\wedge, \quad (4.27)$$

$$dx = H_v dW_v, \quad (4.28)$$

where $R \in \text{SO}(3)$ 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 (4.27) is interpreted in the Stratonovich sense to guarantee the process does not leave $\text{SO}(3)$ [8, 85]. Let the time be discretized by a sequence $\{t_0, t_1, \dots\}$. For convenience, it is assumed that the time step $h \in \mathbb{R}$ is fixed, i.e., $h = t_{k+1} - t_k$ for any k . According to [8, Eqn. 14], the kinematics model can be discretized into

$$R_{k+1} = R_k \exp \left\{ h(\hat{\Omega}_k + \hat{x}_k) + (H_u \Delta W_u)^\wedge \right\}, \quad (4.29)$$

$$x_{k+1} = x_k + H_v \Delta W_v, \quad (4.30)$$

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

$$H_u \Delta W_u \sim \mathcal{N}(0, hG_u), \quad H_v \Delta W_v \sim \mathcal{N}(0, hG_v), \quad (4.31)$$

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

Two types of measurements are considered available at t_k , $k \geq 1$: (i) when the attitude is directly measured, and (ii) when reference directions in the inertial frame, such as the direction of magnetic field or gravity, are measured. First, suppose the attitude is measured by N_a attitude sensors as $Z_i \in \text{SO}(3)$, whose error is distributed by the matrix Fisher distribution. More specifically, given the true attitude $R_t \in \text{SO}(3)$, the measurement error $R_t^T Z_i \in \text{SO}(3)$ follows the matrix Fisher distribution with the parameter $F_{Z_i} \in \mathbb{R}^{3 \times 3}$ 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 [80] with mean direction $R_t^T B_j a_j \in \mathbb{S}^2$ and concentration parameter $\kappa_j > 0$. The parameter $B_j \in \text{SO}(3)$ specifies the constant bias of the direction sensor, and κ_j specifies the concentration of its random noise. The measurements $\left\{ \{Z_i\}_{i=1}^{N_a}, \{z_j\}_{j=1}^{N_v} \right\}$ at time t_k are altogether denoted by \mathcal{Z}_k .

The initial attitude and bias (R_0, x_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. Given the gyroscope measurements $\{\Omega_k\}_{k=0}^\infty$, and the attitude and direction measurements $\{\mathcal{Z}_k\}_{k=1}^\infty$ at discrete time instants, the problem is to find the posterior distribution of $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$, and approximate it to an MFG with parameters $(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$. To find the posterior distribution, two steps are needed: first the uncertainty is propagated from $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$ to $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$ according to the gyroscope kinematics equations (4.29) and (4.30). Second, the prior distribution $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$ is updated by the measurements \mathcal{Z}_{k+1} using Bayes's formula into $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^{k+1}$. These two steps are developed using the MFG in the next two subsections.

4.2.2 Uncertainty Propagation

Suppose $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$. In this subsection, the expectations of (R_{k+1}, x_{k+1}) used in the MLE of MFG are calculated to construct a new MFG describing the uncertainty of (R_{k+1}, x_{k+1}) . The expectations are calculated in two different ways: (i) approximate analytical expressions are developed directly from the gyroscope kinematics, and (ii) they are approximated empirically using deterministically sampled sigma points.

Let us first look at how to develop analytical expressions for required moments. The exponent in (4.29) 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 (4.29).

Lemma 4.3. *Equation (4.29) is almost surely equivalent to*

$$R_{k+1} = R_k e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h)} e^{h(\hat{\Omega}_k + \hat{\mu}_k)}. \quad (4.32)$$

Proof. Equation (4.29) 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 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,

$$R_{k+1} = R_k e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + A} e^{h(\hat{\Omega}_k + \hat{\mu}_k)}, \quad (4.33)$$

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 $\lim_{h \rightarrow 0} \Delta W_u = 0$ almost surely, $A \sim o(h)$. \square

This lemma is helpful in making use of the closed form expression of the exponential map (2.6) for the deterministic components of (4.29). 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, an expression for $\mathbb{E}[R_{k+1}]$ is derived to solve the marginal MLE.

Theorem 4.5. *The expectation of the propagated attitude R_{k+1} is given by*

$$\mathbb{E}[R_{k+1}] = \left\{ \mathbb{E}[R_k] \left(I_{3 \times 3} + \frac{h}{2} (G_u - \text{tr}(G_u) I_{3 \times 3}) \right) + h U_k \mathbb{E} \left[Q_k V_k^T \widehat{P_k v_{R_k}} \right] \right\} e^{h(\hat{\Omega}_k + \hat{\mu}_k)} + O(h^2), \quad (4.34)$$

where $Q_k = U_k^T R_k V_k$, $v_{R_k} = (Q_k S_k - S_k Q_k^T)^\vee$ for MFGI, or $v_{R_k} = (S_k Q_k - Q_k^T S_k)^\vee$ for MFGB.

Proof. First, expand the first exponential term in (4.32) into an infinite sum as

$$e^{h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h)} = \sum_{i=0}^{\infty} \frac{1}{i!} \left\{ h(\hat{x}_k - \hat{\mu}_k) + (H_u \Delta W_u)^\wedge + o(h) \right\}^i. \quad (4.35)$$

Note that (i) ΔW_u is a zero mean Gaussian vector with covariance matrix $h I_{3 \times 3}$, so its odd order moments are zero, and $\mathbb{E}[(H_u \Delta W_u)^\wedge] \sim O(h^n)$; (ii) $o(h)$ in the above equation only has terms of order at least h^2 or $h \Delta W_u$ as shown in (4.33). Combining these two observations, the first order approximation of $\mathbb{E}[R_{k+1}]$ can be written as

$$\mathbb{E}[R_{k+1}] = \left\{ \mathbb{E}[R_k] + h \mathbb{E}[R_k(\hat{x}_k - \hat{\mu}_k)] + \frac{1}{2} \mathbb{E}[R_k((H_u \Delta W_u)^\wedge)^2] \right\} e^{h(\hat{\Omega}_k + \hat{\mu}_k)} + O(h^2). \quad (4.36)$$

Then, since (R_k, x_k) follows MFG,

$$\mathbb{E}[R_k(\hat{x}_k - \hat{\mu}_k)] = \mathbb{E}\left[R_k \widehat{P_k v_{R_k}}\right] = U_k \mathbb{E}\left[Q_k V_k^T \widehat{P_k v_{R_k}}\right].$$

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

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

Substituting the above two equations into (4.36) yields (4.34). \square

With the given $E[R_{k+1}]$, the marginal MLE for the attitude part of MFG can be solved as discussed in Chapter 3.3, which yields the estimates of U_{k+1} , S_{k+1} and V_{k+1} . Define $Q_{k+1} = U_{k+1}^T R_{k+1} V_{k+1}$, and $v_{R_{k+1}} = (Q_{k+1} S_{k+1} - S_{k+1} Q_{k+1}^T)^\vee$ for MFGI, or $v_{R_{k+1}} = (S_{k+1} Q_{k+1} - Q_{k+1}^T S_{k+1})^\vee$ for MFGB 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 3.6 with the moments given as follows.

Theorem 4.6. Let $\tilde{U}, \tilde{V} \in SO(3)$ and $\tilde{S}, \tilde{\tilde{V}}, \tilde{\tilde{S}} \in \mathbb{R}^{3 \times 3}$ be

$$\begin{aligned}\tilde{U} &= U_{k+1}^T U_k, & \tilde{V} &= V_{k+1}^T e^{-h(\hat{\Omega}_k + \hat{\mu}_k)} V_k, & \tilde{S} &= \tilde{U}^T S_{k+1} \tilde{V}, \\ \tilde{\tilde{V}} &= V_{k+1}^T e^{-h(\hat{\Omega}_k + \hat{\mu}_k)} G_u^T V_k, & \tilde{\tilde{S}} &= \tilde{U}^T S_{k+1} \tilde{\tilde{V}}^T.\end{aligned}$$

Also, let $\tilde{v}_R, \tilde{\tilde{v}}_R \in \mathbb{R}^3$, and $\Gamma_Q \in \mathbb{R}^{3 \times 3}$ be

$$(MFGI) \quad \tilde{v}_R = (Q_k \tilde{S}^T - \tilde{S} Q_k^T)^\vee, \quad (4.37a)$$

$$(MFGB) \quad \tilde{v}_R = (\tilde{S}^T Q_k - Q_k^T \tilde{S})^\vee, \quad (4.37b)$$

$$(MFGI) \quad \tilde{\tilde{v}}_R = (Q_k \tilde{\tilde{S}}^T - \tilde{\tilde{S}} Q_k^T)^\vee, \quad (4.38a)$$

$$(MFGB) \quad \tilde{\tilde{v}}_R = (S_{k+1} \tilde{U} Q_k \tilde{\tilde{V}}^T - \tilde{\tilde{V}} Q_k^T \tilde{U}^T S_{k+1})^\vee. \quad (4.38b)$$

$$(MFGI) \quad \Gamma_Q = (\text{tr}(Q_k \tilde{S}^T) I_{3 \times 3} - Q_k \tilde{S}^T) Q_k \quad (4.39a)$$

$$(MFGB) \quad \Gamma_Q = \text{tr}(Q_k^T \tilde{S}) I_{3 \times 3} - Q_k^T \tilde{S} \quad (4.39b)$$

Then, the moments $E[x_{k+1}]$, $E[v_{R_{k+1}}]$, and $E[x_{k+1} x_{k+1}^T]$ required for the conditional MLE

are given by

$$\mathbb{E}[x_{k+1}] = \mu_k, \quad (4.40)$$

$$\mathbb{E}[v_{R_{k+1}}] = 0, \quad (4.41)$$

$$\mathbb{E}[x_{k+1}x_{k+1}^T] = \mathbb{E}[x_kx_k^T] + hG_v. \quad (4.42)$$

Also, the moments $\mathbb{E}[x_{k+1}v_{R_{k+1}}^T]$ and $\mathbb{E}[v_{R_{k+1}}v_{R_{k+1}}^T]$ are

$$(MFGI) \quad \mathbb{E}[x_{k+1}v_{R_{k+1}}^T] = \left[P_k \left(\mathbb{E}[v_{R_k}\tilde{v}_R^T] + \frac{h}{2}\mathbb{E}[v_{R_k}\tilde{v}_R^T] - \frac{h\text{tr}(G_u)}{2}\mathbb{E}[v_{R_k}\tilde{v}_R^T] \right. \right. \\ \left. \left. + h\mathbb{E}[v_{R_k}v_{R_k}^TP_k^TV_k\Gamma_Q^T] \right) + \mu_k \left(\mathbb{E}[\tilde{v}_R^T] + \frac{h}{2}\mathbb{E}[\tilde{v}_R^T] - \frac{h\text{tr}(G_u)}{2}\mathbb{E}[\tilde{v}_R^T] \right. \right. \\ \left. \left. + h\mathbb{E}[v_{R_k}^TP_k^TV_k\Gamma_Q^T] \right) + h\Sigma_{c_k}V_k\mathbb{E}[\Gamma_Q^T] \right] \tilde{U}^T + O(h^2), \quad (4.43a)$$

$$(MFGB) \quad \mathbb{E}[x_{k+1}v_{R_{k+1}}^T] = \left[P_k \left(\mathbb{E}[v_{R_k}\tilde{v}_R^T] - \frac{h\text{tr}(G_u)}{2}\mathbb{E}[v_{R_k}\tilde{v}_R^T] + h\mathbb{E}[v_{R_k}v_{R_k}^TP_k^TV_k\Gamma_Q^T] \right) \right. \\ \left. + \mu_k \left(\mathbb{E}[\tilde{v}_R^T] - \frac{h\text{tr}(G_u)}{2}\mathbb{E}[\tilde{v}_R^T] + h\mathbb{E}[v_{R_k}^TP_k^TV_k\Gamma_Q^T] \right) + h\Sigma_{c_k}V_k\mathbb{E}[\Gamma_Q^T] \right] \tilde{V}^T \\ + \frac{h}{2} \left(\mu_k\mathbb{E}[\tilde{v}_R^T] + P_k\mathbb{E}[v_{R_k}\tilde{v}_R^T] \right) + O(h^2), \quad (4.43b)$$

and

$$(MFGI) \quad \mathbb{E}[v_{R_{k+1}}v_{R_{k+1}}^T] = \tilde{U} \left[\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] + h\mathbb{E}[\Gamma_QV_k^TP_kv_{R_k}\tilde{v}_R^T] + h\mathbb{E}[\tilde{v}_Rv_{R_k}^TP_k^TV_k\Gamma_Q^T] \right. \\ \left. + h\mathbb{E}[\Gamma_QV_k^TG_uV_k\Gamma_Q^T] - h\text{tr}(G_u)\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] \right. \\ \left. + \frac{h}{2}\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] + \frac{h}{2}\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] \right] \tilde{U}^T + O(h^2), \quad (4.44a)$$

$$(MFGB) \quad \mathbb{E}[v_{R_{k+1}}v_{R_{k+1}}^T] = \tilde{V} \left(\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] + h\mathbb{E}[\Gamma_QV_k^TP_kv_{R_k}\tilde{v}_R^T] + h\mathbb{E}[\tilde{v}_Rv_{R_k}^TP_k^TV_k\Gamma_Q^T] \right. \\ \left. + h\mathbb{E}[\Gamma_QV_k^TG_uV_k\Gamma_Q^T] - h\text{tr}(G_u)\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] \right) \tilde{V}^T \\ + \frac{h}{2}(\tilde{V}\mathbb{E}[\tilde{v}_R\tilde{v}_R^T] + \mathbb{E}[\tilde{v}_R\tilde{v}_R^T]\tilde{V}^T) + O(h^2). \quad (4.44b)$$

Proof. The proof of this theorem is a straightforward but tedious extension of Theorem 4.5.

The detailed procedures are given in Appendix E. \square

Table 4.3: Analytical uncertainty propagation for gyroscope kinematics

1: procedure $\mathcal{MG}(t_{k+1}) = \text{ANALYTICAL PROPAGATION}(\mathcal{MG}(t_k), \Omega_k)$
2: Calculate $E[R_{k+1}]$ using (4.34).
3: Obtain $U_{k+1}, S_{k+1}, V_{k+1}$ according to the marginal MLE in Chapter 3.3 using $E[R_{k+1}]$.
4: Calculate the moments in Theorem 4.6.
5: Obtain $\mu_{k+1}, \Sigma_{k+1}, P_{k+1}$ according to Theorem 3.6 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

Note that v_{R_k} , \tilde{v}_R , $\tilde{\tilde{v}}_R$, and Γ_Q are linear in Q_k (see Lemma E.1). Therefore, the expectations on the right hand side of (4.34), (4.43) and (4.44) 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 $E[Q_{ij}]$, $E[Q_{ij}Q_{kl}]$, $E[Q_{ij}Q_{kl}Q_{mn}]$ for $i, j, k, l, m, n \in \{1, 2, 3\}$, which can be evaluated using the formulae given in Appendix B. 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 3.6. In summary, Theorem 4.5 and Theorem 4.6 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 accuracy $O(h^2)$ in moments. The pseudocode for this analytical uncertainty propagation scheme is shown in Table 4.3.

Next, an alternative sampling-based method is presented to propagate the uncertainty using unscented transform. 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 (4.29) and (4.30), 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. To do this, first a canonical form of MFG is given where the mean values are centered, and all the correlations are eliminated.

Lemma 4.4. *Let $(R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Define $Q = U^T R V \in \text{SO}(3)$, and $y = \Sigma_c^{-1/2}(x - \mu - Pv_R) \in \mathbb{R}^n$, then (Q, y) follows $\mathcal{MG}(0, I, 0, I, S, I)$.*

Proof. After change of variables, it can be easily seen that $\text{dyd}Q = \det(\Sigma_c^{-1/2})\text{dxd}R$. Then the probability measure of (Q, y) becomes

$$p(R, x)\text{dxd}R = \frac{1}{c(S)\sqrt{(2\pi)^n}} \exp\left\{-\frac{1}{2}y^T y\right\} \text{etr}(SQ^T) \text{dyd}Q$$

according to (3.1), which finishes the proof. \square

In the above canonical form, the Gaussian part is decoupled from the matrix Fisher part. As such, the sigma points of the canonical MFG are the union of the sigma points for the Gaussian distribution and those for the matrix Fisher distribution [64]. These can be transformed back to the original MFG according to the above theorem as follows.

Definition 4.3. Consider $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Define the $7 + 2n$ sigma points for its canonical distribution as

$$\begin{aligned} (Q, y)_{1,2} &= (\exp(\pm\theta_1\hat{e}_1), [0, \dots, 0]^T), \\ (Q, y)_{3,4} &= (\exp(\pm\theta_2\hat{e}_2), [0, \dots, 0]^T), \\ (Q, y)_{5,6} &= (\exp(\pm\theta_3\hat{e}_3), [0, \dots, 0]^T), \\ (Q, y)_{7,8} &= \left(I_{3 \times 3}, \left[\pm\sqrt{\frac{n}{w_G}}, 0, \dots, 0\right]^T\right), \\ &\vdots \\ (Q, y)_{5+2n, 6+2n} &= \left(I_{3 \times 3}, \left[0, \dots, 0, \pm\sqrt{\frac{n}{w_G}}\right]^T\right), \\ (Q, y)_{7+2n} &= (I_{3 \times 3}, [0, \dots, 0]^T), \end{aligned} \tag{4.45}$$

where $0 \leq \theta_i \leq \pi$ is chosen according to

$$\cos \theta_i = \begin{cases} \sigma + \frac{(1-\sigma)(\log c(S) - s_i)}{s_j + s_k}, & \text{if } s_j + s_k \geq 1, \end{cases} \tag{4.46a}$$

$$\left\{ \sigma + (1-\sigma)(\log c(S) - s_i) + \frac{1}{2} \right\} (s_j + s_k) - \frac{1}{2}, \quad \text{else if } 0 \leq s_j + s_k < 1, \tag{4.46b}$$

for $i, j, k \in \mathcal{I}$. The weights for the first three pairs of sigma points are given by

$$w_i = \frac{1}{4(1-\cos \theta_i)} \left\{ \frac{1}{c(S)} \left(\frac{\partial c(S)}{\partial s_i} - \frac{\partial c(S)}{\partial s_j} - \frac{\partial c(S)}{\partial s_k} \right) + 1 \right\} \quad (4.47)$$

for $i = 1, 2, 3$, where σ in (4.46) is chosen such that $2(w_1 + w_2 + w_3) = w_M$. The weights for the next $2n$ sigma points, namely from the 7-th through $(6+2n)$ -th sigma points are $\frac{w_G}{2n}$, and the weight for the last one is $w_0 = 1 - w_M - w_G$. Let $R_i = U Q_i V^T$, and $x_i = \Sigma_c^{1/2} y_i + \mu + P v_{R_i}$. The sigma points for $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$ are defined as $(R, x)_i$.

In other words, each pair of sigma points are designed to capture the dispersion along each principal axis of MFG. Parameters w_M , w_G and w_0 are used to adjust the weights, respectively for the attitude, the linear components and the sigma point at identity. This selection of sigma points is justified by the following theorem stating that the above sigma points recover the original MFG using the MLE of MFG.

Theorem 4.7. *The marginal-conditional MLE of the MFG is $(\mu, \Sigma, P, U, S, V)$ with the sigma points given in Definition 4.3.*

Proof. The marginal MLE is the same as the MLE of matrix Fisher distribution, and it has been proved in [64, Theorem IV.1] that the sigma points recovers U , S , and V . For the conditional MLE, it is easy to see that $\bar{E}[y] = 0$, $\bar{E}[v_R] = 0$, and $\bar{E}[yv_R^T] = 0$. Then, it can be shown that

$$\begin{aligned} \bar{\text{cov}}(x, v_R) &= P \bar{\text{cov}}(v_R, v_R) \\ \bar{E}[x] &= \mu \\ \bar{\text{cov}}(x, x) &= \Sigma_c + P \bar{\text{cov}}(v_R, v_R) P^T. \end{aligned}$$

According to Theorem 3.6, these prove the conditional MLE with the sigma points recovers the parameters P , μ , and Σ . \square

It should be noted that another unscented transform for the Bingham distribution has been proposed in [35], which can also be applied to the matrix Fisher distribution using the Lie group homomorphism (2.33). Its difference with the first six sigma points in (4.3) is that it does not require the six sigma points have equal density. A benefit of the unscented transform in [35] is that it only uses the first order moment D , so there is no need to solve (2.20) if S is unknown.

Given $(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$, $7 + 2n = 13$ sigma points are selected from the MFG as in Definition 4.3, together with 7 sigma points from the noise $H_u \Delta W_u$ in (4.31) according to the common unscented transform for a Gaussian distribution in \mathbb{R}^3 (for example, see [40, Chapter 9]). These sigma points are propagated to t_{k+1} through the discrete kinematics model (4.29) and (4.30) 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 the MLE introduced in Chapter 3.3. The effect of the noise term $H_v \Delta W_v$ driving the gyro bias in (4.30) is accounted by adding hG_v to the new covariance matrix Σ_{k+1} for x_{k+1} , according to the following Proposition 4.8. The pseudocode for this uncertainty propagation scheme is summarized in Table 4.4.

Theorem 4.8. 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

$$\begin{aligned} f_{R,y}(R, y) &= \int_{x \in \mathbb{R}^n} f_{R,x}(R, x) f_{x'}(y - x) dx \\ &= \frac{1}{c} \int_{x \in \mathbb{R}^n} \text{etr}(FR^T) \exp\left\{-\frac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\right\} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(y - x - \mu')^T (\Sigma')^{-1} (y - x - \mu')\right\} dx \\ &= \frac{1}{c'} \text{etr}(FR^T) \exp\left\{-\frac{1}{2}(y - \mu_c - \mu')^T (\Sigma_c + \Sigma')^{-1} (y - \mu_c - \mu')\right\}, \end{aligned}$$

where c, c' are some normalizing constants, and the last equality is from the addition of two

Table 4.4: Unscented uncertainty propagation for gyroscope kinematics

1:	procedure $\mathcal{MG}(t_{k+1}) = \text{UNSCENTED PROPAGATION}(\mathcal{MG}(t_k), \Omega_k)$
2:	Select sigma points and weights $\{(R, x, w)_i\}_{i=1}^{13}$ from $\mathcal{MG}(t_k)$.
3:	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 [40].
4:	Propagate the sigma points through (4.29) and (4.30) without the noise $H_v \Delta W_v$, i.e.,
	$R_{i,j} = R_i \exp(h(\hat{\Omega}_k + \hat{x}_i) + (H_u \Delta W_u)_j^\wedge), \quad x_{i,j} = x_i,$
	and calculate the weights as $w_{i,j} = w_i w_j$.
5:	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 the MLE of MFG in Chapter 3.3.
6:	Let $\Sigma_{k+1} = \Sigma_{k+1} + hG_v$
7:	Set $\mathcal{MG}(t_{k+1}) = \mathcal{MG}(\mu_{k+1}, \Sigma_{k+1}, P_{k+1}, U_{k+1}, S_{k+1}, V_{k+1})$.
8:	end procedure

independent Gaussian random vectors. Comparing the above equation with (3.1) yields the desired result. \square

4.2.3 Measurement Update

After the uncertainty has been propagated from $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$ to $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$ using the techniques introduced in the previous subsection, this subsection develops an algorithm to update the propagated MFG with the new measurement \mathcal{Z}_{k+1} using Bayes' formula. As the measurement update is assumed to be completed instantaneously at time t_{k+1} , the subscript $k+1$ is omitted throughout this subsection. The variables relevant to the posterior distribution conditioned by measurements are denoted by the superscript $+$.

Suppose the prior distribution of (R, x) before measurement update follows MFG with parameters $(\mu, \Sigma, P, U, S, V)$. By the Bayes' rule and Theorem 3.2 in [64], the posterior

density conditioned on all of the available measurements \mathcal{Z} is

$$p(R, x | \mathcal{Z}) \propto \text{etr} \left(\left(F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T \right) R^T \right) \\ \cdot \exp \left\{ -\frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \right\}, \quad (4.48)$$

where F , μ_c and Σ_c are defined as in Definition 3.1 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. Similar to the previous subsection, a new MFG with parameters $(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ is matched to this density through MLE after calculating the required moments.

Theorem 4.9. Define $F^+ \in \mathbb{R}^{3 \times 3}$ as

$$F^+ = F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T \quad (4.49)$$

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

$$(MFGI) \quad v_R^+ = (Q^+ S^+ - S^+ (Q^+)^T)^\vee \quad (4.50a)$$

$$(MFGB) \quad v_R^+ = (S^+ Q^+ - (Q^+)^T S^+)^{\vee} \quad (4.50b)$$

for $Q^+ = (U^+)^T R V^+ \in \text{SO}(3)$. Then the moments of the posterior density (4.48), namely $E[R | \mathcal{Z}]$, $E[v_R^+ | \mathcal{Z}]$ and $E[v_R^+ (v_R^+)^T | \mathcal{Z}]$ are identical to their counterparts in Theorem 3.5 after replacing U, S, V with U^+, S^+, V^+ , and

$$E[x | \mathcal{Z}] = \mu + P E[v_R | \mathcal{Z}], \quad (4.51)$$

$$E[xx^T | \mathcal{Z}] = \mu\mu^T + \mu E[v_R | \mathcal{Z}]^T P^T + P E[v_R | \mathcal{Z}] \mu^T + P E[v_R v_R^T | \mathcal{Z}] P^T + \Sigma_c, \quad (4.52)$$

$$E[x(v_R^+)^T | \mathcal{Z}] = P E[v_R (v_R^+)^T | \mathcal{Z}], \quad (4.53)$$

where

$$(MFGI) \quad E[v_R | \mathcal{Z}] = \tilde{U}(E[Q^+ | \mathcal{Z}] \tilde{S}^T - \tilde{S}E[Q^+ | \mathcal{Z}]^T)^\vee, \quad (4.54a)$$

$$(MFGB) \quad E[v_R | \mathcal{Z}] = \tilde{V}(\tilde{S}^T E[Q^+ | \mathcal{Z}] - E[Q^+ | \mathcal{Z}]^T \tilde{S})^\vee, \quad (4.54b)$$

$$(MFGI) \quad E[v_R v_R^T | \mathcal{Z}] = \tilde{U}E[\tilde{v}_R^+(\tilde{v}_R^+)^T | \mathcal{Z}] \tilde{U}^T, \quad (4.55a)$$

$$(MFGB) \quad E[v_R v_R^T | \mathcal{Z}] = \tilde{V}E[\tilde{v}_R^+(\tilde{v}_R^+)^T | \mathcal{Z}] \tilde{V}^T, \quad (4.55b)$$

$$(MFGI) \quad E[v_R(Q_R^+)^T | \mathcal{Z}] = \tilde{U}E[\tilde{v}_R^+(v_R^+)^T | \mathcal{Z}], \quad (4.56a)$$

$$(MFGB) \quad E[v_R(Q_R^+)^T | \mathcal{Z}] = \tilde{V}E[\tilde{v}_R^+(v_R^+)^T | \mathcal{Z}], \quad (4.56b)$$

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

$$(MFGI) \quad \tilde{v}_R^+ = (Q^+ \tilde{S}^T - \tilde{S}(Q^+)^T)^\vee, \quad (4.57a)$$

$$(MFGB) \quad \tilde{v}_R^+ = (\tilde{S}^T Q^+ - (Q^+)^T \tilde{S})^\vee. \quad (4.57b)$$

Proof. The expressions for $E[R | \mathcal{Z}]$, $E[v_R^+ | \mathcal{Z}]$, $E[v_R^+(v_R^+)^T | \mathcal{Z}]$, and (4.51)-(4.53) can be obtained by integrating these variables with respect to the density (4.48) similarly as in the proof of Theorem 3.5. Since $Q = U^T R V = \tilde{U} Q^+ \tilde{V}^T$, for MFGB, it can be shown that

$$v_R = (S \tilde{U} Q^+ \tilde{V}^T - \tilde{V}(Q^+)^T \tilde{U}^T S)^\vee = \tilde{V}(\tilde{S}^T Q^+ - (Q^+)^T \tilde{S})^\vee.$$

And for MFGI, the equation becomes

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

from which (4.54a)-(4.56b) follow. \square

Note that $E[\tilde{v}_R^+(\tilde{v}_R^+)^T | \mathcal{Z}]$ and $E[\tilde{v}_R^+(v_R^+)^T | \mathcal{Z}]$ can be expressed as linear combinations of $E[Q_{ij}^+ Q_{kl}^+ | \mathcal{Z}]$ for $i, j, k, l \in \{1, 2, 3\}$, therefore they can be calculated using the second

Table 4.5: Bayesian estimation for attitude and gyroscope bias

```

1: procedure ESTIMATION( $\mathcal{MG}(t_0), \Omega(t), Z(t), z(t)$ )
2:   Let  $k = 0$ .
3:   repeat
4:     Either  $\mathcal{MG}(t_{k+1}) = \text{Analytical Propagation}(\mathcal{MG}(t_k), \Omega(t_k))$  in Table 4.3 or
       $\mathcal{MG}(t_{k+1}) = \text{Unscented Propagation}(\mathcal{MG}(t_k), \Omega(t_k))$  in Table 4.4.
5:      $k = k + 1$ .
6:   until  $Z(t_{k+1})$  or  $z(t_{k+1})$  ia available
7:    $\mathcal{MG}(t_{k+1}) = \text{Measurement Update}(\mathcal{MG}(t_{k+1}), Z(t_{k+1}), z(t_{k+1}))$ .
8:   Obtain the estimates as  $R(t_{k+1}) = U_{k+1}V_{k+1}^T$ ,  $x(t_{k+1}) = \mu_{k+1}$  for  $\mathcal{MG}(t_{k+1})$ .
9:   go to step 3.
10:  end procedure

11: procedure  $\mathcal{MG}^+ = \text{MEASUREMENT UPDATE}(\mathcal{MG}^-, Z, z)$ 
12:   Compute  $F^+$  from (4.49), and calculate its proper SVD as  $U^+S^+(V^+)^T = F^+$ .
13:   Calculate the moments of the posterior density in Theorem 4.9.
14:   Obtain  $\mu^+, \Sigma^+, P^+$  according to Theorem 3.6.
15:   Set  $\mathcal{MG}^+ = \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ 
16: end procedure

```

order moments of the matrix Fisher distribution $Q^+ | \mathcal{Z} \sim \mathcal{M}(S^+)$ given in Appendix B.

Since the attitude part of (4.48) 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 3.6 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 uncertainty represented by MFG can be propagated until an additional measurement is available, based on which the propagated uncertainty 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 4.5. A set of MATLAB codes for the proposed MFG and estimators are available at [117].

4.2.4 Numerical Simulations

In this section, the proposed Bayesian filter based on MFG is compared with the well-established MEKF through simulation studies. Three simulation scenarios are tested (i) the attitude is directly measured; (ii) two inertial reference directions are measured with various accuracy; (iii) the magnetic direction is measured by a satellite orbiting the Earth.

Attitude Measurement In the first benchmark, a rotational motion of a rigid body is considered where the three Euler angles (body-fixed 3-2-1) follow sinusoidal waves with the frequency at 0.35 Hz, and the amplitudes of π , $\pi/2$ and π , respectively. The corresponding average angular speed is 6.17 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 $\sigma_v = 500 \text{ deg/h}/\sqrt{s}$, i.e., $H_v = \sigma_v I_{3 \times 3}$ in (4.27). The white noise of angular velocity is Gaussian (angle random walk) with the isotropic strength $\sigma_u = 10 \text{ deg}/\sqrt{s}$, i.e., $H_u = \sigma_u I_{3 \times 3}$ in (4.28). These two values are greater than those of typical gyroscopes, but they are selected to generate large uncertainties favored by the MFG distribution.

The attitude measurement model for the proposed Bayesian estimator with MFG is given by

$$Z = R_t \delta R, \quad (4.58)$$

where $R_t \in \text{SO}(3)$ is the true attitude, and $\delta R \in \text{SO}(3)$ follows a matrix Fisher distribution with the parameter $S_m = s_m I_{3 \times 3}$. This corresponds to the attitude measurement model introduced in Section 4.2.3 without any constant bias. The measurement equation for MEKF is not written as (4.58). Instead, it is given by

$$Z = R_t \exp(\delta \hat{\theta}), \quad (4.59)$$

where $\delta \theta \in \mathbb{R}^3$ and $\delta \theta \sim \mathcal{N}(0, \sigma_m I_{3 \times 3})$. In order to fairly compare the proposed estimator

based on MFG with MEKF, each estimator is simulated with both of (4.58) and (4.59) after conversion. For example, when (4.58) is used in MEKF, $\log(\delta R)^\vee$, which corresponds to $\delta\theta$ in (4.59), is fitted into a Gaussian distribution. Conversely, if (4.59) is used in the MFG filter, $\exp(\delta\hat{\theta})$, which corresponds to δR in (4.58), is fitted to a matrix Fisher distribution.

Three different levels of concentration are tested: (i) $s_m = 2.4$, $\sigma_m = 0.5$; (ii) $s_m = 12$, $\sigma_m = 0.2$; (iii) $s_m = 200$, $\sigma_m = 0.05$. Therefore, case (i) has the greatest measurement noise among the three cases considered, and case (iii) has the lowest measurement noise. The initial distribution for attitude is the same as Z in (4.58) (or (4.59) for MEKF), the initial distribution for bias is $x_0 \sim \mathcal{N}(0, 0.1^2 I_{3 \times 3})$, and they are independent. The initial attitude and bias estimates are drawn randomly from these distributions.

The sampling frequencies for the gyroscope and the attitude measurements are 150 Hz, and 30 Hz, respectively. Only the MFGI definition is used in the test, however, as the noise and initial conditions are isotropic, the filter based on MFGB is expected to behave very similarly to MFGI. To test the estimators in a statistical sense, under each condition, sixty Monte Carlo simulations were conducted with the simulation time of 60 seconds. The attitude and bias errors are averaged in each simulation, and further they are averaged over sixty simulations. Paired t -tests ($N = 60$, $\alpha = 0.05$) were conducted between the two types of the proposed estimators based on MFG and MEKF. Any significant statistical difference is indicated by $p < 0.05$.

The results for this simulation study are summarized in Table 4.6. It is shown that the accuracy for MFG based filters and MEKF are almost identical, with the differences in the attitude errors and the bias errors appearing mostly in the 3rd to 5th significant digits, if there is any. The statistical differences found through t -tests are mainly caused by different measurement types, i.e., if δR is used, MFG based filters are usually better than MEKF; conversely if $\delta\theta$ is used, MEKF is usually better. This is natural since the measurement model for MFG based filters is formulated with δR , whereas that for MEKF is formulated with $\delta\theta$ [69], and there are some errors in matching a distribution of δR to that of $\delta\theta$, and

Table 4.6: Attitude (deg) and bias (deg/s) errors \pm S.D. with attitude measurements

Measurement		MFGI Analytical (p)	MFGI Unscented (p)	MEKF
(4.59) $\sigma_m=0.5$	attitude	11.94 \pm 0.40 (<0.001)	11.94 \pm 0.40 (<0.001)	11.88 \pm 0.41
	bias	4.6 \pm 1.4 (0.81)	4.6 \pm 1.4 (0.69)	4.6 \pm 1.5
(4.58) $s_m=2.4$	attitude	12.00 \pm 0.49 (<0.001)	12.00 \pm 0.49 (<0.001)	12.13 \pm 0.48
	bias	5.0 \pm 1.7 (0.46)	4.9 \pm 1.7 (0.42)	5.0 \pm 1.8
(4.59) $\sigma_m=0.2$	attitude	7.285 \pm 0.151 (<0.001)	7.285 \pm 0.151 (<0.001)	7.284 \pm 0.152
	bias	3.79 \pm 1.11 (0.41)	3.79 \pm 1.12 (0.47)	3.79 \pm 1.13
(4.58) $s_m=12$	attitude	7.449 \pm 0.182 (0.009)	7.449 \pm 0.182 (0.015)	7.450 \pm 0.181
	bias	3.89 \pm 1.14 (0.09)	3.89 \pm 1.14 (0.033)	3.90 \pm 1.15
(4.59) $\sigma_m=0.05$	attitude	3.54737 \pm 0.04250 (0.64)	3.54741 \pm 0.04249 (0.009)	3.54736 \pm 0.04251
	bias	3.356 \pm 0.900 (0.97)	3.357 \pm 0.903 (0.88)	3.356 \pm 0.910
(4.58) $s_m=200$	attitude	3.55406 \pm 0.04607 (0.21)	3.55414 \pm 0.04606 (0.035)	3.55409 \pm 0.04605
	bias	3.722 \pm 0.939 (0.053)	3.726 \pm 0.940 (0.031)	3.733 \pm 0.940

vice versa. The only exception appears in $s_m = 200$, where though δR is used, the attitude error of MFG unscented filter is still greater than MEKF with the statistical significance of $p = 0.035 < 0.05$. Also, for $s_m = 12$ and $s_m = 200$, the bias error for MFG unscented filter is significantly smaller than MEKF. But these differences are obviously negligible in practice.

Two Direction Measurements In this benchmark, the reference rotational motion of the rigid body and the gyroscope noise settings are chosen the same as the first test with attitude measurements. Two reference directions in the inertial frame are assumed to be measured, given by

$$z_i = R_t^T a_i + v_i \quad (4.60)$$

for $i \in \{1, 2\}$, and they are normalized to have unit lengths. In the above equation, the reference directions 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 $v_1 \sim \mathcal{N}(0, 0.01I_{3 \times 3})$, $v_2 \sim \mathcal{N}(0, \sigma_2^2 I_{3 \times 3})$. 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 (4.49) with $B_1 = B_2 = I_{3 \times 3}$. The gyroscope measurements are sampled at 150 Hz, and the two direction measurements are sampled at

Table 4.7: Attitude (deg) and bias (deg/s) errors (S.D.) for different σ_2

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

30 Hz. The simulation period is 300 s.

It is seen that in the first test with attitude measurements, the MFG filter and MEKF do not show too many differences in estimation accuracy. This is because the attitude uncertainty is still not large enough to distinguish them. In this benchmark, the two estimators are challenged with large uncertainty through two aspects. First, the initial attitude R_0 is set as the true attitude rotated about its first body-fixed axis by 180° , which is completely wrong. The initial attitude uncertainty is set as $\delta R \sim \mathcal{N}(0, 10^{10} I_{3 \times 3})$ 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 is set as $x_0 = [0.2, 0.2, 0.2]^T$ rad/s with the uncertainty $0.1^2 I_{3 \times 3}$ given as the covariance, and the correlation between attitude and bias is zero.

Second, the accuracy σ_2 of the second direction measurement is varied as $\sigma_2^2 \in \{0.01, 0.1, 1, 5, 10, 50, 200, \infty\}$. The first case of $\sigma_2^2 = 0.01$ corresponds to when the second

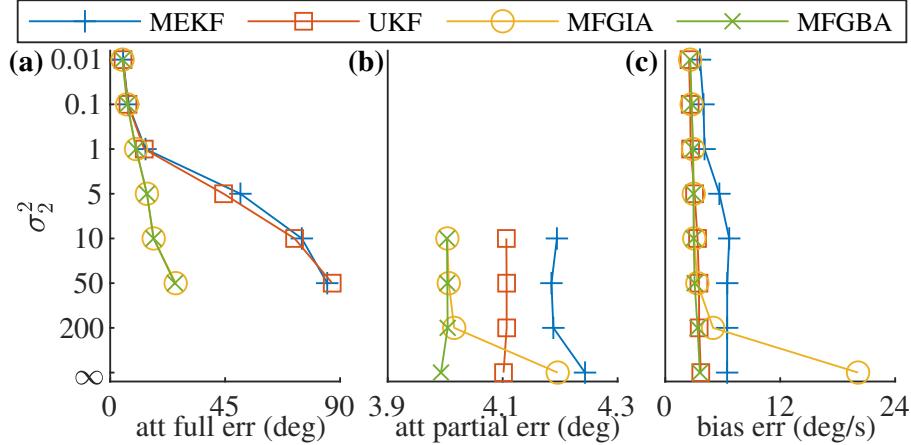


Figure 4.10: Estimation errors for varying accuracies of the second vector measurement: (a) full attitude error; (b) partial attitude error; (c) bias error for MEKF, UKF, MFGIA and MFGBA with varying σ_2^2 . The errors of the unscented MFG filters are similar with the analytical MFG filters, and they are omitted for readability.

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 large uncertainty of σ_2 could lead to large uncertainty of the rotation around the first reference direction, and when $\sigma_2^2 = \infty$, this rotation becomes completely unobservable.

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). 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.

The *full* attitude error is defined as the angle between the true attitude and the estimated attitude. As the second measurement becomes more inaccurate, the full attitude cannot be completely estimated because the rotation around the first reference vector becomes unobservable. Thus, the *partial* attitude error is defined as the angle between $R_t^T a_1$ and $R^T a_1$,

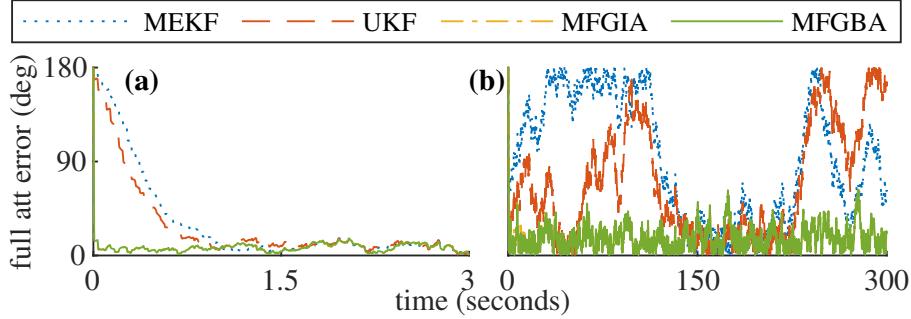


Figure 4.11: (a) Full attitude error of MEKF, UKF, MFGIA and MFGBA for an example simulation ($\sigma_2^2 = 0.1$) during the first three seconds. (b) Full attitude error (0 s to 300 s) for an example simulation ($\sigma_2^2 = 10$). The error of the unscented MFG filters is similar to the analytical MFG filters.

where R_t and R are the true attitude and the estimated attitude, respectively. The partial attitude error only captures the accuracy along the first reference vector. The simulation results are summarized in Table 4.7 and Fig. 4.10.

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. 4.11(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 than 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 filters is much lower than MEKF and UKF (Fig 4.11(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. 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

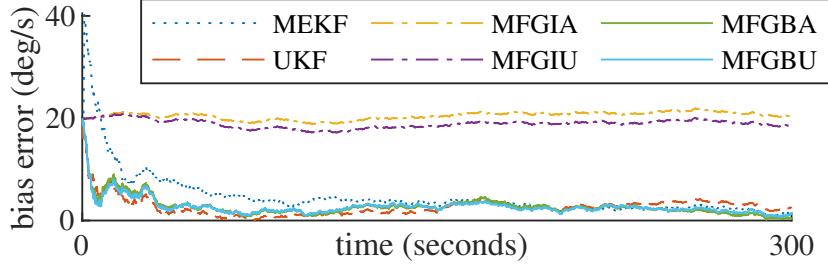


Figure 4.12: Bias error (0 s to 300 s) in an example simulation ($\sigma_2^2 = \infty$).

and bias errors, although their relative difference is still very small. This difference 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 Chapter 3.1.3 begins to emerge. Because the gyroscope bias is resolved in the body-fixed frame, MFGB is more appropriate to model the correlation between attitude and bias than MFGI, this explains why the MFGB filter is more accurate.

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. 4.12, the bias error of an example 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.

Magnetometer Measurement In this benchmark, a more realistic scenario is simulated where a satellite is orbiting around Earth with a single magnetometer and a gyroscope. The orbital elements are given as follows, and three cases are considered for varying inclinations.

$$a = 6916\text{km}, \quad e = 0.02, \quad \Omega = 50^\circ, \quad \omega = 0^\circ, \quad i = 0^\circ, 40^\circ \text{ or } 90^\circ.$$

According to the solution of the two-body problem, the orbital position with respect to the ECI (Earth-centered inertial) frame, namely $r_{ECI} \in \mathbb{R}^3$, is computed at each time step.

A 3U CubeSat with the mass of $m = 3\text{ kg}$ is considered. The inertia matrix is given by

$$J = \text{diag}[0.0325, 0.0325, 0.005]\text{kgm}^2.$$

The true attitude and the true angular velocity, namely (R_{ref}, Ω_{ref}) are generated according to the Euler's equation of motion

$$J\dot{\Omega}_{ref} + \Omega_{ref} \times J\Omega_{ref} = 0,$$

$$\dot{R}_{ref} = R_{ref}\hat{\Omega}_{ref},$$

where the initial condition is chosen as

$$R_{ref}(0) = I_{3 \times 3}, \quad \Omega_{ref}(0) = [0.05, 0.05, 0.02]\text{ rad s}^{-1}.$$

These are numerically integrated with the geometric numerical integrator preserving the structures of $\text{SO}(3)$, referred to as the Lie group variational integrator [66, 67], using the time step of 0.05 s.

It is assumed that the angular velocity measurement is measured by a gyroscope at 20 Hz, i.e., $h = 0.05$ s. The noise parameters for the gyroscope are: (i) angle random walk: $H_u = 0.1I_{3 \times 3}$, and (ii) bias instability: $H_v = 0.0005I_{3 \times 3}$, as indicated in (4.27) and (4.28). The initial bias of the gyroscope is set as $[0.01, 0.01, 0.01]\text{ rad s}^{-1}$.

Next, the magnetometer measurements are generated as follows. Assuming that the simulation begins at the midnight of January 1, 2021, the orbital position in the ECI frame is transformed into the Earth-centered, Earth-fixed (ECEF) frame. This is because the Earth magnetic field is fixed to the ECEF frame. For the given location with respect to the ECEF frame, the MATLAB function `wrldmagm` is called to compute the magnetic field.

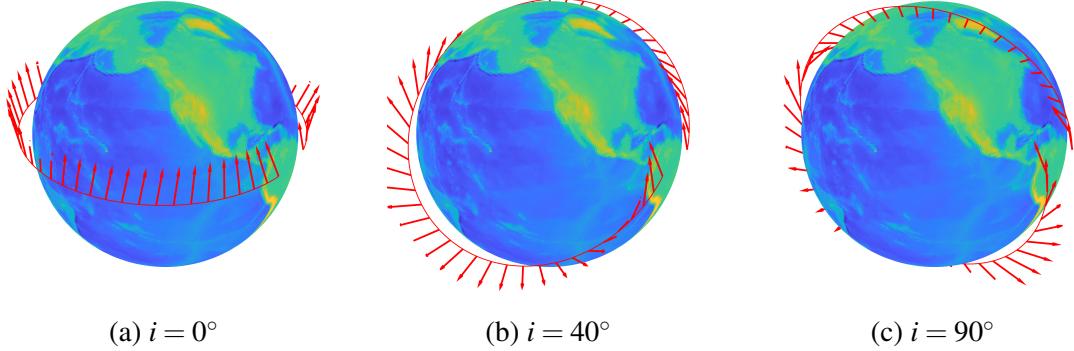


Figure 4.13: Magnetic field on the orbit with different inclinations.

This function implements the NOAA World Magnetic Model presented in [20]. Then, the magnetic field is normalized to a unit-vector, which is transformed back to the ECI frame to obtain the true direction of the magnetic field at $t = t_k$, namely a in (4.49). Figure 4.13 illustrates the NOAA magnetic field model, where the magnetic field along the selected orbits with respect to the ECEF frame is shown.

The measurement from the magnetometer is assumed to be distributed according to the von Mises–Fisher distribution centered at the true magnetic field resolved in the body-fixed frame, namely $R^T b \in \mathbb{S}^2$. The concentration parameter κ is chosen as $\kappa = 200$. The average attitude error of the noisy magnetometer measurement is 5.08° , and it is assumed to be available also at 20 Hz.

Five filters are simulated to estimate the attitude and gyroscope bias of the spacecraft from gyroscope and magnetometer measurements, namely MEKF, two MFGB filters (MFGBA and MFGBU), and two MFGI filters (MFGIA and MFGIU). The filters are initialized by

$$U_0 = \text{diag}[1, -1, -1] \quad S_0 = 10I_{3 \times 3} \quad V_0 = I_{3 \times 3}$$

$$\mu_0 = 0_{3 \times 1} \quad \Sigma_0 = 0.01^2 I_{3 \times 3} \quad P_0 = 0_{3 \times 3}.$$

The initial attitude estimation is the true attitude rotated by 180° about its first body-fixed

Table 4.8: Estimation errors with magnetometer measurements

		MEKF	MFGBA	MFGBU	MFGIA	MFGIU
$i = 0^\circ$	Attitude full error (deg)	82.0	29.4	63.4	90.1	87.0
	Attitude partial error (deg)	2.019	2.017	2.018	2.015	2.020
	Bias error (deg/s)	0.313	0.044	0.045	0.119	1.366
$i = 40^\circ$	Attitude full error (deg)	85.0	8.7	25.7	22.0	82.9
	Attitude partial error (deg)	2.018	2.016	2.019	2.015	2.020
	Bias error (deg/s)	0.321	0.048	0.051	0.176	1.324
$i = 90^\circ$	Attitude full error (deg)	82.8	7.2	20.5	22.8	76.8
	Attitude partial error (deg)	2.018	2.016	2.019	2.014	2.020
	Bias error (deg/s)	0.298	0.053	0.060	0.252	1.178

axis, and the initial bias estimation is off by 0.01 rad s^{-1} in each axis.

Since there is only one direction measurement available from the magnetometer, the rotation of spacecraft about this reference vector is unobservable if the magnetic field is fixed. Fortunately, as the spacecraft circles around Earth, the direction of the magnetic field changes at different locations (see Figure 4.13), and this allows the full attitude to be estimated [37]. However, since the magnetic field changes very slowly, estimating full attitude still remains very challenging. Therefore, similar to the previous test, two attitude errors are defined: the full attitude error (FAE) at time t_k is the angle that rotates $(R_{\text{ref}})_k$ to $(R_{\text{est}})_k$, and the partial attitude error (PAE) is the angle between $(R_{\text{ref}})_k^T b_k$ and $(R_{\text{est}})_k^T b_k$. Note that the partial attitude error does not take into account the rotation about the reference vector b_k .

The estimation errors for the five filters are presented in Table 4.8 and Figure 4.14 to Figure 4.16. In addition, the estimated uncertainties in attitude and bias are presented in Figure 4.14 to Figure 4.16. Looking at the bias estimation, the two MFGB based filters are much more accurate than MEKF, though their estimated uncertainties are similar. Because the bias is represented in the body-fixed frame, MFGB is more appropriate than MFGI to model the attitude-bias correlation. Therefore, MFGB based filters behave better than MFGI based filters, and this can also be seen from that the MFGB filters have lower bias uncertainty. In particular, the MFGIU filter has the worst performance in bias estimation.

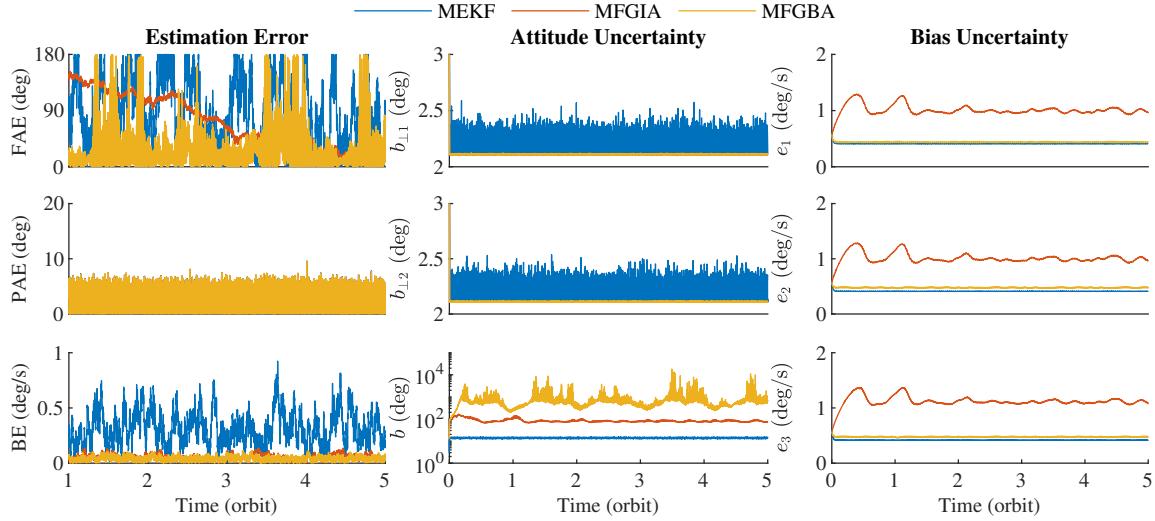


Figure 4.14: Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 0^\circ$. The attitude uncertainty is expressed in the coordinate frame specified by the direction of magnetic field (b), and its two perpendicular vectors ($b_{\perp 1}$ and $b_{\perp 2}$). The attitude uncertainty for MEKF is given by the square root of the diagonals of its covariance; and that for MFG filters is given by the square root of the diagonals of $(\text{tr}(S)I_{3 \times 3} - S)^{-1}$, after transformed into the correct frame. The bias uncertainty is expressed in the body-fixed frame, and is given by the square root of the diagonals of the covariance. The results for MFG unscented filters are omitted for better readability.

The partial attitude error is directly related to the magnetometer measurement, and is corrected in every time step. Hence, it is relatively easy to estimate the partial attitude which is not involved with the rotation about the reference vector. And as a result, all filters tested are able to estimate partial attitude with equally low estimation error. Furthermore, the attitude uncertainties in the two axes perpendicular to the reference magnetic field are consistently low for all filters, though there appears to be more fluctuations for MEKF. Also, the partial attitude error does not depend on the inclination angle i of the orbit.

The full attitude is extremely challenging to estimate due to that the direction of magnetic field b changes very slowly. Among all tested filters, only the MFGB based filters (and in particular MFGBA) are able to give relatively low full attitude error. Compared with the results in [65] where the full attitude can be accurately estimated with a gyroscope and a single magnetometer, the difference here is the inclusion of a non-zero time varying gyroscope bias. Because there is only one slow-varying reference vector, the integration

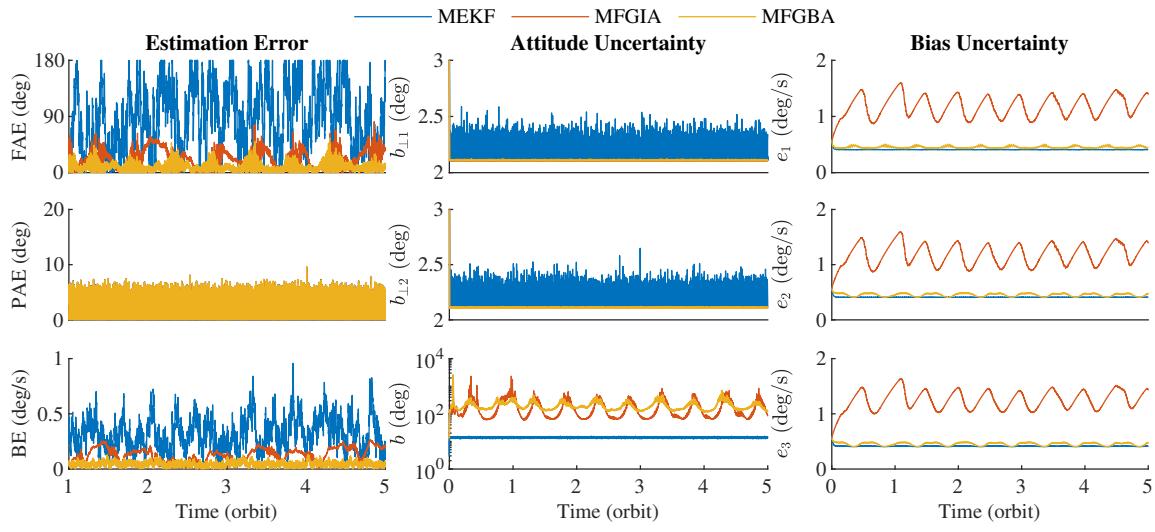


Figure 4.15: Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 40^\circ$.

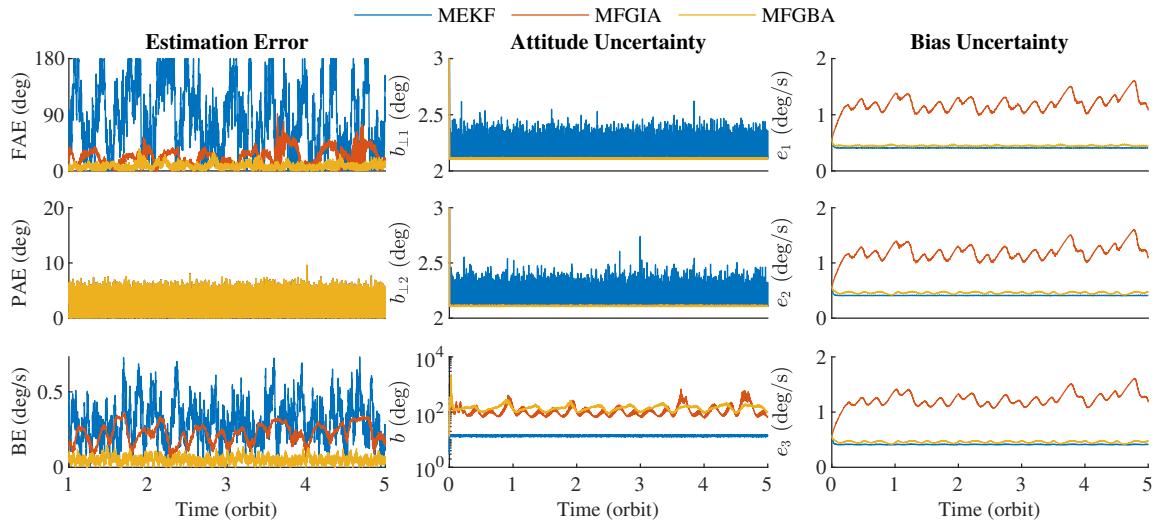


Figure 4.16: Estimation error, attitude uncertainty, and bias uncertainty for various attitude filters when $i = 90^\circ$.

accuracy of angular velocity from gyroscope is of crucial importance for full attitude estimation. However, the inclusion of a time varying gyroscope bias deteriorates the integration accuracy very badly. And therefore, the estimation accuracy of bias has a strong impact on the estimation of full attitude, and this explains why the MFGB based filters have lower full attitude error compared with other filters. It is surprising that MEKF has the lowest attitude uncertainty in the reference magnetic field direction, nonetheless it is the least accurate in full attitude estimation. This may reflect that the Gaussian distribution assumed by MEKF cannot properly model the large attitude dispersion due to the wrapping problem. In addition, the analytical propagation for MFG filters behave consistently better than unscented propagation in full attitude estimation. Also, because the direction of b changes more quickly as the inclination angle i is increased, the full attitude error also becomes lower.

As a conclusion, in this section the MFG defined in Chapter 3 is used to design an attitude estimator that is able to concurrently estimate the attitude and gyroscope bias. Because MFG is defined intrinsically on $\text{SO}(3) \times \mathbb{R}^n$, it is able to handle large attitude uncertainty. Compared with the Gaussian based MEKF, the MFG filter does not show any advantage when the attitude uncertainty is small. However, for challenging cases with large attitude uncertainty, for example, when the initial attitude is completely wrong, the MFG filters exhibit much faster convergence rate than MEKF. Also, if there is a degree of freedom that is not properly observed, for example, when a direction measurement is very inaccurate, or when a single reference direction is slowly varying in the inertial frame, the MFG filters have better accuracy than MEKF. Due to that the gyroscope bias is resolved in the body-fixed frame, the MFGB definition more accurately models the attitude-bias correlation than the MFGI definition. The advantage of MFGB becomes particularly significant when the attitude uncertainty is highly non-isotropic, which is consistent with the discussion in Chapter 3.1.3.

Chapter 5: Bayesian Attitude and Position Estimation with Matrix

Fisher-Gaussian Distribution

In the previous chapter, the MFG is applied in the classic attitude estimation problem, where it is demonstrated through simulation that the filter designed using MFG has better accuracy than MEKF in challenging cases when the attitude has large uncertainty. In inertial navigation, the estimated attitude is used to transform the linear acceleration measured by an accelerometer into the inertial frame, which can be integrated twice into position after the gravity component is subtracted. In this chapter, the filter for attitude estimation is generalized into this inertial navigation setting using MFG. In particular, the linear part of MFG is augmented into \mathbb{R}^{12} , including the linear velocity, position, gyroscope and accelerometer biases.

In Chapter 5.1, the position measurement given by a GNSS receiver is considered in the measurement update step. Then in Chapter 5.2, it is replaced by camera measurements of the locations for unknown features or known landmarks. It is assumed that the camera is an RGB-D or stereo camera, so that projection measurements have already been triangulated into 3D locations. These filters are compared with variants of MEKF [112] through simulation studies, where it is shown that the filters based on MFG have faster convergence rate when the initial heading direction is completely unknown, which is frequently encountered in practical localization problems.

5.1 Loosely Coupled IMU-GNSS Integration

In this section, the attitude estimation problem considered in Chapter 4.2 is extended into a full inertial navigation problem. More specifically, it is assumed that an inertial measurement unit (IMU) is attached to a rigid body, which has a gyroscope measuring the angular velocity, and an accelerometer measuring the linear acceleration, both in the body

fixed frame. The linear acceleration can be transformed into the inertial frame using the integrated attitude, and then it can be integrated twice into position after subtracting the gravity component. It is also assumed that a global navigation satellite system (GNSS) receiver provides position measurements of the rigid body, which are used to correct the attitude and position integrated using the IMU. Since the pseudorange from GNSS is first converted into position measurements, and then used to correct the inertial navigation estimation, this is usually referred to as *loosely* coupled IMU-GNSS integration.

The MEKF for attitude estimation can be generalized to deal with this inertial navigation setting [112]. However, as discussed in previous sections, it cannot handle large attitude uncertainty due to its inherent Gaussian assumption. Similar as in attitude estimation, the MFG is used to design a new estimator, which is suitable for large attitude uncertainty. This is particularly advantageous in some practical problems, for example, when the IMU starts in a building, where the heading direction is usually completely unknown; or when the GNSS receiver cannot receive signal for a long period time, and the uncertainty for inertial navigation accumulates.

The definition of MFG is flexible enough to handle this inertial navigation setting, as its linear part can be arbitrary in \mathbb{R}^n . This makes MFG distinguished from the distribution defined for dual quaternions [71, 72], where the linear part must be position only. In this section, the MFG is used to design a Bayesian filter that estimates attitude, gyroscope bias, position, linear velocity, and accelerometer bias concurrently.

5.1.1 Problem Formulation

The IMU kinematic model is formulated in discrete time as follows.

$$R_{k+1} = R_k \exp \left\{ h(\hat{\Omega}_k + \hat{b}_{g,k}) + H_{gu} \Delta W_{gu} \right\}, \quad (5.1)$$

$$b_{g,k+1} = b_{g,k} + H_{gv} \Delta W_{gv}, \quad (5.2)$$

$$p_{k+1} = p_k + h v_k, \quad (5.3)$$

$$v_{k+1} = v_k - gh + R_k (h(a_k + b_{a,k}) + H_{au}\Delta W_{au}), \quad (5.4)$$

$$b_{a,k+1} = b_{a,k} + H_{av}\Delta W_{av}, \quad (5.5)$$

where $R \in \text{SO}(3)$ is the attitude, $x = [b_g^T, p^T, v^T, b_a^T]^T \in \mathbb{R}^{12}$ are gyroscope bias, position, linear velocity and accelerometer bias, respectively. The position and linear velocity are resolved in the inertial frame. The gravity vector in the inertial frame is denoted by $g \in \mathbb{R}^3$. Next, $\Omega \in \mathbb{R}^3$ is the angular velocity measured by the gyroscope, and $a \in \mathbb{R}^3$ is the linear acceleration measured by the accelerometer, and they are resolved in the body-fixed frame. The angular velocity measurement is affected by two noises: $H_{gu}\Delta W_{gu}$ is a white noise, and b_g is a random walk bias driven by $H_{gv}\Delta W_{gv}$. The noises for accelerometer are similar, given by the white noise $H_{au}\Delta W_{au}$, and the bias driven by $H_{av}\Delta W_{av}$. These noise terms are Gaussian with

$$H_{gu}\Delta W_{gu} \sim \mathcal{N}(0, hG_{gu}), \quad H_{gv}\Delta W_{gv} \sim \mathcal{N}(0, hG_{gv}), \quad (5.6)$$

$$H_{au}\Delta W_{au} \sim \mathcal{N}(0, hG_{au}), \quad H_{av}\Delta W_{av} \sim \mathcal{N}(0, hG_{av}), \quad (5.7)$$

where $G_{gu} = H_{gu}H_{gu}^T$, $G_{gv} = H_{gv}H_{gv}^T$, $G_{au} = H_{au}H_{au}^T$, $G_{av} = H_{av}H_{av}^T$, and they are independent. Finally, the subscript k denote the time step, and $h = t_{k+1} - t_k$ is the constant sampling period for any k .

It is assumed that the position p_k is measured by a GNSS receiver at time t_k , denote by $y_k \in \mathbb{R}^3$. The measurement noise is Gaussian, i.e.,

$$y_k | p_k = p_k + \mathcal{N}(0, \Sigma_y). \quad (5.8)$$

Optionally, the attitude and some reference directions can also be measured as formulated in Chapter 4.2.1. Namely, the attitude can be directly measured by N_a sensors as Z_i , and the measurement noises follow matrix Fisher distributions with parameters F_{Z_i} , for $i = 1, \dots, N_a$. Also, N_v fixed reference directions $a_j \in \mathbb{S}^2$ in inertial frame are measured as $z_j \in \mathbb{S}^2$ in the

body-fixed frame, and the measurement noises follow von Mises-Fisher distribution with the mean direction $R_t^T B_j a_j$, and concentration parameter κ , where $R_t \in \text{SO}(3)$ is the true attitude, for $j = 1, \dots, N_v$. All measurements at time t_k are denoted by \mathcal{Z}_k .

Suppose $(R_0, x_0) = (R_0, b_{g,0}, p_0, v_0, b_{a,0})$ at the initial time t_0 follow an MFG with $n = 12$ and parameters $(\mu_0, \Sigma_0, P_0, U_0, S_0, V_0)$ of appropriate dimensions. Given the IMU measurements $\{\Omega_k, a_k\}_{k=1}^\infty$, and all position, attitude, and direction measurements $\{\mathcal{Z}_k\}_{k=1}^\infty$, the problem is to find the posterior distribution of $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$, and approximate it to an MFG with parameters $(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$. Similar to the attitude estimator, two steps are needed. In the uncertainty propagation step, the posterior distribution $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$ is propagated using the IMU kinematic equations (5.1)-(5.5) to the next time step as $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$. Next, the propagated distribution is updated by the measurements \mathcal{Z}_{k+1} using Bayes' formula into $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^{k+1}$. In the next two subsections, algorithms for these two steps are developed.

5.1.2 Uncertainty Propagation

Suppose $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$. In this subsection, the moments of (R_{k+1}, x_{k+1}) required for the MLE of MFG are calculated using the IMU kinematics equations (5.1)-(5.5). One way of calculating these moments is to propagate deterministically sampled sigma points through the kinematic equations using the unscented transform of MFG, as introduced in Chapter 4.2.2. This approach can be easily generalized to the inertial navigation settings considered in this subsection, and will not be discussed here. Instead, this subsection calculates approximate analytical expressions of the required moments.

First, because (5.1) and (5.2) are exactly the same as the attitude kinematics considered in (4.29) and (4.30), the moment $E[R_{k+1}]$ can be calculated using Theorem 4.5. Then $E[R_{k+1}]$ is used in the marginal MLE of MFG to obtain the parameters U_{k+1} , S_{k+1} , and V_{k+1} . Next, define $Q_{k+1} = U_{k+1}^T R_{k+1} V_{k+1}$, and $v_{R_{k+1}} = (Q_{k+1} S_{k+1} - S_{k+1} Q_{k+1}^T)^\vee$ for MFGL, or $v_{R_{k+1}} = (S_{k+1} Q_{k+1} - Q_{k+1}^T S_{k+1})^\vee$ for MFGB as the intermediate parameters for the

MFG at time t_{k+1} . According to Theorem 3.6, $E[x_{k+1}]$, $E[x_{k+1}v_{R_{k+1}}^T]$, $E[x_{k+1}x_{k+1}^T]$, and $E[v_{R_{k+1}}v_{R_{k+1}}^T]$ need to be calculated. Note that $E[v_{R_{k+1}}v_{R_{k+1}}^T]$ is only involved with the gyroscope kinematics (5.1) and (5.2), so it can be calculated using (4.44) in Theorem 4.6. In the remainder of this subsection, the calculations for $E[x_{k+1}]$, $E[x_{k+1}v_{R_{k+1}}^T]$, and $E[x_{k+1}x_{k+1}^T]$ are developed.

In preparation for these calculations, the parameters μ_k , P_k and Σ_k are written for each component in block form as

$$\mu_k = \begin{bmatrix} \mu_{b_g,k} \\ \mu_{p,k} \\ \mu_{v,k} \\ \mu_{b_a,k} \end{bmatrix}, \quad P_k = \begin{bmatrix} P_{b_g,k} \\ P_{p,k} \\ P_{v,k} \\ P_{b_a,k} \end{bmatrix}, \quad \Sigma_k = \begin{bmatrix} \Sigma_{b_gb_g,k} & \Sigma_{b_gp,k} & \Sigma_{b_gv,k} & \Sigma_{b_gb_a,k} \\ \Sigma_{pb_g,k} & \Sigma_{pp,k} & \Sigma_{pv,k} & \Sigma_{pb_a,k} \\ \Sigma_{vb_g,k} & \Sigma_{vp,k} & \Sigma_{vv,k} & \Sigma_{vb_a,k} \\ \Sigma_{bab_g,k} & \Sigma_{ba_p,k} & \Sigma_{ba_v,k} & \Sigma_{bab_a,k} \end{bmatrix}. \quad (5.9)$$

With these, the calculation for $E[x_{k+1}]$ is presented in the next theorem.

Theorem 5.1. *The moment $E[x_{k+1}] = E[b_{g,k+1}^T, p_{k+1}^T, v_{k+1}^T, b_{a,k+1}^T]^T$ are given by*

$$E[b_{g,k+1}] = E[b_{g,k}] \quad (5.10)$$

$$E[p_{k+1}] = E[p_k] + hE[v_k] \quad (5.11)$$

$$E[v_{k+1}] = E[v_k] + hE[R_k](a_k + \mu_{b_a,k}) + hE[R_k P_{b_a,k} v_{R_k}] - gh, \quad (5.12)$$

$$E[b_{a,k+1}] = E[b_{a,k}] \quad (5.13)$$

where $v_{R_k} = (Q_k S_k - S_k Q_k^T)^\vee$ for MFGI, or $v_{R_k} = (S_k Q_k - Q_k^T S_k)^\vee$ for MFGB, with $Q_k = U_k^T R_k U_k$.

Proof. The first, second, and fourth equations are immediate from (5.2), (5.3), and (5.5), because ΔW_{gv} and ΔW_{av} have zeros means. For the third equation, a similar calculation as

Table 5.1: Uncertainty propagation for IMU kinematics

1: procedure $\mathcal{MG}(t_{k+1}) = \text{UNCERTAINTY PROPAGATION}(\mathcal{MG}(t_k), \Omega_k, a_k)$
2: Calculate $E[R_{k+1}]$ using (4.34) with μ_k, P_k, G_u replaced by $\mu_{b_g,k}, P_{b_g,k}$ in (5.9), and G_{gu} in (5.6), respectively.
3: Obtain $U_{k+1}, S_{k+1}, V_{k+1}$ according to Chapter 3.3 using $E[R_{k+1}]$.
4: Calculate $E[v_{R_{k+1}} v_{R_{k+1}}^T]$ using (4.44), calculate $E[x_{k+1}]$, $E[x_{k+1} v_{R_{k+1}}^T]$ and $E[x_{k+1} x_{k+1}^T]$ with Theorem 5.1, Theorem F.1, and Theorem F.2, respectively.
5: Obtain $\mu_{k+1}, \Sigma_{k+1}, P_{k+1}$ according to Theorem 3.6 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

(3.22) shows that

$$E[R_k b_{a,k}] = E[R_k] \mu_{b_a,k} + E[R_k P_{b_a,k} v_{R_k}],$$

and the desired result is proved. \square

The expressions for $E[x_{k+1} v_{R_{k+1}}^T]$ and $E[x_{k+1} x_{k+1}^T]$ are very tedious, and they are relegated to Appendix F. Note that these two moments, $E[v_{R_{k+1}} v_{R_{k+1}}^T]$, and $E[x_{k+1}]$ can be calculated using up to the third order moments of $Q_k \sim \mathcal{M}(S_k)$, with the accuracy of $O(h^2)$. Then using Theorem 3.6, the parameters μ_{k+1}, P_{k+1} and Σ_{k+1} can be obtained. In summary, the uncertainty $S(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$ at time t_k is propagated 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 accuracy $O(h^2)$ in moments.

The pseudocode for this uncertainty propagation scheme is summarized in Table 5.1.

5.1.3 Measurement Update

The algorithm for propagating the posterior distribution $(R_k, x_k) | \{\mathcal{Z}_m\}_{m=1}^k$ at time t_k into the prior distribution $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$ at time t_{k+1} has been introduced in the previous subsection. In this subsection, the prior distribution $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^k$ is updated into the posterior distribution $(R_{k+1}, x_{k+1}) | \{\mathcal{Z}_m\}_{m=1}^{k+1}$ at time t_{k+1} , with the measurements \mathcal{Z}_{k+1}

using Bayes' formula. Since the update is completed instantaneously at time t_{k+1} , the subscript $k + 1$ is omitted in this subsection, and the variables relevant to the posterior distribution are denoted by the superscript $+$.

Suppose the prior distribution of (R, x) before measurement update follows MFG with parameters $(\mu, \Sigma, P, U, S, V)$. By the Bayes' rule and Theorem 3.2 in [64], the posterior density conditioned on all of the available measurements \mathcal{Z} is

$$p(R, x | \mathcal{Z}) \propto \text{etr} \left(\left(F + \sum_{i=1}^{N_a} Z_i F_i^T + \sum_{j=1}^{N_v} \kappa_j B_j a_j z_j^T \right) R^T \right) \cdot \exp \left\{ -\frac{1}{2} (x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c) \right\} \cdot \exp \left\{ -\frac{1}{2} (Hx - y)^T \Sigma_y^{-1} (Hx - y) \right\}, \quad (5.14)$$

where $H = [0_{3 \times 3}, I_{3 \times 3}, 0_{3 \times 6}]$ selects the position component from x , and F , μ_c and Σ_c are defined as in Definition 3.1 with respect to $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Compared with (4.48) in attitude estimation, (5.14) has an additional term contributed by the position measurement coming from the GNSS receiver.

To deal with this posterior density, it is first shown that (5.14) can be factorized into the marginal component for R , and the conditional component for $x|R$.

Lemma 5.1. *Let \tilde{F} be defined as the F^+ in (4.49), or $\tilde{F} = F$ if there are no attitude or direction measurements. Also, define the “Kalman gain”*

$$K = \Sigma_c H^T (H \Sigma_c H^T + \Sigma_y)^{-1} \in \mathbb{R}^{12 \times 3}. \quad (5.15)$$

Then the posterior density (5.14) can be factorized into two parts:

$$p(R, x | \mathcal{Z}) = p_R(R) p_{x|R}(x | R), \quad (5.16)$$

where the marginal part $p_R(R)$ is given by

$$p_R(R) \propto \text{etr}(\tilde{F}R^T) \exp\left\{-\frac{1}{2}(H\mu_c - y)^T(\Sigma_y + H\Sigma_c H^T)^{-1}(H\mu_c - y)\right\}, \quad (5.17)$$

and the conditional part $f(x|R)$ is given by

$$p_{x|R}(x|R) \propto \exp\left\{-\frac{1}{2}(x - (I - KH)\mu_c - Ky)^T((I - KH)\Sigma_c)^{-1}(x - (I - KH)\mu_c - Ky)\right\}, \quad (5.18)$$

where I is an abbreviation of $I_{12 \times 12}$.

Proof. Using a similar technique that is used to prove the linear Kalman filter, and the Woodbury matrix identity [97], it can be shown that

$$\begin{aligned} & \exp\left\{-\frac{1}{2}(x - \mu_c)^T \Sigma_c^{-1}(x - \mu_c)\right\} \cdot \exp\left\{-\frac{1}{2}(Hx - y)^T \Sigma_y^{-1}(Hx - y)\right\} \\ &= \exp\left\{-\frac{1}{2}(x - (I - KH)\mu_c - Ky)^T((I - KH)\Sigma_c)^{-1}(x - (I - KH)\mu_c - Ky)\right\} \\ & \quad \cdot \exp\left\{-\frac{1}{2}(H\mu_c - y)^T(\Sigma_y + H\Sigma_c H^T)^{-1}(H\mu_c - y)\right\}, \end{aligned}$$

which gives the desired (5.16). \square

Note that the marginal part $p_R(R)$ does not depend on x , so according to the marginal-conditional MLE of MFG, the first step is to approximate $p_R(R)$ into the density function of a matrix Fisher distribution. Since the μ_c is linear in R , $p_R(R)$ has quadratic terms of R in its exponent, making the expectation $E[R]$ corresponding to $p_R(R)$ impractical to be calculated directly from the density function. Instead, $E[R]$ is approximated using the unscented transform of matrix Fisher distribution [35, 64]¹. The idea is that the weighted sigma points $\{R_i, w_i\}_{i=1}^7$ are selected from $\mathcal{M}(\tilde{F})$, corresponding to the first part of (5.17).

¹The unscented transform in [35] is developed for the Bingham distribution, and it can be converted to the matrix Fisher distribution using Theorem 2.2

The weights are reweighed by the second part of (5.17) as

$$w_i^+ = w_i \exp \left\{ -\frac{1}{2} (H\mu_{c,i} - y)^T (\Sigma_y + H\Sigma_c H^T)^{-1} (H\mu_{c,i} - y) \right\}, \quad (5.19)$$

where $\mu_{c,i}$ are calculated using R_i as in (3.2), for $i = 1, \dots, 7$. Then the moment $E[R]$ is approximated using these reweighed sigma points as $\sum_{i=1}^7 w_i^+ R_i$ after normalization. This update procedure comes from particle filters [2] where the deterministic sigma points are replaced by the randomly sampled particles, and has been widely applied in Bayesian filters with nonlinear measurement functions [61].

However, as noted in [61], the approximation of $E[R]$ using this technique can sometimes degenerate, i.e., some or all of w_i^+ can be close to zero due to large discrepancies between $H\mu_{c,i}$ and y . To overcome this problem, the “progressive update” developed in [39] is used here, which has been applied to several stochastic filters [45, 60, 73]. Denote the second term on the right hand side of (5.17) as $f_m(R)$. The idea of progressive update is that instead of multiplying the full $f_m(R_i)$ to the weights, $f_m(R)$ is decomposed into several parts as

$$f_m(R) = f_m(R)^{\lambda_1} f_m(R)^{\lambda_2} \cdots f_m(R)^{\lambda_l},$$

where $\sum_{i=1}^l \lambda_i = 1$, and $\lambda_i > 0$. In the first step, w_i is reweighed by

$$w_i^1 = w_i f_m(R_i)^{\lambda_1},$$

for $i = 1, \dots, 7$, so that no w_i^1 is close to zero if λ_1 is properly chosen. The criterion for proper $\{w_i^1\}$ is that $\frac{\min\{f_m(R_i)^{\lambda_1}\}}{\max\{f_m(R_i)^{\lambda_1}\}} \geq \tau$ for some $\tau \in (0, 1)$, i.e., the smallest and largest w_i^1 are not different by too much. If this is satisfied for $\lambda_1 = 1$, then the process is finished, and $w_i^+ = w_i^1$ for $i = 1, \dots, 7$ are used to calculate $E[R]$, which is matched to a matrix Fisher distribution with parameter F^1 using the MLE in Chapter 2.2.

Otherwise, λ_1 is chosen as $\log \tau / \log \left(\frac{\min\{f_m(R_i)\}}{\max\{f_m(R_i)\}} \right)$, so $\frac{\min\{f_m(R_i)^{\lambda_1}\}}{\max\{f_m(R_i)^{\lambda_1}\}} = \tau$. Then the

reweighted sigma points $\{R_i, w_i^1\}$ are used to calculate the expectation $\sum_{i=1}^7 w_i^1 R_i$, which is matched to a matrix Fisher distribution with the parameter F^1 . Next, in the second step, a new set of sigma points $\{R_i, w_i\}_{i=1}^7$ are selected from $\mathcal{M}(F^1)$, and they are reweighted using $f_m(R_i)^{1-\lambda_1}$. The new weights are tested again whether $\frac{\min\{f_m(R_i)^{1-\lambda_1}\}}{\max\{f_m(R_i)^{1-\lambda_1}\}} \geq \tau$. If this is satisfied, the process is finished, otherwise this is continued. Finally, the last F^l is set as the parameter F^+ of the matrix Fisher distribution that approximates (5.17).

The motivation for progressive update is that the update in each step is not too aggressive that the posterior distribution is drastically shifted from the prior. And sample degeneration is avoided so that no sample has weight close to zero after each step. This progressive update scheme for the posterior parameter F^+ is summarized in Table 5.2. It should be noted that if the unscented transform in [64] is used to select sigma points in step 13, the moment $E[R]$ must be first matched to a matrix Fisher distribution. Whereas if the unscented transform in [35] is used, $E[R]$ can be directly used and there is no need to solve (2.20) which is very computationally demanding.

Next, the conditional part of (5.16), i.e., the $p_{x|R}(x|R)$ in (5.18) is considered. As $p_R(R)$ in (5.17) has been approximated by the density function of the matrix Fisher distribution with parameter F^+ , the posterior density (5.16) is now approximated by

$$p(R, x|\mathcal{Z}) \approx c \cdot \text{etr}(F^+ R^T) p_{x|R}(x|R), \quad (5.20)$$

where c is a constant independent of (R, x) . Let $F^+ = U^+ S^+ (V^+)^T$ be its proper singular value decomposition, define $v_R^+ = (Q^+ S^+ - S^+ (Q^+)^T)^\vee$ for MFGI, or $v_R^+ = (S^+ Q^+ - (Q^+)^T S^+)^\vee$ for MFGB, with $Q^+ = (U^+)^T S^+ V^+$. Then to match the above density function to an MFG, the conditional MLE in Theorem 3.6 needs to be used, which requires the moments $E[x|\mathcal{Z}]$, $E[x(v_R^+)^T|\mathcal{Z}]$, and $E[xx^T|\mathcal{Z}]$ to be calculated. These expectations are evaluated with respect to (5.20), and are calculated in the next theorem.

Theorem 5.2. *The moments $E[x|\mathcal{Z}]$ and $E[x(v_R^+)^T|\mathcal{Z}]$ with respect to the density function*

Table 5.2: Attitude progressive update

```

1: procedure  $F^+ = \text{ATTITUDE UPDATE}(F^-, Z, z, y)$ 
2:   Compute  $\tilde{F} = F^+$  from (4.49).
3:   Select sigma points [35, 64] from  $\mathcal{M}(\tilde{F})$  as  $\{R_i, w_i\}_{i=1}^7$ .
4:   Let  $\lambda = 1$ .
5:   while  $\lambda > 0$  do
6:     For  $i = 1, \dots, 7$ , calculate  $f_m(R_i)^\lambda$ , where  $f_m(R)$  is the second term on the right
      hand side of (5.17).
7:     if  $\min\{f_m(R_i)^\lambda\} / \max\{f_m(R_i)^\lambda\} < \tau$  then
8:       Let  $\alpha = \log \tau / \log \left( \frac{\min\{f_m(R_i)\}}{\max\{f_m(R_i)\}} \right)$ .
9:       For  $i = 1, \dots, 7$ , calculate  $f_m(R_i)^\alpha$ .
10:      For  $i = 1, \dots, 7$ , update the weights as  $w_i^+ = w_i f_m(R_i)^\alpha$ .
11:      Normalize the weights as  $w_i^+ = w_i^+ / \sum_{i=1}^7 w_i^+$ .
12:      Calculate  $E[R] = \sum_{i=1}^7 w_i^+ R_i$ .
13:      Select sigma points  $\{R_i, w_i\}_{i=1}^7$  from the matrix Fisher distribution that has
          the moment  $E[R]$ .
14:      Set  $\lambda = \lambda - \alpha$ .
15:    else
16:      Set  $\lambda = 0$ .
17:    end if
18:  end while
19:  For  $i = 1, \dots, 7$ , update the weights as  $w_i^+ = w_i f_m(R_i)^\lambda$ .
20:  Normalize the weights as  $w_i^+ = w_i^+ / \sum_{i=1}^7 w_i^+$ .
21:  Calculate  $E[R] = \sum_{i=1}^7 w_i^+ R_i$ .
22:  Obtain  $F^+$  using the MLE of matrix Fisher distribution in Chapter 2.2 from  $E[R]$ .
23: end procedure

```

(5.20) are given by

$$\mathbb{E}[x|\mathcal{Z}] = (I - KH)(\mu + P\mathbb{E}[v_R|\mathcal{Z}]) + Ky, \quad (5.21)$$

$$\mathbb{E}[x(v_R^+)^T|\mathcal{Z}] = (I - KH)P\mathbb{E}[v_R(v_R^+)^T|\mathcal{Z}]. \quad (5.22)$$

And the moment $\mathbb{E}[xx^T|\mathcal{Z}]$ is

$$\begin{aligned} \mathbb{E}[xx^T|\mathcal{Z}] &= (I - KH)\Sigma_c + (I - KH)\left(\mu\mu^T + \mu\mathbb{E}[v_R|\mathcal{Z}]^TP^T + P\mathbb{E}[v_R|\mathcal{Z}]\mu^T \right. \\ &\quad \left.+ P\mathbb{E}[v_Rv_R^T|\mathcal{Z}]P^T\right)(I - KH)^T + (I - KH)(\mu + P\mathbb{E}[v_R|\mathcal{Z}])y^TK^T \\ &\quad + Ky(\mu^T + \mathbb{E}[v_R^T|\mathcal{Z}]P^T)(I - KH)^T + Kyy^TK^T. \end{aligned} \quad (5.23)$$

In the above equations, I is an abbreviation of $I_{12 \times 12}$. Also, $\mathbb{E}[v_R|\mathcal{Z}]$, $\mathbb{E}[v_Rv_R^T|\mathcal{Z}]$, and $\mathbb{E}[v_R(v_R^+)^T|\mathcal{Z}]$ are calculated the same as in Theorem 4.9.

Proof. These equations are obtained by first integrating against the Gaussian density $f_{x|R}(x|R)$ in (5.20), as in (3.22), and noting that $\mathbb{E}[v_R^+|\mathcal{Z}] = 0$. \square

Using $\mathbb{E}[x|\mathcal{Z}]$, $\mathbb{E}[x(\mu_R^+)^T|\mathcal{Z}]$, and $\mathbb{E}[xx^T|\mathcal{Z}]$ in the above theorem, the conditional MLE of MFG can be used to obtain the posterior parameters μ^+ , P^+ , and Σ^+ . In summary, this subsection develops an algorithm to update the prior distribution $(R, x) \sim \mathcal{MG}(\mu, \Sigma, P, U, S, V)$ into the posterior distribution $(R, x)|\mathcal{Z} \sim \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ conditioned by the measurement \mathcal{Z} . Together with the uncertainty propagation scheme in Chapter 5.1.2, they constitute a recursive Bayesian filter for a loosely coupled IMU-GNSS system that simultaneously estimates attitude, linear velocity, position, gyroscope bias and accelerometer bias. The pseudocode is summarized in Table 5.3.

5.1.4 Numerical Simulations

In this subsection, the recursive Bayesian filter developed for IMU-GNSS navigation using MFG is compared with the filter in [112] which is based on MEKF through numerical

Table 5.3: Bayesian estimation for IMU-GNSS navigation

```

1: procedure ESTIMATION( $\mathcal{MG}(t_0), \Omega(t), a(t), \mathcal{Z}(t)$ )
2:   Let  $k = 0$ .
3:   repeat
4:      $\mathcal{MG}(t_{k+1}) = \text{Uncertainty Propagation}(\mathcal{MG}(t_k), \Omega(t_k), a(t_k))$  in Table 5.1.
5:      $k = k + 1$ .
6:   until  $y(t_{k+1})$  ia available
7:    $\mathcal{MG}(t_{k+1}) = \text{Measurement Update}(\mathcal{MG}(t_{k+1}), Z(t_{k+1}), z(t_{k+1}), y(t_{k+1}))$ .
8:   Obtain the estimates as  $R(t_{k+1}) = U_{k+1}V_{k+1}^T$ ,  $x(t_{k+1}) = \mu_{k+1}$  for  $\mathcal{MG}(t_{k+1})$ .
9:   go to step 3.
10:  end procedure

11: procedure  $\mathcal{MG}^+ = \text{MEASUREMENT UPDATE}(\mathcal{MG}^-, Z, z, y)$ 
12:    $F^+ = \text{Attitude Update}(U^- S^- (V^-)^T, Z, z, y)$  in Table 5.2.
13:   Let the pSVD of  $F^+$  be  $F^+ = U^+ S^+ (V^+)^T$ .
14:   Calculate the moments of the posterior density in Theorem 5.2.
15:   Obtain  $\mu^+, \Sigma^+, P^+$  according to Theorem 3.6.
16:   Set  $\mathcal{MG}^+ = \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ .
17: end procedure

```

simulations. In this simulation study, a vehicle is assumed to move in a circle with radius 20 m on the ground. The speed of the vehicle fluctuates from 0 m/s to 10 m/s as a sinusoidal function with the frequency at 0.2 Hz. The resulting average free acceleration is 4.88 m/s². The vehicle simultaneously undergoes a rotational motion where the three Euler angles (body-fixed 3-2-1) follow sinusoidal functions with the frequency at 0.35 Hz, and the amplitudes of π , $\pi/2$ and π , respectively. The corresponding average angular speed is 6.17 rad/s.

The IMU is assumed to be sampled at 200 Hz. The gyroscope white noise is isotropic with $\sigma_{gu} = 1 \text{ deg}/\sqrt{\text{s}}$, i.e., $H_{gu} = \sigma_{gu} I_{3 \times 3}$ in (5.1). And its bias random walk noise is isotropic with $\sigma_{gv} = 50 \text{ deg}/\text{h}/\sqrt{\text{s}}$, i.e., $H_{gv} = \sigma_{gv} I_{3 \times 3}$ in (5.2). Similarly, the accelerometer white noise is isotropic with $\sigma_{au} = 0.01 \text{ m}/\text{s}/\sqrt{\text{s}}$, and its bias random walk noise is $\sigma_{av} = 0.333 \text{ m}/\text{s}^2/\sqrt{\text{h}}$, i.e., $H_{au} = \sigma_{au} I_{3 \times 3}$ in (5.4) and $H_{av} = \sigma_{av} I_{3 \times 3}$ in (5.5). The GNSS receiver is assumed to be sampled at 10 Hz, with the noise $\Sigma_y = \sigma_y^2 I_{3 \times 3}$, where $\sigma_y = 1 \text{ m}$. There are no attitude or direction measurements.

The initial position is given as the first measurement from the GNSS receiver, and its uncertainty is set the same as Σ_y . The initial velocity, gyroscope and accelerometer biases are assumed to be zero, with the uncertainty $0.01^2 I_{3 \times 3}$. Two cases for the initial attitude are tested. In the first case, the initial attitude is randomly chosen from $R_t(0) \exp\{\delta\hat{\theta}\}$, where $\delta\theta \sim \mathcal{N}(0, 0.1^2 I_{3 \times 3})$. The initial attitude uncertainty for MEKF is set as $0.1^2 I_{3 \times 3}$, and for the MFG filters is $S_0 = 50I_{3 \times 3}$. This indicates an initial condition with relatively small attitude uncertainty. To challenge the filters with large uncertainty, in the second case, it is assumed that the gravity direction is given by the first accelerometer measurement, assuming zero free acceleration. And the rotation around gravity is set as the true direction rotated by 180 deg, indicating a completely wrong heading. This scenario is frequently encountered in practical navigation problems when the initial heading direction cannot be estimated, for example in a building where the magnetometer reading is unreliable. The initial attitude uncertainties in the directions perpendicular to gravity, i.e., $s_1 + s_2$ and $s_1 + s_3$ for MFG, and the first two diagonal entries of the covariance matrix for MEKF, are determined from the accuracy of the accelerometer. And the initial uncertainty in the heading direction is set for MFG as $s_2 = s_3 = 0$, and for MEKF as 1000 for the third diagonal entry of the covariance matrix. All initial correlation terms are set as zero.

The simulation is run for 60 seconds. A hundred Monte Carlo runs with respect to the random noises are carried out for each initial attitude condition. The estimation error for attitude is defined as the angle needed to rotate the estimated attitude to the true attitude, and for other Euclidean quantities are defined as the Euclidean distance. The errors are first averaged across all time steps in a single simulation, and further averaged across 100 simulations. Paired t-test is conducted between MEKF and the two MFG filters, with the significance level set as $\alpha = 0.001$, to indicate any statistical significance.

The results for the case with small attitude error is summarized in Table 5.4. It is shown that the MEKF and two MFG filters have negligible differences in all estimations, and no statistical significance is found. This result is consistent with the small uncertainty in

Table 5.4: Estimation errors with small initial attitude error \pm S.D. (p -value)

	MEKF	MFGI	MFGB
attitude error (deg)	3.21 ± 0.43	3.23 ± 0.39 (0.207)	3.24 ± 0.38 (0.166)
gyroscope bias error (deg/s)	0.37 ± 0.11	0.36 ± 0.11 (0.777)	0.36 ± 0.11 (0.860)
position error (m)	0.585 ± 0.034	0.586 ± 0.033 (0.163)	0.586 ± 0.033 (0.259)
velocity error (m/s)	0.568 ± 0.039	0.570 ± 0.036 (0.169)	0.569 ± 0.036 (0.233)
accelerometer bias error (m/s^2)	0.045 ± 0.016	0.045 ± 0.015 (0.777)	0.045 ± 0.015 (0.793)

Table 5.5: Estimation errors with large initial attitude error \pm S.D. (p -value)

	MEKF	MFGI	MFGB
attitude error (deg)	23.3 ± 21.7	6.91 ± 0.95 (<0.001)	7.23 ± 1.25 (<0.001)
gyroscope bias error (deg/s)	1.63 ± 1.31	0.36 ± 0.12 (<0.001)	0.36 ± 0.11 (<0.001)
position error (m)	1.54 ± 0.99	0.63 ± 0.04 (<0.001)	0.63 ± 0.04 (<0.001)
velocity error (m/s)	2.23 ± 1.59	0.72 ± 0.06 (<0.001)	0.73 ± 0.08 (<0.001)
accelerometer bias error (m/s^2)	0.074 ± 0.043	0.044 ± 0.015 (<0.001)	0.044 ± 0.015 (<0.001)

attitude estimation as presented in Table 4.6, and is expected because the MFG with small uncertainty can be well approximated by a Gaussian distribution.

Next, the result with completely wrong initial heading direction is shown in Table 5.5. In the presence of large initial attitude uncertainty, the two MFG filters have significantly better accuracy in all estimations than MEKF. The error trajectory of a single simulation is presented in Figure 5.1. It is seen that the major advantage of MFG filters is that the attitude error converges much faster than MEKF. Although the estimation errors for other quantities are initially small for all filters, the wrong estimation of attitude makes the error grow very large for MEKF later. This is because the integrated velocity and position from (5.4) and (5.3) contain large error due to the inaccuracy of attitude, but the MEKF still maintains small variance of these integrated values, giving falsely high confidence to them in the measurement update step.

This is more clearly seen in Figure 5.2, where the estimated standard deviations of the attitude and position are shown. It is seen that although the MEKF has very large attitude error, its uncertainty is on the same level of MFG filters. Looking at the attitude standard deviation around the gravity direction (the third row in Figure 5.2a), it converges

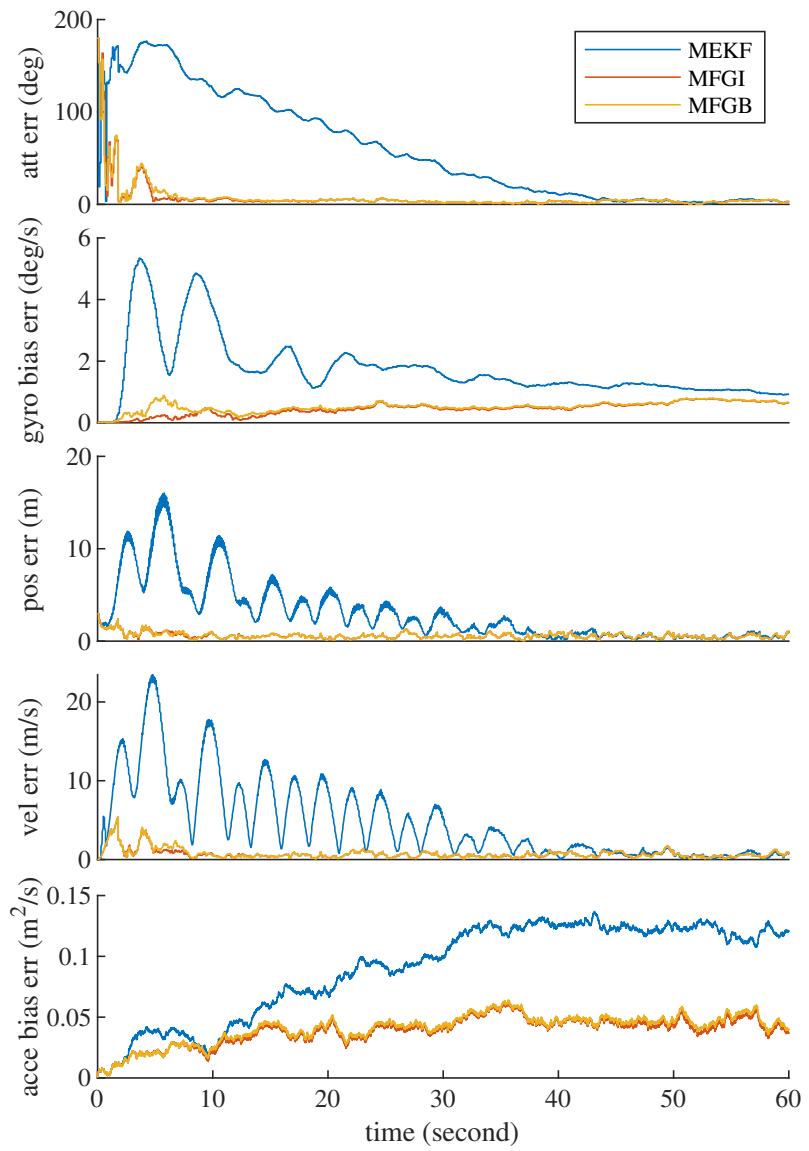
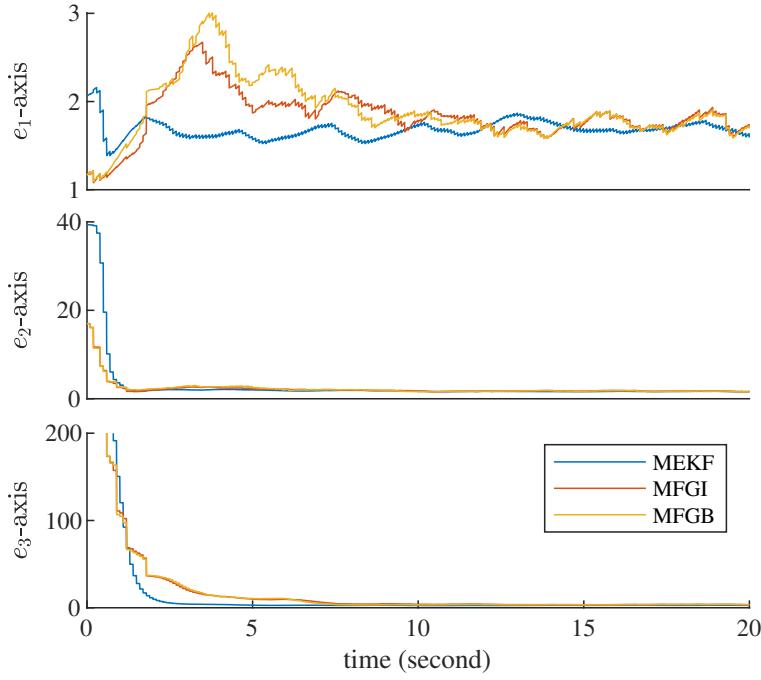
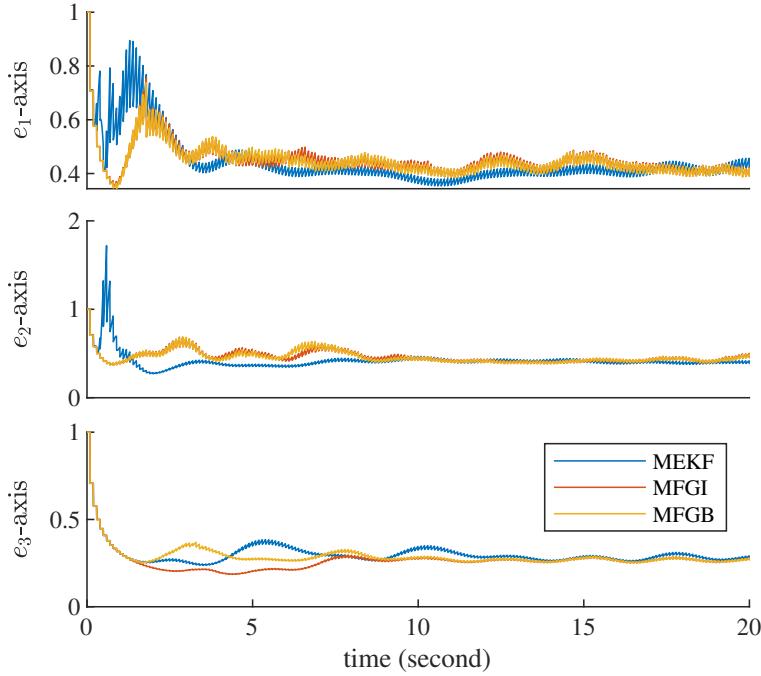


Figure 5.1: Estimation errors with large initial attitude error.



(a) Attitude Uncertainty



(b) Position Uncertainty

Figure 5.2: Estimation standard deviations with large initial error. The attitude standard deviation for MFG is the square root of the diagonals of $U(\text{tr}(S)I_{3 \times 3} - S)^{-1}U^T$, and for MEKF is the square root of the diagonals of the covariance matrix transformed into the inertial frame using the estimated attitude. And the position standard deviation is the square root of the corresponding entry in the covariance matrix.

exponentially for MEKF, which is typical for an EKF. On the other hand, it converges slower for the MFG filters, possibly indicating that MFG has better accuracy in modeling large attitude uncertainty, as the attitude dose not converge very fast as indicated in Figure 5.1. Note that in this IMU-GNSS navigation simulation, the attitude is not directly observed, and it is corrected by its correlation with the position. This suggests that the MFG not only has better accuracy in attitude uncertainty, but also in the correlation between attitude and position, when the attitude uncertainty is large.

5.2 Visual-Inertial Odometry

In this section, the same inertial navigation setting as in the previous section is studied. But instead of using direct position measurements from GNSS receivers, it is assumed that the 3D coordinates of some unknown features or known landmarks in the inertial frame are measured by cameras. This is usually referred to as a visual-inertial system. In this section, two cases are distinguished: (i) if the features captured by the camera are landmarks whose 3D coordinates are known in advance (possibly with noises), it is referred to as visual-inertial *navigation*; (ii) if instead there is no prior information of the features, it is referred to as visual-inertial *odometry*. To simplify analysis, instead of studying projection measurements on the image plane, it is assumed that the camera is in RGB-D or stereo configuration, so that the 3D coordinates of the feature or landmark in the body-fixed frame are directly measured. And the measured location is directly used to correct any integration errors from the inertial system.

As reviewed in Chapter 1.2.3, the VIO algorithms can be roughly classified into two categories: filtering and optimization. The optimization based VIO algorithms have achieved better accuracy, but it is usually more computationally demanding. This section will primarily focus on the filtering approach. The reason is that in the optimization method, the uncertainty of pose is not modeled, and the uncertainty for pre-integrated IMU pose is only for a short period of time which is very accurate, so the MFG which is designed to model

large uncertainty is not very useful. On the other hand, the filtering method has to model the uncertainty of pose directly, where large uncertainty is frequently encountered.

More specifically, this section will first study a general problem of estimating a pose and its uncertainty that aligns two 3D point sets with known correspondences. This is usually one of the steps in point cloud registration [44]. It is first assumed that the 3D coordinates of one point set are perfectly known (map), and the other point set has noises (measurement). Then based on the proposed algorithm, noises for the map point set are also considered. The uncertainty of the estimated pose is modeled by MFG to deal with possibly large attitude uncertainties. Together with the uncertainty propagation method proposed in Chapter 5.1.2, these constitute a visual-inertial navigation algorithm when known landmarks in the inertial frame are measured. Finally, this algorithm is extended to a visual-inertial odometry algorithm where unknown features are measured without any prior information.

5.2.1 Pose Estimation Without Map Uncertainty

First, this subsection studies the problem of finding the pose that aligns two point sets using MFG. Suppose the coordinates of landmarks in the inertial frame are given as $\{p_i\}_{i=1}^N$. These are measured by a camera in the body-fixed frame as $\{p'_i\}_{i=1}^N$. It is assumed that the correspondences are known, i.e., p'_i is the measurement of p_i , for $i = 1, \dots, N$. The pose of the rigid body is denoted by $(R, t) \in \text{SE}(3)$, where $\text{SE}(3)$ is the three dimensional Euclidean group. The measurements are affected by noises which are assumed to be Gaussian with zero mean and covariance matrices A_i , i.e.,

$$p'_i = R^T(p_i - t) + \mathcal{N}(0, A_i). \quad (5.24)$$

To make the calculation easier, the noise A_i is assumed to be isotropic, i.e., $A_i = \sigma_i^2 I_{3 \times 3}$. The objective is to find an MFG that describes the uncertainty of the pose (R, t) .

The negative log-likelihood function of (R, t) given the point measurements can be

written as

$$f(p'_1, \dots, p'_N | R, t) = \sum_{i=1}^N (R^T(p_i - t) - p'_i)^T A_i^{-1} (R^T(p_i - t) - p'_i). \quad (5.25)$$

The conventional method to find (R, t) is to solve the maximum likelihood estimation by posting the above equation into a least square problem, and finding the minimum using Gauss-Newton or Levenberg–Marquardt algorithm. Denote the Jacobian matrix of the residual error with respect to (R, t) as $J_r \in \mathbb{R}^{3N \times 6}$ in the last step of the optimization, then the covariance of the estimated pose is approximated by $(J_r^T J_r)^{-1}$. However, this method is expected to only deal with small uncertainty, and in this subsection, (5.25) is approximated by the exponent of an MFG as in (3.1).

As in chapter 5.1.3, the idea is to decompose (5.25) into two parts: the marginal likelihood of R , and the conditional likelihood of $x|R$. And then the MLE for MFG can be used to recover the parameters from these two likelihood functions. The approach is motivated by the conventional SVD method [3, 11], where the rotation R is first found by centering the two point sets. This decomposition is formulated in the following theorem.

Theorem 5.3. Define $T \in \mathbb{R}^{(N-1) \times N}$ as

$$T = \begin{bmatrix} \frac{N-1}{N} & \cdots & -\frac{1}{N} & -\frac{1}{N} \\ \vdots & \ddots & \vdots & \vdots \\ -\frac{1}{N} & \cdots & \frac{N-1}{N} & -\frac{1}{N} \end{bmatrix}, \quad (5.26)$$

and $\mathbf{T} = T \otimes I_{3 \times 3} \in \mathbb{R}^{3(N-1) \times 3N}$. Denote $\mathbf{p}' \in \mathbb{R}^{3N \times 1}$ and $\mathbf{p}_m \in \mathbb{R}^{3N \times 1}$ as the column concatenation of $\{p'_i\}_{i=1}^3$ and $\{R^T(p_i - t)\}_{i=1}^3$, respectively. Let $\mathbf{A} = \text{diag}(A_1, \dots, A_N) \in \mathbb{R}^{3N \times 3N}$ be a block diagonal matrix. Also, let $\mathbf{Y} = \left[\frac{1}{N} I_{3 \times 3}, \dots, \frac{1}{N} I_{3 \times 3} \right] \in \mathbb{R}^{3N \times 3}$ be N copies

of identity matrices divided by N in a row. Define $C \in \mathbb{R}^{3 \times 3}$, and $\mathbf{D} \in \mathbb{R}^{3 \times 3N}$ as

$$C = \mathbf{Y}\mathbf{A}\mathbf{Y}^T - \mathbf{Y}\mathbf{A}\mathbf{T}^T(\mathbf{T}\mathbf{A}\mathbf{T}^T)^{-1}\mathbf{T}\mathbf{A}\mathbf{Y}^T, \quad (5.27)$$

$$\mathbf{D} = \mathbf{Y}\mathbf{A}\mathbf{T}^T(\mathbf{T}\mathbf{A}\mathbf{T}^T)^{-1}\mathbf{T} - \mathbf{Y}. \quad (5.28)$$

Let \mathbf{D} be written in block form as $\mathbf{D} = [D_1, \dots, D_N]$, where $D_i \in \mathbb{R}^{3 \times 3}$ for $i = 1, \dots, N$. Then, the likelihood (5.25) can be decomposed as

$$f(p'_1, \dots, p'_N | R, t) = f_R(R) + f_{t|R}(t | R), \quad (5.29)$$

where the marginal likelihood $f_R(R)$ is

$$f_R(R) = (\mathbf{T}\mathbf{p}' - \mathbf{T}\mathbf{p}_m)^T (\mathbf{T}\mathbf{A}\mathbf{T}^T)^{-1} (\mathbf{T}\mathbf{p}' - \mathbf{T}\mathbf{p}_m). \quad (5.30)$$

And the conditional likelihood $f_{t|R}(t | R)$ is

$$f_{t|R}(t | R) = (t - \mu_{t|R})^T (R\mathbf{C}R^T)^{-1} (t - \mu_{t|R}), \quad (5.31)$$

where

$$\mu_{t|R} = R \sum_{i=1}^N D_i (p'_i - R^T p_i). \quad (5.32)$$

Proof. By definition, (5.25) can be written as

$$f(p'_1, \dots, p'_N | R, t) = (\mathbf{p}' - \mathbf{p}_m)^T \mathbf{A}^{-1} (\mathbf{p}' - \mathbf{p}_m).$$

So it remains to check that

$$f(p'_1, \dots, p'_N | R, t) - f_R(R) = (\mathbf{p}' - \mathbf{p}_m)^T \left(\mathbf{A}^{-1} - \mathbf{T}^T (\mathbf{T}\mathbf{A}\mathbf{T}^T)^{-1} \mathbf{T} \right) (\mathbf{p}' - \mathbf{p}_m) \quad (5.33)$$

equates $f_{x|R}(x|R)$. Let $\mathbf{T}_f = \begin{bmatrix} \mathbf{T} \\ \mathbf{Y} \end{bmatrix} \in \mathbb{R}^{3N \times 3N}$. It is straightforward to check that \mathbf{T}_f is full rank and hence invertible. So $\mathbf{A}^{-1} = \mathbf{T}_f^T (\mathbf{T}_f \mathbf{A} \mathbf{T}_f^T)^{-1} \mathbf{T}_f$, and

$$\begin{aligned} \mathbf{A}^{-1} - \mathbf{T}^T (\mathbf{T} \mathbf{A} \mathbf{T}^T)^{-1} \mathbf{T} &= \mathbf{T}_f^T (\mathbf{T}_f \mathbf{A} \mathbf{T}_f^T)^{-1} \mathbf{T}_f - \mathbf{T}^T (\mathbf{T} \mathbf{A} \mathbf{T}^T)^{-1} \mathbf{T} \\ &= \begin{bmatrix} \mathbf{T}^T & \mathbf{Y}^T \end{bmatrix} \begin{bmatrix} \mathbf{T} \mathbf{A} \mathbf{T}^T & \mathbf{T} \mathbf{A} \mathbf{Y}^T \\ \mathbf{Y} \mathbf{A} \mathbf{T}^T & \mathbf{Y} \mathbf{A} \mathbf{Y}^T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{T} \\ \mathbf{Y} \end{bmatrix} - \mathbf{T}^T (\mathbf{T} \mathbf{A} \mathbf{T}^T)^{-1} \mathbf{T} \\ &= \mathbf{D}^T \mathbf{C}^{-1} \mathbf{D}, \end{aligned}$$

where the last equality uses the block inversion formula [97]. Note that $\mathbf{T} \mathbf{Y}^T = 0$, so

$$\sum_{i=1}^N D_i = N \mathbf{D} \mathbf{Y}^T = N \mathbf{Y} \mathbf{A} \mathbf{T}^T (\mathbf{T} \mathbf{A} \mathbf{T}^T)^{-1} \mathbf{T} \mathbf{Y}^T - N \mathbf{Y} \mathbf{Y}^T = -N \mathbf{Y} \mathbf{Y}^T = -I_{3 \times 3}.$$

Next, it can be shown that

$$\begin{aligned} \mathbf{D}(\mathbf{p}' - \mathbf{p}_m) &= \sum_{i=1}^N D_i(p'_i - R^T p_i + R^T t) \\ &= -R^T t - \sum_{i=1}^N D_i R^T p_i + \sum_{i=1}^N D_i p'_i \\ &= -R^T \left(t - R \sum_{i=1}^N D_i (p'_i - R^T p_i) \right). \end{aligned}$$

Substitute these into (5.33) yields $f(p'_1, \dots, p'_N | R, t) - f_R(R) = f_{x|R}(x|R)$. \square

With this decomposition of the likelihood (5.25), the marginal part $f_R(R)$ can be first approximated by a matrix Fisher density, then the conditional part $f_{x|R}(x|R)$ can be used in Theorem 3.6 to recover the parameters μ , Σ , and P of the MFG.

Marginal Likelihood First, an algorithm to approximate the marginal likelihood (5.30) is developed. Define $q_i, q'_i \in \mathbb{R}^3$ for $i = 1, \dots, N-1$ as

$$q_i = p_i - \frac{1}{N} \sum_{i=1}^N p_i, \quad q'_i = p'_i - \frac{1}{N} \sum_{i=1}^N p'_i, \quad (5.34)$$

i.e., they are $\{p_i\}$ and $\{p'_i\}$ centered by their geometric centers. Define $\mathbf{q}' \in \mathbb{R}^{3(N-1) \times 1}$ and $\mathbf{q}_m \in \mathbb{R}^{3(N-1) \times 1}$ as the concatenation of $\{q'_i\}_{i=1}^{N-1}$ and $\{R^T q_i\}_{i=1}^{N-1}$ respectively. It is straightforward to verify that $\mathbf{q}' = \mathbf{T}\mathbf{p}'$, and $\mathbf{q}_m = T\mathbf{p}_m$. Note that the translation t in \mathbf{p}_m does not appear in the centered \mathbf{q}_m anymore, which is the key point that $f_R(R)$ does not depend on t . With these notations, the marginal likelihood (5.30) can be written as

$$f_R(R) = (\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{T}\mathbf{A}\mathbf{T}^T)^{-1} (\mathbf{q}' - \mathbf{q}_m). \quad (5.35)$$

Now finding the $R \in \text{SO}(3)$ that maximizes $f_R(R)$ becomes very similar to the classic Wahba's problem [82, 110]. In other words, q'_i can be regarded as direction measurements of q_i , for $i = 1, \dots, N-1$, and R can be found by aligning these two sets of directions. However, the difficulty is that $\{q'_i\}$ are correlated, since $\mathbf{T}\mathbf{A}\mathbf{T}^T$ in general has nonzero off-diagonal terms. Two cases are considered here: (i) all measurements have equal covariance matrices, i.e. $A_1 = \dots = A_N$, and (ii) the covariance matrices are different. It is firstly shown that for equal covariance matrices, a linear transform can be applied to $\{q_i\}_{i=1}^{N-1}$ again to make them uncorrelated, so R can be found using an algorithm similar to Theorem III.2 in [64].

Theorem 5.4. *The matrix $TT^T \in \mathbb{R}^{(N-1) \times (N-1)}$ is invertible. Let $TT^T = U^T DU$ be the eigen-decomposition of TT^T . For $i = 1, \dots, N-1$, let*

$$r_i = \sum_{j=1}^{N-1} U_{ij} q_i \quad r'_i = \sum_{j=1}^{N-1} U_{ij} q'_i, \quad (5.36)$$

where U_{ij} is the (i, j) -th entry of U . If $A_1 = \dots = A_N \triangleq A = \sigma^2 I_{3 \times 3}$, then

$$f_R(R) = (\mathbf{r}' - \mathbf{r}_m)^T (D \otimes A)^{-1} (\mathbf{r}' - \mathbf{r}_m), \quad (5.37)$$

where \mathbf{r}' and \mathbf{r}_m are the concatenation of $\{r'_i\}_{i=1}^{N-1}$ and $\{R^T r_i\}_{i=1}^{N-1}$ respectively.

Proof. The invertibility of TT^T can be verified by direct calculation. Also, it can be easily shown that $\mathbf{r}' = (U \otimes I_{3 \times 3})\mathbf{q}'$, and $\mathbf{r}_m = (U \otimes I_{3 \times 3})\mathbf{q}_m$. Note that $(U \otimes I_{3 \times 3})^{-1} = U^T \otimes I_{3 \times 3} = (U \otimes I_{3 \times 3})^T$, so $\mathbf{q}' = (U \otimes I_{3 \times 3})^T \mathbf{r}'$, and $\mathbf{q}_m = (U \otimes I_{3 \times 3})^T \mathbf{r}_m$. Then by (5.35), $f_R(R)$ can be written as

$$f_R(R) = (\mathbf{r}' - \mathbf{r}_m)^T (U \otimes I_{3 \times 3}) (\mathbf{TAT}^T)^{-1} (U \otimes I_{3 \times 3})^T (\mathbf{r}' - \mathbf{r}_m).$$

Because $\mathbf{TAT}^T = (T \otimes I_{3 \times 3})(I_{N \times N} \otimes A)(T \otimes I_{3 \times 3})^T$, the covariance matrix for \mathbf{r}' becomes

$$\begin{aligned} & (U \otimes I_{3 \times 3}) (\mathbf{TAT}^T)^{-1} (U \otimes I_{3 \times 3})^T \\ &= ((U \otimes I_{3 \times 3})(T \otimes I_{3 \times 3})(I_{N \times N} \otimes A)(T \otimes I_{3 \times 3})^T (U \otimes I_{3 \times 3})^T)^{-1} \\ &= ((UT \otimes I_{3 \times 3})(I_{N \times N} \otimes A)(UT \otimes I_{3 \times 3})^T)^{-1} \\ &= (UTT^T U^T \otimes A)^{-1} \\ &= (D \otimes A)^{-1}. \end{aligned}$$

And (5.37) is obtained. \square

Since $D = \text{diag}(d_1, \dots, d_{N-1})$ is a diagonal matrix, only the diagonal blocks of $D \otimes A$ is nonzero. Therefore, the marginal likelihood (5.37) can be treated as measuring $N-1$ directions r_i in the inertial frame with noises. More specifically, the measurements r'_i can be written as

$$r'_i = R^T r_i + \mathcal{N}(0, d_i \sigma^2 I_{3 \times 3}), \quad (5.38)$$

for $i = 1, \dots, N - 1$. To approximate (5.37) into a matrix Fisher density, the Gaussian distribution of r'_i is first approximated by a von Mises–Fisher distribution on \mathbb{S}^2 according to the following lemma.

Lemma 5.2. *The density function of $\frac{r'_i}{\|r'_i\|}$ in (5.38) can be approximated by*

$$p\left(\frac{r'_i}{\|r'_i\|}\right) \approx c \exp\left\{\kappa_i \frac{(r'_i)^T R^T r_i}{\|r'_i\| \|r_i\|}\right\} \quad (5.39)$$

where c is a normalizing constant. The concentration parameter κ_i is solved from

$$\coth \kappa_i - \frac{1}{\kappa_i} = \frac{2}{\gamma} \phi(\gamma) + \left(1 - \frac{1}{\gamma^2}\right) (2\Phi(\gamma) - 1), \quad (5.40)$$

where $\gamma = \|r_i\| / (\sqrt{d_i} \sigma)$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density and cumulative density functions of the one dimensional normal distribution.

Proof. Because r'_i follows a Gaussian distribution, its direction $r'_i / \|r'_i\|$ follows a projected normal distribution [80]. And its mean resultant length [100] can be calculated as on the right hand side of (5.40). Based on the maximum likelihood estimation [80] for a von Mises–Fisher distribution, $r'_i / \|r'_i\|$ can be matched to a von Mises–Fisher distribution with mean direction $R^T r_i / \|r_i\|$, and the concentration parameter solved from (5.40). The density function is given in (5.39). \square

Based on the above lemma, the marginal likelihood $f_R(R)$ in (5.37) can be further approximated by

$$\begin{aligned} \exp\left\{-\frac{1}{2}f_R(R)\right\} &\approx c \sum_{i=1}^{N-1} \exp\left\{\kappa_i \frac{(r'_i)^T R^T r_i}{\|r'_i\| \|r_i\|}\right\} \\ &= c \cdot \text{etr}\left(\left(\sum_{i=1}^{N-1} \kappa_i \frac{r_i (r'_i)^T}{\|r_i\| \|r'_i\|}\right) R^T\right) \triangleq c \cdot \text{etr}(FR^T), \end{aligned} \quad (5.41)$$

for some constant c . In other words, the marginal likelihood $f_R(R)$ is matched to a matrix

Fisher density with the parameter $F = \sum_{i=1}^{N-1} \kappa_i \frac{r_i(r'_i)^T}{\|r_i\| \|r'_i\|}$.

Nevertheless, the approximation in (5.41) requires that all measurements have the same isotropic noise $A = \sigma^2 I_{3 \times 3}$, which seems too stringent. To relax this requirement such that the measurements can have different levels of noises $A_i = \sigma_i^2 I_{3 \times 3}$, the technique of *importance sampling* is used. Similar to Chapter 5.1.3, the idea is to use unscented transform to select sigma points from a matrix Fisher distribution, and reweigh these sigma points to calculate the expectation $E[R]$ according to the density function (5.35). To use importance sampling, a *proposal* matrix Fisher distribution is needed that is close to the density function (5.35). A good choice is simply treating $\{q'_i\}_{i=1}^{N-1}$ as uncorrelated direction measurements, and using Theorem 5.2 and (5.41) to find a matrix Fisher distribution.

Let this matrix Fisher distribution be denoted by $\mathcal{M}(F_{\text{prop}})$, and its density function be denoted by $f_{\text{prop}}(R)$. Select sigma points [35, 64] $\{(R_j, w_j)\}_{j=1}^7$ from the proposal distribution $\mathcal{M}(F_{\text{prop}})$, and reweigh these sigma points as

$$w_j^+ = w_j \cdot \frac{\exp\left\{-\frac{1}{2}f_R(R_j)\right\}}{f_{\text{prop}}(R_j)}. \quad (5.42)$$

Then the reweighed sigma points $\{(R_j, w_j^+)\}_{j=1}^7$ can be regarded as weighted samples from the density (5.35). And the expectation $E[R]$ can be calculated as $E[R] = \sum_{j=1}^7 w_j^+ R_j$ after normalization, which can be matched to a matrix Fisher distribution using the MLE as introduced in Chapter 2.2.

Note that the update of weights in (5.42) is similar to that in (5.19), so the progressive update procedure in Chapter 5.1.3 can also be used here to improve accuracy. In addition, even if all measurements have the same covariance matrix, (5.41) is still not perfect due to the approximation of projected normal distribution as von Mises–Fisher distribution in Theorem 5.2. So using this importance sampling approach with (5.41) as the proposal distribution is also expected to improve accuracy even for the case of same covariance matrix. The pseudocode to match (5.30) to a matrix Fisher density is given in Table 5.6.

Table 5.6: Attitude Estimation From Marginal Likelihood

```

1: procedure  $F = \text{ATTITUDE ESTIMATION}(\{p_i\}, \{p'_i\})$ 
2:   Calculate  $\{q_i\}$  and  $\{q'_i\}$  using (5.34).
3:   if  $A_1 = \dots = A_N$  then
4:     Calculate  $\{r_i\}$  and  $\{r'_i\}$  using (5.36).
5:     Calculate  $F_{\text{prop}}$  as in (5.41) using Theorem 5.2.
6:   else
7:     Calculate  $F_{\text{prop}}$  as in (5.41) using Theorem 5.2, with  $\{r_i\}$  and  $\{r'_i\}$  replaced by
        $\{q_i\}$  and  $\{q'_i\}$ , and  $d_i\sigma^2 I_{3 \times 3}$  in (5.38) replaced by the  $i$ -th diagonal block of  $\mathbf{TAT}^T$ .
8:   end if
9:   Select sigma points [35, 64] from  $\mathcal{M}(F_{\text{prop}})$  as  $\{R_j, w_j\}_{j=1}^7$ .
10:  Let  $\lambda = 1$ .
11:  while  $\lambda > 0$  do
12:    For  $j = 1, \dots, 7$ , calculate  $f_m(R_j)^\lambda$ , with  $f_m(R) = \exp\left\{-\frac{1}{2}f_R(R)\right\}/f_{\text{prop}}(R)$ ,
        where  $f_R(R)$  is in (5.35), and  $f_{\text{prop}}$  is the density function of  $\mathcal{M}(F_{\text{prop}})$ .
13:    if  $\min\{f_m(R_j)^\lambda\} / \max\{f_m(R_j)^\lambda\} < \tau$  then
14:      Let  $\alpha = \log \tau / \log \left( \frac{\min\{f_m(R_j)\}}{\max\{f_m(R_j)\}} \right)$ .
15:      For  $j = 1, \dots, 7$ , calculate  $f_m(R_j)^\alpha$ .
16:      For  $j = 1, \dots, 7$ , update the weights as  $w_j^+ = w_j f_m(R_j)^\alpha$ .
17:      Normalize the weights as  $w_j^+ = w_j^+ / \sum_{j=1}^7 w_j^+$ .
18:      Calculate  $E[R] = \sum_{i=1}^7 w_i^+ R_i$ .
19:      Select sigma points  $\{R_j, w_j\}_{j=1}^7$  from the matrix Fisher distribution that has
        the moment  $E[R]$ .
20:      Set  $\lambda = \lambda - \alpha$ .
21:    else
22:      Set  $\lambda = 0$ .
23:    end if
24:  end while
25:  For  $j = 1, \dots, 7$ , update the weights as  $w_j^+ = w_j f_m(R_j)^\lambda$ .
26:  Normalize the weights as  $w_j^+ = w_j^+ / \sum_{j=1}^7 w_j^+$ .
27:  Calculate  $E[R] = \sum_{j=1}^7 w_j^+ R_j$ .
28:  Obtain  $F$  using the MLE of matrix Fisher distribution in Chapter 2.2 from  $E[R]$ .
29: end procedure

```

In summary, after matching the marginal likelihood (5.30) to a matrix Fisher distribution with parameter F , the full likelihood (5.25) has been approximated by

$$\exp\left\{-\frac{1}{2}f(p'_1, \dots, p'_N|R, t)\right\} \approx c \cdot \text{etr}(FR^T) \exp\left\{-\frac{1}{2}f_{t|R}(t|R)\right\}, \quad (5.43)$$

where c is a normalizing constant, and $f_{t|R}(t|R)$ is the conditional likelihood (5.31). In the next subsection, the conditional likelihood (5.31) is further used to match (5.43) to an MFG density.

Conditional Likelihood As before, the conditional likelihood is handled by the conditional MLE of MFG in Theorem 3.6. After obtaining F from the marginal likelihood, let $F = USV^T$ be its pSVD, and define $v_R = (QS - SQ^T)^\vee$ for MFGL, or $v_R = (SQ - Q^TS)^\vee$ for MFGB, with $Q = U^TRV$. To use the theorem, the following moments need to be calculated: $E[t]$, $E[tv_R^T]$, and $E[tt^T]$. These moments are according to the density function (5.43). As shown in (5.32), clearly $E[tt^T]$ involves the fourth order of R , which currently do not have closed form formulae. To simplify these moments, the follow lemma is very useful.

Lemma 5.3. *Given $A_i = \sigma_i^2 I_{3 \times 3}$ for $i = 1, \dots, N$, C in (5.27) and \mathbf{D} in (5.28) satisfy the following equations:*

$$RCR^T = C, \quad (5.44)$$

$$RD_iR^T = D_i. \quad (5.45)$$

Proof. Note that $R\mathbf{Y} = \mathbf{Y}(I_{N \times N} \otimes R)$, and $(I_{N \times N} \otimes R)\mathbf{A}(I_{N \times N} \otimes R)^T = \text{diag}(RA_1R^T, \dots, RA_NR^T) = \mathbf{A}$. For C , it can be shown that

$$R\mathbf{Y}\mathbf{A}\mathbf{Y}^T R^T = \mathbf{Y}(I_{N \times N} \otimes R)\mathbf{A}(I_{N \times N} \otimes R)^T \mathbf{Y}^T = \mathbf{Y}\mathbf{A}\mathbf{Y}^T.$$

Also, note that $(I_{(N-1) \times (N-1)} \otimes R)\mathbf{T} = \mathbf{T}(I_{N \times N} \otimes R)$, and $(I_{N \times N} \otimes R)^T \mathbf{A}(I_{N \times N} \otimes R) = \mathbf{A}$, so

the following equality holds:

$$\mathbf{TAT}^T = (I_{(N-1) \times (N-1)} \otimes R)^T \mathbf{TAT}^T (I_{(N-1) \times (N-1)} \otimes R).$$

Therefore, it can be shown that

$$\begin{aligned} & R\mathbf{YAT}^T(\mathbf{TAT}^T)^{-1}\mathbf{TAY}^TR^T \\ &= R\mathbf{YAT}^T(I_{(N-1) \times (N-1)} \otimes R)^T(\mathbf{TAT}^T)^{-1}(I_{(N-1) \times (N-1)} \otimes R)\mathbf{TAY}^TR^T \\ &= R\mathbf{YA}(I_{N \times N} \otimes R)^T\mathbf{T}^T(\mathbf{TAT}^T)^{-1}\mathbf{T}(I_{N \times N} \otimes R)\mathbf{AY}^TR \\ &= R\mathbf{Y}(I_{N \times N} \otimes R)^T\mathbf{AT}^T(\mathbf{TAT}^T)^{-1}\mathbf{TA}(I_{N \times N} \otimes R)\mathbf{Y}^TR \\ &= \mathbf{YAT}^T(\mathbf{TAT}^T)^{-1}\mathbf{TAY}^T, \end{aligned}$$

which proves (5.44). Equation (5.45) can be proved using similar calculations. \square

With the above lemma, the conditional likelihood (5.31) with its conditional mean (5.32) is now simplified as

$$f_{t|R}(t|R) = \left(t - \sum_{i=1}^N D_i(Rp'_i - p_i) \right)^T C^{-1} \left(t - \sum_{i=1}^N D_i(Rp'_i - p_i) \right). \quad (5.46)$$

Then the moments required for the conditional MLE of MFG are evaluated as follows:

Theorem 5.5. *The moments $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ according to the density (5.43) are*

$$E[t] = \sum_{i=1}^N D_i(E[R]p'_i - p_i), \quad (5.47)$$

$$E[tv_R^T] = \sum_{i=1}^N D_i E[Rp'_i v_R^T], \quad (5.48)$$

$$E[tt^T] = C + \sum_{i=1}^N \sum_{j=1}^N D_i \left(E[Rp'_i(p'_j)^T R^T] - E[R]p'_i p_j^T - p_i(p'_j)^T E[R]^T + p_i p_j^T \right) D_j^T. \quad (5.49)$$

Proof. These equations can be easily derived by integrating the left hand side with respect to the Gaussian density (5.46) first, as in (3.22). \square

It is notable that the complexity for $E[tt^T]$ in (5.49) is $O(N^2)$, which becomes computationally demanding if the number of points is large. This cannot be avoided for general cases, but in next theorem, it is shown that if all measurements have the same covariance matrix, the expressions for $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ can be greatly simplified.

Theorem 5.6. Let $p_c = \frac{1}{N} \sum_{i=1}^N p_i$, and $p'_c = \frac{1}{N} \sum_{i=1}^N p'_i$. If $A_1 = \dots = A_N \triangleq A = \sigma^2 I_{3 \times 3}$, then $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ according to the density (5.43) are

$$E[t] = p_c - E[R] p'_c \quad (5.50)$$

$$E[tv_R^T] = -E[Rp'_c v_R^T] \quad (5.51)$$

$$E[tt^T] = \frac{\sigma^2}{N} I_{3 \times 3} + E[Rp'_c (p'_c)^T R^T] - p_c (p'_c)^T E[R]^T - E[R] p'_c p_c^T + p_c p_c^T. \quad (5.52)$$

Proof. Let $Y = \frac{1}{N}[1, \dots, 1] \in \mathbb{R}^{1 \times N}$ so that $\mathbf{Y} = Y \otimes I_{3 \times 3}$. Note that $YT^T = 0$, so it can be shown that

$$\mathbf{YAT}^T = (Y \otimes I_{3 \times 3})(I_{N \times N} \otimes A)(T^T \otimes I_{3 \times 3}) = YT^T \otimes A = 0.$$

So $C = \mathbf{YAY}^T = \frac{1}{N}A = \frac{\sigma^2}{N}I_{3 \times 3}$, and $\mathbf{D} = -\mathbf{Y}$, i.e., $D_i = -\frac{1}{N}I_{3 \times 3}$ for $i = 1, \dots, N$. Substitute these into (5.46), it is simplified to

$$f_{t|R}(t|R) = (t - (p_c - Rp'_c))^T \left(\frac{\sigma^2}{N} I_{3 \times 3} \right)^{-1} (t - (p_c - Rp'_c)), \quad (5.53)$$

and the desired moments can be easily derived. \square

After calculating the moments $E[t]$, $E[tv_R^T]$, and $E[tt^T]$, they can be used in the conditional MLE for MFG to obtain the parameters μ , Σ , and P . So the likelihood function (5.25) is matched to an MFG. The pseudocode is summarized in Table 5.7.

Table 5.7: Pose estimation without map noises

1:	procedure $\mathcal{MG} = \text{POSE ESTIMATION}(\{p_i\}, \{p'_i\})$
2:	Obtain $F = \text{Attitude Estimation}(\{p_i\}, \{p'_i\})$ in Table 5.6.
3:	Let $F = USV^T$ be its pSVD.
4:	Calculate $E[v_R v_R^T]$ as in (3.21) using S .
5:	if $A_1 = \dots = A_N$ then
6:	Calculate $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ in Theorem 5.6.
7:	else
8:	Calculate $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ in Theorem 5.5.
9:	end if
10:	Obtain μ, Σ, P according to Theorem 3.6, with $E[t]$, $E[tv_R^T]$, $E[tt^T]$, $E[v_R v_R^T]$, and $E[v_R] = 0$.
11:	Set $\mathcal{MG} = \mathcal{MG}(\mu, \Sigma, P, U, S, V)$.
12:	end procedure

Posterior Density Up to Table (5.7), an algorithm to find the distribution of the pose (R, t) modeled by MFG is given to align the map point set $\{p_i\}$ and measurement point set $\{p'_i\}$. If this algorithm is used in a filtering algorithm, the pose also has a prior distribution. Here a more general problem is considered. Let $y = [t^T, a^T]^T \in \mathbb{R}^n$, suppose (R, y) follows an MFG with the parameter $(\mu, \Sigma, P, U, S, V)$ before measurements, where a includes other Euclidean quantities, such as linear velocity and sensor biases as in the IMU-GNSS navigation. According to the Bayes' formula, the posterior density for (R, y) after incorporating the measurements $\{p'_i\}$ is

$$\begin{aligned}
 p(R, y | p'_1, \dots, p'_N) &\propto p(p'_1, \dots, p'_N | R, t) p(R, y) \\
 &= \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (R^T (p_i - Hy) - p'_i)^T A_i^{-1} (R^T (p_i - Hy) - p'_i) \right\} \\
 &\quad \cdot \text{etr}(FR^T) \exp \left\{ -\frac{1}{2} (y - \mu_c)^T \Sigma_c^{-1} (y - \mu_c) \right\}, \tag{5.54}
 \end{aligned}$$

where the first exponential term on the right hand side is the likelihood (5.25), with $H = [I_{3 \times 3}, 0_{3 \times (n-3)}]$ selecting t from y . The intermediate parameters F , μ_c , and Σ_c are defined with respect to $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$. Next, matching the posterior density $p(R, y | p'_1, \dots, p'_N)$

to an MFG is studied.

In Theorem 5.3, the likelihood $f(p'_1, \dots, p'_N | R, t)$ has been factorized into the marginal $f_R(R)$ and the conditional $f_{t|R}(t|R)$, which can be directly applied to the posterior density.

$$p(R, y | p'_1, \dots, p'_N) \propto \exp\{FR^T\} \exp\left\{-\frac{1}{2}(\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{TAT}^T)^{-1}(\mathbf{q}' - \mathbf{q}_m)\right\} \\ \cdot \exp\left\{-\frac{1}{2}(y - \mu_c)^T \Sigma_c^{-1}(y - \mu_c)\right\} \exp\left\{-\frac{1}{2}(Hy - \mu_{t|R})^T C^{-1}(Hy - \mu_{t|R})\right\} \quad (5.55)$$

with $\mu_{t|R}$ given as in (5.46). The last two terms has the same structure as in (5.14), so they can be calculated similarly as in Lemma 5.1. More specifically, define

$$K = \Sigma_c H^T (H\Sigma_c H^T + C)^{-1}. \quad (5.56)$$

Then (5.55) can be further factorized into

$$p(R, y | p'_1, \dots, p'_N) \propto \exp\{FR^T\} \exp\left\{-\frac{1}{2}(\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{TAT}^T)^{-1}(\mathbf{q}' - \mathbf{q}_m)\right\} \\ \cdot \exp\left\{-\frac{1}{2}(H\mu_c - \mu_{t|R})^T (C + H\Sigma_c H^T)^{-1}(H\mu_c - \mu_{t|R})\right\} \\ \cdot \exp\left\{-\frac{1}{2}(y - (I - KH)\mu_c - K\mu_{t|R})^T ((I - KH)\Sigma_c)^{-1}(y - (I - KH)\mu_c - K\mu_{t|R})\right\}, \quad (5.57)$$

where I is an abbreviation of $I_{n \times n}$. The first three terms on the right hand side of (5.57) do not depend on y , and they can be regarded as the marginal posterior density for R . The same algorithm as in Table 5.2 can be used to approximate it to a matrix Fisher distribution. Let the approximated parameter be denoted by F^+ , then (5.57) is approximated by

$$p(R, y | p'_1, \dots, p'_N) \propto \exp\{F^+ R^T\} \cdot p_{y|R}(y|R), \quad (5.58)$$

where $p_{y|R}(y|R)$ is the last term in (5.57).

Let $F^+ = U^+ S^+ (V^+)^T$ be its pSVD, define $v_R^+ = (Q^+ S^+ - S^+(Q^+)^T)^\vee$ for MFGI, or

$v_R^+ = (S^+Q^+ - (Q^+)^TS^+)^{\vee}$ for MFGB, with $Q^+ = (U^+)^TS^+V^+$. The function $p_{y|R}(y|R)$ is Gaussian for y , and it can be regarded as the conditional posterior density for $y|R$. To use the conditional MLE of MFG, the moments $E[y]$, $E[y(v_R^+)^T]$, and $E[yy^T]$ according to the density (5.58) need to be calculated. To distinguish these moments with those moments according to the prior $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$, they are denoted by $E[y|\mathcal{Z}]$, $E[y(v_R^+)^T|\mathcal{Z}]$, and $E[yy^T|\mathcal{Z}]$, where $\mathcal{Z} = \{p_i'\}_{i=1}^N$ are all measurements.

Theorem 5.7. *The moments $E[y|\mathcal{Z}]$, $E[y(v_R^+)^T|\mathcal{Z}]$ according to the density (5.58) can be calculated as*

$$E[y|\mathcal{Z}] = (I - KH)\mu + KE[\mu_{t|R}|\mathcal{Z}], \quad (5.59)$$

$$E[y(v_R^+)^T|\mathcal{Z}] = (I - KH)PE[v_R(v_R^+)^T|\mathcal{Z}] + KE[\mu_{t|R}(v_R^+)^T|\mathcal{Z}]. \quad (5.60)$$

And the moment $E[yy^T|\mathcal{Z}]$ is

$$\begin{aligned} E[yy^T|\mathcal{Z}] &= (I - KH)\Sigma_c + (I - KH)\left(\mu\mu^T + \mu E[v_R|\mathcal{Z}]^T P^T + PE[v_R|\mathcal{Z}]\mu^T\right. \\ &\quad \left.+ PE[v_R v_R^T|\mathcal{Z}] P^T\right)(I - KH)^T + (I - KH)\left(\mu E[\mu_{t|R}^T|\mathcal{Z}] + PE[v_R \mu_{t|R}^T|\mathcal{Z}]\right)K^T \\ &\quad + K\left(E[\mu_{t|R}|\mathcal{Z}]\mu^T + E[\mu_{t|R} v_R^T|\mathcal{Z}] P^T\right)(I - KH)^T + KE[\mu_{t|R} \mu_{t|R}^T|\mathcal{Z}]K^T. \end{aligned} \quad (5.61)$$

In the above equations, I is an abbreviation of $I_{n \times n}$. The moments $E[v_R|\mathcal{Z}]$, $E[v_R v_R^T|\mathcal{Z}]$, and $E[v_R(v_R^+)^T|\mathcal{Z}]$ are calculated the same as in Theorem 4.9. Also, the moments $E[\mu_{t|R}|\mathcal{Z}]$, $E[\mu_{t|R}(v_R^+)^T|\mathcal{Z}]$, and $E[\mu_{t|R} \mu_{t|R}^T|\mathcal{Z}]$ are the same as $E[t]$, $E[tv_R^T]$ and $E[tt^T] - C$ in Theorem 5.5 and Theorem 5.6 depending on the respective conditions of the theorem, with v_R replaced by v_R^+ . Also, if $A_1 = \dots = A_N$, $E[\mu_{t|R} v_R^T|\mathcal{Z}]$ is given as

$$E[\mu_{t|R} v_R^T|\mathcal{Z}] = p_c E[v_R^T|\mathcal{Z}] - E[Rp'_c v_R^T|\mathcal{Z}], \quad (5.62)$$

Table 5.8: Measurement update without map noises

```

1: procedure  $\mathcal{MG}^+ = \text{MEASUREMENT UPDATE}(\mathcal{MG}^-, \{p_i\}, \{p'_i\})$ 
2:   Let  $F^- = U^- S^- (V^-)^T$ .
3:   Calculate  $F^+ = \text{Attitude Update}(F^-, \{p_i\}, \{p'_i\})$ .
4:   Let the pSVD of  $F^+$  be  $F^+ = U^+ S^+ (V^+)^T$ .
5:   Calculate the moments of the posterior density in Theorem 5.7.
6:   Obtain  $\mu^+, \Sigma^+$ , and  $P^+$  using the conditional MLE in Theorem 3.6.
7:   Set  $\mathcal{MG}^+ = \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$ .
8: end procedure

9: procedure  $F^+ = \text{ATTITUDE UPDATE}(F^-, \{p_i\}, \{p'_i\})$ 
10:  Select sigma points [35, 64] from  $\mathcal{M}(F^-)$  as  $\{R_j, w_j\}_{j=1}^7$ .
11:  Let  $\lambda = 1$ .
12:  while  $\lambda > 0$  do
13:    For  $j = 1, \dots, 7$ , calculate  $f_m(R_j)^\lambda$ , where  $f_m(R)$  is the second and third terms
        on the right hand side of (5.57).
14:    if  $\min\{f_m(R_j)^\lambda\} / \max\{f_m(R_j)^\lambda\} < \tau$  then
15:      Let  $\alpha = \log \tau / \log \left( \frac{\min\{f_m(R_j)\}}{\max\{f_m(R_j)\}} \right)$ .
16:      For  $j = 1, \dots, 7$ , calculate  $f_m(R_j)^\alpha$ .
17:      For  $j = 1, \dots, 7$ , update the weights as  $w_j^+ = w_j f_m(R_j)^\alpha$ .
18:      Normalize the weights as  $w_j^+ = w_j^+ / \sum_{i=j}^7 w_i^+$ .
19:      Calculate  $E[R] = \sum_{j=1}^7 w_j^+ R_j$ .
20:      Select sigma points  $\{R_j, w_j\}_{j=1}^7$  from the matrix Fisher distribution that has
        the moment  $E[R]$ .
21:      Set  $\lambda = \lambda - \alpha$ .
22:    else
23:      Set  $\lambda = 0$ .
24:    end if
25:  end while
26:  For  $j = 1, \dots, 7$ , update the weights as  $w_j^+ = w_j f_m(R_j)^\lambda$ .
27:  Normalize the weights as  $w_j^+ = w_j^+ / \sum_{i=j}^7 w_i^+$ .
28:  Calculate  $E[R] = \sum_{j=1}^7 w_j^+ R_j$ .
29:  Obtain  $F^+$  using the MLE of matrix Fisher distribution in Chapter 2.2 from  $E[R]$ .
30: end procedure

```

otherwise it is

$$\mathbb{E} [\mu_{t|R} v_R^T | \mathcal{Z}] = \sum_{i=1}^N D_i (\mathbb{E} [Rp_i' v_R^T | \mathcal{Z}] - p_i \mathbb{E} [v_R^T | \mathcal{Z}]). \quad (5.63)$$

Proof. The proof is a very straightforward extension of Theorem 5.2, with y in (5.18) replaced by $\mu_{t|R}$. \square

With these moments, the conditional MLE of μ^+ , Σ^+ , and P^+ can be obtained using Theorem 3.6 as before. In summary, the posterior density (5.54) has been matched to an MFG with parameters $(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$, and the pseudocode is given in Table 5.8.

5.2.2 Numerical Simulations Without Map Uncertainty

In this subsection, the algorithm in Table 5.7 that estimates the uncertainty of pose using an MFG is compared with the conventional least square approach. Namely, an Levenberg–Marquardt (LM) algorithm is used to directly optimize the likelihood (5.25), and the covariance of the pose estimate is given as $(J_r^T J_r)^{-1}$, where J_r is the Jacobian in the last step of the optimization iteration.

First, the algorithm in Table 5.7 without pose prior is tested. The true attitude is set as $R_t = \exp \left\{ \frac{\pi}{2} \hat{e}_3 \right\}$, and the true translation is $t_t = [5, 0, 0]^T$. The measurement noise is set as $A_1 = \dots = A_N = 0.5^2 I_{3 \times 3}$, and the number of points are varied as $N \in \{4, 6, 8, 10, 12, 14, 16, 18, 20\}$. Smaller number of points indicates more uncertainty in the estimation of pose. The map points $\{p_i\}_{i=1}^N$ are randomly scattered as $p_i = t_t + \xi_i$, where $\xi_i \in \mathbb{R}^3$ follows a three dimensional uniform distribution on $[0, 1]^3$. For the algorithm in Table 5.7, the attitude estimate is given as UV^T , and the translation estimate is given as μ . A thousand Monte Carlo simulations are carried out for each N with respect to the random measurement noises. Paired t -test with significance level at $\alpha = 0.001$ are used to detect significant differences between the two algorithms.

The estimation errors of the two algorithms are listed in Table 5.9. Unfortunately, the

Table 5.9: Estimation error of the pose that aligns two point sets

estimator	MFG		
<i>N</i>	20	18	16
att err (deg)	30.6±17.3(p<0.001)	29.47±15.6(p<0.001)	32.9±18.7(p<0.001)
pos err	0.38±0.18(p<0.001)	0.38±0.19(p=0.590)	0.43±0.23(p=0.905)
estimator	LM optimization		
<i>N</i>	20	18	16
att err (deg)	28.7±14.1	28.4±13.9	31.8±17.9
pos err	0.37±0.18	0.38±0.19	0.43±0.23
estimator	MFG		
<i>N</i>	14	12	10
att err (deg)	35.1±20.4(p<0.001)	43.7±26.2(p<0.001)	47.8±29.6(p<0.001)
pos err	0.46±0.24(p=0.223)	0.52±0.27(p=0.006)	0.53±0.27(p<0.001)
estimator	LM optimization		
<i>N</i>	14	12	10
att err (deg)	34.1±19.1	42.1±25.2	45.6±27.7
pos err	0.46±0.25	0.53±0.29	0.54±0.29
estimator	MFG		
<i>N</i>	8	6	4
att err (deg)	49.5±29.4(p=0.002)	57.1±35.6(p=0.372)	77.8±42.9(p=0.707)
pos err	0.58±0.29(p<0.001)	0.67±0.33(p<0.001)	0.92±0.36(p<0.001)
estimator	LM optimization		
<i>N</i>	8	6	4
att err (deg)	48.5±28.8	56.7±35.5	77.7±43.0
pos err	0.61±0.32	0.72±0.39	1.04±0.49

attitude estimation from the MFG based method is worse than the LM optimization, and their difference is significant when $N > 8$. This is not surprising because the optimization method finds the maximum likelihood estimation directly, and should have better accuracy if the algorithm converges properly. The translation (position) estimation for MFG is worse than the LM method when $N = 20$, but it becomes better when $N < 12$. The reason for this is currently unknown, but it may indicate that the gradient based optimization is not very effective when the uncertainty is large, i.e., the Hessian or the Fisher information is small.

The estimation of uncertainty is also compared between the two methods. Let $p(R, t) = \exp\left\{-\frac{1}{2}f(p'_1, \dots, p'_N | R, t)\right\}$ be the real un-normalized probability density function of (R, t) , and denote the normalizing constant by c . Also, let the density function of the MFG as the output of Table 5.7 be denoted by $f_{\text{MFG}}(R, t)$, and the density function of the Gaussian distribution from the LM optimization be denoted by $f_{\text{LM}}(R, t)$. The KL divergence $KL(p || p_{\text{MFG}})$ and $KL(p || p_{\text{LM}})$ is a metric on how different the true density p is from the estimated p_{MFG} and p_{LM} , and they are defined as

$$KL(p || p_{\text{MFG}}) = \int_{R \in \text{SO}(3)} \int_{x \in \mathbb{R}^3} \frac{1}{c} p \frac{\log(p/c)}{\log(p_{\text{MFG}})} dx dR,$$

$$KL(p || p_{\text{LM}}) = \int_{R \in \text{SO}(3)} \int_{x \in \mathbb{R}^3} \frac{1}{c} p \frac{\log(p/c)}{\log(p_{\text{LM}})} dx dR.$$

These are compared to see which method has a better approximation of the true density $p(R, t)$. To evaluate the KL divergence, a hundred million random samples $\{R_j, t_j\}$ are drawn from the Gaussian distribution of the LM optimization, which is treated as the proposal distribution in importance sampling for the true density function. Then, these samples are weighed by $w_j = p(R_j, t_j)/p_{\text{LM}}(R_j, t_j)$, and the weights are normalized by $w_j = w_j / \sum_j w_j$.

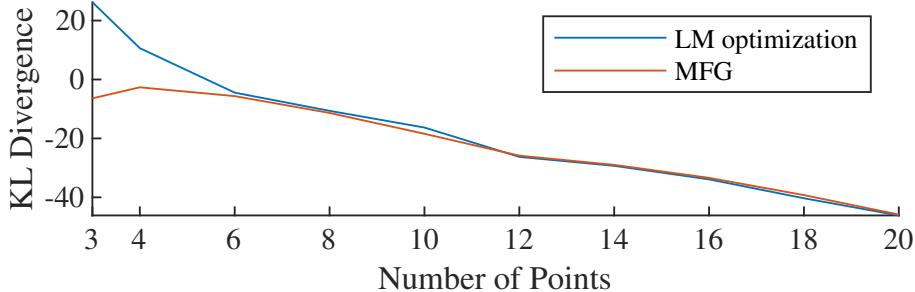


Figure 5.3: KL divergence $d(p||p_{\text{MFG}})$ and $d(p||p_{\text{LM}})$ of the MFG and LM-optimization methods.

The KL divergence are evaluated using these weighed sampled as

$$d(p||p_{\text{MFG}}) \triangleq KL(p||p_{\text{MFG}}) + \log(c) = \sum_j w_j \log \left(p(R_j, t_j) / p_{\text{MFG}}(R_j, t_j) \right),$$

$$d(p||p_{\text{LM}}) \triangleq KL(p||p_{\text{LM}}) + \log(c) = \sum_j w_j \log \left(p(R_j, t_j) / p_{\text{LM}}(R_j, t_j) \right).$$

Because only the difference between the KL divergence of the two methods matters, the constant $\log(c)$ which cannot be evaluated is deleted from the calculation.

The results are presented in Figure 5.3. Because the constant $\log(c)$ is not calculated, the KL divergence in the figure is sometimes negative. It can be seen that when $N > 10$, the KL divergence for the LM method is slightly lower, indicating a slightly better estimation of the uncertainty. When $N \leq 10$, because the uncertainty in pose increases, the KL-divergence of the MFG becomes smaller. And when $N = 4$ or 3 , it becomes notably smaller than the LM method, indicating MFG is more accurate when dealing with very large uncertainty in the pose.

Next, the pose estimation algorithms are challenged with a singular configuration when the attitude has one degree of freedom that is not properly observed. More specifically, the true attitude is set as $R_t = \exp \left\{ \frac{\pi}{2} \hat{e}_1 \right\} \exp \left\{ \frac{\pi}{2} \hat{e}_2 \right\}$, and the true translation is set as $t_t = [0, 0, 0]^T$. There are ten map points, scattered as $p_i = [5, 0, 0]^T + \xi_i$, where $\xi_i \in \mathbb{R}^3$ follows a three dimensional uniform distribution on $[-1, 1] \times [-a, a]^2$, and a is varied

Table 5.10: Estimation error \pm S.D. (p -value) of the pose for a degenerative configuration

estimator	MFG			
a	1	0.1	0.01	0
att err (deg)	3.93 \pm 1.79 (0.448)	24.0 \pm 21.7 (0.642)	82.4 \pm 51.4 (0.065)	90.1 \pm 0.2 (0.279)
pos err	0.28 \pm 0.15 (0.339)	0.36 \pm 0.20 (<0.001)	0.14 \pm 0.06 (<0.001)	0.11 \pm 0.08 (<0.001)
estimator	LM optimization			
a	1	0.1	0.01	0
att err (deg)	3.93 \pm 1.79	24.0 \pm 21.8	80.7 \pm 51.2	90.1 \pm 0.1
pos err	0.28 \pm 0.15	0.39 \pm 0.22	0.39 \pm 0.22	0.38 \pm 0.20

in $\{1, 0.1, 0.01, 0\}$. As a is decreased, the map points are closer to the e_1 -axis of the inertial frame, making the rotation about this axis gradually becomes unobservable. The measurement noise is set as $A_1 = \dots = A_N = 0.1^2 I_{3 \times 3}$.

The estimation errors of the MFG and LM methods are listed in Table 5.10. It is shown that the two methods have very similar accuracy of attitude errors, and the error becomes very large when $a = 0.01$ or 0 , indicating that the attitude cannot be fully observed due to the placement of the map points. On the other hand, the translation accuracy of the MFG method is much better than the LM optimization when $a = 0.01$ or 0 . Inspection of the Jacobian $J_r^T J_r$ indicates that it becomes almost singular, due to the unobservability of attitude for the LM optimization. Since $J_r^T J_r$ is used to calculate the step size in the LM-optimization, this possibly explains why it has worse accuracy. Interestingly, when a is close to zero, the translation accuracy for MFG is even better than that when a is relatively large.

Finally, the measurement update algorithm in Table 5.8 is combined with the uncertainty propagation in Table 5.1 into a recursive Bayesian filter. This filter is compared with the MEKF based method in a visual-inertial navigation setting. A vehicle is supposed to move in a circle with radius of 3 m. The height of the trajectory is fluctuating between 0 m and 1 m as a sinusoidal function with the frequency at 0.785 Hz. The resulting average free acceleration is 0.51 m/s^2 . The first body-fixed axis of the vehicle points forward along the tangent direction of the trajectory, and the second body-fixed axis is confined in the horizontal plane, pointing towards the center of the circle. The resulting average angular velocity

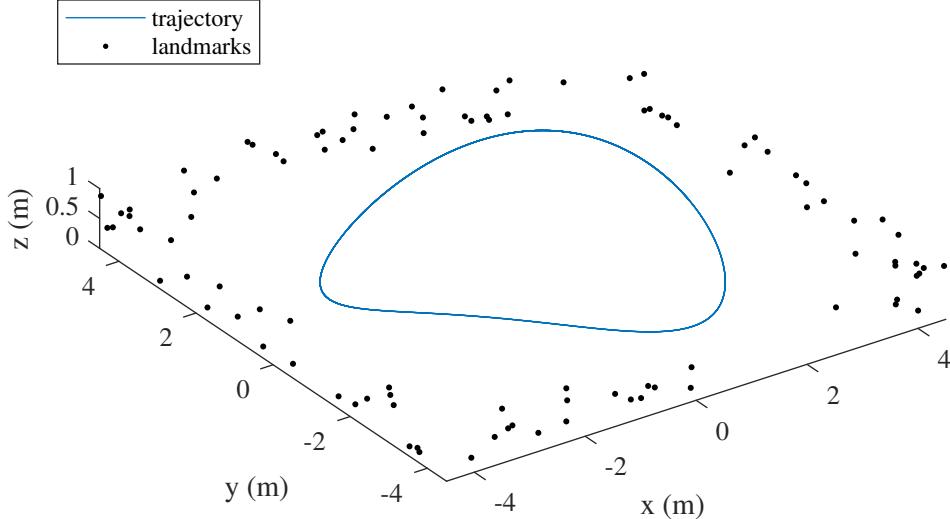


Figure 5.4: Trajectory and landmarks for visual-inertial navigation.

is 24.7 deg/s. A hundred landmarks are scattered around the trajectory in a rectangular shape. The trajectory of the vehicle and locations of landmarks are depicted in Figure 5.4. A landmark is measured by the camera only if it is within 3 m of the vehicle.

The white noise for the gyroscope is set as $0.1 \text{ deg}/\sqrt{\text{s}}$, and for the accelerometer is $0.1 \text{ m/s}/\sqrt{\text{s}}$. It is assumed that the gyroscope and accelerometer do not have random walk biases, i.e., H_{gv} in (5.2) and H_{av} in (5.5) are zero, so the biases are not estimated in the filter. The gyroscope is sampled at 200 Hz. The position measurement noise of the camera is set identically as $\sigma^2 I_{3 \times 3}$, with $\sigma = 1 \text{ m}$. And the sampling frequency for the camera is at 10 Hz. The simulation lasts for 40 s.

Two initial conditions are tested. For the first case, the initial attitude uncertainty for the MFG filter is $S_0 = 50I_{3 \times 3}$, and for MEKF is $0.1^2 I_{3 \times 3}$. The initial velocity and position uncertainties for both filters are set as $0.1^2 I_{3 \times 3}$. The initial attitude, velocity, and position are their true values plus a random noise sampled from those initial uncertainty distributions. This corresponds to an initial condition with relatively small uncertainty. Next, for the second case, the initial conditions for velocity and position are set the same as in the first case. But the initial attitude is rotated from the true attitude around the inertial vertical axis by 180 deg, simulating the scenario when the initial heading direction is completely unknown. The

Table 5.11: Estimation error for visual-inertial navigation without map noises

initial condition	small uncertainty		large uncertainty	
	MFG	MEKF	MFG	MEKF
att err (deg)	0.97±0.21 (p<0.001)	0.90±0.20	1.62±0.23 (p<0.001)	75.0±38.5
vel err (m/s)	0.208±0.012 (p<0.001)	0.205±0.013	0.235±0.013 (p<0.001)	1.59±0.54
pos err (m)	0.173±0.011 (p<0.001)	0.172±0.011	0.184±0.010 (p<0.001)	2.37±1.03

attitude uncertainty for MFG is set as $S_0 = 0.01I_{3\times 3}$, and for MEKF is $1000I_{3\times 3}$, indicating a very large initial attitude uncertainty. A hundred Monte Carlo simulations are carried out for both initial conditions, and paired *t*-test is used to find any significant difference with the significant level at $\alpha = 0.001$.

For the MFG filter, only definition MFGI is tested, because all linear components, i.e., velocity and position, are defined in the inertial frame. The estimation errors for MFG and MEKF are summarized in Table 5.11. It demonstrates that MEKF has slightly better accuracy than MFG for the small initial uncertainty case, which is consistent with the results in Table 5.9 for static pose estimation. But their relative differences are very small compared to the magnitude of error. On the other hand, for large initial uncertainty, the MFG filter behaves much better than MEKF. The trajectory of the estimation error for one particular simulation is presented in Figure 5.5. It is seen that the MEKF suffers from extremely slow convergence from the initially wrong attitude. Even worse, the wrong attitude estimation for MEKF contributes to the large errors in velocity and position. This is very similar to the simulation for IMU-GNSS navigation in Figure 5.1. And the reason for this slow convergence can be attributed to that the Gaussian distribution of MEKF does not properly describe the initial large attitude uncertainty.

5.2.3 Pose Estimation With Map Uncertainty

In Chapter 5.2.1, an algorithm is developed to model the uncertainty of the pose $(R, t) \in \text{SE}(3)$ that aligns two point sets. In particular, the map point set is assumed to be perfectly known, and the measurement point set is affected by random noises. However, in some

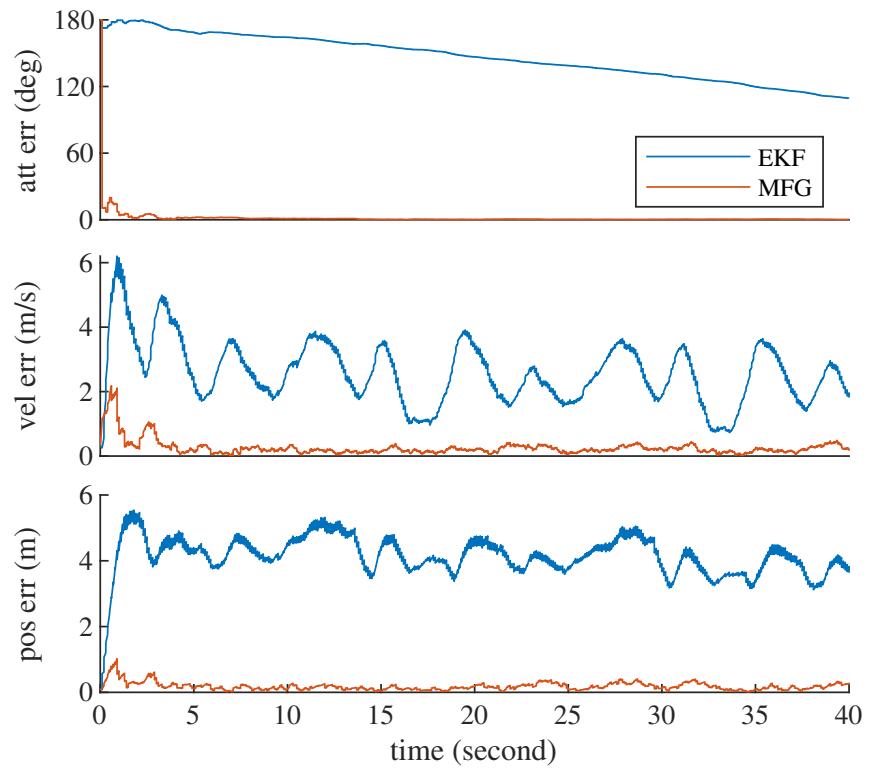


Figure 5.5: Error trajectory of a particular simulation for visual-inertial navigation without map noises.

cases the noise of map point set also needs to be modeled. For example, the 3D coordinates of triangulated features from a stereo camera have larger noises when the feature is further away from the camera, so it may be beneficial to also model this difference in uncertainty after the features are added to the map. In this subsection, the MFG is used to model the pose where the map point set also suffers from noises.

Suppose there are N landmarks, whose coordinates in the inertial frame are denoted by $\{p_i\}_{i=1}^N$. The coordinates are randomly distributed by $p_i \sim \mathcal{N}(\bar{p}_i, B_i)$. These map points are measured by a camera in the body fixed frame as $\{p'_i\}_{i=1}^N$, affected by Gaussian noises with zero mean and covariance matrices $A_i = \sigma_i^2 I_{3 \times 3}$. Note that only the noise of measurements A_i is assumed to be isotropic. If there is no prior information for the pose (non-informative prior), the posterior density function for the pose and map, according to the Bayes' formula, is given by

$$p(R, t, \{p_i\} | \{p'_i\}) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (R^T(p_i - t) - p'_i)^T A_i^{-1} (R^T(p_i - t) - p'_i) \right\} \\ \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (p_i - \bar{p}_i)^T B_i^{-1} (p_i - \bar{p}_i) \right\}. \quad (5.64)$$

The first exponential term in the above equation is the likelihood (5.25) considered in Chapter 5.2.1, and the last term is the prior density function for map points. The objective is to model the pose $(R, t) | \{p'_i\}$ with an MFG, and to find the posterior Gaussian distribution for each map point $p_i | \{p'_i\}$.

Because the coordinates of map points p_i are now random variables, the algorithm developed in Chapter 5.2.1 cannot be directly used. But in the next lemma, it is shown that (5.64) can be decomposed into two parts, and one of them has exactly the same structure as (5.25), so the algorithm in Chapter 5.2.1 can be applied.

Lemma 5.4. *Let $\delta p_i = p_i - \bar{p}_i \in \mathbb{R}^3$ for $i = 1, \dots, N$. Equation (5.64) can be decomposed*

into

$$p(R, t, \{p_i\} | \{p'_i\}) = p(R, t | \{p'_i\}) p(\{\delta p_i\} | R, t, \{p'_i\}), \quad (5.65)$$

where $p(R, t | \{p'_i\})$ is

$$\exp \left\{ -\frac{1}{2} \sum_{i=1}^N (R^T (\bar{p}_i - t) - p'_i)^T (A_i + R^T B_i R)^{-1} (R^T (\bar{p}_i - t) - p'_i) \right\}. \quad (5.66)$$

Let $K_i = (A_i^{-1} + B_i^{-1})^{-1} A_i^{-1} \in \mathbb{R}^{3 \times 3}$, then $p(\{\delta p_i\} | R, t, \{p'_i\})$ is

$$\exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\delta p_i - K_i (R p'_i + t - \bar{p}_i))^T (A_i^{-1} + B_i^{-1}) (\delta p_i - K_i (R p'_i + t - \bar{p}_i)) \right\}. \quad (5.67)$$

Proof. Because $A_i = \sigma_i^2 I_{3 \times 3}$, for $R \in \text{SO}(3)$, $R A_i^{-1} R^T = A_i^{-1}$. For $i = 1, \dots, N$, the exponent of (5.64) can be written as

$$\begin{aligned} & (R^T (p_i - t) - p'_i)^T A_i^{-1} (R^T (p_i - t) - p'_i) + (p_i - \bar{p}_i)^T B_i^{-1} (p_i - \bar{p}_i) \\ &= (\bar{p}_i + \delta p_i - t - R p'_i)^T A_i^{-1} (\bar{p}_i + \delta p_i - t - R p'_i) + \delta p_i^T B_i^{-1} \delta p_i \\ &= (R p'_i + t - \bar{p}_i)^T A_i^{-1} (R p'_i + t - \bar{p}_i) - 2 \delta p_i^T A_i^{-1} (R p'_i + t - \bar{p}_i) + \delta p_i^T (A_i^{-1} + B_i^{-1}) \delta p_i. \end{aligned}$$

The last two terms can be further arranged into a quadratic form as

$$\begin{aligned} & \delta p_i^T (A_i^{-1} + B_i^{-1}) \delta p_i - 2 \delta p_i^T A_i^{-1} (R p'_i + t - \bar{p}_i) \\ &= (\delta p_i - K_i (R p'_i + t - \bar{p}_i))^T (A_i^{-1} + B_i^{-1}) (\delta p_i - K_i (R p'_i + t - \bar{p}_i)) \\ &\quad - (R p'_i + t - \bar{p}_i)^T A_i^{-1} (A_i^{-1} + B_i^{-1})^{-1} A_i^{-1} (R p'_i + t - \bar{p}_i). \end{aligned}$$

Using the Woodbury matrix identity [97], $(R p'_i + t - \bar{p}_i)^T A_i^{-1} (R p'_i + t - \bar{p}_i)$ and the last term

in the above equation can be combined into

$$\begin{aligned}
& (Rp'_i + t - \bar{p}_i)^T A_i^{-1} (Rp'_i + t - \bar{p}_i) - (Rp'_i + t - \bar{p}_i)^T A_i^{-1} (A_i^{-1} + B_i^{-1})^{-1} A_i^{-1} (Rp'_i + t - \bar{p}_i) \\
&= (Rp'_i + t - \bar{p}_i)^T \left[A_i^{-1} - A_i^{-1} (A_i^{-1} + B_i^{-1})^{-1} A_i^{-1} \right] (Rp'_i + t - \bar{p}_i) \\
&= (Rp'_i + t - \bar{p}_i)^T (A_i + B_i)^{-1} (Rp'_i + t - \bar{p}_i) \\
&= (R^T (\bar{p}_i - t) - p'_i)^T (A_i + R^T B_i R)^{-1} (R^T (\bar{p}_i - t) - p'_i).
\end{aligned}$$

And the desired decomposition is derived. \square

The density $p(R, t | \{p'_i\})$ in (5.66) has exactly the same structure as (5.25), except the noise is increased from A_i to $A_i + R^T B_i R$. So using Theorem 5.3 and (5.35), $p(R, t | \{p'_i\})$ can be further decomposed into

$$\begin{aligned}
p(R, t | \{p'_i\}) &= \exp \left\{ -\frac{1}{2} (\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{T} \mathbf{A} \mathbf{T}^T)^{-1} (\mathbf{q}' - \mathbf{q}_m) \right\} \\
&\quad \cdot \exp \left\{ -\frac{1}{2} (t - \mu_{t|R})^T (R C R^T)^{-1} (t - \mu_{t|R}) \right\} \triangleq p(R | \{p'_i\}) p(t | R, \{p'_i\}),
\end{aligned} \tag{5.68}$$

where \mathbf{q}' and \mathbf{q}_m are defined in (5.34), with p_i replaced by \bar{p}_i . The covariance matrix $\mathbf{A} = \text{diag}(A_1 + RB_1R^T, \dots, A_N + RB_NR^T)$, and $\mathbf{T}, \mu_{t|R}$ are defined in Theorem 5.3. The first exponential term $p(R | \{p'_i\})$ can be matched to a matrix Fisher density using the algorithm in Table 5.6, with A_i replaced by $A_i + B_i$. Note that although the definition of \mathbf{A} has RB_iR^T , it can be simply replaced by B_i in the proposal distribution for importance sampling, as the proposal distribution only needs to be close to the target distribution.

Let the approximated parameter of the matrix Fisher distribution be denoted by F . And let $F = USV^T$ be its pSVD, define $v_R = (QS - SQ^T)^\vee$ for MFGI, or $v_R = (SQ - Q^T S)^\vee$ for MFGB, with $Q = U^T RV$ as before. The next step is to study the density function

$$p(t | R, \{p'_i\}) p(\{\delta p_i\} | R, t, \{p'_i\})$$

conditioned by R . The next lemma deals with this density.

Lemma 5.5. Let $\mathbf{x} = [p_1^T, \dots, p_N^T, t^T]^T \in \mathbb{R}^{3N+3}$, and $\boldsymbol{\mu}_{x|R} = [\mu_{p_1|R}^T, \dots, \mu_{p_N|R}^T, \mu_{t|R}^T]^T \in \mathbb{R}^{3N+3}$, where $\mu_{p_i|R} = (I_{3 \times 3} - K_i)\bar{p}_i + K_i(Rp'_i + \mu_{t|R})$. Also, let $\Sigma_{x|R}$ be defined as

$$\boldsymbol{\Sigma}_{x|R} = \begin{bmatrix} A_1^{-1} + B_1^{-1} & \cdots & 0_{3 \times 3} & -A_1^{-1} \\ \vdots & \ddots & \vdots & \vdots \\ 0_{3 \times 3} & \cdots & A_N^{-1} + B_N^{-1} & -A_N^{-1} \\ -A_1^{-1} & \cdots & -A_N^{-1} & (RCR^T)^{-1} + \sum_{i=1}^N A_i^{-1} (A_i^{-1} + B_i^{-1})^{-1} A_i^{-1} \end{bmatrix}^{-1}. \quad (5.69)$$

Then the conditional density can be written as

$$p(t|R, \{p'_i\})p(\{\delta p_i\}|R, t, \{p'_i\}) = \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{x|R})^T \boldsymbol{\Sigma}_{x|R}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{x|R}) \right\}. \quad (5.70)$$

Proof. The mean of δp_i in (5.67) can be arranged into

$$K_i(Rp'_i + t - \bar{p}_i) = K_i(Rp'_i + \mu_{t|R} - \bar{p}_i) + K_i(t - \mu_{t|R}).$$

Let $\mu_{\delta p_i|R} = K_i(Rp'_i + \mu_{t|R} - \bar{p}_i)$, then the exponent of $p(\{\delta p_i\}|R, t, \{p'_i\})$ becomes

$$\begin{aligned} & (\delta p_i - K_i(Rp'_i + t - \bar{p}_i))^T (A_i^{-1} + B_i^{-1}) (\delta p_i - K_i(Rp'_i + t - \bar{p}_i)) \\ &= \left(\begin{bmatrix} I_{3 \times 3} & -K_i \end{bmatrix} \begin{bmatrix} \delta p_i - \mu_{\delta p_i|R} \\ t - \mu_{t|R} \end{bmatrix} \right)^T (A_i^{-1} + B_i^{-1}) \left(\begin{bmatrix} I_{3 \times 3} & -K_i \end{bmatrix} \begin{bmatrix} \delta p_i - \mu_{\delta p_i|R} \\ t - \mu_{t|R} \end{bmatrix} \right) \\ &= \begin{bmatrix} p_i - \mu_{p_i|R} \\ t - \mu_{t|R} \end{bmatrix}^T \begin{bmatrix} A_i^{-1} + B_i^{-1} & -A_i^{-1} \\ -A_i^{-1} & A_i^{-1}(A_i^{-1} + B_i^{-1})A_i^{-1} \end{bmatrix} \begin{bmatrix} p_i - \mu_{p_i|R} \\ t - \mu_{t|R} \end{bmatrix}. \end{aligned}$$

And (5.70) can be proved by summing up the above equations for $i = 1, \dots, N$, and $(t - \mu_{t|R})^T (RCR^T)^{-1} (t - \mu_{t|R})$. \square

The conditional density for the map coordinates $\{p_i\}$ and translation t is written into a Gaussian density in (5.70), with the mean and variance dependent on R . Because the objective is to match an MFG to $(R, t) | \{p'_i\}$, and Gaussian distributions to $p_i | \{p'_i\}$, only the covariance matrices for t and each of p_i are needed. Namely, the diagonal blocks of $\Sigma_{x|R}$ in (5.69) need to be evaluated, whereas the off-diagonal blocks representing correlations are neglected. Using the block inversion formula, the i -th 3-by-3 diagonal block of $\Sigma_{x|R}$ representing the conditional covariance matrix for p_i is

$$\Sigma_{p_i|R} = B_i - B_i \left[(A_i + B_i)^{-1} (A_i + B_i - RCR^T) (A_i + B_i)^{-1} \right] B_i. \quad (5.71)$$

And the last 3-by-3 diagonal block of $\Sigma_{x|R}$ representing the conditional covariance matrix for t is

$$\Sigma_{t|R} = RCR^T. \quad (5.72)$$

In other words, when matching $(R, t) | \{p'_i\}$ to an MFG, the map coordinates $\{p_i\}$ are marginalized; and when matching $p_i | \{p'_i\}$ to a Gaussian distribution, the pose (R, t) and all other $\{p_j\}$ with $i \neq j$ are marginalized.

Before giving necessary moments for the conditional MLE of $t|R$, and the first two moments for p_i , note that the expressions of C and \mathbf{D} in (5.27) and (5.28) are now defined using $\mathbf{A} = \text{diag}(A_1 + RB_1R^T, \dots, A_N + RB_NR^T)$ in (5.66) in this subsection. So a similar lemma as Lemma 5.3 is needed to simplify calculations.

Lemma 5.6. Define $\tilde{C} = RCR^T \in \mathbb{R}^{3 \times 3}$, and $\tilde{\mathbf{D}} \in \mathbb{R}^{3 \times 3N}$ such that the i -th 3-by-3 column block of $\tilde{\mathbf{D}}$ is $\tilde{D}_i = RD_iR^T$. If the expressions of C and \mathbf{D} in (5.27) and (5.28) are defined using $\mathbf{A} = \text{diag}(A_1 + RB_1R^T, \dots, A_N + RB_NR^T)$, then \tilde{C} and $\tilde{\mathbf{D}}$ have the same expressions as C and \mathbf{D} with $\mathbf{A} = \text{diag}(A_1 + B_1, \dots, A_N + B_N)$.

Proof. This lemma is a straightforward extension of Lemma 5.3. \square

With the above lemma, the expressions for \tilde{C} and $\tilde{\mathbf{D}}$ do not dependent on R anymore.

Also, the conditional mean $\mu_{t|R}$ is simplified as follows

$$\mu_{t|R} = \sum_{i=1}^N \tilde{D}_i (Rp'_i - \bar{p}_i). \quad (5.73)$$

With these simplifications, the conditional mean $\mu_{t|R}$ and covariance matrix $\Sigma_{t|R}$ have exactly the same expression as in (5.46), so the moments $E[t]$, $E[tv_R^T]$, and $E[tt^T]$ according to (5.70), that are necessary for the conditional MLE, can be calculated as in Theorem 5.5 or Theorem 5.6, with C , D_i , and p_i replaced by \tilde{C} , \tilde{D}_i , and \bar{p}_i , depending on whether $A_1 + B_1 = \dots = A_N + B_N$. Also, the moments $E[p_i]$ and $E[p_ip_i^T]$ are given as follows.

Theorem 5.8. *The moments $E[p_i]$ and $E[p_ip_i^T]$ according to (5.70) can be calculated as*

$$E[p_i] = (I_{3 \times 3} - K_i)\bar{p}_i + K_i(E[R]p'_i + E[\mu_{t|R}]), \quad (5.74)$$

$$\begin{aligned} E[p_ip_i^T] &= \Sigma_{p_i|R} + (I_{3 \times 3} - K_i)\bar{p}_i\bar{p}_i^T(I_{3 \times 3} - K_i)^T + (I_{3 \times 3} - K_i)\bar{p}_i(E[R]p'_i + E[\mu_{t|R}])^T K_i^T \\ &\quad + K_i \left(E[Rp'_i(p'_i)^T R^T] + E[Rp'_i \mu_{t|R}^T] + E[\mu_{t|R}(p'_i)^T R^T] + E[\mu_{t|R} \mu_{t|R}^T] \right) K_i^T \\ &\quad + K_i(E[R]p'_i + E[\mu_{t|R}])\bar{p}_i^T(I_{3 \times 3} - K_i)^T, \end{aligned} \quad (5.75)$$

where $E[\mu_{t|R}] = E[t]$, $E[\mu_{t|R}\mu_{t|R}^T] = E[tt^T] - \tilde{C}$. Also, if $A_1 + B_1 = \dots = A_N + B_N$, then

$$E[\mu_{t|R}(p'_i)^T R^T] = \bar{p}_c(p'_i)^T E[R]^T - E[Rp'_c(p'_i)^T R^T], \quad (5.76)$$

otherwise

$$E[\mu_{t|R}(p'_i)^T R^T] = \sum_{j=1}^N \tilde{D}_j \left(E[Rp'_j(p'_i)^T R^T] - \bar{p}_j(p'_i)^T E[R]^T \right). \quad (5.77)$$

Proof. These moments can be calculated by integrating the left hand side with respect to the conditional density (5.70), and noting that the marginal distribution of a multi-variate Gaussian distribution is also Gaussian with the corresponding mean and covariance matrix.

Table 5.12: Pose and map coordinates estimation with map noises

```

1: procedure ( $\mathcal{MG}, \{\mathcal{N}_i^+\}$ ) = POSE-MAP ESTIMATION( $\{p'_i\}, \{\mathcal{N}_i(\bar{p}_i, B_i)\}$ )
2:   Obtain  $F = \text{Attitude Estimation}(\{\bar{p}_i\}, \{p'_i\})$  in Table 5.6 with  $A_i$  replaced by  $A_i + B_i$ ,
   and  $\exp\left\{-\frac{1}{2}f_R(R)\right\}$  by the first term on the right hand side of (5.68).
3:   Let  $F = USV^T$  be its pSVD.
4:   Calculate  $E[v_R v_R^T]$  as in (3.21) using  $S$ .
5:   if  $A_1 + B_1 = \dots = A_N + B_N$  then
6:     Calculate  $E[t], E[tv_R^T],$  and  $E[tt^T]$  in Theorem 5.6, with  $\frac{\sigma^2}{N}I_{3 \times 3}$  replaced by
    $\frac{A+B}{N}$ , and  $p_i$  replaced by  $\bar{p}_i$ .
7:   else
8:     Calculate  $E[t], E[tv_R^T],$  and  $E[tt^T]$  in Theorem 5.5, with  $C, D_i$ , and  $p_i$  replaced
   by  $\tilde{C}, \tilde{D}_i$ , and  $\bar{p}_i$ .
9:   end if
10:  Obtain  $\mu, \Sigma, P$  according to Theorem 3.6, with  $E[t], E[tv_R^T], E[tt^T], E[v_R v_R^T]$ , and
    $E[v_R] = 0$ .
11:  Set  $\mathcal{MG} = \mathcal{MG}(\mu, \Sigma, P, U, S, V)$ .
12:  For  $i = 1, \dots, N$ , calculate  $E[p_i]$  and  $E[p_i p_i^T]$  in Theorem 5.8.
13:  Set  $\mathcal{N}_i^+$  as  $\mathcal{N}(E[p_i], E[p_i p_i^T] - E[p_i]E[p_i]^T)$ .
14: end procedure

```

Then the results are further integrated with respect to the marginal distribution for R , which has already been matched to a matrix Fisher distribution with parameter F . \square

With the moments $E[p_i]$ and $E[p_i p_i^T]$, the distribution for p_i can be matched to a Gaussian distribution with the same first and second order moments. In summary, the posterior density (5.64) has been approximated by an MFG for the pose (R, t) , and N Gaussian distributions for the map coordinates $\{p_i\}_{i=1}^N$. The pseudocode is summarized in Table 5.12.

Next, then case when there is a prior density for the pose (R, t) is also considered, so that the algorithm can be applied in a recursive filter. Similar to Chapter 5.2.1, let $y = [t^T, a^T]^T \in \mathbb{R}^n$, suppose (R, y) follows an MFG with the parameter $(\mu, \Sigma, P, U, S, V)$ before measurements, where a includes other Euclidean quantities, such as linear velocity and sensor biases. According to the Bayes' formula, the posterior density for the pose (R, y)

and map coordinates $\{p_i\}$ after incorporating the measurements $\{p'_i\}$ is

$$p(R, y, \{p_i\} | \{p'_i\}) = \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (R^T(p_i - Hy) - p'_i)^T A_i^{-1} (R^T(p_i - Hy) - p'_i) \right\} \\ \cdot \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (p_i - \bar{p}_i)^T B_i^{-1} (p_i - \bar{p}_i) \right\} \cdot \text{etr}(FR^T) \exp \left\{ -\frac{1}{2} (y - \mu_c)^T \Sigma_c^{-1} (y - \mu_c) \right\}, \quad (5.78)$$

where F, μ_c, Σ_c are defined with respect to $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$. According to Lemma 5.4, Lemma 5.5, and (5.68), the above density can be decomposed into

$$p(R, y, \{p_i\} | \{p'_i\}) = \text{etr}(FR^T) \exp \left\{ -\frac{1}{2} (\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{TAT}^T)^{-1} (\mathbf{q}' - \mathbf{q}_m) \right\} \\ \cdot \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_{x|R})^T \Sigma_{x|R}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{x|R}) \right\} \exp \left\{ -\frac{1}{2} (y - \mu_c)^T \Sigma_c^{-1} (y - \mu_c) \right\}. \quad (5.79)$$

As before, this density function is factorized into a marginal density for R , and a conditional density for $t|R$, and $\{p_i\}|R$.

Theorem 5.9. Let $z = [\mathbf{x}^T, a^T]^T \in \mathbb{R}^{3N+n}$, $H = [I_{(3N+3) \times (3N+3)}, 0_{(3N+3) \times (n-3)}]$, and $G = [0_{n \times 3N}, I_{n \times n}]$. Define $\boldsymbol{\mu}_{z|R}$ as

$$\boldsymbol{\mu}_{z|R} = \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} \left(H^T \Sigma_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + G^T \Sigma_c^{-1} \boldsymbol{\mu}_c \right), \quad (5.80)$$

and $\Sigma_{z|R}$ as

$$\Sigma_{z|R} = \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1}. \quad (5.81)$$

Also, let $\boldsymbol{\mu}_{c,t}$ and $\Sigma_{c,t}$ be the appropriate blocks in $\boldsymbol{\mu}_c$ and Σ_c for t . Then the posterior density

$p(R, y, \{p_i\} | \{p'_i\})$ can be written as

$$\begin{aligned} p(R, y, \{p_i\} | \{p'_i\}) &= \text{etr}(FR^T) \exp \left\{ -\frac{1}{2}(\mathbf{q}' - \mathbf{q}_m)^T (\mathbf{TAT}^T)^{-1}(\mathbf{q}' - \mathbf{q}_m) \right\} \\ &\cdot \exp \left\{ -\frac{1}{2}(\mu_{t|R} - \mu_{c,t})^T (\Sigma_{t|R} + \Sigma_{c,t})^{-1}(\mu_{t|R} - \mu_{c,t}) \right\} \exp \left\{ -\frac{1}{2}(z - \mu_{z|R})^T \Sigma_{z|R}^{-1}(z - \mu_{z|R}) \right\}. \end{aligned} \quad (5.82)$$

Proof. The proof is given in Appendix G. \square

The key point is that the first three terms on the right hand side of (5.82) do not depend on z , so they can be matched to a matrix Fisher distribution using the progressive update method in Table 5.2. As before, let the approximated parameter of the matrix Fisher distribution be denoted by F^+ . And let $F^+ = U^+S^+(V^+)^T$ be its pSVD, define $v_R = (Q^+S^+ - S^+(Q^+)^T)^\vee$ for MFGI, or $v_R = (S^+Q^+ - (Q^+)^TS^+)^\vee$ for MFGB, with $Q^+ = (U^+)^TR^+V^+$. Similar to the case without pose prior, the next step is to calculate the moments for $y|R$ with $\{p_i\}$ marginalized out; and the moments for p_i , with (R, y) and other $\{p_j\}, i \neq j$, marginalized out. These moments are according to the posterior density (5.82), and to distinguish with those according to the prior distribution $\mathcal{MG}(\mu, \Sigma, P, U, S, V)$, they are denoted by $E[\cdot | \mathcal{Z}]$, where $\mathcal{Z} = \{p'_i\}_{i=1}^N$ are all measurements. The calculations for these moments are relegated to Theorem G.1 in Appendix G.

Using the moments $E[y|\mathcal{Z}]$, $E[y(v_R^+)^T|\mathcal{Z}]$, and $E[yy^T|\mathcal{Z}]$, the posterior density for (R, t) is matched to an MFG using the conditional MLE in Theorem 3.6. And using the moments $E[p_i|\mathcal{Z}]$ and $E[p_ip_i^T|\mathcal{Z}]$, the posterior density for p_i is matched to a Gaussian distribution, for $i = 1, \dots, N$. The pseudocode is summarized in Table 5.13.

5.2.4 Numerical Simulations With Map Uncertainty

The proposed method in the previous subsection that uses an MFG to model the pose that aligns two points set with noises is compared with an LM optimization method in this subsection.

Table 5.13: Measurement update with map noises

1:	procedure $(\mathcal{MG}^+, \{\mathcal{N}_i^+\}) = \text{MEASUREMENT UPDATE}(\mathcal{MG}^-, \{\mathcal{N}_i^-(\bar{p}_i, B_i)\}, \{p'_i\})$
2:	Let $F^- = U^- S^- (V^-)^T$.
3:	Calculate $F^+ = \text{Attitude Update}(F^-, \{\bar{p}_i\}, \{p'_i\})$ in Table 5.8, where $f_m(R)$ is the second and third term on the right hand side of (5.82).
4:	Let the pSVD of F^+ be $F^+ = U^+ S^+ (V^+)^T$.
5:	Calculate $E[y \mathcal{Z}]$, $E[y(v_R^+)^T \mathcal{Z}]$, and $E[yy^T \mathcal{Z}]$ in Theorem G.1.
6:	Obtain μ^+ , Σ^+ , and P^+ using the conditional MLE in Theorem 3.6.
7:	Set $\mathcal{MG}^+ = \mathcal{MG}(\mu^+, \Sigma^+, P^+, U^+, S^+, V^+)$.
8:	For $i = 1, \dots, N$, calculate $E[p_i \mathcal{Z}]$ and $E[p_i p_i^T \mathcal{Z}]$ in Theorem G.1.
9:	Set \mathcal{N}_i^+ as $\mathcal{N}(E[p_i \mathcal{Z}], E[p_i p_i^T \mathcal{Z}] - E[p_i \mathcal{Z}] E[p_i \mathcal{Z}]^T)$.
10:	end procedure

First, the estimation algorithm in Table 5.12 without pose prior is tested. The true attitude is set as $R_t = \exp\left\{\frac{\pi}{2}\hat{e}_3\right\}$, and the true translation is $t_t = [5, 0, 0]^T$. There are 10 map points, scattered as $\bar{p}_i = t_t + \xi_i$, where $\xi_i \in \mathbb{R}^3$ is a three dimensional uniform distribution on $[0, 1]^3$. The prior noise of p_i are grouped into two levels, N_1 of them have small noise as $B_i = 0.1^2 I_{3 \times 3}$, and $10 - N_1$ of them have large noise as $B_i = 1^2 I_{3 \times 3}$. The measurement noise is set as $A_i = 0.1^2 I_{3 \times 3}$ for all $i = 1, \dots, 10$. The number N_1 is varied in $\{0, 2, 4, 6, 8, 10\}$. For each N_1 , a thousand Monte Carlo simulations are carried out. Paired t -tests are used to compare the attitude, translation, and map coordinate errors of the MFG algorithm and LM optimization.

The estimation errors are summarized in Table 5.14. Generally speaking, the MFG algorithm has slightly worse accuracy in pose estimation, especially when $N_1 = 4$ and 6, indicating it may have degraded performance when the map prior has mixed uncertainty levels. But when $N_1 = 0$, the translation error for MFG is significantly smaller than LM optimization, which is consistent with the results in Table 5.9. Namely, when the pose has large uncertainty, the MFG is better at estimating the translation. Next, looking at the map coordinate errors, in general the MFG algorithm is more accurate, especially for those points with large prior noises, where the differences are significant for all cases. The reason for this better performance of MFG is to be investigated in future works, but a guess is that

Table 5.14: Pose and map coordinates errors with map uncertainty

estimator	MFG		
N_1	10	8	6
att err (deg)	9.98 ± 4.21 ($p=0.061$)	10.7 ± 4.9 ($p=0.213$)	13.4 ± 5.9 ($p<0.001$)
pos err	0.173 ± 0.082 ($p=0.240$)	0.198 ± 0.098 ($p=0.048$)	0.237 ± 0.112 ($p<0.001$)
map err (N_1)	0.1237 ± 0.0165 ($p<0.001$)	0.126 ± 0.020 ($p<0.001$)	0.130 ± 0.024 ($p<0.001$)
map err ($10 - N_1$)	-	0.187 ± 0.057 ($p<0.001$)	0.205 ± 0.050 ($p<0.001$)
estimator	LM optimization		
N_1	10	8	6
att err (deg)	9.96 ± 4.20	10.7 ± 4.9	13.0 ± 5.8
pos err	0.174 ± 0.083	0.197 ± 0.099	0.230 ± 0.113
map err (N_1)	0.1241 ± 0.0165	0.127 ± 0.020	0.131 ± 0.024
map err ($10 - N_1$)	-	0.191 ± 0.058	0.209 ± 0.050
estimator	MFG		
N_1	4	2	0
att err (deg)	16.5 ± 7.9 ($p<0.001$)	50.6 ± 30.9 ($p<0.001$)	77.1 ± 41.1 ($p=0.056$)
pos err	0.305 ± 0.147 ($p<0.001$)	0.755 ± 0.372 ($p=0.005$)	1.05 ± 0.38 ($p<0.001$)
map err (N_1)	0.138 ± 0.030 ($p=0.033$)	0.149 ± 0.046 ($p<0.001$)	-
map err ($10 - N_1$)	0.225 ± 0.056 ($p<0.001$)	0.501 ± 0.189 ($p<0.001$)	0.661 ± 0.173 ($p<0.001$)
estimator	LM optimization		
N_1	4	2	0
att err (deg)	16.0 ± 7.5	48.6 ± 29.5	76.4 ± 40.8
pos err	0.294 ± 0.149	0.772 ± 0.455	1.17 ± 0.55
map err (N_1)	0.138 ± 0.030	0.153 ± 0.048	-
map err ($10 - N_1$)	0.230 ± 0.056	0.518 ± 0.216	0.708 ± 0.196

the gradient based optimization may not be very effective for high dimensional variable space, which is 66 in this simulation study; or that the parameters for the LM method is not properly tuned for this high dimensional space.

Next, the measurement update algorithm in Table 5.13 is combined with the uncertainty propagation for IMU in Table 5.1 into a Bayesian filter for visual-inertial navigation. And this filter is compared with an MEKF based SLAM algorithm which also estimates the map point coordinates. In this simulation study, a vehicle is supposed to follow the same trajectory and rotational motion introduced in the simulation in Chapter 5.2.2. The distribution of landmarks, noise parameters for the IMU, sampling frequencies for the IMU and camera are also chosen to be the same. There are in total 400 landmarks, and the measurement noise of the camera is set as $A_i = 0.05^2 I_{3 \times 3}$. Only the landmarks that are within 3 m from the vehicle are measured by the camera. The simulation lasts for forty seconds.

Three cases are tested. The first case is a *normal* condition: the prior noise for all map coordinates are $B_i = 0.1^2 I_{3 \times 3}$. The initial attitude, velocity, and position uncertainty for MEKF is set as $0.1^2 I_{3 \times 3}$; and the initial attitude uncertainty for MFG is $S_0 = 50 I_{3 \times 3}$. The initial attitude, velocity, and position estimates are set as their true values plus random noises sampled from those initial uncertainty distributions. Namely, the normal condition corresponds to small prior map noises, and initial conditions with relatively small uncertainty. In the second case, the initial conditions are the same as the normal case, but the prior map noise is $B_i = 1^2 I_{3 \times 3}$, i.e., this corresponds to large map uncertainty. In the third case, the map uncertainty is the same as the normal case. But the initial attitude is rotated from the true attitude around the inertial vertical axis by 180 deg, simulating the scenario when the initial heading direction is completely unknown. The initial attitude uncertainty for MFG is $S_0 = 0.01 I_{3 \times 3}$, and for MEKF is $1000 I_{3 \times 3}$. One hundred Monte Carlo simulations are run for each of the three cases, and paired t -test ($\alpha = 0.001$) is used to detect statistical significance.

The estimation results for the two filters are summarized in Table 5.15. For the first two

Table 5.15: Estimation error for visual-inertial navigation with map noises

case		normal	
filter		MFG	MEKF
att err (deg)		0.65 ± 0.25 ($p < 0.001$)	0.73 ± 0.32
vel err (m/s)		0.078 ± 0.005 ($p = 0.004$)	0.079 ± 0.005
pos err (m)		0.042 ± 0.015 ($p < 0.001$)	0.048 ± 0.020
mean map err (m)		0.044 ± 0.014 ($p < 0.001$)	0.051 ± 0.019
case		large map uncertainty	
filter		MFG	MEKF
att err (deg)		4.36 ± 2.07 ($p < 0.001$)	5.12 ± 2.08
vel err (m/s)		0.158 ± 0.054 ($p < 0.001$)	0.179 ± 0.056
pos err (m)		0.314 ± 0.149 ($p = 0.006$)	0.355 ± 0.152
mean map err (m)		0.384 ± 0.182 ($p = 0.002$)	0.441 ± 0.185
case		large initial uncertainty	
filter		MFG	MEKF
att err (deg)		1.07 ± 0.22 ($p < 0.001$)	101.64 ± 0.23
vel err (m/s)		0.082 ± 0.004 ($p < 0.001$)	1.610 ± 0.005
pos err (m)		0.041 ± 0.014 ($p < 0.001$)	5.033 ± 0.014
mean map err (m)		0.043 ± 0.013 ($p < 0.001$)	6.167 ± 0.010

cases, the MFG filter is slightly more accurate than MEKF, but their differences relative to the magnitude of errors are small. This may be attributed to that the MFG filter has slightly better accuracy in map coordinate estimation as seen in Table 5.14, and the subsequent pose estimation relies on the map coordinates estimated in previous steps. For the third case when the initial heading direction is completely reversed, the MFG filter performs much better than MEKF. The estimation error for one particular simulation is presented in Figure 5.6, and the estimated trajectory of the vehicle and final map point coordinates are shown in Figure 5.7. Similar to the simulation without prior map noises (Figure 5.4), the MEKF suffers from convergence problems. Moreover, since MEKF also estimates map coordinates in this simulation, they are also adversely affected by the wrong attitude estimation. These large errors in the estimated map coordinates further contribute to large pose errors in the subsequent estimation, making the filter unable to converge at all. Therefore, as seen in Figure 5.6, the estimation error for MEKF becomes periodic as the vehicle circles around, i.e., it has smaller errors near where it starts because of the smaller map error. This signifies

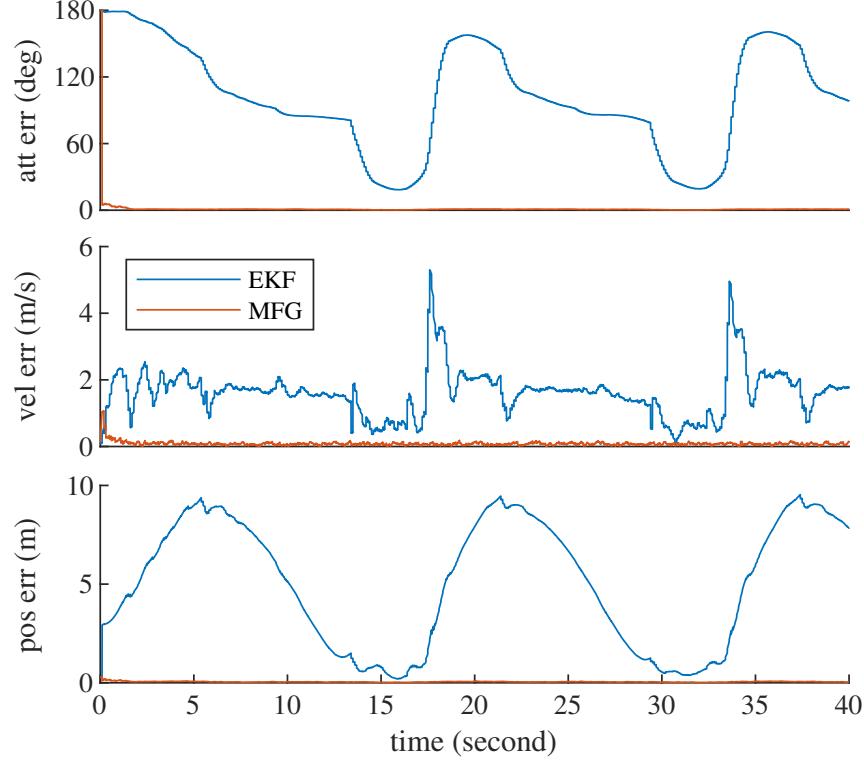


Figure 5.6: Estimation error of visual-inertial navigation with map noises.

the advantage of MFG filter that has very fast convergence speed thanks to its better accuracy when modeling large attitude uncertainty, and linearization free measurement update.

Finally, to test the proposed algorithm in visual-inertial odometry settings where there is no prior map information, the unknown map point is constructed using its measurement in the body-fixed frame by the camera. More specifically, suppose the pose (R, t) has a density function $p_{R,t}(R, t)$, and the map point p_i is measured as $p'_i = R^T(p_i - t)$ in the body-fixed frame, with a density function $p_{p'_i}(p'_i)$. Then using convolution, it can be shown that the probability density function of p_i is

$$p_{p_i}(p_i) = \int_{R \in \text{SO}(3)} \int_{t \in \mathbb{R}^3} p_{R,t}(R, t) p_{p'_i}(R^T(p_i - t)) dt dR. \quad (5.83)$$

In particular, if (R, t) follows an MFG, and p'_i follows a Gaussian distribution, the density of

— true trajectory	• true landmarks
— MEKF trajectory	• MEKF landmarks
— MFG trajectory	• MFG landmarks

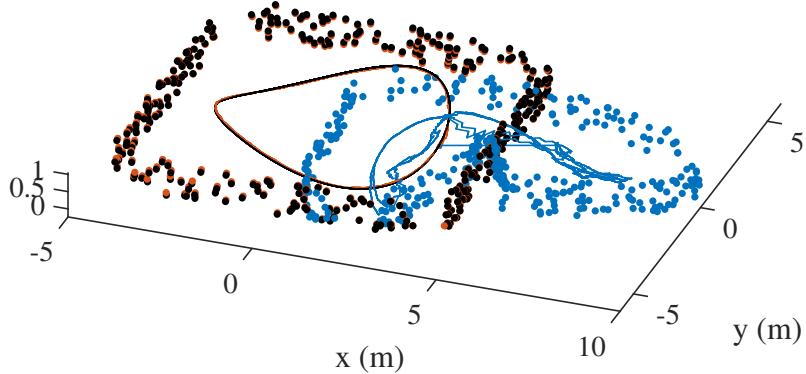


Figure 5.7: Trajectory and landmarks of visual-inertial navigation with map noises.

p_i becomes very complicated. But its first and second order moments can be calculated as

$$\mathbb{E}[p_i] = \mathbb{E}[R] p'_i + \mathbb{E}[t] \quad (5.84)$$

$$\mathbb{E}[p_i p_i^T] = \mathbb{E}[R p'_i (p'_i)^T R^T] + \mathbb{E}[R p'_i t^T] + \mathbb{E}[t (p'_i)^T R^T] + \mathbb{E}[t t^T]. \quad (5.85)$$

So the distribution of p_i can be matched to $\mathcal{N}(\mathbb{E}[p_i], \mathbb{E}[p_i p_i^T] - \mathbb{E}[p_i] \mathbb{E}[p_i]^T)$. But unfortunately, using Gaussian to model the distribution of p_i is very inaccurate when R has large uncertainty. For example, if R has a uniform distribution on $\text{SO}(3)$, i.e., when $S = 0_{3 \times 3}$, the shape of the density for p_i is a spherical thin shell centered around t , which cannot be approximated by a Gaussian distribution.

As a consequence, for the third simulation case when the attitude error is initialized as $S = 0.01I_{3 \times 3}$, the resulting uncertainty of the map coordinates cannot be accurately modeled by Gaussian distributions. In order to circumvent this problem but still test the proposed algorithm in visual-inertial odometry settings, in the next simulation, only the features measured in the first second are assumed to have a map prior, and subsequent map coordinates are built incrementally from the camera measurements, modeled by Gaussian

Table 5.16: Estimation error for visual-inertial odometry with map noises

case	normal	
filter	MFG	MEKF
att err (deg)	1.16 ± 0.55 ($p=0.066$)	1.25 ± 0.66
vel err (m/s)	0.086 ± 0.012 ($p=0.054$)	0.088 ± 0.016
pos err (m)	0.103 ± 0.055 ($p=0.310$)	0.108 ± 0.063
mean map err (m)	0.109 ± 0.054 ($p=0.221$)	0.115 ± 0.061
case	large map uncertainty	
filter	MFG	MEKF
att err (deg)	4.47 ± 2.07 ($p<0.001$)	5.18 ± 2.07
vel err (m/s)	0.161 ± 0.054 ($p<0.001$)	0.180 ± 0.056
pos err (m)	0.321 ± 0.150 ($p=0.009$)	0.360 ± 0.150
mean map err (m)	0.393 ± 0.183 ($p=0.003$)	0.448 ± 0.184
case	large initial uncertainty	
filter	MFG	MEKF
att err (deg)	1.56 ± 0.55 ($p<0.001$)	179.0 ± 0.24
vel err (m/s)	0.090 ± 0.013 ($p<0.001$)	2.328 ± 0.005
pos err (m)	0.098 ± 0.056 ($p<0.001$)	9.958 ± 0.019
mean map err (m)	0.104 ± 0.053 ($p<0.001$)	11.55 ± 0.015

distributions with the first and second moments.

The estimation errors for the three cases are summarized in Table 5.16. For the normal, and large map uncertainty cases, the MFG filter is slightly more accurate than MEKF, but most differences are statistically insignificant, indicating that the MFG and Gaussian distributions work similarly well for these cases when the attitude uncertainty is relatively small. However, coming to the third case where the initial attitude uncertainty is very large, the MEKF again does not converge. This is indicated in Figure 5.8, where the estimation errors for a particular simulation is presented; and in Figure 5.9, where the trajectory of the vehicle and the estimated feature positions are shown. In this simulation, because there are no map priors for the features measured after one second, the trajectory and map coordinates estimated by MEKF are exactly rotated around the vertical axis by 180 deg, i.e., they are not constrained by the map priors as in Figure 5.7. This is usually considered to be successful in visual-inertial odometry algorithms since the shape of the trajectory is accurately estimated. On the other hand, for the MFG filter, because the attitude successfully converges within

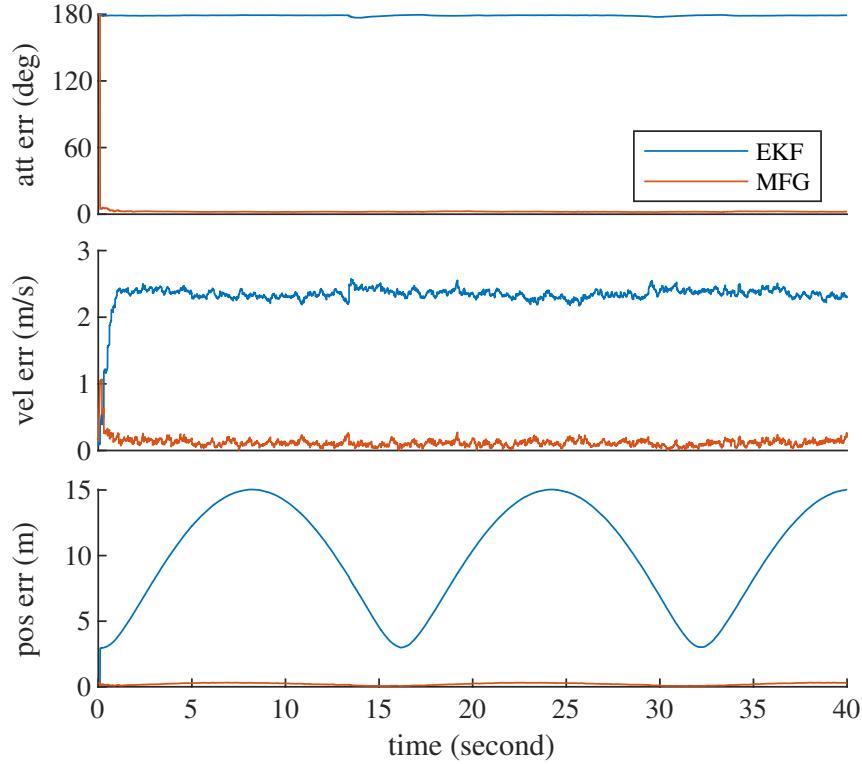


Figure 5.8: Estimation error of visual-inertial odometry with map noises.

one second when the map priors are available for the measured features, the subsequent trajectory and map coordinates follow around their respective true values.

As a summary, the MFG filter when applied to visual-inertial navigation or odometry problems, has similar accuracy with MEKF based algorithms when the initial attitude uncertainty is small. Nevertheless, if the initial attitude uncertainty is very large, for example when the heading direction is completely unknown, the MFG filter has a much faster convergence rate from a wrong initial heading direction.

5.2.5 Applications to the VCU-RVI Dataset

The algorithms proposed in Chapter 5.2.1 and Chapter 5.2.3 are also tested with a real world IMU-RGBD dataset for visual-inertial odometry performance. The VCU_RVI dataset is used [132], which is the only available dataset that has synchronized IMU and RGB-D camera measurements, together with ground truth from a motion capture system. The data

— true trajectory	• true landmarks
— MEKF trajectory	• MEKF landmarks
— MFG trajectory	• MFG landmarks

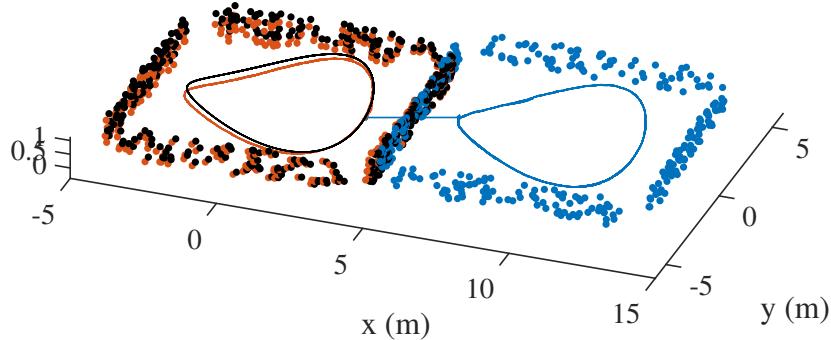


Figure 5.9: Trajectory and features of visual-inertial odometry with map noises.

are collected with a BMI1055 IMU at 100Hz, and a Structure Core RGB-D camera that provides the 3D coordinates of captured features in the camera’s body-fixed frame at 30Hz. The five trials “lab_bumper1” to “lab_bumper5” are used, where the IMU and camera are attached to a wheeled robot, moving around an indoor environment installed with speed bumps to provide 3D motions.

Next, some implementation details of the algorithms are described. The map is composed of a set of previously detected points, including their coordinates in the inertial frame (and their uncertainties for the algorithm in Chapter 5.2.3), their ORB features [91], their viewing directions (a unit vector connecting the camera and the position of the point in inertial frame), and time indices denoting the last frame in which they were detected. The visual features are detected using the MATLAB `detectORBFeatures` function, and extracted into ORB features. After valid points are detected for the current frame, they are matched to the map. Before matching, the search range is refined to those map points satisfying both the following two conditions: (i) the distance from the camera to the point is between 0.5m and 5m; (ii) the difference of viewing directions is within 60deg. The distance and viewing directions are calculated using the predicted position of the camera using IMU. Then, the

ORB features from the current frame, and the ORB features from the reduced map are matched using the MATLAB `matchFeatures` function, with the `unique` parameter set as true, which means no feature will be matched with more than one features. RANSAC is then applied to reduce mismatches. The captured points that have been successfully matched to the map are used as the measurements in the current frame to correct the pose, and the map coordinates if the algorithm in Chapter 5.2.3 is used.

After matching, all valid points detected in the current frame, which are not matched to any point in the map, are transformed into the inertial frame using the estimated pose, and are added to the map. Also, for the map points which are matched to the map in the current frame, their time indices are renewed to be the index of the current frame. Then, all map points with time indices more than 10 frames older than the current frame, are deleted from the map. In other words, loop closure is not considered in this test.

The algorithm proposed in Chapter 5.2.1, where the map coordinates are fixed without any uncertainty, is denoted by “MFGd”, and the corresponding MEKF is denoted by “MEKFd”. The algorithm in Chapter 5.2.3, where the map coordinates are random variables with uncertainties, is denoted by “MFGr”, and the corresponding MEKF is denoted by “MEKFr”. The definition MFGI is used in the MFG filters. The white noise for the gyroscope is set as $0.1 \text{ deg}/\sqrt{\text{s}}$, and for the accelerometer is $0.1 \text{ m/s}/\sqrt{\text{s}}$. It is assumed that the gyroscope and accelerometer have zero biases, so the bias is not estimated in the filter. The noise for the camera measurements is set as $A_i = 0.002^2 I_{3 \times 3}$. The attitude, velocity, and position of the robot is initialized as true values, and the initial uncertainties for them are set as $0.1^2 I_{3 \times 3}$.

The estimation results for the five trials: “lab_bumper1” to “lab_bumper5” are presented in Figure 5.10 to Figure 5.14, and the final estimation errors are listed in Table 5.17. It is seen that the tested four filtering algorithms have very similar estimation accuracy for the tested five trials. In particular, the algorithms with map uncertainty behave better in the first and fifth trials, while worse in the other three, than the algorithms with fixed map

Table 5.17: Estimation errors for the VCU_RVI dataset

trials		MFKFd	MFGd	MEKFr	MFGGr
lab_bumper1	att err (deg)	9.89	9.75	9.16	8.53
	pos err (m)	0.399	0.395	0.264	0.281
lab_bumper2	att err (deg)	3.33	2.73	4.05	6.09
	pos err (m)	0.069	0.095	0.069	0.076
lab_bumper3	att err (deg)	9.41	9.41	10.55	13.37
	pos err (m)	0.676	0.677	0.672	0.738
lab_bumper4	att err (deg)	10.26	10.22	13.40	18.38
	pos err (m)	0.919	0.863	0.975	0.984
lab_bumper5	att err (deg)	7.95	8.68	6.21	6.47
	pos err (m)	0.635	0.663	0.610	0.561

coordinates. The performance of the MFG filter and MEKF when the map is fixed is almost indistinguishable; whereas if the map coordinates have uncertainty, the MFG filter seems to be a little worse than MEKF.

As discussed in Chapter 5.2.3, the uncertainty of map coordinates cannot be modeled by Gaussian distributions when the attitude uncertainty is large. Therefore, for this odometry problem, all tested algorithms, including the MFG filters, cannot deal with wrong initial attitude as in the simulation studies. Also, this testing with real world dataset is still too preliminary to draw any conclusions, as the parameters for the filters are not well tuned, and the sensor biases are not estimated. These will be pursued in future works.

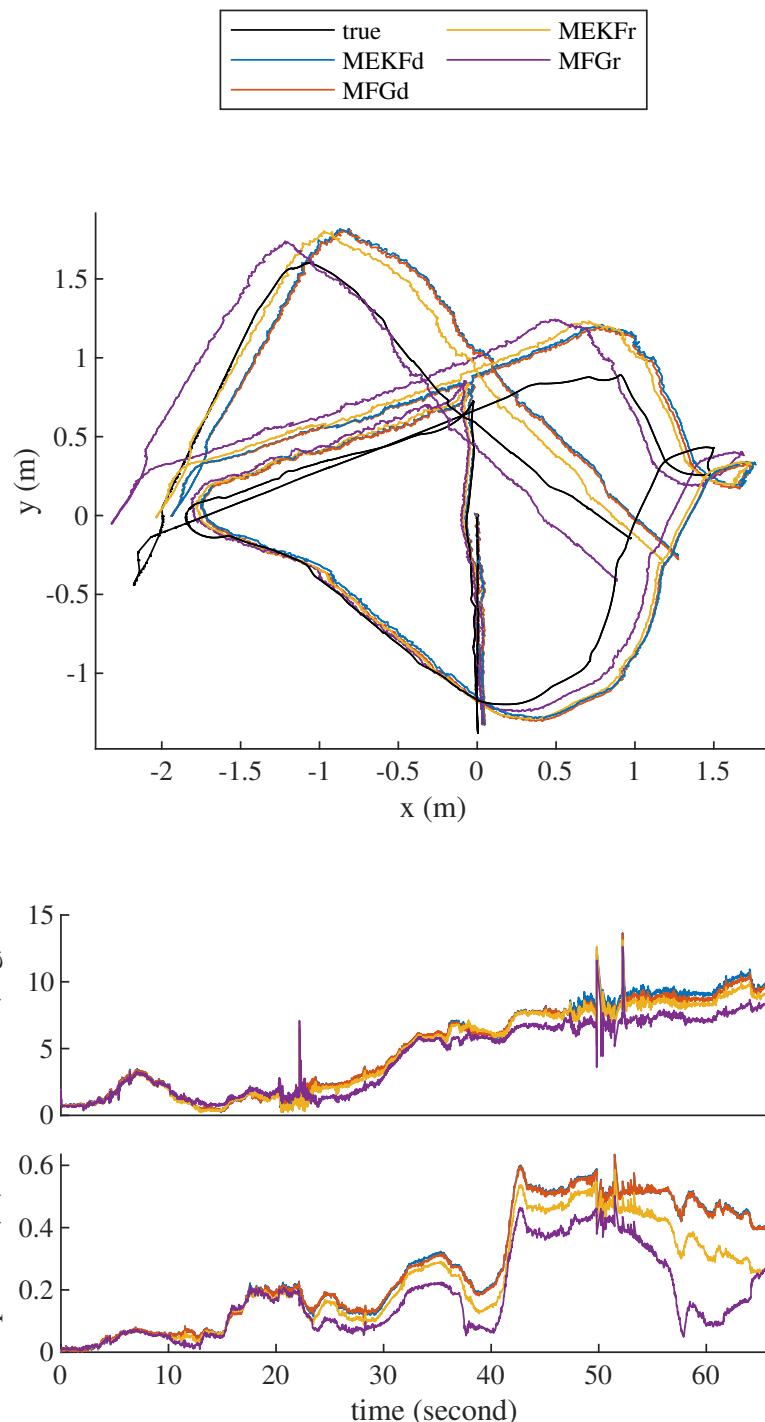


Figure 5.10: Estimation errors for the “lab_bumper1” trial. The top row is the true and estimated trajectories.

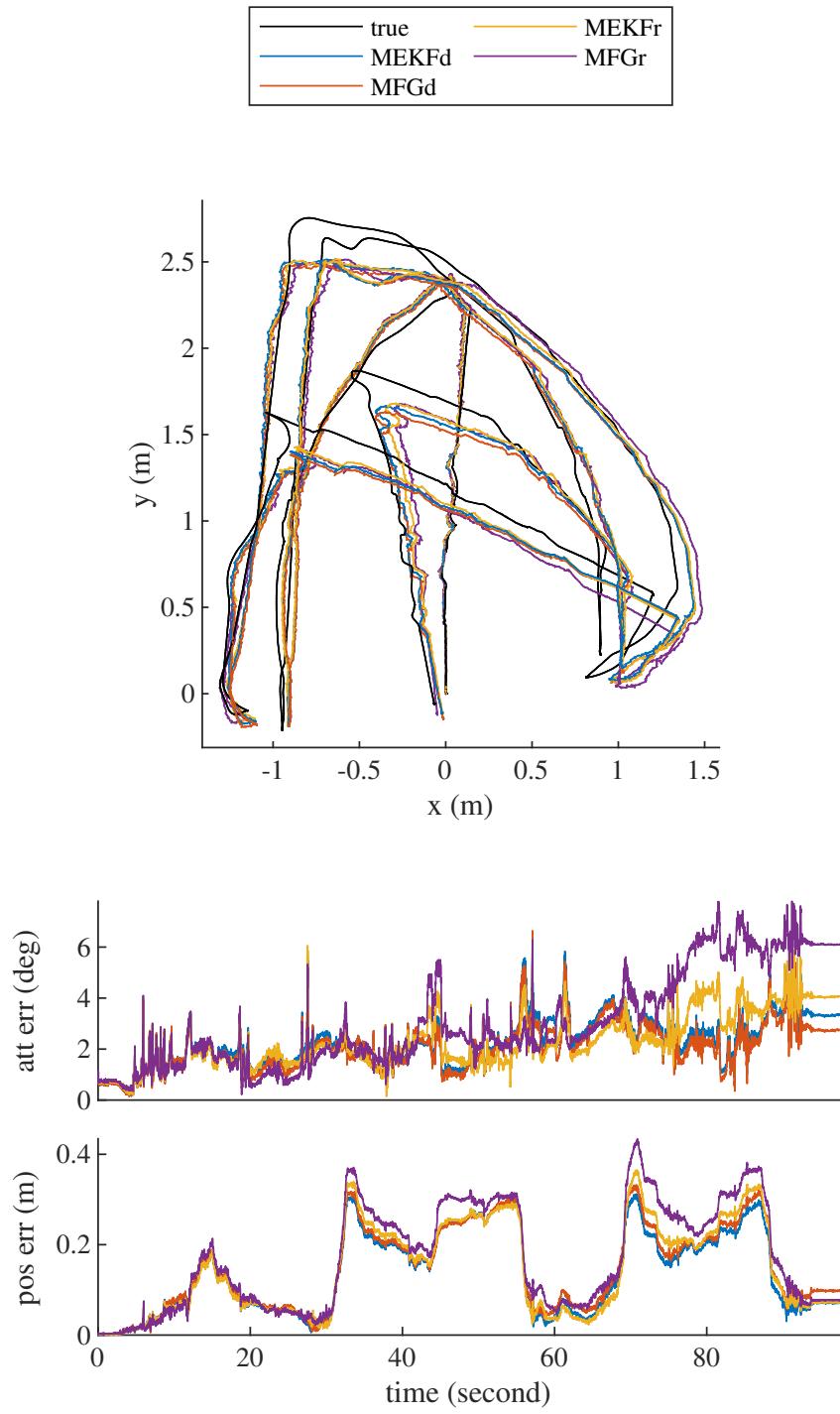


Figure 5.11: Estimation errors for the “lab_bumper2” trial. The top row is the true and estimated trajectories.

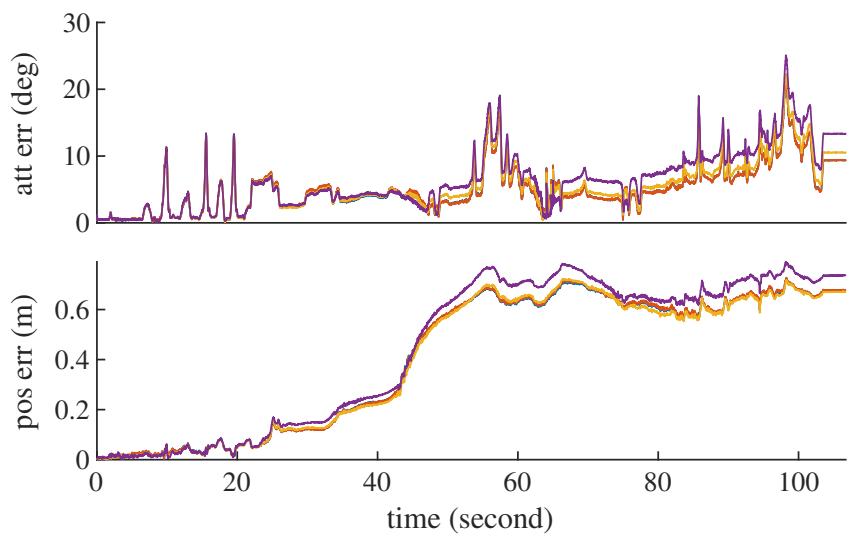
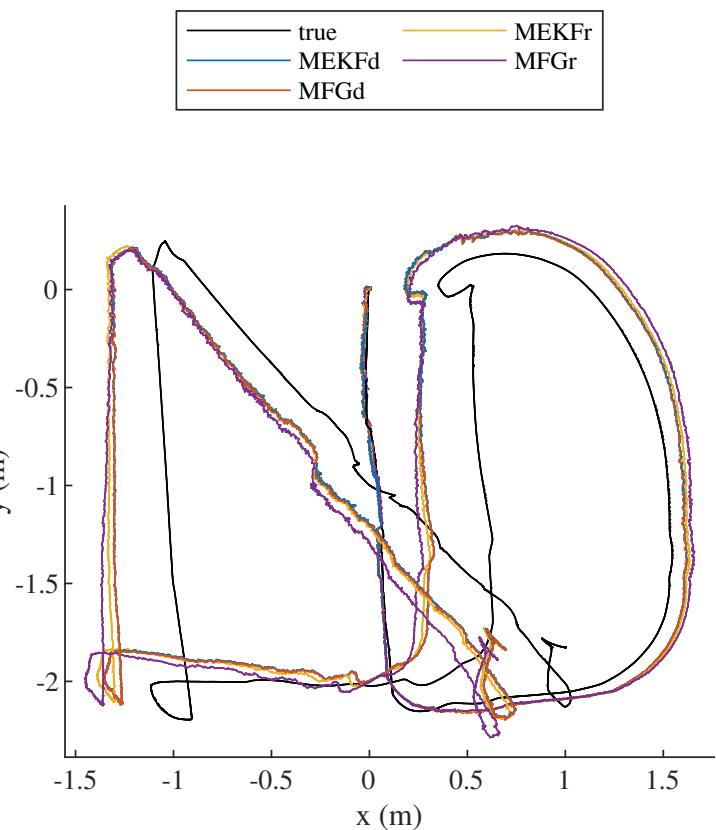


Figure 5.12: Estimation errors for the “lab_bumper3” trial. The top row is the true and estimated trajectories.

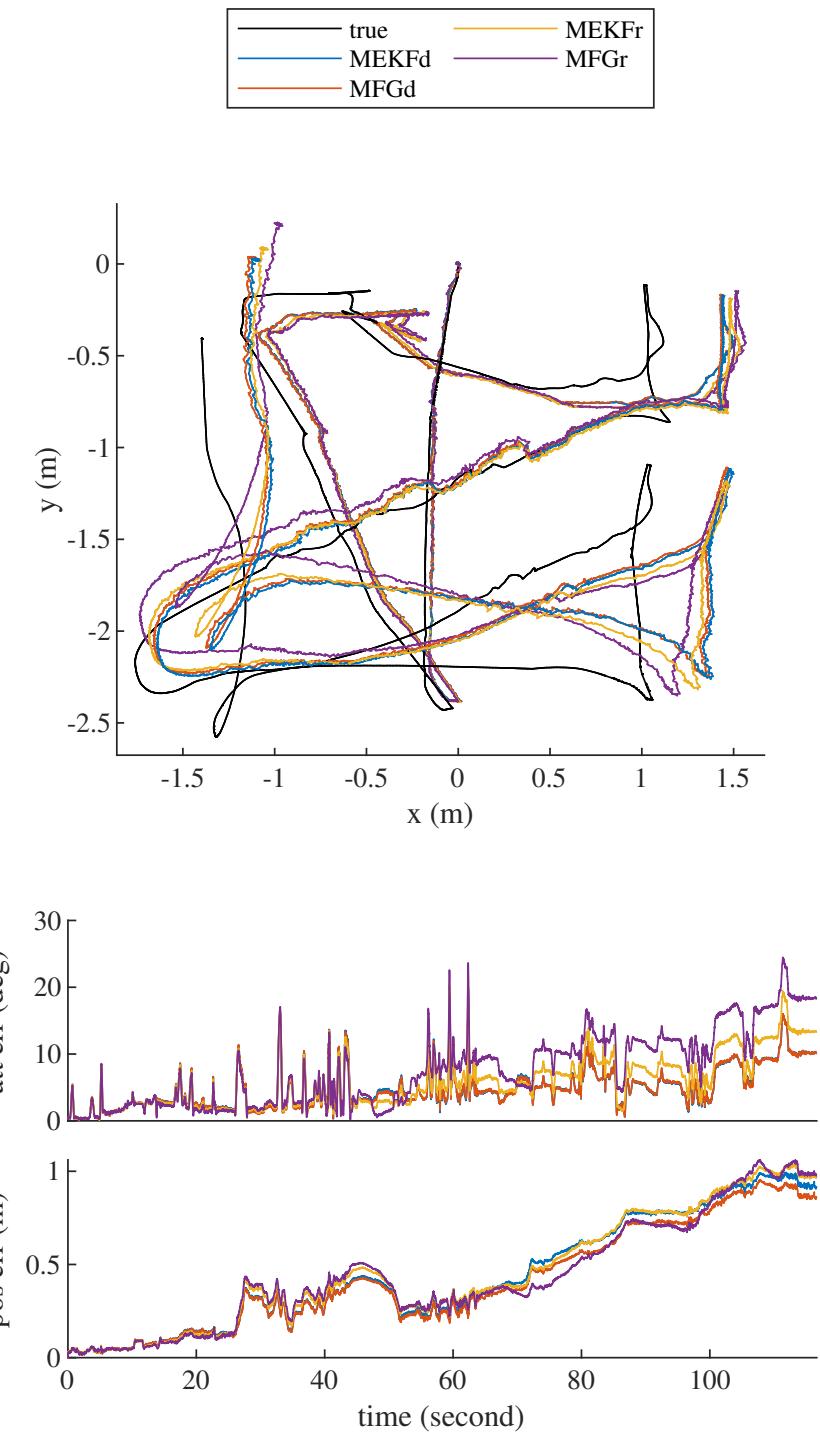


Figure 5.13: Estimation errors for the “lab_bumper4” trial. The top row is the true and estimated trajectories.

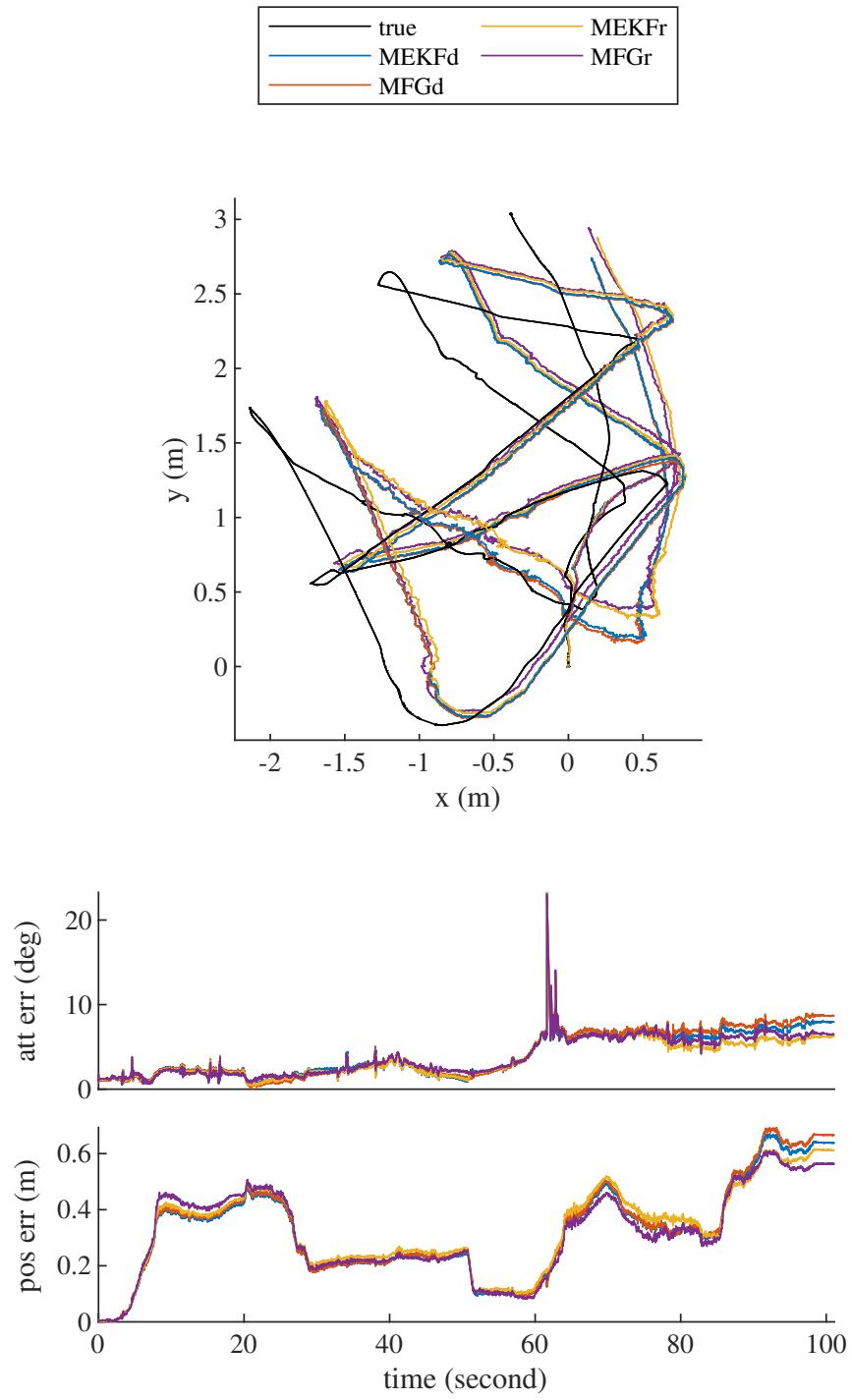


Figure 5.14: Estimation errors for the “lab_bumper5” trial. The top row is the true and estimated trajectories.

Chapter 6: Conclusions

6.1 Summary of Contributions

In this dissertation, the matrix Fisher distribution defined on $\text{SO}(3)$ is used to model the uncertainty of 3D rotations for a rigid body. To apply the matrix Fisher distribution in practical estimation problems, a new recursive algorithm to calculates its higher order moments, and a new approximation for a highly concentrated case are firstly introduces. Next, the matrix Fisher–Gaussian (MFG) distribution is constructed on $\text{SO}(3) \times \mathbb{R}^n$ to model the correlation between random rotations and variables in the Euclidean space. The MFG generalizes the application of matrix Fisher distribution into a wide variety of classic estimation problems, where the attitude must be estimated concurrently with other quantities in Euclidean space, such as sensor biases, position, and landmark locations, etc. In particular, recursive Bayesian filters are developed for (i) attitude estimation with a gyroscope and direction sensors, (ii) loosely coupled IMU-GNSS navigation, and (iii) visual-inertial odometry and navigation. Compared with the widely used Gaussian distribution, the algorithms developed in this dissertation achieve faster convergence speed and better accuracy for scenarios with large attitude uncertainty, for example, when the initial attitude is completely unknown, when there is a degree of freedom that is not properly observed, or when a sensor fails for a long period of time.

6.2 Future Directions

There are a few directions to further apply the techniques in directional statistics to estimation problems in practical engineering. First, the pose estimation algorithms that aligns two point sets proposed in Chapter 5.2.1 and Chapter 5.2.3 can be combined with a recursive algorithms for point cloud registration [4, 11, 93], where the correspondences between

the two points sets are unknown. To deal with large initial attitude uncertainty for the visual odometry algorithm in Chapter 5.2.3, the density in (5.83) for the map point can be decomposed into the distribution for the distance from the camera on \mathbb{R}^1 , and the direction on \mathbb{S}^2 , so the large dispersion for direction can be more accurately modeled by a density function on the sphere. In addition, the uncertainty for map points can be more carefully constructed by taking the triangulation process into account, so the uncertainty for depth becomes more dispersed when the feature is further away from the camera.

Although the matrix Fisher distribution models large attitude uncertainty more accurately, it is still restrictive in the sense that it is Gaussian shaped, i.e., it is uni-modal and symmetric around the mean. To deal with more flexible uncertainties on $\text{SO}(3)$, analogs of polynomial chaos expansions can be developed for the matrix Fisher distribution. This involves constructing a basis of orthogonal polynomials, analogous to the Gauss Hermite polynomial, with respect to the density function of matrix Fisher distribution. This has been done on the unit circle for the wrapped normal distribution [49, 105], but still remains to be investigated on $\text{SO}(3)$. Besides, the orthogonal polynomial can be used to design more accurate quadrature rules through deterministic sampling then the unscented transform introduced in [35, 64], which will make the measurement update step of the filters designed in this dissertation more accurate.

Bibliography

- [1] 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.
- [2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [3] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-D point sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.
- [4] R. Arun Srivatsan, M. Xu, N. Zevallos, and H. Choset, “Probabilistic pose estimation using a Bingham distribution-based linear filter,” *The International Journal of Robotics Research*, vol. 37, no. 13-14, pp. 1610–1631, 2018.
- [5] 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.
- [6] ——, “The invariant extended Kalman filter as a stable observer,” *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2016.
- [7] ——, “Invariant Kalman filtering,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 237–257, 2018.
- [8] ——, “Stochastic observers on Lie groups: a tutorial,” in *Conference on Decision and Control*, 2018, pp. 1264–1269.
- [9] P. Batista, C. Silvestre, and P. Oliveira, “A GES attitude observer with single vector observations,” *Automatica*, vol. 48, no. 2, pp. 388–395, 2012.
- [10] B. M. Bell and F. W. Cathey, “The iterated Kalman filter update as a Gauss-Newton method,” *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, 1993.
- [11] P. J. Besl and N. D. McKay, “A method for registration of 3-D shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [12] C. Bingham, “An antipodally symmetric distribution on the sphere,” *The Annals of Statistics*, pp. 1201–1225, 1974.
- [13] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, “Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback,” *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

- [14] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, “Robust visual inertial odometry using a direct EKF-based approach,” in *International Conference on Intelligent Robots and Systems*, 2015, pp. 298–304.
- [15] M. Brossard, A. Barrau, P. Chauchat, and S. Bonnabel, “Associating uncertainty to extended poses for on Lie group IMU preintegration with rotating Earth,” *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 998–1015, 2021.
- [16] M. Brossard, S. Bonnabel, and A. Barrau, “Invariant Kalman filtering for visual inertial SLAM,” in *International Conference on Information Fusion*, 2018, pp. 2021–2028.
- [17] H.-J. Bunge, *Texture analysis in materials science: mathematical methods*. Elsevier, 2013.
- [18] Y. Chikuse, *Statistics on special manifolds*. Springer, 2003.
- [19] G. S. Chirikjian, *Stochastic models, information theory, and Lie groups, volume 2: Analytic methods and modern applications*. Springer, 2011.
- [20] A. Chulliat, W. Brown, P. Alken, C. Beggan, M. Nair, G. Cox, A. Woods, S. Macmillan, B. Meyer, and M. Paniccia, “The US/UK world magnetic model for 2020-2025: Technical report,” National Oceanic and Atmospheric Administration, Tech. Rep., 2020.
- [21] J. L. Crassidis and F. L. Markley, “Attitude estimation using modified Rodrigues parameters,” in *Flight Mechanics/Estimation Theory Symposium*, 1996, pp. 71–83.
- [22] ——, “Unscented filtering for spacecraft attitude estimation,” *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 4, pp. 536–542, 2003.
- [23] 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.
- [24] 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.
- [25] T. D. Downs, “Orientation statistics,” *Biometrika*, vol. 59, no. 3, pp. 665–676, 1972.
- [26] H. Dym and M. HP, *Fourier series and integrals*. Academic Press, 1972.
- [27] J. L. Farrell, “Attitude determination by Kalman filtering,” *Automatica*, vol. 6, no. 3, pp. 419–430, 1970.
- [28] R. A. Fisher, “Dispersion on a sphere,” *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 217, no. 1130, pp. 295–305, 1953.

- [29] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “On-manifold preintegration for real-time visual–inertial odometry,” *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.
- [30] C. Forster, M. Pizzoli, and D. Scaramuzza, “SVO: Fast semi-direct monocular visual odometry,” in *International Conference on Robotics and Automation*, 2014, pp. 15–22.
- [31] E. Foxlin, “Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter,” in *Virtual Reality Annual International Symposium*, 1996, pp. 185–194.
- [32] E. Gai, K. Daly, J. Harrison, and L. Lemos, “Star-sensor-based satellite attitude/attitude rate estimator,” *Journal of Guidance, Control, and Dynamics*, vol. 8, no. 5, pp. 560–565, 1985.
- [33] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, “Efficient Bingham filtering based on saddlepoint approximations,” in *International Conference on Multisensor Fusion and Information Integration for Intelligent Systems*, 2014, pp. 1–7.
- [34] ——, “A new probability distribution for simultaneous representation of uncertain position and orientation,” in *International Conference on Information Fusion*, 2014, pp. 1–7.
- [35] ——, “Unscented orientation estimation based on the Bingham distribution,” *IEEE Transactions on Automatic Control*, vol. 61, no. 1, pp. 172–177, 2015.
- [36] J. Glover and L. P. Kaelbling, “Tracking the spin on a ping pong ball with the quaternion Bingham filter,” in *International Conference on Robotics and Automation*, 2014, pp. 4133–4140.
- [37] 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.
- [38] B. C. Hall, *Lie groups, Lie algebras, and representations: an elementary introduction*. Springer, 2003.
- [39] U. D. Hanebeck, “PGF 42: Progressive Gaussian filtering with a twist,” in *International Conference on Information Fusion*, 2013, pp. 1103–1110.
- [40] A. J. Haug, *Bayesian estimation and tracking: a practical guide*. John Wiley & Sons, 2012.
- [41] R. Hermann and A. Krener, “Nonlinear controllability and observability,” *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, 1977.
- [42] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, “Consistency analysis and improvement of vision-aided inertial navigation,” *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, 2013.

- [43] G. Huang, “Visual-inertial navigation: A concise review,” in *International Conference on Robotics and Automation*, 2019, pp. 9572–9582.
- [44] X. Huang, G. Mei, J. Zhang, and R. Abbas, “A comprehensive survey on point cloud registration,” *arXiv preprint arXiv:2103.02690*, 2021.
- [45] Y. Huang, Y. Zhang, N. Li, and L. Zhao, “Gaussian approximate filter with progressive measurement update,” in *Conference on Decision and Control*, 2015, pp. 4344–4349.
- [46] M. Idan, “Estimation of Rodrigues parameters from vector observations,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 578–586, 1996.
- [47] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, “Information fusion in navigation systems via factor graph based incremental smoothing,” *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013.
- [48] 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.
- [49] B. A. Jones and M. Balducci, “Stochastic expansions including data on the unit circle,” *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 12, pp. 2741–2746, 2019.
- [50] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, “iSAM2: Incremental smoothing and mapping using the Bayes tree,” *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.
- [51] T. Karvonen, S. Bonnabel, E. Moulines, and S. Särkkä, “On stability of a class of filters for nonlinear stochastic systems,” *SIAM Journal on Control and Optimization*, vol. 58, no. 4, pp. 2023–2049, 2020.
- [52] 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.
- [53] J. T. Kent, A. M. Ganeiber, and K. V. Mardia, “A new method to simulate the Bingham and related distributions in directional data analysis with applications,” *arXiv preprint arXiv:1310.8110*, 2013.
- [54] 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.
- [55] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” in *International Symposium on Mixed and Augmented Reality*, 2007, pp. 225–234.
- [56] P. Koev and A. Edelman, “The efficient evaluation of the hypergeometric function of a matrix argument,” *Mathematics of Computation*, vol. 75, no. 254, pp. 833–846, 2006.

- [57] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.
- [58] A. Kume and A. T. Wood, “Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants,” *Biometrika*, vol. 92, no. 2, pp. 465–476, 2005.
- [59] ——, “On the derivatives of the normalising constant of the Bingham distribution,” *Statistics & Probability Letters*, vol. 77, no. 8, pp. 832–837, 2007.
- [60] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, “Nonlinear measurement update for estimation of angular systems based on circular distributions,” in *American Control Conference*, 2014, pp. 5694–5699.
- [61] ——, “Recursive Bayesian filtering in circular state spaces,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, pp. 70–87, 2016.
- [62] 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.
- [63] T. Lee, “Bayesian attitude estimation with approximate matrix Fisher distributions on SO(3),” in *Conference on Decision and Control*, 2018, pp. 5319–5325.
- [64] ——, “Bayesian attitude estimation with the matrix Fisher distribution on SO(3),” *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3377–3392, 2018.
- [65] ——, “Spacecraft attitude estimation with a single magnetometer using matrix Fisher distributions on SO(3),” in *AIAA Scitech 2019 Forum*, 2019, p. 1173.
- [66] T. Lee, M. Leok, and N. H. McClamroch, “Lie group variational integrators for the full body problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196, no. 29-30, pp. 2907–2924, 2007.
- [67] ——, “Lie group variational integrators for the full body problem in orbital mechanics,” *Celestial Mechanics and Dynamical Astronomy*, vol. 98, no. 2, pp. 121–144, 2007.
- [68] T. Lee, M. Leok, N. H. McClamroch, and A. Sanyal, “Global attitude estimation using single direction measurements,” in *American Control Conference*, 2007, pp. 3659–3664.
- [69] 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.
- [70] C. A. León, J.-C. Massé, and L.-P. Rivest, “A statistical model for random rotations,” *Journal of Multivariate Analysis*, vol. 97, no. 2, pp. 412–430, 2006.
- [71] K. Li, F. Pfaff, and U. D. Hanebeck, “Geometry-driven stochastic modeling of SE(3) states based on dual quaternion representation,” in *International Conference on Industrial Cyber Physical Systems*, 2019, pp. 254–260.

- [72] ——, “Unscented dual quaternion particle filter for SE(3) estimation,” *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 647–652, 2020.
- [73] ——, “Progressive von Mises–Fisher filtering using isotropic sample sets for nonlinear hyperspherical estimation,” *Sensors*, vol. 21, no. 9, p. 2991, 2021.
- [74] M. Li and A. I. Mourikis, “Improving the accuracy of EKF-based visual-inertial odometry,” in *International Conference on Robotics and Automation*, 2012, pp. 828–835.
- [75] ——, “High-precision, consistent EKF-based visual-inertial odometry,” *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.
- [76] H. Liu, S. Fang, Z. Zhang, D. Li, K. Lin, and J. Wang, “Mfdnet: Collaborative poses perception and matrix Fisher distribution for head pose estimation,” *IEEE Transactions on Multimedia*, 2021.
- [77] T. Lupton and S. Sukkarieh, “Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions,” *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61–76, 2011.
- [78] 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.
- [79] K. V. Mardia, G. Hughes, C. C. Taylor, and H. Singh, “A multivariate von Mises distribution with applications to bioinformatics,” *Canadian Journal of Statistics*, vol. 36, no. 1, pp. 99–109, 2008.
- [80] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009.
- [81] 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.
- [82] F. L. Markley, “Attitude determination using vector observations and the singular value decomposition,” *Journal of the Astronautical Sciences*, vol. 36, no. 3, pp. 245–258, 1988.
- [83] ——, “Attitude error representations for Kalman filtering,” *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 2, pp. 311–317, 2003.
- [84] ——, “Attitude estimation or quaternion estimation?” *The Journal of the Astronautical Sciences*, vol. 52, no. 1, pp. 221–238, 2004.
- [85] ——, “Attitude filtering on SO(3),” *The Journal of the Astronautical Sciences*, vol. 54, no. 3-4, pp. 391–413, 2006.

- [86] S. Matthies, “Form effects in the description of the orientation distribution function (ODF) of texturized materials by model components,” *Physica Status Solidi (b)*, vol. 112, no. 2, pp. 705–716, 1982.
- [87] M. Moakher, “Means and averaging in the group of rotations,” *SIAM Journal on Matrix Analysis and Applications*, vol. 24, no. 1, pp. 1–16, 2002.
- [88] R. Mohler and C. Hwang, “Nonlinear data observability and information,” *Journal of the Franklin Institute*, vol. 325, no. 4, pp. 443–464, 1988.
- [89] D. Mohlin, J. Sullivan, and G. Bianchi, “Probabilistic orientation estimation with matrix Fisher distributions,” in *International Conference on Neural Information Processing Systems*, 2020, pp. 4884–4893.
- [90] A. I. Mourikis, S. I. Roumeliotis *et al.*, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *International Conference on Robotics and Automation*, 2007, pp. 3565–3572.
- [91] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: a versatile and accurate monocular SLAM system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [92] R. Mur-Artal and J. D. Tardós, “Visual-inertial monocular SLAM with map reuse,” *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, 2017.
- [93] A. Myronenko and X. Song, “Point set registration: Coherent point drift,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [94] D. I. Nikolayev and T. I. Savoylov, “Normal distribution on the rotation group $\text{SO}(3)$,” *Textures and Microstructures*, vol. 29, pp. 201–233, 1997.
- [95] T. Papadopoulo and M. I. Lourakis, “Estimating the Jacobian of the singular value decomposition: Theory and applications,” in *European Conference on Computer Vision*, 2000, pp. 554–570.
- [96] X. Pennec, “Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements,” *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, p. 127, 2006.
- [97] K. B. Petersen, M. S. Pedersen *et al.*, “The matrix cookbook,” 2012.
- [98] A. Pewsey and E. García-Portugués, “Recent advances in directional statistics,” *Test*, vol. 30, no. 1, pp. 1–58, 2021.
- [99] 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.

- [100] B. Presnell and P. Rumcheva, “The mean resultant length of the spherically projected normal distribution,” *Statistics & Probability Letters*, vol. 78, no. 5, pp. 557–563, 2008.
- [101] T. Qin, P. Li, and S. Shen, “Vins-mono: A robust and versatile monocular visual-inertial state estimator,” *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [102] Y. Qiu, D. J. Nordman, and S. B. Vardeman, “A wrapped trivariate normal distribution and Bayes inference for 3-D rotations,” *Statistica Sinica*, pp. 897–917, 2014.
- [103] J. Reis, P. Batista, P. Oliveira, and C. Silvestre, “Nonlinear observer on $\text{SO}(3)$ for attitude estimation on rotating Earth using single vector measurements,” *IEEE Control Systems Letters*, vol. 3, no. 2, pp. 392–397, 2018.
- [104] D. Scaramuzza and F. Fraundorfer, “Visual odometry [tutorial],” *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.
- [105] C. Schmid and K. J. DeMars, “Angular correlation using Rogers-Szegő-chaos,” *Mathematics*, vol. 8, no. 2, p. 171, 2020.
- [106] T. Sei and A. Kume, “Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method,” *Statistics and Computing*, vol. 25, no. 2, pp. 321–332, 2015.
- [107] M. D. Shuster, “A survey of attitude representations,” *Navigation*, vol. 8, no. 9, pp. 439–517, 1993.
- [108] ——, “Constraint in attitude estimation part I: Constrained estimation,” *The Journal of the Astronautical Sciences*, vol. 51, no. 1, pp. 51–74, 2003.
- [109] ——, “Constraint in attitude estimation part II: Unconstrained estimation,” *The Journal of the Astronautical Sciences*, vol. 51, no. 1, pp. 75–101, 2003.
- [110] M. D. Shuster and S. D. Oh, “Three-axis attitude determination from vector observations,” *Journal of Guidance, Control, and Dynamics*, vol. 4, no. 1, pp. 70–77, 1981.
- [111] S. T. Smith, “Covariance, subspace, and intrinsic Cramér-Rao bounds,” *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.
- [112] J. Solà, “Quaternion kinematics for the error-state Kalman filter,” *arXiv preprint arXiv:1711.02508*, 2017.
- [113] J. Stuelpnagel, “On the parametrization of the three-dimensional rotation group,” *SIAM review*, vol. 6, no. 4, pp. 422–430, 1964.
- [114] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment—a modern synthesis,” in *International Workshop on Vision Algorithms*, 1999, pp. 298–372.

- [115] J. Trumpf, R. Mahony, T. Hamel, and C. Lageman, “Analysis of non-linear attitude observers for time-varying reference measurements,” *IEEE Transactions on Automatic Control*, vol. 57, no. 11, pp. 2789–2800, 2012.
- [116] R. von Mises, “Über die “ganzzahligkeit” der atomgewicht und verwandte fragen,” *Physikal. Z.*, vol. 19, pp. 490–500, 1918.
- [117] W. Wang and T. Lee, “Matrix Fisher-Gaussian distribution,” [Online]. Available: <https://github.com/fdcl-gwu/Matrix-Fisher-Gaussian-Code>, 2019.
- [118] W. Wang, K. Gamagedara, and T. Lee, “On the observability of attitude with single direction measurements,” *arXiv preprint arXiv:2008.13067*, 2020.
- [119] ——, “On the observability of attitude with single direction measurements,” *IEEE Transactions on Automatic Control*, 2022, accepted.
- [120] W. Wang and T. Lee, [Online]. Available: https://github.com/fdcl-gwu/MF_Moments, 2020.
- [121] ——, “Matrix Fisher-Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ for attitude estimation with a gyro bias,” in *American Control Conference*, 2020, pp. 4429–4434.
- [122] ——, “Spectral Bayesian estimation for general stochastic hybrid systems,” *Automatica*, vol. 117, p. 108989, 2020.
- [123] ——, “Spectral uncertainty propagation for generalized stochastic hybrid systems with applications to a bouncing ball,” in *American Control Conference*, 2020, pp. 1803–1808.
- [124] ——, “Bingham-Gaussian distribution on $\mathbb{S}^3 \times \mathbb{R}^n$ for unscented attitude estimation,” in *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2021, pp. 1–7.
- [125] ——, “Higher-order central moments of matrix Fisher distribution on $\text{SO}(3)$,” *Statistics & Probability Letters*, vol. 169, p. 108983, 2021.
- [126] ——, “Matrix Fisher-Gaussian distribution on $\text{SO}(3) \times \mathbb{R}^n$ and Bayesian attitude estimation,” *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2175–2191, 2021.
- [127] ——, “Spacecraft attitude and gyro-bias estimation with a single magnetometer on $\text{SO}(3) \times \mathbb{R}^n$,” in *AIAA Scitech 2022 Forum*, 2022.
- [128] ——, “Uncertainty propagation for general stochastic hybrid systems on compact Lie groups,” *SIAM Journal on Applied Dynamical Systems*, 2022, accepted.
- [129] A. T. Wood, “Estimation of the concentration parameters of the Fisher matrix distribution on $\text{SO}(3)$ and the Bingham distribution on \mathbb{S}^q , $q \geq 2$,” *Australian Journal of Statistics*, vol. 35, no. 1, pp. 69–79, 1993.

- [130] T.-H. Wu, E. Kaufman, and T. Lee, “Globally asymptotically stable attitude observer on $\text{SO}(3)$,” in *Conference on Decision and Control*, 2015, pp. 2164–2168.
- [131] Y. Yin, Y. Cai, H. Wang, and B. Chen, “FisherMatch: Semi-supervised rotation regression via entropy-based filtering,” in *Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 164–11 173.
- [132] H. Zhang, L. Jin, and C. Ye, “The VCU-RVI benchmark: Evaluating visual inertial odometry for indoor navigation applications with an RGB-D camera,” in *International Conference on Intelligent Robots and Systems*, 2020, pp. 6209–6214.
- [133] 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.

Appendix A: A Special Case in Theorem 2.4

In this appendix, a special case which is embedded in the recursion of Theorem (2.4) is untwisted into a non-recursive formula except for an integer coefficient.

Theorem A.1. *If $s_1 \neq s_2 \neq |s_3|$ and $i \neq j$, then*

$$\begin{aligned} \left. \left(\frac{\partial^{2n} s'_i}{\partial T_{ij}^{2n}} + \frac{\partial^{2n} s'_j}{\partial T_{ij}^{2n}} \right) \right|_{T=0} &= \frac{a(n)}{(s_i + s_j)^{2n-1}}, \\ \left. \left(\frac{\partial^{2n} s'_i}{\partial T_{ij}^{2n}} - \frac{\partial^{2n} s'_j}{\partial T_{ij}^{2n}} \right) \right|_{T=0} &= \frac{a(n)}{(s_i - s_j)^{2n-1}}, \end{aligned} \quad (\text{A.1})$$

where $a(n)$ is an integer-valued function of n .

Before proving this theorem, a lemma for differentiating U' and V' is needed.

Lemma A.1. *Under the conditions of Theorem 2.4, if any of k or l is distinct from i or j , then*

$$\left. \frac{\partial^n U'_{kl}}{\partial T_{ij}^n} \right|_{T=0} = 0, \quad \left. \frac{\partial^n V'_{kl}}{\partial T_{ij}^n} \right|_{T=0} = 0, \quad (\text{A.2})$$

for all $n > 0$.

Proof. Suppose $i = 1, j = 2$, and let $\begin{bmatrix} s_1 & T_{12} \\ 0 & s_2 \end{bmatrix} = \begin{bmatrix} U'_{11} & U'_{12} \\ U'_{21} & U'_{22} \end{bmatrix} \begin{bmatrix} s'_1 & 0 \\ 0 & s'_2 \end{bmatrix} \begin{bmatrix} V'_{11} & V'_{12} \\ V'_{21} & V'_{22} \end{bmatrix}^T$ be the singular value decomposition of the upper left 2-by-2 diagonal block of $S + T$. Then by direct calculation, $S + T$ has singular value decomposition

$$\begin{bmatrix} s_1 & T_{12} & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix} = \begin{bmatrix} U'_{11} & U'_{12} & 0 \\ U'_{21} & U'_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s'_1 & 0 & 0 \\ 0 & s'_2 & 0 \\ 0 & 0 & s_3 \end{bmatrix} \begin{bmatrix} V'_{11} & V'_{12} & 0 \\ V'_{21} & V'_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}^T.$$

This means if the subscripts of U' or V' contain 3 which is different from $\{1, 2\}$, the

corresponding element of U' or V' is either zero or one, whose derivatives with respect to T_{12} of any order evaluated at $T = 0$ is zero. Other cases can be shown similarly. \square

Next, the proof for Theorem A.1 is provided.

Proof. By direct calculation using (2.49), is can be shown that

$$\begin{aligned}\frac{\partial(s'_i + s'_j)}{\partial T_{ij}} &= U'_{ii}V'_{ji} + U'_{ij}V'_{jj} \triangleq X_1, \\ \frac{\partial(s'_i - s'_j)}{\partial T_{ij}} &= U'_{ii}V'_{ji} - U'_{ij}V'_{jj} \triangleq X_2,\end{aligned}$$

for $X_1, X_2 \in \mathbb{R}$. Next, the second order differentiation is calculated using (2.50) as

$$\begin{aligned}\frac{\partial X_1}{\partial T_{ij}} &= (U'_{ii}V'_{jj} - U'_{ij}V'_{ji})(\Omega_{U_{ij}}^{ij} + \Omega_{V_{ij}}^{ij}) + U'_{ik}V'_{ji}\Omega_{U_{ki}}^{ij} - U'_{ii}V'_{jk}\Omega_{V_{ki}}^{ij} + U'_{ik}V'_{jj}\Omega_{U_{kj}}^{ij} - U'_{ij}V'_{jk}\Omega_{V_{kj}}^{ij}, \\ \frac{\partial X_2}{\partial T_{ij}} &= (U'_{ii}V'_{jj} + U'_{ij}V'_{ji})(\Omega_{V_{ij}}^{ij} - \Omega_{U_{ij}}^{ij}) + U'_{ik}V'_{ji}\Omega_{U_{ki}}^{ij} - U'_{ii}V'_{jk}\Omega_{V_{ki}}^{ij} - U'_{ik}V'_{jj}\Omega_{U_{kj}}^{ij} + U'_{ij}V'_{jk}\Omega_{V_{kj}}^{ij},\end{aligned}$$

where $k \in \{1, 2, 3\}$ and $i \neq j \neq k$. The crucial observation is that if we further differentiate the above two equations with respect to T_{ij} and evaluate at $T = 0$, then the last four terms in each subequation yield terms involving either U'_{ik} , V'_{jk} , or their derivatives with respect to T_{ij} . However, Lemma A.1 indicates that all these terms vanish after evaluated at $T = 0$. This means when calculating (A.1), the last four terms in each of the above subequations can be simply omitted. Therefore, by (2.51) is can be shown that

$$\begin{aligned}\frac{\partial X_1}{\partial T_{ij}} &= (U'_{ii}V'_{jj} - U'_{ij}V'_{ji})(\Omega_{U_{ij}}^{ij} + \Omega_{V_{ij}}^{ij}) = \frac{(U'_{ii}V'_{jj} - U'_{ij}V'_{ji})^2}{s'_i + s'_j} \triangleq \frac{Y_1^2}{s'_i + s'_j}, \\ \frac{\partial X_2}{\partial T_{ij}} &= (U'_{ii}V'_{jj} + U'_{ij}V'_{ji})(-\Omega_{U_{ij}}^{ij} + \Omega_{V_{ij}}^{ij}) = \frac{(U'_{ii}V'_{jj} + U'_{ij}V'_{ji})^2}{s'_i - s'_j} \triangleq \frac{Y_2^2}{s'_i - s'_j}.\end{aligned}$$

If the above two equations are differentiated with respect to T_{ij} again, we obtain the derivatives of $s_i \pm s_j$ which have been calculated, and the derivatives of Y_1, Y_2 which are calculated

as follows.

$$\begin{aligned}\frac{\partial Y_1}{\partial T_{ij}} &= -\frac{(V'_{ii}V'_{ji} + U'_{ij}V'_{jj})(U'_{ii}V'_{jj} - U'_{ij}V'_{ji})}{s'_i + s'_j} = -\frac{X_1 Y_1}{s'_i + s'_j}, \\ \frac{\partial Y_2}{\partial T_{ij}} &= -\frac{(U'_{ii}V'_{ji} - U'_{ij}V'_{ji})(U'_{ii}V'_{jj} + U'_{ij}V'_{ji})}{s'_i - s'_j} = -\frac{X_2 Y_2}{s'_i - s'_j},\end{aligned}$$

where any term having the subscript k has been omitted.

Now, a very special branch of smaller recursion is distilled from the large recursion stated in Theorem 2.4, which reads

$$\begin{aligned}\frac{\partial(s'_i + s'_j)}{\partial T_{ij}} &= X_1, & \frac{\partial X_1}{\partial T_{ij}} &= \frac{Y_1^2}{s'_i + s'_j}, & \frac{\partial Y_1}{\partial T_{ij}} &= -\frac{X_1 Y_1}{s'_i + s'_j}, \\ \frac{\partial(s'_i - s'_j)}{\partial T_{ij}} &= X_2, & \frac{\partial X_2}{\partial T_{ij}} &= \frac{Y_2^2}{s'_i - s'_j}, & \frac{\partial Y_2}{\partial T_{ij}} &= -\frac{X_2 Y_2}{s'_i - s'_j}.\end{aligned}\quad (\text{A.3})$$

Since the recursive structure for $s'_i + s'_j$ and $s'_i - s'_j$ are the same, they can be written in the same form as

$$\frac{\partial s'_{ij}}{\partial T_{ij}} = X; \quad \frac{\partial X}{\partial T_{ij}} = \frac{Y^2}{s'_{ij}}; \quad \frac{\partial Y}{\partial T_{ij}} = -\frac{XY}{s'_{ij}},$$

where s'_{ij} denotes either $s'_i + s'_j$ or $s'_i - s'_j$. Finally, to prove the theorem, it is claimed that

$$\frac{\partial^{2n} s'_{ij}}{\partial T_{ij}^{2n}} = \frac{a_0 Y^{2n} + a_1 Y^{2n-2} X^2 + \cdots + a_{n-1} Y^2 X^{2n-2}}{s'_{ij}^{2n}},$$

where a_m is an integer coefficient for $m = 0, \dots, n-1$. This claim is proved by induction. When $n = 1$, it is clearly seen from (A.3). Suppose for case n it is true, then differentiate T_{ij} one more time, and after some rearrangement, it is shown that

$$\frac{\partial^{2n+1} s'_{ij}}{\partial T_{ij}^{2n+1}} = \frac{a'_0 Y^{2n} X + a'_1 Y^{2n-2} X^3 + \cdots + a'_{n-1} Y^2 X^{2n-1}}{\partial s'_{ij}^{2n}},$$

where $a'_m = (-4n + 2m + 1)a_m + 2(m+1)a_{m+1}$ for $m = 0, \dots, n-2$, and $a'_{n-1} = (-2n -$

$1)a_{n-1}$. Then, differentiate T_{ij} once again, it is shown that

$$\frac{\partial^{2n+2} s'_{ij}}{\partial T_{ij}^{2n+2}} = \frac{a''_0 Y^{2n+2} + a''_1 Y^{2n} X^2 + \cdots + a''_n Y^2 X^{2n}}{s'^{2n+1}_{ij}},$$

where $a''_0 = a'_0$, $a''_m = (-4n + 2m - 2)a'_{m-1} + (2m + 1)a'_m$ for $m = 1, \dots, n-1$, and $a''_n = (-2n - 2)a'_{n-1}$. This finishes the proof for the claim. Now, note that when evaluated at $T = 0$, we have $Y = 1, X = 0$. Thus $\partial^{2n} s'_{ij}/\partial T_{ij}^{2n} \Big|_{T=0} = a_0/s'^{2n-1}_{ij}$, which finishes the proof for Theorem A.1 by noting that a_0 is an integer function of n . \square

It should be noted that the idea in the proof of Theorem A.1 cannot be generalized to the large recursion stated in Theorem 2.4, due to the omitted terms in the derivatives of X and Y . Even in this simpler case, giving a non-recursive expression for $a(n)$ is complicated, as seen from the recursive formula given for $a_m(n)$ in the proof.

Appendix B: The Second and Third Order Moment of Matrix Fisher Distribution

In this Appendix, non-recursive expressions of the second and third order moments are provided based on the development in Chapter 2.4.

B.1 The Second Order Moments

Based on Lemma 2.5, the second order moment of the form $E[Q_{ii}Q_{jj}]$ can be calculated as

$$E[Q_{ii}Q_{jj}] = \frac{1}{c(S)} \frac{\partial^2 c(S)}{\partial s_i \partial s_j}, \quad (\text{B.1})$$

for any $i, j \in \{1, 2, 3\}$. Next, using the recursion developed in Theorem 2.4, the second order moment of the forms $E[Q_{ij}Q_{ij}]$ and $E[Q_{ij}Q_{ji}]$ can be calculated as

$$E[Q_{ij}Q_{ij}] = \frac{1}{c(S)} \left(-\frac{\partial c(S)}{\partial s_i} \frac{s_i}{s_j^2 - s_i^2} + \frac{\partial c(S)}{\partial s_j} \frac{s_j}{s_j^2 - s_i^2} \right), \quad (\text{B.2a})$$

$$E[Q_{ij}Q_{ji}] = \frac{1}{c(S)} \left(-\frac{\partial c(S)}{\partial s_i} \frac{s_j}{s_j^2 - s_i^2} + \frac{\partial c(S)}{\partial s_j} \frac{s_i}{s_j^2 - s_i^2} \right), \quad (\text{B.2b})$$

for any $i \neq j \in \{1, 2, 3\}$, and $|s_i| \neq |s_j|$.

If $s_i = s_j \neq 0$, (B.2) can be evaluated by taking the limit $s_j \rightarrow s_i$, and their explicit expressions are

$$E[Q_{ij}Q_{ij}] = \frac{1}{c(S)} \frac{1}{2s_i} \left(\frac{\partial c(S)}{\partial s_i} + s_i \left(\frac{\partial^2 c(S)}{\partial s_i^2} - \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \right) \right), \quad (\text{B.3a})$$

$$E[Q_{ij}Q_{ji}] = \frac{1}{c(S)} \frac{1}{2s_i} \left(-\frac{\partial c(S)}{s_i} + s_i \left(\frac{\partial^2 c(S)}{\partial s_i^2} - \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \right) \right). \quad (\text{B.3b})$$

Similarly, if $s_i = -s_j \neq 0$, their expressions become

$$\mathbb{E}[Q_{ij}Q_{ij}] = \frac{1}{c(S)} \frac{1}{2s_i} \left(\frac{\partial c(S)}{\partial s_i} + s_i \left(\frac{\partial^2 c(S)}{\partial s_i^2} + \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \right) \right), \quad (\text{B.4a})$$

$$\mathbb{E}[Q_{ij}Q_{ji}] = \frac{1}{c(S)} \frac{1}{2s_i} \left(\frac{\partial c(S)}{\partial s_i} - s_i \left(\frac{\partial^2 c(S)}{\partial s_i^2} + \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \right) \right). \quad (\text{B.4b})$$

If $s_i = s_j = 0$, the above equations can also be evaluated by taking the limit $s_i \rightarrow 0$, as

$$\mathbb{E}[Q_{ij}Q_{ij}] = \frac{1}{c(S)} \frac{\partial^2 c(S)}{\partial s_i^2}, \quad (\text{B.5a})$$

$$\mathbb{E}[Q_{ij}Q_{ji}] = -\frac{1}{c(S)} \frac{\partial^2 c(S)}{\partial s_i \partial s_j}. \quad (\text{B.5b})$$

Given these results and Theorem 2.5, the second order derivatives of the normalizing constant can be easily evaluated by solving a linear system using the first order derivatives. This result is summarized in the following theorem.

Theorem B.1. *Let $\partial^2 c(S) \in \mathbb{R}^{3 \times 3}$ be a matrix whose elements are the second order derivatives $\partial^2 c(S)/\partial s_i \partial s_j$, $i, j \in \{1, 2, 3\}$. Then $\partial^2 c(S)$ satisfies a linear system $A \cdot \text{vec}(\partial^2 c(S)) = b$, where $A \in \mathbb{R}^{9 \times 9}$ is constant, and $b \in \mathbb{R}^9$ only involves S , $c(S)$, and the first order derivatives of $c(S)$.*

Proof. According to (B.1), it can be shown that

$$\frac{\partial^2 c(S)}{\partial s_i^2} = c(S) \mathbb{E}[Q_{ii}^2] = c(S) (1 - \mathbb{E}[Q_{ij}^2] - \mathbb{E}[Q_{ik}^2]),$$

where $i \neq j \neq k$. If $s_j \neq s_k$, then by (B.2) and the above equation, $\partial^2 c(S)/\partial s_i^2$ can be written as an expression involved with S , $c(S)$, and the first order derivatives of $c(S)$. If $s_j = s_k \neq 0$, or $s_j = -s_k \neq 0$, or $s_j = s_k = 0$, substitute (B.3) to (B.5) into the above equation. Then the coefficient of $\partial^2 c(S)/\partial s_i^2$, $\partial^2 c(S)/\partial s_i \partial s_j$, and $\partial^2 c(S)/\partial s_i \partial s_k$ on the right hand side are constants. After moving them to the left hand side, the right hand side only has S , $c(S)$, and the first order derivatives of $c(S)$. These provides three equations in total.

Next, by (B.1) and (2.11), it can be shown that

$$\frac{\partial^2 c(S)}{\partial s_i \partial s_j} = c(S) E[Q_{ii} Q_{kk}] = c(S) (E[Q_{kk}] + E[Q_{ij} Q_{ji}]),$$

for $i \neq j \neq k$. Substitute (2.20) and one from (B.2) to (B.5) into the above equation, and move $\partial^2 c(S)/\partial s_i^2$, $\partial^2 c(S)/\partial s_i \partial s_j$ from the right hand side to the left hand side, the remaining six equations are obtained. \square

B.2 The Third Order Moments

Based on Lemma 2.5, the third order moments of the form $E[Q_{ii} Q_{jj} Q_{kk}]$ are

$$E[Q_{ii} Q_{jj} Q_{kk}] = \frac{1}{c(S)} \frac{\partial^3 c(S)}{\partial s_i \partial s_j \partial s_k}, \quad (\text{B.6})$$

for $i, j, k \in \{1, 2, 3\}$. Next, moments of the form $E[Q_{ii} Q_{jk} Q_{kj}]$ are

$$E[Q_{ii} Q_{jk} Q_{kj}] = \frac{1}{c(S)} \left(\frac{\partial^2 c(S)}{\partial s_i \partial s_k} \frac{s_k}{s_k^2 - s_j^2} - \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \frac{s_j}{s_k^2 - s_j^2} \right), \quad (\text{B.7a})$$

$$E[Q_{ii} Q_{jk} Q_{kj}] = \frac{1}{c(S)} \left(\frac{\partial^2 c(S)}{\partial s_i \partial s_k} \frac{s_j}{s_k^2 - s_j^2} - \frac{\partial^2 c(S)}{\partial s_i \partial s_j} \frac{s_k}{s_k^2 - s_j^2} \right), \quad (\text{B.7b})$$

for $i \neq j \neq k$ and $|s_j| \neq |s_k|$. Moments of the form $E[Q_{jj} Q_{jk} Q_{kj}]$ are

$$E[Q_{jj} Q_{jk} Q_{kj}] = \frac{1}{c(S)} \left(\frac{\partial^2 c(S)}{\partial s_j \partial s_k} \frac{s_k}{s_k^2 - s_j^2} - \frac{\partial^2 c(S)}{\partial s_j^2} \frac{s_j}{s_k^2 - s_j^2} \right. \\ \left. - \frac{\partial c(S)}{\partial s_j} \frac{s_j^2 + s_k^2}{(s_j^2 - s_k^2)^2} + \frac{\partial c(S)}{\partial s_k} \frac{2s_j s_k}{(s_j^2 - s_k^2)^2} \right), \quad (\text{B.8a})$$

$$E[Q_{jj} Q_{jk} Q_{kj}] = \frac{1}{c(S)} \left(\frac{\partial^2 c(S)}{\partial s_j \partial s_k} \frac{s_j}{s_k^2 - s_j^2} - \frac{\partial^2 c(S)}{\partial s_j^2} \frac{s_k}{s_k^2 - s_j^2} \right. \\ \left. - \frac{\partial c(S)}{\partial s_j} \frac{2s_j s_k}{(s_j^2 - s_k^2)^2} + \frac{\partial c(S)}{\partial s_k} \frac{s_j^2 + s_k^2}{(s_j^2 - s_k^2)^2} \right), \quad (\text{B.8b})$$

for $j \neq k$ and $|s_j| \neq |s_k|$. Finally, moments of the form $E[Q_{ij}Q_{jk}Q_{ki}]$ are

$$E[Q_{ij}Q_{jk}Q_{ki}] = \frac{1}{c(S)} \left(\frac{\partial c(S)}{\partial s_i} \frac{s_j s_k}{(s_i^2 - s_j^2)(s_i^2 - s_k^2)} + \frac{\partial c(S)}{\partial s_j} \frac{s_i s_k}{(s_j^2 - s_i^2)(s_j^2 - s_k^2)} + \frac{\partial c(S)}{\partial s_k} \frac{s_i s_j}{(s_k^2 - s_i^2)(s_k^2 - s_j^2)} \right), \quad (\text{B.9a})$$

$$E[Q_{ij}Q_{jk}Q_{ik}] = \frac{1}{c(S)} \left(\frac{\partial c(S)}{\partial s_i} \frac{s_i s_j}{(s_j^2 - s_i^2)(s_k^2 - s_i^2)} + \frac{\partial c(S)}{\partial s_j} \frac{s_j^2}{(s_i^2 - s_j^2)(s_k^2 - s_j^2)} + \frac{\partial c(S)}{\partial s_k} \frac{s_j s_k}{(s_j^2 - s_k^2)(s_i^2 - s_k^2)} \right), \quad (\text{B.9b})$$

for $i \neq j \neq k$ and $|s_i| \neq |s_j| \neq |s_k|$.

Appendix C: Non-uniqueness Parameters of MFG

In this appendix, it is shown that how the MFG can be parameterized differently if S has repeated values. First, Definition 2.1 is augmented by the following uniqueness condition: *the first nonzero element of each column of U' is positive* [54]. This condition ensures that the columns of U and V cannot undergo simultaneous sign changes. Then the equivalence of MFG parameterizations is given in the following theorem.

Theorem C.1. *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

2I. $\exists \theta_1, \theta_2 \in \mathbb{R}$ such that $\tilde{U} = UT_1$, $\tilde{V} = VT_2$ where $T_1 = \exp(\theta_1 \hat{e}_1)$ and $T_2 = \exp(\theta_2 \hat{e}_1)$, $[\tilde{P}_{:,2}, \tilde{P}_{:,3}] = [P_{:,2}, P_{:,3}] \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix}$ where $P_{:,i}$ is the i -th column of P , and $\Sigma = \tilde{\Sigma}$.

2B. $\exists \theta_1, \theta_2 \in \mathbb{R}$ 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 $\exists T \in \text{SO}(3)$ 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 $s_1 = s_2 = -s_3 \neq 0$, let $L = \text{diag}(1, 1, -1)$, then
- 7I. $\exists T \in \text{SO}(3)$ such that $\tilde{U} = UT$, $\tilde{V} = VLTL$, $\tilde{P} = PT$, and
 $\tilde{\Sigma} = \Sigma + P[T(\text{tr}(S)I_{3 \times 3} - S)T^T - (\text{tr}(S)I_{3 \times 3} - S)]P^T$.
- 7B. $\exists T \in \text{SO}(3)$ such that $\tilde{U} = ULTL$, $\tilde{V} = VT$, $\tilde{P} = PT$, and
 $\tilde{\Sigma} = \Sigma + P[T(\text{tr}(S)I_{3 \times 3} - S)T^T - (\text{tr}(S)I_{3 \times 3} - S)]P^T$.
8. if $s_1 \neq s_2 = -s_3 \neq 0$, then
- 8I. $\exists \theta \in \mathbb{R}$ such that $\tilde{U} = UT$, $\tilde{V} = VT^T$, $\tilde{P} = PT$, and $\tilde{\Sigma} = \Sigma + P[T(\text{tr}(S)I_{3 \times 3} - S)T^T - (\text{tr}(S)I_{3 \times 3} - S)]P^T$, where $T = \exp(\theta \hat{e}_1)$.
- 8B. $\exists \theta \in \mathbb{R}$ such that $\tilde{U} = UT^T$, $\tilde{V} = VT$, $\tilde{P} = PT$, and $\tilde{\Sigma} = \Sigma + P[T(\text{tr}(S)I_{3 \times 3} - S)T^T - (\text{tr}(S)I_{3 \times 3} - S)]P^T$, where $T = \exp(\theta \hat{e}_1)$.

For cases 2), 3) and 8), “I” means the condition holds for MFGI, and “B” means the condition holds for MFBG.

Proof. By Lemma 3.1, the two MFGs are equivalent if and only if $F = \tilde{F}$, $\mu_c = \tilde{\mu}_c$ for all $R \in \text{SO}(3)$, and $\Sigma = \tilde{\Sigma}_c$.

The conditions provided in the theorem begin with $S = \tilde{S}$. We first consider additional conditions to make them equivalent to $F = \tilde{F}$. Since a matrix cannot have two sets of different singular values, $F = \tilde{F}$ implies $S = \tilde{S}$. For case 1), $S = \tilde{S}$ trivially implies $F = \tilde{F}$ as they are all zeros. For case 6), $F = \tilde{F}$ if and only if their proper SVDs are the same, due to the augmented uniqueness condition for SVD when the singular values are different. In short, $F = \tilde{F}$ if and only if $S = \tilde{S}$ for cases 1) and 6). However, when F has repeated singular values, U and V are only unique up to a rotation. We consider three possibilities regarding the multiplicity of S .

(i) If $s_1 \neq s_2 = s_3 = 0$, corresponding to case 2) in the theorem, then $F = \tilde{F}$ if and only if $U \text{diag}(s_1, 0, 0)V^T = \tilde{U} \text{diag}(s_1, 0, 0)\tilde{V}^T$. This means the first column of U and V is unique,

while other columns can be arbitrarily chosen as long as $U, V \in \text{SO}(3)$, which is the same as the conditions on U, V in case 2).

(ii) If $s_1 = s_2$ and(or) $s_2 = s_3 \neq 0$, which corresponds to case 3) to 5) in the theorem, then the left and right singular vectors (columns of U and V) w.r.t the repeated singular value form subspaces of \mathbb{R}^3 . And $F = \tilde{F}$ if and only if the left and right singular vectors w.r.t the repeated singular value are rotated by the same rotation matrix, while the singular vectors w.r.t the non-repeated singular value are the same. The above condition on U, V is the same as those given in case 3) to 5).

(iii) If $s_2 = -s_3 \neq 0$, corresponding to cases 7) and 8) in the theorem. Let $S' = SL \triangleq \text{diag}(s'_1, s'_2, s'_3)$, then $USV^T = US'LV^T = ULS'V^T$. Since $s'_2 = s'_3$, for all $\theta \in \mathbb{R}$ and $T = \exp(\theta \hat{e}_1)$, it can be proved for MFGI that

$$US'LV^T = UTS'T^T LV^T = UTS'LLT^T LV^T = UTS(VLTL)^T = UTS(VT^T)^T,$$

and for MFGB that

$$ULS'V^T = ULTS'T^T V^T = ULTLLS'T^T V^T = (ULTL)S(VT)^T = UT^T S(VT)^T,$$

This shows the sufficiency and necessity for the conditions on U, V in case 8). For case 7), T can be arbitrarily chosen from $\text{SO}(3)$, but in general $LTL \neq T^T$.

Next, we consider the equivalent conditions for $\mu_c = \tilde{\mu}_c$ and $\Sigma_c = \tilde{\Sigma}_c$. For case 1), the matrix Fisher density parts reduce to one, $\mu_c = \mu$, and $\Sigma_c = \Sigma$, so the remaining Gaussian density parts are equivalent if and only if their means and covariance matrices are the same, as shown in Lemma 3.1. Next, the three possibilities regarding the multiplicity of S are considered individually as follows.

For (i), let $Q = U^T RV$, since $\tilde{U} = UT_1$, $\tilde{V} = VT_2$ and $S = \tilde{S}$, it can be proved for MFGI

that

$$\begin{aligned}
\tilde{\mu}_c &= \tilde{\mu} + \tilde{P}(T_1^T U^T R V T_2 S - S T_2^T V^T R^T U T_1)^\vee \\
&= \tilde{\mu} + \tilde{P} T_1^T (U^T R V T_2 S T_1^T - T_1 S T_2^T V^T R^T U)^\vee \\
&= \tilde{\mu} + \tilde{P} T_1^T (Q S - S Q^T)^\vee \\
&= \tilde{\mu} + \tilde{P} T_1^T [0, -s_1 Q_{31}, s_1 Q_{21}]^T.
\end{aligned} \tag{C.1}$$

And similarly, for MFGB, the above equation becomes

$$\begin{aligned}
\tilde{\mu}_c &= \tilde{\mu} + \tilde{P}(S T_1^T U^T R V T_2 - T_2^T V^T R^T U T_1 S)^\vee \\
&= \tilde{\mu} + \tilde{P} T_2^T (T_2 S T_1^T U^T R V - V^T R^T U T_1 S T_2^T)^\vee \\
&= \tilde{\mu} + \tilde{P} T_2^T (S Q - Q^T S)^\vee \\
&= \tilde{\mu} + \tilde{P} T_2^T [0, s_1 Q_{13}, -s_1 Q_{12}]^T.
\end{aligned} \tag{C.2}$$

Note that Q and R have a one-to-one correspondence, so $\mu_c = \tilde{\mu}_c$ for all $R \in \text{SO}(3)$ is equivalent to $\mu_c = \tilde{\mu}_c$ for all $Q \in \text{SO}(3)$. It is clear the conditions on μ and P in case 2) are sufficient for this. On the other hand, if $\mu_c = \tilde{\mu}_c$ for all $Q \in \text{SO}(3)$, substituting $Q = I_{3 \times 3}$ into (C.1) and (C.2) proves $\mu = \tilde{\mu}$; substituting $Q = \exp\{\frac{\pi}{2} e_2\}$ and $Q = \exp\{\frac{\pi}{2} e_3\}$ into (C.1) and (C.2) proves $[\tilde{P}_{:,2}, \tilde{P}_{:,3}] = [P_{:,2}, P_{:,3}] \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix}$. Finally, denote T' as either $\begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix}$ or $\begin{bmatrix} \cos \theta_2 & -\sin \theta_2 \\ \sin \theta_2 & \cos \theta_2 \end{bmatrix}$ then

$$\tilde{P}(\text{tr}(S)I_{3 \times 3} - S)(\tilde{P})^T = s_1 \begin{bmatrix} P_{:,2} & P_{:,3} \end{bmatrix} T'(T')^T \begin{bmatrix} P_{:,2}^T \\ P_{:,3}^T \end{bmatrix} = P(\text{tr}(S)I_{3 \times 3} - S)P^T. \tag{C.3}$$

so $\Sigma_c = \tilde{\Sigma}_c$ if and only if $\Sigma = \tilde{\Sigma}$.

For (ii), a similar calculation as in (C.1) and (C.2) shows $\tilde{\mu}_c = \tilde{\mu} + \tilde{P}T^T v_R$. The sufficiency of the conditions on μ and P in case 3) to 5) follows immediately. For the necessary direction, substituting $Q = I_{3 \times 3}$ into the above equation proves $\mu = \tilde{\mu}$; substituting $Q = \exp\left\{\frac{\pi}{2}\hat{e}_1\right\}$, $Q = \exp\left\{\frac{\pi}{2}\hat{e}_2\right\}$ and $Q = \exp\left\{\frac{\pi}{2}\hat{e}_3\right\}$ proves $\tilde{P} = PT$ since $s_i + s_j > 0$ for any $i \neq j$ in case 3) to 5). Finally, note that the repeated diagonal entries of S and $\text{tr}(S)I_{3 \times 3} - S$ share the same indices, thus $T(\text{tr}(S)I_{3 \times 3} - S)T^T = \text{tr}(S)I_{3 \times 3} - S$. This proves $\Sigma_c = \tilde{\Sigma}_c$ if and only if $\Sigma = \tilde{\Sigma}$. The same argument in this paragraph also proves case 6) in the theorem by letting $T = I_{3 \times 3}$.

For (iii), a similar calculation as in (C.1) and (C.2) shows $\tilde{\mu}_c = \tilde{\mu} + \tilde{P}T^T v_R$. The sufficiency of the conditions on μ and P in case 7) and 8) is clear. For necessity, substituting $Q = I_{3 \times 3}$ proves $\mu = \tilde{\mu}$; substituting $Q = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $Q = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, and $Q = \exp\left\{\frac{\pi}{2}\hat{e}_3\right\}$ proves $\tilde{P} = PT$, because $s_2 - s_3 > 0$, $s_1 - s_3 > 0$, and $s_1 + s_2 > 0$ in these two cases. Finally, $\tilde{\Sigma}_c$ is

$$\tilde{\Sigma}_c = \tilde{\Sigma} + PT(\text{tr}(S)I - S)T^T P^T. \quad (\text{C.4})$$

However, different from (ii), $T(\text{tr}(S)I_{3 \times 3} - S)T^T \neq \text{tr}(S)I_{3 \times 3} - S$. And as a result, $\Sigma_c = \tilde{\Sigma}_c$ if and only if $\tilde{\Sigma} = \Sigma + P [T(\text{tr}(S)I - S)T^T - (\text{tr}(S)I - S)] P^T$. \square

In other words, Theorem C.1 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. This is intuitively straightforward by looking at how the correlation P is constructed in Theorem 3.2. 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 $s_2 = -s_3$, $T(\text{tr}(S)I_{3 \times 3} - S)T^T \neq \text{tr}(S)I_{3 \times 3} - S$.

Appendix D: Deterministic Attitude Observability

This appendix presents parallel results for Theorem 4.3 and Theorem 4.4 under deterministic settings. Specifically, the attitude kinematics (4.1) and (4.2) are formulated without noise as

$$\dot{R} = \hat{\omega}(t)R, \quad (\text{D.1})$$

$$\dot{R} = R\hat{\Omega}(t). \quad (\text{D.2})$$

Similarly, the single reference direction (4.3) and (4.4) are also measured without noise:

$$x = R^T a, \quad (\text{D.3})$$

$$y = Rb. \quad (\text{D.4})$$

Then the deterministic version of Theorem 4.3 and Theorem 4.4 is formulated as [41]:

Theorem D.1. *The system (D.1) and (D.3), and the system (D.2) and (D.4) are weakly locally observable. Conversely, the system (D.1) and (D.4), and the system (D.2) and (D.3) are unobservable.*

Proof. The proof is based on the results of [41], and we adopt notations therein without reintroducing them here for brevity. Without loss of generality, we assume $a = b = e_1$.

For (D.1), $\mathcal{F}^0(R)$ is spanned by $\{\hat{e}_i R\}_{i=1}^3$. For any $R \in \text{SO}(3)$ and $\hat{\eta} \in \mathfrak{so}(3)$, the Lie derivative of $x(R)$ along $\hat{\eta}R$ is

$$(L_{(\hat{\eta}R)}x)(R) = \frac{d}{dt} \Big|_{t=0} R^T \exp(t\hat{\eta})^T a = R^T \hat{\eta}^T a.$$

Let $\tilde{\mathcal{G}} = \{L_{(\hat{e}_2 R)}x, L_{(\hat{e}_3 R)}x\} \subset \mathcal{G}$. For any $R_0 \in \text{SO}(3)$, define a local coordinate $\theta \in \mathbb{R}^3$ with

$R(\theta) = \exp(\hat{\theta})R_0$. Then $dR = \widehat{d\theta}R_0$, and it can be shown that

$$\begin{aligned} d(L_{(\hat{e}_2 R_0)}x)(R_0) &= R_0^T \widehat{d\theta}^T \hat{e}_2^T e_1 = R_0^T \hat{e}_3 d\theta, \\ d(L_{(\hat{e}_3 R_0)}x)(R_0) &= R_0^T \widehat{d\theta}^T \hat{e}_3^T e_1 = -R_0^T \hat{e}_2 d\theta. \end{aligned}$$

Thus, $d\tilde{\mathcal{G}}(R_0) = R_0^T [\hat{e}_3, -\hat{e}_2] d\theta$. Since $\text{rank}[\hat{e}_3, -\hat{e}_2] = 3$, it follows that the dimension of $d\mathcal{G}(R_0)$ is three. Therefore, the system (D.1) and (D.3) is weakly locally observable [41, Theorem 3.2].

Next, for any $R \in \text{SO}(3)$ and $\hat{\eta}_1, \hat{\eta}_2 \in \mathfrak{so}(3)$, it can be shown that

$$d(L_{(\hat{\eta}_1 R)}(L_{(\hat{\eta}_2 R)}y))(R_0) = \hat{\eta}_1 \hat{\eta}_2 \widehat{d\theta} R_0 b = -\hat{\eta}_1 \hat{\eta}_2 \widehat{R_0 b} d\theta.$$

As such, any higher-order Lie derivative of y along (D.1) would include the factor $\widehat{R_0 b}$, which has rank two. Thus, the dimension of $d\mathcal{G}(R_0)$ is at most two. Because the system (D.1) is locally controllable, the system of (D.1) and (D.4) is unobservable according to [41, Theorem 3.12]. The remaining cases with (D.2) can be shown similarly. \square

Appendix E: Proof for Theorem 4.6

Proof of Theorem 4.6. Equation (4.40) and (4.42) follow from the independence of x_k and ΔW_v , and the linearity of expectation. Equation (4.41) is from the fact that $E[Q_{k+1}] = U_{k+1}^T E[R_{k+1}] V_{k+1}$ and S_{k+1} are both diagonal. Next, we prove (4.43) and (4.44).

In (4.32), denote the exponent of the stochastic part as $\xi \triangleq h(x_k - \mu_k) + H_u \Delta W_u \in \mathbb{R}^3$, and the deterministic part as $\delta R_k \triangleq e^{h(\hat{\Omega}_k + \hat{\mu}_k)} \in \text{SO}(3)$ such that

$$R_{k+1} = R_k e^{\xi + o(h)} \delta R_k.$$

First consider the MFGI definition. For any $A \in \mathbb{R}^{3 \times 3}$, let $\Lambda(A) \in \mathbb{R}^3$ be defined as $\Lambda(A) = (U_{k+1}^T A V_{k+1} S_{k+1} - S_{k+1} V_{k+1}^T A^T U_{k+1})^\vee$. For example, $v_{R_{k+1}} = \Lambda(R_{k+1})$. Further, ΔW_v in (4.30) is independent of R_k and ξ . Therefore,

$$E[x_{k+1} v_{R_{k+1}}^T] = \sum_{i=0}^{\infty} \frac{1}{i!} E[x_k \Lambda(R_k (\hat{\xi} + o(h))^i \delta R_k)^T],$$

after expanding $e^{\xi + o(h)}$. Similar with (4.36), the first order approximation is

$$\begin{aligned} E[x_{k+1} v_{R_{k+1}}^T] &= E[x_k \Lambda(R_k \delta R_k)^T] + h E[x_k \Lambda(R_k (\hat{x}_k - \hat{\mu}_k) \delta R_k)^T] \\ &\quad + \frac{1}{2} E[x_k \Lambda(R_k ((H_u \Delta W_u)^\wedge)^2 \delta R_k)^T] + O(h^2). \end{aligned} \quad (\text{E.1})$$

Now we calculate the three terms on the right hand side of (E.1). Similar with (3.22), the first term is

$$E[x_k \Lambda(R_k \delta R_k)^T] = \mu_k E[\Lambda(R_k \delta R_k)]^T + P_k E[v_{R_k} \Lambda(R_k \delta R_k)^T],$$

where

$$\begin{aligned}\mathbb{E}[\Lambda(R_k \delta R_k)] &= \tilde{U}(\mathbb{E}[Q_k] \tilde{S}^T - \tilde{S} \mathbb{E}[Q_k]^T)^\vee = \tilde{U} \mathbb{E}[\tilde{v}_R], \\ \mathbb{E}[v_{R_k} \Lambda(R_k \delta R)^T] &= \mathbb{E}[(Q_k S_k - S_k Q_k^T)^\vee ((Q_k \tilde{S}^T - \tilde{S} Q_k^T)^\vee)^T] \tilde{U}^T = \mathbb{E}[v_{R_k} \tilde{v}_R^T] \tilde{U}^T.\end{aligned}$$

Next, it can be shown that

$$\Lambda(R_k(\hat{x}_k - \hat{\mu}_k) \delta R_k) = \tilde{U} (\text{tr}(Q_k \tilde{S}^T) I - Q_k \tilde{S}^T) Q_k V_k^T (x_k - \mu_k) = \tilde{U} \Gamma_Q V_k^T (x_k - \mu_k).$$

Therefore, the second term of (E.1) is

$$\begin{aligned}\mathbb{E}[x_k \Lambda(R_k(\hat{x}_k - \hat{\mu}_k) \delta R_k)^T] &= \mathbb{E}[x_k (x_k - \mu_k)^T V_k \Gamma_Q^T] \tilde{U}^T \\ &= \mathbb{E}[(\Sigma_{c_k} + \mu_k v_{R_k}^T P_k^T + P_k v_{R_k} v_{R_k}^T P_k^T) V_k \Gamma_Q^T] \tilde{U}^T.\end{aligned}$$

Due to the independence of ΔW_u with other random variables, the third term of (E.1) is given by

$$\mathbb{E}[x_k \Lambda(R_k((H_u \Delta W_u)^\wedge)^2 \delta R_k)^T] = h \mathbb{E}[x_k \Lambda(R_k G_u \delta R_k)^T] - h \text{tr}(G_u) \mathbb{E}[x_k \Lambda(R_k \delta R_k)^T].$$

Let $\tilde{V} = V_{k+1}^T \delta R_k^T G_u^T V_k$ and $\tilde{S} = \tilde{U}^T S_k \tilde{V}$. The first term of the above is

$$\begin{aligned}\mathbb{E}[x_k \Lambda(R_k G_u \delta R_k)^T] &= \mu \left((\mathbb{E}[Q_k] \tilde{S}^T - \tilde{S} \mathbb{E}[Q_k]^T)^\vee \right)^T \tilde{U}^T \\ &\quad + P_k \mathbb{E}[v_{R_k} \left((Q \tilde{S}^T - \tilde{S} Q^T)^\vee \right)^T] \tilde{U}^T = \mu \mathbb{E}[\tilde{v}_R^T] \tilde{U}^T + P \mathbb{E}[v_R \tilde{v}_R^T] \tilde{U}^T.\end{aligned}$$

Substituting these three terms into (E.1), equation (4.43a) is derived. The calculation for MFGB (4.43b) is similar, and it is omitted for brevity. But it should be noted that $\tilde{V} \notin \text{SO}(3)$, so (4.38b) cannot be written into a compact form like (4.38a).

Finally, we present the proof for (4.44). Similar to the derivation of (E.1), after expanding

the exponential term, the following equation can be obtained

$$\mathbb{E} \left[v_{R_{k+1}} v_{R_{k+1}}^T \right] = \sum_{i,j=0}^{\infty} \frac{1}{i!j!} \mathbb{E} \left[\Lambda(R_k(\hat{\xi} + o(h))^i \delta R_k) \Lambda(R_k(\hat{\xi} + o(h))^j \delta R_k)^T \right].$$

After collecting the first order terms, it becomes

$$\begin{aligned} \mathbb{E} \left[v_{R_{k+1}} v_{R_{k+1}}^T \right] &= \mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k \delta R)^T \right] + h \mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k(\hat{x}_k - \hat{\mu}_k) \delta R_k)^T \right] \\ &\quad + h \mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k(\hat{x}_k - \hat{\mu}_k) \delta R_k)^T \right]^T \\ &\quad + \frac{1}{2} \mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k((H_u \Delta W_u)^\wedge)^2 \delta R_k)^T \right] \\ &\quad + \frac{1}{2} \mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k((H_u \Delta W_u)^\wedge)^2 \delta R_k)^T \right]^T \\ &\quad + \mathbb{E} \left[\Lambda(R_k(H_u \Delta W_u)^\wedge \delta R_k) \Lambda(R_k(H_u \Delta W_u)^\wedge \delta R_k)^T \right] + O(h^2), \end{aligned} \quad (\text{E.2})$$

where the first term on the right hand side is

$$\mathbb{E} \left[\Lambda(R_k \delta R) \Lambda(R_k \delta R)^T \right] = \tilde{U} \mathbb{E} \left[(Q_k \tilde{S}^T - \tilde{S} Q_k)^\vee ((Q_k \tilde{S}^T - \tilde{S} Q_k)^\vee)^T \right] \tilde{U}^T = \tilde{U} \mathbb{E} \left[\tilde{v}_R \tilde{v}_R^T \right] \tilde{U}^T,$$

and the second term is

$$\begin{aligned} \mathbb{E} \left[\Lambda(R_k \delta R_k) \Lambda(R_k(\hat{x}_k - \hat{\mu}_k) \delta R_k)^T \right] &= \tilde{U} \mathbb{E} \left[(Q_k \tilde{S}^T - \tilde{S} Q_k^T)^\vee v_{R_k}^T P_k^T V_k \Gamma_Q^T \right] \tilde{U}^T \\ &= \tilde{U} \mathbb{E} \left[\tilde{v}_R v_{R_k}^T P^T V_k \Gamma_Q^T \right] \tilde{U}^T, \end{aligned}$$

Similarly, the fourth term on the right hand side of (E.2) is

$$\mathbb{E} \left[\Lambda(R_k \delta R_k) \Lambda(R_k((H_u \Delta W_u)^\wedge)^2 \delta R_k)^T \right] = h \tilde{U} \mathbb{E} \left[\tilde{v}_R \tilde{v}_R^T \right] \tilde{U}^T - \text{tr}(G_u) h \tilde{U} \mathbb{E} \left[\tilde{v}_R \tilde{v}_R^T \right] \tilde{U}^T,$$

and the sixth term is

$$\mathbb{E} \left[\Lambda(R_k(H_u \Delta W_u)^\wedge \delta R_k) \Lambda(R_k(H_u \Delta W_u)^\wedge \delta R_k)^T \right] = h \tilde{U} \mathbb{E} \left[\Gamma_Q V_k^T G_u V_k \Gamma_Q^T \right] \tilde{U}^T.$$

Substituting these into (E.2) yields (4.44a). The calculation for MFGB (4.44b) is similar, and it is omitted for brevity. \square

Note that the expression for Γ_Q of MFGB (4.39b) only has the first order of Q , but the expression for Γ_Q of MFGI (4.39a) has the second order of Q . Fortunately, it is shown in the next lemma that (4.39a) can also be reduced to the first order of Q , so that the right hand side of (4.43a) and (4.44a) only involve at most the third order moments of Q .

Lemma E.1. Γ_Q in (4.39a) can be simplified as follows:

$$\begin{aligned}
(\Gamma_Q)_{11} &= \tilde{S}_{22}Q_{33} + \tilde{S}_{33}Q_{22} - \tilde{S}_{23}Q_{32} - \tilde{S}_{32}Q_{23} \\
(\Gamma_Q)_{12} &= \tilde{S}_{23}Q_{31} + \tilde{S}_{31}Q_{23} - \tilde{S}_{21}Q_{33} - \tilde{S}_{33}Q_{21} \\
(\Gamma_Q)_{13} &= \tilde{S}_{21}Q_{32} + \tilde{S}_{32}Q_{21} - \tilde{S}_{22}Q_{31} - \tilde{S}_{31}Q_{22} \\
(\Gamma_Q)_{21} &= \tilde{S}_{13}Q_{32} + \tilde{S}_{32}Q_{13} - \tilde{S}_{12}Q_{33} - \tilde{S}_{33}Q_{12} \\
(\Gamma_Q)_{22} &= \tilde{S}_{11}Q_{33} + \tilde{S}_{33}Q_{11} - \tilde{S}_{13}Q_{31} - \tilde{S}_{31}Q_{13} \\
(\Gamma_Q)_{23} &= \tilde{S}_{12}Q_{31} + \tilde{S}_{31}Q_{12} - \tilde{S}_{11}Q_{32} - \tilde{S}_{32}Q_{11} \\
(\Gamma_Q)_{31} &= \tilde{S}_{12}Q_{23} + \tilde{S}_{23}Q_{12} - \tilde{S}_{13}Q_{22} - \tilde{S}_{22}Q_{13} \\
(\Gamma_Q)_{32} &= \tilde{S}_{13}Q_{21} + \tilde{S}_{21}Q_{13} - \tilde{S}_{11}Q_{23} - \tilde{S}_{23}Q_{11} \\
(\Gamma_Q)_{33} &= \tilde{S}_{11}Q_{22} + \tilde{S}_{22}Q_{11} - \tilde{S}_{12}Q_{21} - \tilde{S}_{21}Q_{12}.
\end{aligned} \tag{E.3}$$

Proof. This is directly calculated using the fact that for any $Q \in \text{SO}(3)$, $Q_{ij} = (Q^{-1})_{ji} = (-1)^{i+j}M_{ij}$, where M_{ij} is the minor of Q_{ij} . \square

Appendix F: Propagating Moments for IMU Kinematics

This appendix completes the calculations of $E[x_{k+1}v_{R_{k+1}}^T]$ and $E[x_{k+1}x_{k+1}^T]$ in Chapter 5.1.2, which are propagated from $(R_k, x_k) \sim \mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$ through the IMU kinematic equations (5.1)-(5.5).

The notation Λ in the proof of (4.6) is used in this appendix. Specifically, for any $A \in \mathbb{R}^{3 \times 3}$, it is defined as

$$(MFGI) \quad \Lambda(A) = (U_{k+1}^T A V_{k+1} S_{k+1} - S_{k+1} V_{k+1}^T A^T U_{k+1})^\vee,$$

$$(MFGB) \quad \Lambda(A) = (S_{k+1} U_{k+1}^T A V_{k+1} - V_{k+1}^T A^T U_{k+1} S_{k+1})^\vee.$$

Theorem F.1. *The moment $E[x_{k+1}v_{R_{k+1}}^T]$ is calculated in the following equation:*

$$\begin{aligned} E[x_{k+1}v_{R_{k+1}}^T] &= \left(1 - \frac{\text{tr}(G_{gu})h}{2}\right) E[x_{k+1}\Lambda(R_k \delta R_k)^T] + h E[x_{k+1}\Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k)^T] \\ &\quad + \frac{h}{2} E[x_{k+1}\Lambda(R_k G_{gu} \delta R_k)^T] + O(h^2), \end{aligned} \quad (\text{F.1})$$

where $\delta R_k = \exp\{h(\hat{\Omega}_k + \hat{\mu}_{b_{g,k}})\}$. Define $\tilde{U}, \tilde{V} \in \text{SO}(3)$, $\tilde{S}, \tilde{V}, \tilde{\tilde{S}} \in \mathbb{R}^{3 \times 3}$, $\tilde{v}, \tilde{\tilde{v}} \in \mathbb{R}^3$ and $\Gamma_Q \in \mathbb{R}^{3 \times 3}$ as in Theorem 4.6. Then, the first term $E[x_{k+1}\Lambda(R_k \delta R_k)^T]$ in (F.1) can be decomposed into blocks $E[b_{g,k+1}\Lambda(R_k \delta R_k)^T]$, $E[p_{k+1}\Lambda(R_k \delta R_k)^T]$, $E[v_{k+1}\Lambda(R_k \delta R_k)^T]$, and $E[b_{a,k+1}\Lambda(R_k \delta R_k)^T]$. The moment $E[b_{g,k+1}\Lambda(R_k \delta R_k)^T]$ is

$$(MFGI) \quad E[b_{g,k+1}\Lambda(R_k \delta R_k)^T] = \left(\mu_{b_g,k} E[\tilde{v}_R]^T + P_{b_g,k} E[v_{R_k} \tilde{v}_R^T]\right) \tilde{U}^T, \quad (\text{F.2a})$$

$$(MFGB) \quad E[b_{g,k+1}\Lambda(R_k \delta R_k)^T] = \left(\mu_{b_g,k} E[\tilde{v}_R]^T + P_{b_g,k} E[v_{R_k} \tilde{v}_R^T]\right) \tilde{V}^T. \quad (\text{F.2b})$$

The moment $\mathbb{E} \left[p_{k+1} \Lambda (R_k \delta R_k)^T \right]$ is

$$(MFGI) \quad \mathbb{E} \left[p_{k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{p,k} \mathbb{E} [\tilde{v}_R]^T + P_{p,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{U}^T + h \left(\mu_{v,k} \mathbb{E} [\tilde{v}_R]^T + P_{v,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{U}^T, \quad (\text{F.3a})$$

$$(MFGB) \quad \mathbb{E} \left[p_{k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{p,k} \mathbb{E} [\tilde{v}_R]^T + P_{p,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{V}^T + h \left(\mu_{v,k} \mathbb{E} [\tilde{v}_R]^T + P_{v,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{V}^T. \quad (\text{F.3b})$$

The moment $\mathbb{E} \left[v_{k+1} \Lambda (R_k \delta R_k)^T \right]$ is

$$(MFGI) \quad \mathbb{E} \left[v_{k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{v,k} \mathbb{E} [\tilde{v}_R]^T + P_{v,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] - hg \mathbb{E} [\tilde{v}_R]^T \right) \tilde{U}^T + h \left(\mathbb{E} [R_k (a_k + \mu_{b_a,k}) \tilde{v}_R^T] + \mathbb{E} [R_k P_{b_a,k} v_{R_k} \tilde{v}_R^T] \right) \tilde{U}^T, \quad (\text{F.4a})$$

$$(MFGB) \quad \mathbb{E} \left[v_{k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{v,k} \mathbb{E} [\tilde{v}_R]^T + P_{v,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] - hg \mathbb{E} [\tilde{v}_R]^T \right) \tilde{V}^T + h \left(\mathbb{E} [R_k (a_k + \mu_{b_a,k}) \tilde{v}_R^T] + \mathbb{E} [R_k P_{b_a,k} v_{R_k} \tilde{v}_R^T] \right) \tilde{V}^T \quad (\text{F.4b})$$

And the moment $\mathbb{E} \left[b_{a,k+1} \Lambda (R_k \delta R_k)^T \right]$ is

$$(MFGI) \quad \mathbb{E} \left[b_{a,k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{a_g,k} \mathbb{E} [\tilde{v}_R]^T + P_{a_g,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{U}^T, \quad (\text{F.5a})$$

$$(MFGB) \quad \mathbb{E} \left[b_{a,k+1} \Lambda (R_k \delta R_k)^T \right] = \left(\mu_{a_g,k} \mathbb{E} [\tilde{v}_R]^T + P_{a_g,k} \mathbb{E} [v_{R_k} \tilde{v}_R^T] \right) \tilde{V}^T. \quad (\text{F.5b})$$

Next, the second term $\mathbb{E} \left[x_{k+1} \Lambda (R_k (\hat{b}_{g,k} - \hat{\mu}_{b_g,k}) \delta R_k)^T \right]$ in (F.1) can be approximated by

$$(MFGI) \quad \mathbb{E} \left[x_{k+1} \Lambda (R_k (\hat{b}_{g,k} - \hat{\mu}_{b_g,k}) \delta R_k)^T \right] = \left(\Sigma_{c,xb_g,k} V_k \mathbb{E} [\Gamma_Q^T] + \mu_k \mathbb{E} [v_{R_k}^T P_{b_g,k}^T V_k \Gamma_Q^T] + P_k \mathbb{E} [v_{R_k} v_{R_k}^T P_{b_g,k}^T V_k \Gamma_Q^T] \right) \tilde{U}^T + O(h), \quad (\text{F.6a})$$

$$(MFGB) \quad \mathbb{E} \left[x_{k+1} \Lambda (R_k (\hat{b}_{g,k} - \hat{\mu}_{b_g,k}) \delta R_k)^T \right] = \left(\Sigma_{c,xb_g,k} V_k \mathbb{E} [\Gamma_Q^T] + \mu_k \mathbb{E} [v_{R_k}^T P_{b_g,k}^T V_k \Gamma_Q^T] + P_k \mathbb{E} [v_{R_k} v_{R_k}^T P_{b_g,k}^T V_k \Gamma_Q^T] \right) \tilde{V}^T + O(h), \quad (\text{F.6b})$$

where $\Sigma_{c,xb_g,k}$ is the first three columns of Σ_c of $\mathcal{MG}(\mu_k, \Sigma_k, P_k, U_k, S_k, V_k)$.

Finally, the third term $E[x_{k+1}\Lambda(R_k G_{gu} \delta R_k)^T]$ in (F.1) is

$$(MFGI) \quad E[x_{k+1}\Lambda(R_k G_{gu} \delta R_k)^T] = \left(\mu_k E[\tilde{v}_R]^T + P_k E[v_{R_k} \tilde{v}_R^T] \right) \tilde{U}^T + O(h), \quad (\text{F.7a})$$

$$(MFGB) \quad E[x_{k+1}\Lambda(R_k G_{gu} \delta R_k)^T] = \mu_k E[\tilde{v}_R]^T + P_k E[v_{R_k} \tilde{v}_R^T] + O(h) \quad (\text{F.7b})$$

Proof. Equation (F.1) can be derived exactly the same as (E.1) in the proof of Theorem 4.5, noting that ΔW_{gu} is independent on (R_k, x_k) , and $E[((H_{gu}\Delta W_{gu})^\wedge)^2] = h(G_u - \text{tr}(G_u) I_{3 \times 3})$. Note that $\Lambda(R_k \delta R_k) = \tilde{U} \tilde{v}_R$ for MFGI, and $\Lambda(R_k \delta R_k) = \tilde{V} \tilde{v}_R$ for MFGB. Using a similar calculation as (3.22), it can be shown that

$$E[x_k \tilde{v}_R^T] = \mu_k E[\tilde{v}_R]^T + P_k E[v_{R_k} \tilde{v}_R^T]. \quad (\text{F.8})$$

Then (F.2), (F.3), and (F.5) are immediate from (5.2), (5.3), and (5.5). For (F.4), it can be shown that

$$E[R_k(a_k + b_{a,k})\tilde{v}^T] = E[R_k(a_k + \mu_{b_{a,k}})\tilde{v}^T] + E[R_k P_{b_{a,k}} v_{R_k} \tilde{v}^T],$$

and (F.4) is derived.

Next, since the second term in (F.1) already has h , (F.6) only needs to be calculated up to $O(h)$. Therefore, (F.6) can be approximated using

$$E[x_{k+1}\Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k)^T] = E[x_k\Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k)^T] + O(h). \quad (\text{F.9})$$

Note that $\Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k)$ can be simplified as

$$(MFGI) \quad \Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k) = \tilde{U} \Gamma_Q V_k^T (b_{g,k} - \mu_{b_{g,k}}),$$

$$(MFGB) \quad \Lambda(R_k(\hat{b}_{g,k} - \hat{\mu}_{b_{g,k}})\delta R_k) = \tilde{V} \Gamma_Q V_k^T (b_{g,k} - \mu_{b_{g,k}}).$$

Also, it can be shown that

$$\mathbb{E} [x_k (b_{g,k} - \mu_{b_{g,k}})^T V_k \Gamma_Q^T] = \mathbb{E} \left[\left(\Sigma_{c,x b_{g,k}} + \mu_k v_{R_k}^T P_{b_g}^T + P_k v_{R_k} v_{R_k}^T P_{b_g,k}^T \right) V_k \Gamma_Q^T \right],$$

from which (F.6) follows.

Finally, the third term in (F.1) also has h , so (F.7) only needs to be calculated up to $O(h)$, and x_{k+1} can be replaced by x_k as in (F.9). Because $\Lambda(R_k G_{gu} \delta R_k) = \tilde{U} \tilde{\nu}_R$ for MFGI, and $\Lambda(R_k G_{gu} \delta R_k) = \tilde{\nu}_R$ for MFGB. (F.7) follows by the same calculation in (F.8). \square

Next, the calculation for $\mathbb{E} [x_{k+1} x_{k+1}^T]$ is given in the following theorem.

Theorem F.2. *Let $\mathbb{E} [x_{k+1} x_{k+1}^T] \in \mathbb{R}^{12 \times 12}$ be written in blocks as in (5.9). Then the blocks without v_{k+1} are given by*

$$\mathbb{E} [b_{g,k+1} b_{g,k+1}^T] = \mathbb{E} [b_{g,k} b_{g,k}^T] + h G_{gv}, \quad (\text{F.10})$$

$$\mathbb{E} [b_{g,k+1} p_{k+1}^T] = \mathbb{E} [b_{g,k} p_k^T] + h \mathbb{E} [b_{g,k} v_k^T], \quad (\text{F.11})$$

$$\mathbb{E} [b_{g,k+1} b_{a,k+1}^T] = \mathbb{E} [b_{g,k} b_{a,k}^T], \quad (\text{F.12})$$

$$\mathbb{E} [p_{k+1} p_{k+1}^T] = \mathbb{E} [p_k p_k^T] + h (\mathbb{E} [p_k v_k^T] + \mathbb{E} [v_k p_k^T]) + h^2 \mathbb{E} [v_k v_k^T], \quad (\text{F.13})$$

$$\mathbb{E} [p_{k+1} b_{a,k+1}^T] = \mathbb{E} [p_k b_{a,k}^T] + h \mathbb{E} [v_k b_{a,k}^T], \quad (\text{F.14})$$

$$\mathbb{E} [b_{a,k+1} b_{a,k+1}^T] = \mathbb{E} [b_{a,k} b_{a,k}^T] + h G_{av}, \quad (\text{F.15})$$

where the expectations on the right hand side can be calculated using the blocks of $\mathbb{E} [x_k x_k^T]$.

Next, the block $\mathbb{E} [v_{k+1} b_{g,k+1}^T]$ is given by

$$\begin{aligned} \mathbb{E} [v_{k+1} b_{g,k+1}^T] &= \mathbb{E} [v_k b_{g,k}^T] - hg \mathbb{E} [b_{g,k}^T] + h (\mathbb{E} [R_k] a_k + \mathbb{E} [R_k P_{b_a,k} v_{R_k}]) \mu_{b_{g,k}}^T \\ &\quad + h (\mathbb{E} [R_k (a_k + \mu_{b_a,k}) v_{R_k}^T] + \mathbb{E} [R_k P_{b_a,k} v_{R_k} v_{R_k}^T]) P_{b_{g,k}}^T \\ &\quad + h \mathbb{E} [R_k] (\Sigma_{c,b_a b_{g,k}} + \mu_{b_a,k} \mu_{b_{g,k}}^T). \end{aligned} \quad (\text{F.16})$$

The block $\mathbb{E}[v_{k+1}p_{k+1}^T]$ is

$$\begin{aligned}
\mathbb{E}[v_{k+1}p_{k+1}^T] &= \mathbb{E}[v_k p_k^T] + h\mathbb{E}[v_k v_k^T] - hg\mathbb{E}[p_k^T] - h^2g\mathbb{E}[v_k^T] \\
&\quad + h(\mathbb{E}[R_k]a_k + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])\mu_{p,k}^T + h\mathbb{E}[R_k](\Sigma_{c,b_a p,k} + \mu_{b_a,k}\mu_{p,k}^T) \\
&\quad + h(\mathbb{E}[R_k(a_k + \mu_{b_a,k})v_{R_k}^T] + \mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T])P_{p,k}^T \\
&\quad + h^2(\mathbb{E}[R_k]a_k + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])\mu_{v,k}^T + h\mathbb{E}[R_k](\Sigma_{c,b_a v,k} + \mu_{b_a,k}\mu_{v,k}^T) \\
&\quad + h^2(\mathbb{E}[R_k(a_k + \mu_{b_a,k})v_{R_k}^T] + \mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T])P_{v,k}^T. \tag{F.17}
\end{aligned}$$

The block $\mathbb{E}[v_{k+1}v_{k+1}^T]$ is

$$\begin{aligned}
\mathbb{E}[v_{k+1}v_{k+1}^T] &= \mathbb{E}[v_k v_k^T] + h^2gg^T + h\mathbb{E}[R_k G_{au} R_k^T] - h(g\mathbb{E}[v_k]^T + \mathbb{E}[v_k]g^T) \\
&\quad - h^2g(\mathbb{E}[R_k](a_k + \mu_{b_a,k}) + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])^T \\
&\quad - h^2(\mathbb{E}[R_k](a_k + \mu_{b_a,k}) + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])g^T \\
&\quad + h(\mathbb{E}[R_k]a_k + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])\mu_{v,k}^T + h\mathbb{E}[R_k](\Sigma_{c,b_a v,k} + \mu_{b_a,k}\mu_{v,k}^T) \\
&\quad + h(\mathbb{E}[R_k(a_k + \mu_{b_a,k})v_{R_k}^T] + \mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T])P_{v,k}^T \\
&\quad + h\mu_{v,k}(\mathbb{E}[R_k]a_k + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])^T + h(\Sigma_{c,vb_a,k} + \mu_{v,k}\mu_{b_a,k}^T)\mathbb{E}[R_k]^T \\
&\quad + hP_{v,k}(\mathbb{E}[R_k(a_k + \mu_{b_a,k})v_{R_k}^T] + \mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T])^T \\
&\quad + h^2\mathbb{E}[R_k(\Sigma_{c,b_a b_a,k} + a_k a_k^T + a_k \mu_{b_a,k}^T + \mu_{b_a,k} a_k^T + \mu_{b_a,k} \mu_{b_a,k}^T)R_k^T] \\
&\quad + h^2\mathbb{E}[R_k((a_k + \mu_{b_a,k})v_{R_k}^T P_{b_a,k}^T + P_{b_a,k} v_{R_k} (a_k + \mu_{b_a,k})^T)R_k^T] \\
&\quad + h^2\mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T P_{b_a,k}^T R_k^T]. \tag{F.18}
\end{aligned}$$

And finally, the block $\mathbb{E}[v_{k+1}b_{a,k+1}^T]$ is

$$\begin{aligned}
\mathbb{E}[v_{k+1}b_{a,k+1}^T] &= \mathbb{E}[v_k b_{a,k}^T] - hg\mathbb{E}[b_{a,k}^T] + h(\mathbb{E}[R_k]a_k + \mathbb{E}[R_k P_{b_a,k} v_{R_k}])\mu_{b_a,k}^T \\
&\quad + h(\mathbb{E}[R_k(a_k + \mu_{b_a,k})v_{R_k}^T] + \mathbb{E}[R_k P_{b_a,k} v_{R_k} v_{R_k}^T])P_{b_a,k}^T \\
&\quad + h\mathbb{E}[R_k](\Sigma_{c,b_a b_a,k} + \mu_{b_a,k}\mu_{b_a,k}^T). \tag{F.19}
\end{aligned}$$

Proof. Equation (F.10)-(F.15) can be easily derived from (5.1)-(5.5). For (F.16), it can be calculated as

$$E[v_{k+1}b_{g,k+1}^T] = E[v_k b_{g,k}^T] + hE[R_k a_k b_{g,k}^T] + hE[R_k b_{a,k} b_{g,k}^T] - hgE[b_{g,k}^T],$$

where the second and third terms can be expanded into

$$E[R_k a_k b_{g,k}^T] = E[R_k] a_k \mu_{b_g,k}^T + E[R_k a_k v_{R_k}^T] P_{b_g,k}^T,$$

and

$$\begin{aligned} E[R_k b_{a,k} b_{g,k}^T] &= E[R_k] (\Sigma_{c,b_a b_g,k} + \mu_{b_a,k} \mu_{b_g,k}^T) + E[R_k \mu_{b_a,k} v_{R_k}^T] P_{b_g,k}^T \\ &\quad + E[R_k P_{b_a,k} v_{R_k}] \mu_{b_g,k}^T + E[R_k P_{b_a,k} v_{R_k} v_{R_k}^T] P_{b_g,k}^T. \end{aligned}$$

Then (F.16) can be proved after collecting the terms in the above two equations. The proofs for (F.17)-(F.19) are similar but extremely tedious, and they are omitted for brevity. \square

Note that the term $E[R_k P_{b_a,k} v_{R_k} v_{R_k}^T P_{b_a,k}^T R_k^T]$ in (F.18) needs the fourth order moments of $E[Q_k]$, which do not have closed form formula. But its accuracy is on the order of $O(h^2)$, and can be omitted if the sampling period h is small enough.

Appendix G: Posterior Density With Map Uncertainty

This appendix deals with the calculations related to the posterior density function (5.78).

First, the proof for Theorem 5.9 is given.

Proof of Theorem 5.9. The last two terms in (5.79) can be combined into

$$\begin{aligned}
& (\mathbf{x} - \boldsymbol{\mu}_{x|R})^T \boldsymbol{\Sigma}_{x|R}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{x|R}) + (y - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} (y - \boldsymbol{\mu}_c) \\
&= (Hz - \boldsymbol{\mu}_{x|R})^T \boldsymbol{\Sigma}_{x|R}^{-1} (Hz - \boldsymbol{\mu}_{x|R}) + (Gz - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} (Gz - \boldsymbol{\mu}_c) \\
&= z^T \left(H^T \boldsymbol{\Sigma}_{x|R}^{-1} H + G^T \boldsymbol{\Sigma}_c^{-1} G \right) z - z^T \left(H^T \boldsymbol{\Sigma}_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + G^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c \right) + \boldsymbol{\mu}_{x|R}^T \boldsymbol{\Sigma}_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c \\
&= (z - \boldsymbol{\mu}_{z|R})^T \boldsymbol{\Sigma}_{z|R}^{-1} (z - \boldsymbol{\mu}_{z|R}) - \boldsymbol{\mu}_{z|R}^T \boldsymbol{\Sigma}_{z|R}^{-1} \boldsymbol{\mu}_{z|R} + \boldsymbol{\mu}_{x|R}^T \boldsymbol{\Sigma}_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c.
\end{aligned}$$

Substitute this into (5.79), it remains to show that

$$(\boldsymbol{\mu}_{t|R} - \boldsymbol{\mu}_{c,t})^T (\boldsymbol{\Sigma}_{t|R} + \boldsymbol{\Sigma}_{c,t})^{-1} (\boldsymbol{\mu}_{t|R} - \boldsymbol{\mu}_{c,t}) = \boldsymbol{\mu}_{x|R}^T \boldsymbol{\Sigma}_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c - \boldsymbol{\mu}_{z|R}^T \boldsymbol{\Sigma}_{z|R}^{-1} \boldsymbol{\mu}_{z|R}.$$

Note that $HH^T = I_{3N+3}$, and $GG^T = I_{n \times n}$, so the right hand side can be expanded into

$$\begin{aligned}
& \boldsymbol{\mu}_{x|R}^T \boldsymbol{\Sigma}_{x|R}^{-1} \boldsymbol{\mu}_{x|R} + \boldsymbol{\mu}_c^T \boldsymbol{\Sigma}_c^{-1} \boldsymbol{\mu}_c - \boldsymbol{\mu}_{z|R}^T \boldsymbol{\Sigma}_{z|R}^{-1} \boldsymbol{\mu}_{z|R} \\
&= (H^T \boldsymbol{\mu}_{x|R})^T \left[H^T \boldsymbol{\Sigma}_{x|R}^{-1} H - H^T \boldsymbol{\Sigma}_{x|R}^{-1} H \left(H^T \boldsymbol{\Sigma}_{x|R}^{-1} H + G^T \boldsymbol{\Sigma}_c^{-1} G \right)^{-1} H^T \boldsymbol{\Sigma}_{x|R}^{-1} H \right] (H^T \boldsymbol{\mu}_{x|R}) \\
&\quad + (G^T \boldsymbol{\mu}_c)^T \left[G^T \boldsymbol{\Sigma}_c^{-1} G - G^T \boldsymbol{\Sigma}_c^{-1} G \left(H^T \boldsymbol{\Sigma}_{x|R}^{-1} H + G^T \boldsymbol{\Sigma}_c^{-1} G \right)^{-1} G^T \boldsymbol{\Sigma}_c^{-1} G \right] (G^T \boldsymbol{\mu}_c) \\
&\quad - 2(H^T \boldsymbol{\mu}_{x|R})^T H^T \boldsymbol{\Sigma}_{x|R}^{-1} H \left(H^T \boldsymbol{\Sigma}_{x|R}^{-1} H + G^T \boldsymbol{\Sigma}_c^{-1} G \right)^{-1} G^T \boldsymbol{\Sigma}_c^{-1} G (G^T \boldsymbol{\mu}_c). \tag{G.1}
\end{aligned}$$

The first term on the right hand side of (G.1) can be simplified using block inversion formula,

and Woodbury matrix identity as

$$\begin{aligned}
& (H^T \mu_{x|R})^T \left[H^T \Sigma_{x|R}^{-1} H - H^T \Sigma_{x|R}^{-1} H \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} H^T \Sigma_{x|R}^{-1} H \right] (H^T \mu_{x|R}) \\
&= \mu_{x|R}^T \Sigma_{x|R}^{-1} H \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} G^T \Sigma_c^{-1} G H^T \mu_{x|R} \\
&= \mu_{t|R}^T \Sigma_{t|R}^{-1} \left(\Sigma_{t|R}^{-1} + \Sigma_{c,t}^{-1} \right)^{-1} \Sigma_{c,t}^{-1} \mu_{t|R} \\
&= \mu_{t|R}^T (\Sigma_{t|R} + \Sigma_{c,t})^{-1} \mu_{t|R},
\end{aligned}$$

where the first equality is a straightforward but extremely tedious application of the block inversion formula to $\Sigma_{z|R}$, which has to be omitted here. And the second equality is because

$$\Sigma_{x|R}^{-1} H \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} G^T \Sigma_c^{-1} = \begin{bmatrix} 0_{3N \times 3} & 0_{3N \times n} \\ \Sigma_{t|R}^{-1} \left(\Sigma_{t|R}^{-1} + \Sigma_{c,t}^{-1} \right)^{-1} \Sigma_{c,t}^{-1} & 0_{3 \times n} \end{bmatrix}.$$

Similarly, the second term on the right hand side of (G.1) can be simplified as

$$\begin{aligned}
& (G^T \mu_c)^T \left[G^T \Sigma_c^{-1} G - G^T \Sigma_c^{-1} G \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} G^T \Sigma_c^{-1} G \right] (G^T \mu_c) \\
&= \mu_{c,t}^T (\Sigma_{t|R} + \Sigma_{c,t})^{-1} \mu_{c,t},
\end{aligned}$$

and the third term can be simplified as

$$\begin{aligned}
& (H^T \mu_{x|R})^T H^T \Sigma_{x|R}^{-1} H \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} G^T \Sigma_c^{-1} G (G^T \mu_c) \\
&= \mu_{t|R}^T (\Sigma_{t|R} + \Sigma_{c,t})^{-1} \mu_{c,t}.
\end{aligned}$$

Combining these three terms yields the desired equality. \square

Next, the moments for $y|R$ and $p_i|R$ according to the posterior density (5.82) are given in the following theorem.

Theorem G.1. Define $K_y \in \mathbb{R}^{n \times n}$ as

$$K_y = \left(\begin{bmatrix} \tilde{C}^{-1} & 0_{3 \times (n-3)} \\ 0_{(n-3) \times 3} & 0_{(n-3) \times (n-3)} \end{bmatrix} + \Sigma_c^{-1} \right)^{-1} \Sigma_c^{-1}. \quad (\text{G.2})$$

Then the moment $\mathbb{E}[y|\mathcal{Z}]$ can be calculated as

$$\mathbb{E}[y|\mathcal{Z}] = (I - K_y) \begin{bmatrix} \mathbb{E}[\mu_{t|R}|\mathcal{Z}] \\ 0_{(n-3) \times 1} \end{bmatrix} + K_y(\mu + P\mathbb{E}[v_R|\mathcal{Z}]). \quad (\text{G.3})$$

The moment $\mathbb{E}[y(v_R^+)^T|\mathcal{Z}]$ is

$$\mathbb{E}[y(v_R^+)^T|\mathcal{Z}] = (I - K_y) \begin{bmatrix} \mathbb{E}[\mu_{t|R}(v_R^+)^T|\mathcal{Z}] \\ 0_{(n-3) \times 1} \end{bmatrix} + K_y P \mathbb{E}[v_R(v_R^+)^T|\mathcal{Z}]. \quad (\text{G.4})$$

And the moment $\mathbb{E}[yy^T|\mathcal{Z}]$ is

$$\begin{aligned} \mathbb{E}[yy^T|\mathcal{Z}] &= \left(\begin{bmatrix} \tilde{C}^{-1} & 0_{3 \times (n-3)} \\ 0_{(n-3) \times 3} & 0_{(n-3) \times (n-3)} \end{bmatrix} + \Sigma_c^{-1} \right)^{-1} \\ &+ (I - K_y) \begin{bmatrix} \mathbb{E}[\mu_{t|R}\mu_{t|R}^T|\mathcal{Z}] & 0_{3 \times (n-3)} \\ 0_{(n-3) \times 3} & 0_{(n-3) \times (n-3)} \end{bmatrix} (I - K_y)^T \\ &+ (I - K_y) \begin{bmatrix} \mathbb{E}[\mu_{t|R}|\mathcal{Z}]\mu^T + \mathbb{E}[\mu_{t|R}v_R^T|\mathcal{Z}]P^T \\ 0_{(n-3) \times n} \end{bmatrix} K_y^T \\ &+ K_y \begin{bmatrix} \mathbb{E}[\mu_{t|R}|\mathcal{Z}]\mu^T + \mathbb{E}[\mu_{t|R}v_R^T|\mathcal{Z}]P^T \\ 0_{(n-3) \times n} \end{bmatrix}^T (I - K_y)^T \\ &+ K_y (\mu\mu^T + \mu\mathbb{E}[v_R^T|\mathcal{Z}]P^T + P\mathbb{E}[v_R|\mathcal{Z}]\mu^T + P\mathbb{E}[v_Rv_R^T|\mathcal{Z}]P^T) K_y^T. \end{aligned} \quad (\text{G.5})$$

In the above three equations, I is abbreviated for $I_{n \times n}$, $\mathbb{E}[v_R|\mathcal{Z}]$, $\mathbb{E}[v_R(v_R^+)^T|\mathcal{Z}]$ and

$\mathbb{E}[v_R v_R^T | \mathcal{Z}]$ are calculated as in Theorem 4.9. Also, $\mathbb{E}[\mu_{t|R} | \mathcal{Z}]$, $\mathbb{E}[\mu_{t|R}(v_R^+)^T | \mathcal{Z}]$, $\mathbb{E}[\mu_{t|R} v_R^T | \mathcal{Z}]$, and $\mathbb{E}[\mu_{t|R} \mu_{t|R}^T | \mathcal{Z}]$ are calculated as in Theorem 5.7, with p_i , C , D_i replaced by \bar{p}_i , \tilde{C} , \tilde{D}_i .

Next, the moment $\mathbb{E}[p_i | \mathcal{Z}]$ is

$$\mathbb{E}[p_i | \mathcal{Z}] = (I_{3 \times 3} - K_i) \bar{p}_i + K_i (R p'_i + \mathbb{E}[t | \mathcal{Z}]). \quad (\text{G.6})$$

And the moment $\mathbb{E}[p_i p_i^T | \mathcal{Z}]$ is

$$\begin{aligned} \mathbb{E}[p_i p_i^T | \mathcal{Z}] &= B_i - B_i [(A_i + B_i)^{-1} (A_i + B_i + (\tilde{C}^{-1} + \Sigma_{c,t}^{-1})^{-1}) (A_i + B_i)^{-1}] B_i \\ &\quad + (I_{3 \times 3} - K_i) \bar{p}_i \bar{p}_i^T (I_{3 \times 3} - K_i)^T + (I_{3 \times 3} - K_i) \bar{p}_i \left((p'_i)^T \mathbb{E}[R | \mathcal{Z}]^T + \mathbb{E}[t | \mathcal{Z}]^T \right) K_i^T \\ &\quad + K_i \left(\mathbb{E}[R p'_i (p'_i)^T R^T | \mathcal{Z}] + \mathbb{E}[R p'_i t^T | \mathcal{Z}] + \mathbb{E}[t (p'_i)^T R^T | \mathcal{Z}] + \mathbb{E}[\mu'_{t|R} (\mu'_{t|R})^T | \mathcal{Z}] \right) \\ &\quad + K_i (\mathbb{E}[R | \mathcal{Z}] p'_i + \mathbb{E}[t | \mathcal{Z}]) \bar{p}_i^T (I_{3 \times 3} - K_i)^T. \end{aligned} \quad (\text{G.7})$$

In the above two equations, $\mathbb{E}[t | \mathcal{Z}]$ is the first three columns of $\mathbb{E}[y | \mathcal{Z}]$, and $\mathbb{E}[\mu'_{t|R} (\mu'_{t|R})^T | \mathcal{Z}]$ is the first 3-by-3 diagonal block of the last four term on the right hand side of (G.5). Finally, let $K_t \in \mathbb{R}^{3 \times 3}$ be the first 3-by-3 diagonal block of K_y , then

$$\mathbb{E}[t (p'_i)^T R^T | \mathcal{Z}] = (I_{3 \times 3} - K_t) \mathbb{E}[\mu_{t|R} (p'_i)^T R^T | \mathcal{Z}] + K_t \mathbb{E}[\mu_{c,t} (p'_i)^T R^T | \mathcal{Z}], \quad (\text{G.8})$$

where $\mathbb{E}[\mu_{t|R} (p'_i)^T R^T | \mathcal{Z}]$ is given in Theorem 5.8.

Proof. The key idea is to calculate the diagonal blocks of the covariance matrix $\Sigma_{z|R}$ and the conditional mean $\mu_{z|R}$ in Theorem 5.9, corresponding to y and p_i . First, by directly applying block inversion formula to $\Sigma_{z|R}$, the diagonal block for y is calculated as

$$\Sigma'_{y|R} = \left(\begin{bmatrix} \tilde{C}^{-1} & 0_{3 \times (n-3)} \\ 0_{(n-3) \times 3} & 0_{(n-3) \times (n-3)} \end{bmatrix} + \Sigma_c^{-1} \right)^{-1},$$

and the block for p_i is

$$\Sigma'_{p_i|R} = B_i - B_i \left[(A_i + B_i)^{-1} (A_i + B_i + (\tilde{C}^{-1} + \Sigma_{c,t}^{-1})^{-1}) (A_i + B_i)^{-1} \right] B_i.$$

Next, the conditional mean can be written as

$$\begin{aligned} \mu_{z|R} &= \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} \left(H^T \Sigma_{x|R}^{-1} \mu_{x|R} + G^T \Sigma_c^{-1} \mu_c \right) \\ &= H^T \mu_{x|R} + \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} (G^T \Sigma_c^{-1} \mu_c - G^T \Sigma_c^{-1} G H^T \mu_{x|R}) \\ &= H^T \mu_{x|R} + \left(H^T \Sigma_{x|R}^{-1} H + G^T \Sigma_c^{-1} G \right)^{-1} G^T \Sigma_c^{-1} \left(\mu_c - \begin{bmatrix} \mu_{t|R} \\ 0_{(n-3) \times 1} \end{bmatrix} \right). \end{aligned}$$

So the component for y is

$$\mu'_{y|R} = (I - K_y) \begin{bmatrix} \mu_{t|R} \\ 0_{(n-3) \times 1} \end{bmatrix} + K_y \mu_c,$$

and the component for p_i is

$$\mu'_{p_i|R} = \mu_{p_i|R} + \begin{bmatrix} K_i & 0_{3 \times (n-3)} \end{bmatrix} K_y \left(\mu_c - \begin{bmatrix} \mu_{t|R} \\ 0_{(n-3) \times 1} \end{bmatrix} \right).$$

Using block inversion formula, note that

$$K_y = \begin{bmatrix} (\tilde{C}^{-1} + \Sigma_{c,t}^{-1})^{-1} \Sigma_{c,t}^{-1} & 0_{3 \times n} \\ \cdot & I_{n \times n} \end{bmatrix}.$$

So $\mu'_{p_i|R}$ can be simplified as

$$\begin{aligned}\mu'_{p_i|R} &= \mu_{p_i|R} + K_i K_t (\mu_{c,t} - \mu_{t|R}) \\ &= (I - F_i) \bar{p}_i + K_i (R p'_i + \mu_{t|R}) + K_i K_t (\mu_{c,t} - \mu_{t|R}) \\ &= (I - F_i) \bar{p}_i + K_i (R p'_i + \mu'_{t|R}),\end{aligned}$$

where $\mu'_{t|R}$ is the first three columns of $\mu'_{y|R}$. The above calculations are used when integrating the left hand side of the moments with respect to the last term of (5.82), and evaluating the resulting moments with respect to $\mathcal{M}(F^+)$, which is the matrix Fisher distribution matched to the first three terms on the right hand side of (5.82). \square