JOURNAL OF ALGEBRAIC STATISTICS
Vol. 2, No. 1, 2011, 75-97
ISSN 1309-3452 – www.jalgstat.com

Bayesian Fusion on Lie Groups

Kevin C. Wolfe [1], Michael Mashner [1], Gregory S. Chirikjian[1,*]
[1] *Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland, USA*

Abstract. An increasing number of real-world problems involve the measurement of data, and the computation of estimates, on Lie groups. Moreover, establishing confidence in the resulting estimates is important. This paper therefore seeks to contribute to a larger theoretical framework that generalizes classical multivariate statistical analysis from Euclidean space to the setting of Lie groups. The particular focus here is on extending Bayesian fusion, based on exponential families of probability densities, from the Euclidean setting to Lie groups. The definition and properties of a new kind of Gaussian distribution for connected unimodular Lie groups are articulated, and explicit formulas and algorithms are given for finding the mean and covariance of the fusion model based on the means and covariances of the constituent probability densities. The Lie groups that find the most applications in engineering are rotation groups and groups of rigid-body motions. Orientational (rotation-group) data and associated algorithms for estimation arise in problems including satellite attitude, molecular spectroscopy, and global geological studies. In robotics and manufacturing, quantifying errors in the position and orientation of tools and parts are important for task performance and quality control. Developing a general way to handle problems on Lie groups can be applied to all of these problems. In particular, we study the issue of how to 'fuse' two such Gaussians and how to obtain a new Gaussian of the same form that is 'close to' the fused density.This is done at two levels of approximation that result from truncating the Baker-Campbell-Hausdorff formula with different numbers of terms. Algorithms are developed and numerical results are presented that are shown to generate the equivalent fused density with good accuracy.
2000 Mathematics Subject Classifications: 46L53, 62F15, 54H15
Key Words and Phrases: Bayesian fusion, Lie groups, parametric distributions, belief propagation

---

1. Introduction

In this paper we extend concepts and computations from Bayesian belief propagation to the case when the belief state is an element of a Lie group, and all corresponding probability densities are functions on that group. In particular, we focus on *connected unimodular matrix Lie groups*. Henceforth, when referring to Lie groups, these are the Lie groups being addressed.

*Corresponding author.

Email addresses: `kevin.wolfe@jhu.edu` (K. Wolfe), `mashner@jhu.edu` (M. Mashner)
`gregc@jhu.edu` (G. Chirikjian)

1.1. Literature Review

The concept of probability densities on Lie groups arise in practical settings such as rotational Brownian motion of rigid molecules in solution [20, 16, 28, 8], and this has led to more theoretical studies of Brownian motion and heat flow on the rotation group and other Lie groups [5, 6, 17]. The discussion that follows is concerned with a general connected unimodular Lie group, G, with rotations and rigid-body motions serving as important examples.

Given two probability densities, f_1 and f_2 that take their arguments in G, there are several natural operations that result in new probability densities. One such operation is the convolution,

$$(f_1 * f_2)(g) \doteq \int_G f_1(h)\, f_2(h^{-1} \circ g)\, dh\,.$$

Another is fusion

$$f_{1,2}(g) \doteq \frac{f_1(g)\, f_2(g)}{\int_G f_1(h)\, f_2(h)\, dh}\,.$$

Various concepts of mean and covariance of probability densities on Lie groups have been defined in the literature over the past half century, as described in [4, 9, 10, 7, 1]. A natural question to ask within a given definition of mean and covariance of f_1 and f_2, is: "what are the means and covariances of $f_1 * f_2$ and $f_{1,2}$?" The former is answered in [26, 27] in the context of the concept of mean and covariance defined there, and used later in this paper. Such "propagation" formulas have been used (either explicitly or implicitly) in applications ranging from mobile robotics [23, 25, 11] and robot arms [24, 18, 19] to biomolecular conformational motions [2, 22]. Our goal here is to do the same for $f_{1,2}$ in the context of a particular kind of exponential family on G.

In the literature, a number of exponential families on compact Lie groups have been defined [13, 12, 14, 15, 21]. These are defined so as to have nice properties under conditioning, and are typically of the form

$$\rho(g; \{\beta(\lambda)\}) = \alpha(\{\beta(\lambda)\}) \cdot \exp\left(\sum_{\lambda} \mathrm{tr}[\beta(\lambda) U(g,\lambda)]\right)$$

where $U(g;\lambda)$ is an irreducible unitary representation, which has the property

$$U(g_1 \circ g_2; \lambda) = U(g_1;\lambda) U(g_2;\lambda),$$

and $\{\beta(\lambda)\}$ is a set of weighting functions enumerated by $\lambda \in \hat{G}$, which is the space of all such λ values, and is called the unitary dual of G. When considering noncompact Lie groups, the additional condition that the probability density functions (pdfs) decay in spatial dimensions that extend to infinity is required. This can be problematic to incorporate into the above form.

Moreover, these forms do not provide intuitive properties under convolution. An exponential family introduced in [26, 27] that behaves well under convolution is reviewed later in the paper. Though this does not have the exact form closure under conditioning or fusion, it is shown how very good approximations of the fusion can be obtained which do have the same form as the original distributions.

1.2. Overview of Paper

Section 2 reviews Bayesian fusion of belief states described in terms of probability densities, one of which represents a prior, and the other of which is a corrector based on observations, in the case where the domain of the probability densities is a Euclidean space. Section 3 formulates the general problem of Bayesian fusion when the domain of the probability densities of interest is a Lie group. Section 4 focuses on a particular version of the problem in which the probability densities of interest are parametric in nature. Numerical validation of the approach proposed in Section 4 is presented in Section 5 for $SO(3)$. Nomenclature used throughout the paper can found in Appendix C.

2. Review of Bayesian Fusion in $\mathbb{R}^d$

Given a conditional probability density $f(\mathbf{y}\,|\,\mathbf{x})$ and a marginal $f(\mathbf{x})$, Bayes rule states that

$$f(\mathbf{x}\,|\,\mathbf{y}) = \frac{1}{f(\mathbf{y})} f(\mathbf{y}\,|\,\mathbf{x}) f(\mathbf{x}). \tag{1}$$

If $\mathbf{y}$ is a fixed value (e.g., an observation) then $f(\mathbf{y})$ can be treated as a constant. This result is nonparametric (i.e., it is true for all kinds of probability densities).

For so-called exponential families such as Gaussians, computations are greatly facilitated. Recall that a Gaussian distribution on $\mathbb{R}^d$ has the form

$$f(\mathbf{x};\boldsymbol{\mu},\Sigma) = (2\pi)^{-d/2}|\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]. \tag{2}$$

A well-known property of Gaussians is that they are closed under convolution:

$$f(\mathbf{x};\boldsymbol{\mu}_1,\Sigma_1) * f(\mathbf{x};\boldsymbol{\mu}_2,\Sigma_2) = f(\mathbf{x};\boldsymbol{\mu}_1+\boldsymbol{\mu}_2,\Sigma_1+\Sigma_2),$$

where the notation $f_1(\mathbf{x}) * f_2(\mathbf{x})$ means the same thing as

$$(f_1 * f_2)(\mathbf{x}) = \int_{\mathbf{y}\in\mathbb{R}^d} f_1(\mathbf{y}) f_2(\mathbf{x}-\mathbf{y})\, d\mathbf{y}.$$

Whereas convolution, and its extension to the context of Lie groups, is a central operation in works such as [26, 27, 18, 1, 2], the properties of convolution are not the focus of the current discussion.

If $f_1(\mathbf{x}) \doteq f(\mathbf{x})$ is Gaussian and $f(\mathbf{y}\,|\,\mathbf{x})$ is Gaussian (to within a constant scale factor), then their product will also be Gaussian to within a scale factor. Let

$$f_2(\mathbf{x}) \doteq \frac{f(\mathbf{y}\,|\,\mathbf{x})}{\int_{\mathbf{y}} f(\mathbf{y}\,|\,\mathbf{x}) d\mathbf{y}}.$$

Then at the core of (parametric) Baysian calculations following from (1) is the generation of the new Gaussian

$$f_{1,2}(\mathbf{x}) \doteq \frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\int_{\mathbf{x}} f_1(\mathbf{x})f_2(\mathbf{x})d\mathbf{x}}. \tag{3}$$

The denominator is just a constant. If the mean and covariance of the numerator are extracted from the exponential in the numerator, it is easy to obtain $f_{1,2}(\mathbf{x}) = f(\mathbf{x}; \boldsymbol{\mu}_{1,2}, \Sigma_{1,2})$.

More generally, if $f_i(\mathbf{x}) = f(\mathbf{x}; \boldsymbol{\mu}_1, \Sigma_i)$, we can find $(\boldsymbol{\mu}_{1,2,\ldots,n}, \Sigma_{1,2,\ldots,n})$ where

$$f_{1,2,\ldots,n}(\mathbf{x}) \doteq \alpha \prod_{i=1}^{n} f_i(\mathbf{x}) \tag{4}$$

(here α is a normalizing constant), by simply observing that

$$\prod_{i=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right] = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right]$$

and recollecting terms in the sum in the form $(\mathbf{x}-\boldsymbol{\mu}_{1,2,\ldots,n})^T \Sigma_{1,2,\ldots,n}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{1,2,\ldots,n})$. This gives

$$\Sigma_{1,2,\ldots,n}^{-1} = \sum_{i=1}^{n} \Sigma_i^{-1} \tag{5}$$

and

$$\boldsymbol{\mu}_{1,2,\ldots,n} = \Sigma_{1,2,\ldots,n}\left(\sum_{i=1}^{n} \Sigma_i^{-1}\boldsymbol{\mu}_i\right). \tag{6}$$

These equations lend themselves to recursive implementation, as

$$\Sigma_{1,2,\ldots,n}^{-1} = \Sigma_{1,2,\ldots,n-1}^{-1} + \Sigma_n^{-1} \tag{7}$$

and

$$\boldsymbol{\mu}_{1,2,\ldots,n} = \Sigma_{1,2,\ldots,n}\left(\Sigma_{1,2,\ldots,n-1}^{-1}\boldsymbol{\mu}_{1,2,\ldots,n-1} + \Sigma_n^{-1}\boldsymbol{\mu}_n\right). \tag{8}$$

In what follows, we will examine how to formulate similar formulas for data in Lie groups.

3. Bayesian Fusion of Observations in Connected Unimodular Lie Groups

In this section, Lie groups are viewed as a domain in which data is measured and on which probability densities are defined. Our goal is to extend formulas such as (5)–(8) from the Euclidean setting to this Lie-group setting.

3.1. Notation and Terminology

Let G be a *connected* Lie group, let $\mathcal{G}$ be the corresponding Lie algebra, and let d denote their dimension. In many applications, the Lie groups of interest are $SO(N)$, the special orthogonal group of $N \times N$ matrices, and $SE(N)$, the special Euclidean group consisting of $(N+1) \times (N+1)$ homogeneous transformation matrices of the form

$$g = \begin{bmatrix} R & \mathbf{p} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

where $g \in SE(N)$, $R \in SO(N)$, $\mathbf{p} \in \mathbb{R}^d$, and $\mathbf{0}$ is the zero vector in $\mathbb{R}^d$. Both are connected for all $N > 1$. The dimensions of these groups are $d = N(N-1)/2$ for $SO(N)$ and $d = N(N+1)/2$ for $SE(N)$. For matrix Lie groups, the exponential map

$$\exp : \mathcal{G} \longrightarrow G \tag{9}$$

is simply the matrix exponential. For elements of G for which this map can be uniquely inverted, the logarithm map is defined. Let G' denote this set. Then

$$\log : G' \longrightarrow \mathcal{G}'$$

and (9) holds for $\mathcal{G}'$ and G' as well.

The exponential map for $SO(N)$ applied to an open ball of radius π centered at the origin of the Lie algebra $so(N)$ produces all of $SO(N)$ minus a set of measure zero. This slightly depleted subset of $SO(N)$, $SO(N)'$, maps bijectively back to the open ball in $so(N)$, $so(N)'$, under the logarithm map. A similar result follows for $SE(N)$. Since all of our results are robust to changes on sets of measure zero, we will make no distinction between the whole groups, G, and their depleted subsets, G', that map bijectively with a region in the corresponding Lie algebras.

Given a basis $\{E_i\}$ for the Lie algebra $\mathcal{G}$, it is possible to identify an arbitrary element $X \in \mathcal{G}$ with a vector $\mathbf{x} \in \mathbb{R}^d$ by defining the "vee" operator $\vee : \mathcal{G} \to \mathbb{R}^d$ by making the identification $(E_i)^\vee = \mathbf{e}_i$, the i^{th} natural unit basis vector. It will happen frequently that we will apply the $\vee$ to $\log g$. This will be written in shorthand as

$$\mathbf{v}(g) \doteq (\log g)^\vee. \tag{10}$$

For example, if $X \in so(3)$ such that

$$X = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}$$

then $(X)^\vee = [x_1, x_2, x_3]^T$. The "vee" operators for SE(3) and SE(2) can be found in Appendix A. Similarly, a "hat" operator $\wedge : \mathbb{R}^d \to \mathcal{G}$ can be defined as the inverse of the "vee" operator such that $(\mathbf{e}_i)^\wedge = E_i$.

A *unimodular* Lie group, $(G, \circ)$, consisting of a continuum of elements, G, is one that possesses a bi-invariant integration measure, dg, such that the concept of probability densities $f(g)$ make sense in that the the integral

$$\int_G f(g)dg = \int_G f(g_0 \circ g)dg = \int_G f(g \circ g_0)dg = 1.$$

for any fixed $g_0 \in G$, where $\circ$ is the group operation. Our discussion will be restricted to matrix Lie groups that are both connected and unimodular, with $SO(N)$ and $SE(N)$ being of particular interest from the standpoint of applications.

3.2. Problem Formulation

Given two probability densities $f_1(g)$ and $f_2(g)$, the goal of fusion in this context is simply the calculation of a third probability density function $f_{1,2}(g)$ that minimizes a cost such as

$$C = \int_G \left| f_{1,2}(g) \; - \; \frac{f_1(g)f_2(g)}{\int_G f_1(h)f_2(h)dh} \right| dg. \tag{11}$$

Ideally, one would like the cost to be zero, as in the case of Gaussians in $\mathbb{R}^d$, but there is no a priori reason to believe that this should be possible.

In previous works [27, 18], the mean and covariance of a pdf on a connected unimodular Lie group for which the exponential map is surjective (depleted by the set of measure zero on which the logarithm map becomes singular) were respectively defined as $\tilde{\mu} \in G$ and $\tilde{\Sigma} = \tilde{\Sigma}^T \in \mathbb{R}^{d \times d}$ such that

$$\int_G \mathbf{v}(\tilde{\mu}^{-1} \circ g)f(g)dg = \mathbf{0} \tag{12}$$

and

$$\tilde{\Sigma} = \int_G \mathbf{v}(\tilde{\mu}^{-1} \circ g)[\mathbf{v}(\tilde{\mu}^{-1} \circ g)]^T f(g)dg. \tag{13}$$

This nonparametric definition is most useful when the probability density function is concentrated (in the sense that $\|\tilde{\Sigma}\|$ is small), and symmetric around the mean (in the sense that $f(\mu \circ g) = f(g^{-1} \circ \tilde{\mu}^{-1})$). The problem of interest is then to solve (11) in terms of the resulting $(\tilde{\mu}, \tilde{\Sigma})$ given $(\tilde{\mu}_1, \tilde{\Sigma}_1)$ and $(\tilde{\mu}_2, \tilde{\Sigma}_2)$ corresponding to $f_1(g)$ and $f_2(g)$. In the following section, a concept of a Gaussian distribution that satisfies these properties is defined and used.

4. Parametric Bayesian Fusion of Observations in Connected Unimodular Matrix Lie Groups

A Gaussian distribution on $\mathbb{R}^d$ can be defined equivalently in terms of its parametric form, or as the solution to a diffusion equation evaluated at a specific value of time. This is not true in more general settings, including the case of connected unimodular matrix

Lie groups. Moreover, the mean and covariance defined in (12) and (13) may not be the most natural ways to parameterize a concept of Gaussians on Lie groups. In this section an exponential family is defined and an algorithm for fusion is presented that mimics the Euclidean case. The Baker-Campbell-Hausdorff formula is used at different levels of truncation to obtain approximate results that are accurate under different ranges of conditions.

4.1. Concentrated Gaussians on Lie Groups

A Gaussian distribution on G can be defined as [26, 27]

$$f(g; \mu, \Sigma) \doteq \alpha \, \exp\left\{-\frac{1}{2}[\mathbf{v}(\mu^{-1} \circ g)]^T \Sigma^{-1} \mathbf{v}(\mu^{-1} \circ g)\right\}. \tag{14}$$

Here α is a normalizing constant to ensure that the distribution is a pdf, and $g = \exp(X)$ is defined for all values in the Lie algebra that map to the depleted version of G. For the set of measure zero in G that is outside of this depleted version of G, the function $f(\cdot)$ is defined to have a value of zero. It should be noted that the μ and Σ found in (14) may not be the same as the mean and covariance defined in (12) and (13), respectively. However, when $\|\Sigma\|$ is small, then so too will be $\|\tilde{\Sigma}\|$, and it can be shown that in this case $\tilde{\Sigma} \to \Sigma$, and likewise $\tilde{\mu} = \mu$.

The exponential family in (14) by design has the property that

$$\prod_{i=1}^{n} f(g; \mu_i, \Sigma_i) = \left(\prod_{i=1}^{n} \alpha_i\right) \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}[\mathbf{v}(\mu_i^{-1} \circ g)]^T \Sigma_i^{-1} \mathbf{v}(\mu_i^{-1} \circ g)\right\}. \tag{15}$$

However, due to the nonlinearity of the exponential and logarithm maps, there is no hope to obtain an exact closed-from for $f(g; \mu_{1,2,\ldots,n}, \Sigma_{1,2,\ldots,n})$ that is proportional to $\prod_{i=1}^{n} f(g; \mu_i, \Sigma_i)$. Yet, if the μ_i's are sufficiently clustered in the sense that $\delta(\mu_i, \mu_j) = O(\epsilon)^{\dagger}$, and the Σ_i's are sufficiently concentrated in the sense that $\|\Sigma_i\| = O(\epsilon^2)$ where $\epsilon \in \mathbb{R}_{\geq 0}$ is a sufficiently small positive number, then various levels of approximation can be made.

At the core of these approximations will be the realization that $\mu_{1,2,\ldots,n} = \hat{\mu}_{1,2,\ldots,n} \circ \epsilon_{1,2,\ldots,n}$ and each $\mu_i = \hat{\mu}_{1,2,\ldots,n} \circ \epsilon_i$ where $\hat{\mu}_{1,2,\ldots,n}, \epsilon_{1,2,\ldots,n}, \epsilon_i \in G$, $\hat{\mu}_{1,2,\ldots,n}$ is an initial estimate of $\mu_{1,2,\ldots,n}$, and $\epsilon_{1,2,\ldots,n}$ and ϵ_i are small in the sense that they can be approximated accurately taking linear or quadratic terms in the Taylor series defining the exponential map. When they are so small that linear terms are sufficient, this will result in a "first-order theory." When quadratic terms are required, this will lead to a "second-order theory."

There are many ways to define $\hat{\mu}_{1,2,\ldots,n}$ such that it is an initial estimate of the "mean of the means". For this work, an initial estimate inspired by (5)–(6) for the product of

$^{\dagger}\delta(\cdot\,,\,\cdot)$ is a metric such as those found in [3]. For $SO(N)$, a natural metric is $\delta(\mu_i, \mu_j) = \|\log(\mu_i^{-1} \circ \mu_j)\|$ where $\|\cdot\|$ is the Frobenius norm of the resulting skew-symmetric matrix.

Gaussians on $\mathbb{R}^d$ is used

$$\hat{\mu}_{1,2,\ldots,n} = \exp\left\{\left[\left(\sum_{i=1}^{n}\Sigma_i^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\Sigma_i^{-1}\mathbf{v}(\mu_i)\right)\right]^{\wedge}\right\}. \tag{16}$$

Since

$$\log(\mu_i^{-1}\circ g) = \log(\epsilon_i^{-1}\circ\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g),$$

the Baker-Campbell-Hausdorff (BCH) formula

$$\log(e^X e^Y) = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] + [Y,[Y,X]]) + \frac{1}{24}[X,[Y,[Y,X]]] + \ldots \tag{17}$$

can be used to expand this out with $\epsilon_i^{-1} = e^X$ and $\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g = e^Y$. Or, put another way, $X = -\log\epsilon_i$ and $Y = \log(\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g)$.

After expansion, recollecting terms in the exponent of (15), and matching them with the analogous expansion of

$$-\frac{1}{2}[\mathbf{v}(\mu_{1,2,\ldots,n}^{-1}\circ g)]^T\Sigma_{1,2,\ldots,n}^{-1}\mathbf{v}(\mu_{1,2,\ldots,n}^{-1}\circ g)$$

under the assumption that

$$\mu_{1,2,\ldots,n} = \hat{\mu}_{1,2,\ldots,n}\circ\epsilon_{1,2,\ldots,n}$$

where again $\epsilon_{1,2,\ldots,n}$ is a small motion, provides a way to fuse Gaussians that are not too far away from each other and not too spread out. The results of these expansions using different approximation levels are provided in the following subsections. A more detailed derivation is provided in Appendix B.

The results that follow do not provide $\mu_{1,2,\ldots,n}$ or $\Sigma_{1,2,\ldots,n}$ (which are the values that minimize (11)), rather they provide approximations of these values. Therefore, let $\mu_{1,2,\ldots,n}^{(k)}$ and $\Sigma^{(k)}$ be the kth-order approximations of $\mu_{1,2,\ldots,n}$ and $\Sigma_{1,2,\ldots,n}$, respectively. This will allow $\epsilon_{1,2,\ldots,n}^{(k)}$ to be defined such that $\mu_{1,2,\ldots,n}^{(k)} = \hat{\mu}_{1,2,\ldots,n}\circ\epsilon_{1,2,\ldots,n}^{(k)}$.

4.1.1. First-Order Theory

By "first-order theory" we are referring to terms in the BCH expansion of

$$\sum_{i=1}^{n}[\mathbf{v}(\epsilon_i^{-1}\circ\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g)]^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i^{-1}\circ\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g) \tag{18}$$

that are at most linear in ϵ_i and at most quadratic in $(\hat{\mu}_{1,2,\ldots,n}^{-1}\circ g)$. Using this criteria provides conditions for $\epsilon_{1,2,\ldots,n}^{(1)}$ and $\Sigma_{1,2,\ldots,n}^{(1)}$ given ϵ_i's and Σ_i's. These conditions are

$$\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1}\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)}) = \sum_{i=1}^{n}\Sigma_i^{-1}\mathbf{v}(\epsilon_i) \tag{19}$$

and

$$\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1} - \left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1} \mathrm{ad}(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)})) = \sum_{i=1}^{n} \left(\Sigma_i^{-1} - \Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\right). \quad (20)$$

A review of the $\mathrm{ad}(\cdot)$ operator used here can be found in Appendix A. It is important to note that because of the quadratic nature of the equation that leads to (20) and the presumed symmetry of $\Sigma_{1,2,\ldots,n}^{(1)}$, we are only concerned with the symmetric part of the resulting matrices in (20). These constraints can then be recast as

$$\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1} \mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)}) - \sum_{i=1}^{n} \Sigma_i^{-1}\mathbf{v}(\epsilon_i) = \mathbf{0} \quad (21)$$

and

$$M_{1,2,\ldots,n}^{(1)} + \left(M_{1,2,\ldots,n}^{(1)}\right)^{T} - \sum_{i=1}^{n} \left(M_i + M_i^T\right) = \mathbf{0} \quad (22)$$

where

$$M_i = \Sigma_i^{-1} - \Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i)) \quad \text{and} \quad M_{1,2,\ldots,n}^{(1)} = \left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1} - \left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1} \mathrm{ad}(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)})).$$

Simultaneously solving these constraints analytically may be possible, however due to the nonlinear nature of the equations, numerical methods for obtaining $\epsilon_{1,2,\ldots,n}^{(1)}$ and $\Sigma_{1,2,\ldots,n}^{(1)}$ are used in the examples given in Section 5.

4.1.2. Second-Order Theory

The criteria for the retaining terms in the "second-order" BCH approximation of (18) is that they be at most quadratic in ϵ_i and at most quadratic in $\hat{\mu}_{1,2,\ldots,n}^{-1} \circ g$. This results in two constraint equations that are analogous to (21) and (22),

$$\left[\mathrm{ad}\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)})\right)\right]^{T} \left(\Sigma_{1,2,\ldots,n}^{(2)}\right)^{-1} \mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)}) - 2\left(\Sigma_{1,2,\ldots,n}^{(2)}\right)^{-1} \mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)}) = \sum_{i=1}^{n} \left(\mathrm{ad}(\mathbf{v}(\epsilon_i))^{T}\Sigma_i^{-1}\mathbf{v}(\epsilon_i) - 2\Sigma_i^{-1}\mathbf{v}(\epsilon_i)\right) \quad (23)$$

and

$$\left[\mathcal{M}_{1,2,\ldots,n}^{(2)} + \left(\mathcal{M}_{1,2,\ldots,n}^{(2)}\right)^{T}\right]_{jk} + \frac{1}{3}\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)})\right)^{T} \left(\Sigma_{1,2,\ldots,n}^{(2)}\right)^{-1} \mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)}) = \sum_{i=1}^{n} \left(\left[\mathcal{M}_i + \mathcal{M}_i^T\right]_{jk} + \frac{1}{3}\mathbf{v}(\epsilon_i)^{T}\Sigma_i^{-1}\mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_i)\right) \quad (24)$$

for $1 \leq j, k \leq n$ where

$$\mathcal{M}_i = \Sigma_i^{-1} - \Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{6}\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{4}\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))$$

and

$$\begin{aligned}\mathcal{M}^{(2)}_{1,2,\ldots,n} = \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} - \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right) \\ + \frac{1}{6}\left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)\mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right) \\ + \frac{1}{4}\left[\mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)\right]^T \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}(\mathbf{v}(\epsilon_i)).\end{aligned}$$

Here $[\,\cdot\,]_{jk}$ refers to the element in the jth row and the kth column of the matrix in the brackets.

Again, we recast (23) and (24) as

$$\begin{aligned}\left[\mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)\right]^T \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n}) - 2\left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n}) \\ - \sum_{i=1}^{n}\left(\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i) - 2\Sigma_i^{-1}\mathbf{v}(\epsilon_i)\right) = \mathbf{0}\end{aligned} \quad (25)$$

and

$$\begin{aligned}\left[\mathcal{M}^{(2)}_{1,2,\ldots,n} + \left(\mathcal{M}^{(2)}_{1,2,\ldots,n}\right)^T\right]_{jk} + \frac{1}{3}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)^T \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n}) \\ - \sum_{i=1}^{n}\left(\left[\mathcal{M}_i + \mathcal{M}_i^T\right]_{jk} + \frac{1}{3}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_i)\right) = 0.\end{aligned} \quad (26)$$

As in the first-order case, an analytical solution to these constraints may be possible. However, for the examples provided in Section 5 they are solved numerically.

5. Numerical Approximations for Fusion on $SO(3)$

Given μ_1, μ_2, Σ_1, and Σ_2, the constraints for obtaining first-order and second-order approximations of $\mu_{1,2}$ and $\Sigma_{1,2}$ have been established in Section 4. However, the effectiveness of these approximations is still uncertain. One way to quantify their effectiveness is to look at C in (11).

The integral in (11) is not easily computed analytically. Nevertheless, we can numerically evaluate the costs using a discretized version. For C the discretized version is given by

$$C' = \sum_{i=1}^{N}\left|\frac{f'(e^{X_i};\mu^{(k)}_{1,2},\Sigma^{(k)}_{1,2})}{\zeta_1} - \frac{f'(e^{X_i};\mu_1,\Sigma_1)f'(e^{X_i};\mu_2,\Sigma_2)}{\zeta_2}\right| \, |J\left((X_i)^\vee\right)| \, \Delta(X_i)^\vee \quad (27)$$

such that

$$\zeta_1 = \sum_{i=1}^{N} f'(e^{Y_i}; \mu_{1,2}, \Sigma_{1,2}) \left| J\left((Y_i)^\vee\right) \right| \Delta(Y_i)^\vee$$

and

$$\zeta_2 = \sum_{i=1}^{N} f'(e^{Y_i}; \mu_1, \Sigma_1) f'(e^{Y_i}; \mu_2, \Sigma_2) \left| J\left((Y_i)^\vee\right) \right| \Delta(Y_i)^\vee,$$

where $|J\left((X_i)^\vee\right)|$ is the determinant of the Jacobian relating the group element to its associated exponential coordinates, $\Delta X_i^\vee$ is the volume of the voxel at $(X_i)^\vee$, and N is the number of voxels considered. If (27) is evaluated on a regularly spaced Cartesian grid in exponential coordinates, then $\Delta(X_i)^\vee$ is a constant. Note that $f'(\cdot)$ is not $f(\cdot)$ as defined in (14); rather, $f'(\cdot) = f(\cdot)/\alpha$.

We can explore the values over which the first-order and second-order constraints are valid by using them to determine $\mu_{1,2}^{(k)}$ and $\Sigma_{1,2}^{(k)}$ given various μ_i's and Σ_i's. Two examples are given below for $SO(3)$. In these examples, the first-order constraint equations (21) and (22) were simultaneously solved by minimizing the sum of the squares of all of the elements on the left-hand side of these equations. If this minimization reaches a value of zero, we consider the constraints to have been solved.

Using the $\hat{\mu}_{1,2,\ldots,n}$ from (16) and

$$\hat{\Sigma}_{1,2,\ldots,n} = \left(\sum_{i=1}^{n} \Sigma_i^{-1} \right)^{-1} \tag{28}$$

as initial values for the minimization procedure helps to ensure that the minimization reaches zero. A similar procedure was used to solve (25) and (26) for the second-order approximations.

For the first example

$$\mu_1 = \exp\left\{ \frac{\gamma}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}^\wedge \right\} \quad \text{and} \quad \mu_2 = \exp\left\{ \frac{\gamma}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}^\wedge \right\}$$

$$\Sigma_1 = \xi \cdot R_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} R_1^T \quad \text{and} \quad \Sigma_2 = \xi \cdot R_2 \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} R_2^T$$

where R_1 and R_2 are arbitrary rotation matrices (i.e., $R_1, R_2 \in SO(3)$). Two scale factors γ and ξ are used vary the μ_i's and Σ_i's, respectively. γ is used to "separate" the two means; ξ is used to "spread out" the distributions. Figures 1, 2, and 3 show the value of C' for a range of values of γ and ξ for the first example at different orders of approximation. Figure 1 presents C' using a so called "zeroth-order" approximation where $\mu_{1,2,\ldots,n}^{(0)} = \hat{\mu}_{1,2,\ldots,n}$ and $\Sigma_{1,2,\ldots,n}^{(0)} = \hat{\Sigma}_{1,2,\ldots,n}$ from (16) and (28). Figures 2 and 3 demonstrate C' for the

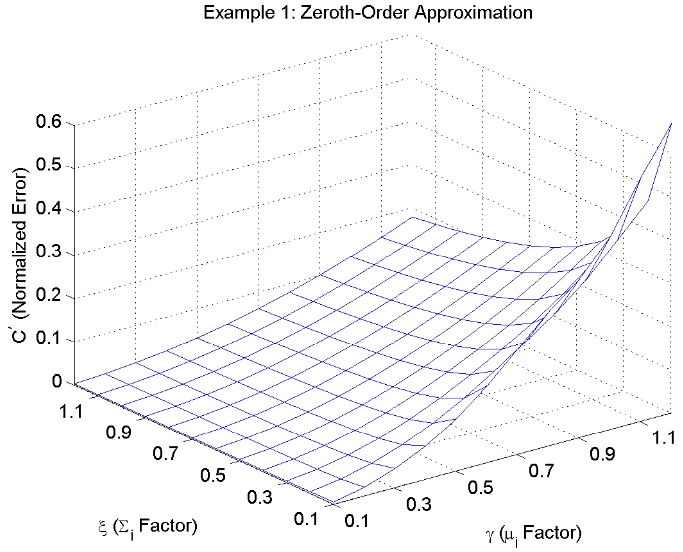

Figure 1: Normalized error, C', plotted versus γ and ξ for the first example using a zeroth-order approximation.

first-order and second-order approximations, respectively. C' represents a normalized error; a percent error can be obtained by multiplying C' by 100.

For the second example

$$\mu_1 = \exp\left\{\frac{\gamma}{\sqrt{2}}\begin{pmatrix} 1 \\ -\frac{1}{2} \\ \frac{\sqrt{3}}{2}\end{pmatrix}^{\wedge}\right\} \quad \text{and} \quad \mu_2 = \exp\left\{\frac{\gamma}{\sqrt{2}}\begin{pmatrix} -1 \\ \frac{1}{2} \\ -\frac{\sqrt{3}}{2}\end{pmatrix}^{\wedge}\right\}$$

$$\Sigma_1 = \xi \cdot R_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0.5 \end{bmatrix} R_1^T \quad \text{and} \quad \Sigma_2 = \xi \cdot R_2 \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} R_2^T$$

where the R_i's are not those used in the first example. Using these μ_i's and Σ_i's, values of C' are given for various values of γ and ξ in Figures 4, 5, and 6.

6. Conclusions

A method for approximating Bayesian fusion on connected unimodular matrix Lie groups has been presented for the case when the means are clustered sufficiently closely and covariances are sufficiently small. This work relies on the Baker-Campbell-Hausdorff expansion of the product of exponentials of Lie algebra elements. Conditions for both first-order and second-order approximations were developed. As expected, of the three approximations used, the second-order approximations resulted in the lowest error over the largest range of scale factors for both of the examples explored in Section 5. It is also important to note that both the first and second-order approximations result in lower

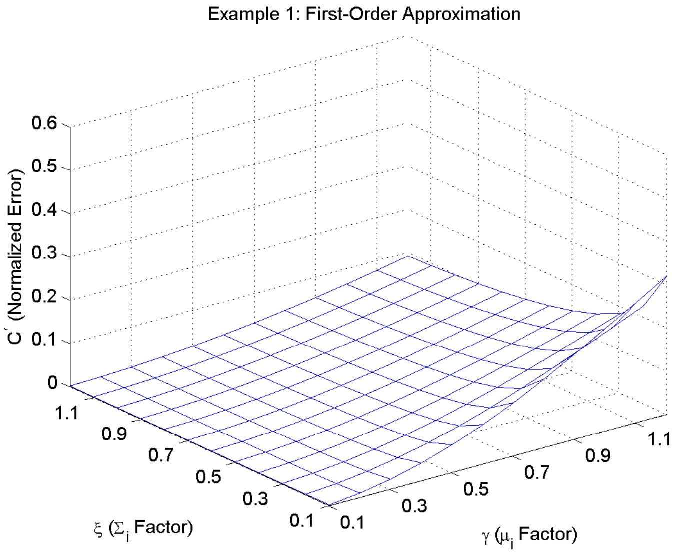

Figure 2: Normalized error, C', plotted versus γ and ξ for the first example using a first-order approximation.

error than the zeroth-order approximation which was based on the product of Gaussians taken on $\mathbb{R}^d$.

The means of the two numerical examples used were chosen in an attempt to characterize very different scenarios. In the first example, the vectors used to define the two means were taken so that they were perpendicular or $(\mathbf{v}(\mu_1))^T \mathbf{v}(\mu_2) = 0$. For the second example, the vectors used to define the two means were taken so they had opposite sense or $\mathbf{v}(\mu_1) = -\mathbf{v}(\mu_1)$. The fact that these approximations perform well in each of these two cases provides a reasonable expectation that they will perform well over all of $SO(3)$.

While the numerical examples presented focused on $SO(3)$, the results in Section 4 generalized to any connected unimodular Lie group. In particular, these methods could easily be used with other motion groups such as $SE(2)$ and $SE(3)$.

Acknowledgements

This research was partially supported under an appointment to the Department of Homeland Security (DHS) Scholarship and Fellowship Program, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-AC05-06OR23100.

This work was also supported in part by NSF grant IIS-0915542 RI: Small: Robotic Inspection, Diagnosis, and Repair.

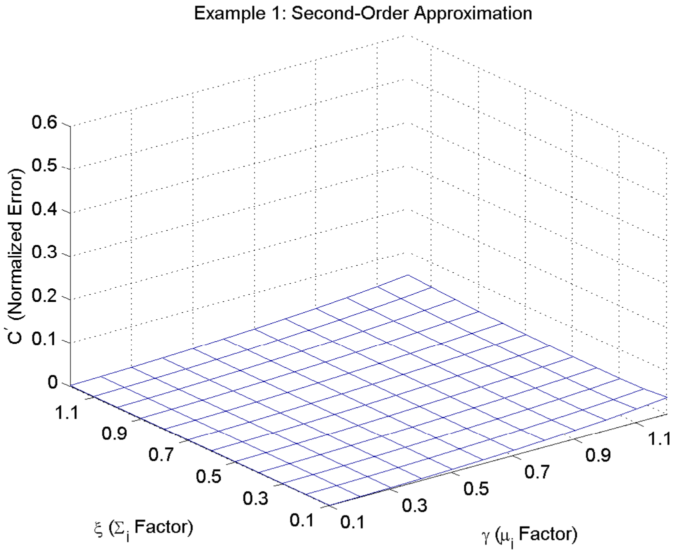

Figure 3: Normalized error, C', plotted versus γ and ξ for the first example using a second-order approximation.

A. Appendix: The Lie Bracket and Adjoint Matrix, $\mathrm{ad}(X)$

For two elements of a Lie algebra, $\mathcal{G}$, the quantity

$$[X, Y] = XY - YX \tag{29}$$

is known as the *Lie bracket* of $X, Y \in \mathcal{G}$. Based on its definition in (29), it is clear that the Lie bracket is linear in both arguments:

$$[aX_1 + bX_2, Y] = a[X_1, Y] + b[X_2, Y] \quad \text{and} \quad [X, aY_1 + bY_2] = a[X, Y_1] + b[X, Y_2].$$

It is also easily verified that the Lie bracket is antisymmetric:

$$[X, Y] = -[Y, X].$$

Based on the definition of the "vee" operator discussed in Section 3.1, we can define an adjoint function $\mathrm{ad}(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ where d is the dimension of the Lie algebra so that[‡]

$$[X, Y]^\vee \doteq \mathrm{ad}(X^\vee)Y^\vee. \tag{30}$$

From the definitions given in (10) and (30) it follows that

$$[X, Y]^\vee = \mathrm{ad}(\mathbf{v}(x))\mathbf{v}(y)$$

where $x = e^X$ and $y = e^Y$. The use of $\vee$ and $\mathrm{ad}(\cdot)$ for different Lie algebras should not be a source of confusion as their meaning can be obtained through their arguments and

[‡]Note that often the adjoint is defined such that its arguments are elements of a Lie algebra (i.e., X as opposed to $X^\vee$) however the definition given in (30) is used here to simplify expressions.

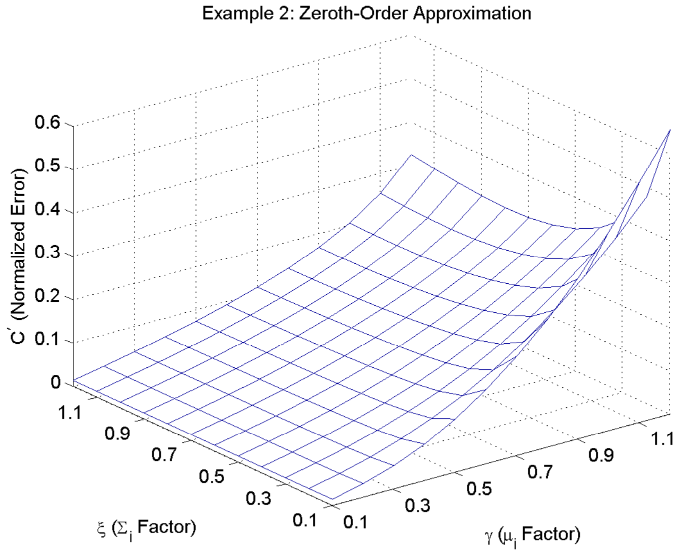

Figure 4: Normalized error, C', plotted versus γ and ξ for the second example using a zeroth-order approximation.

usage. This allows the Baker-Campbell-Hausdorff formula given in (17) to be rewritten as

$$\begin{aligned}\left(\log(e^X e^Y)\right)^\vee =& \mathbf{v}(x) + \mathbf{v}(y) + \frac{1}{2}\mathrm{ad}(\mathbf{v}(x))\mathbf{v}(y) + \frac{1}{12}\big(\mathrm{ad}(\mathbf{v}(x))\mathrm{ad}(\mathbf{v}(x))\mathbf{v}(y) \\ &+ \mathrm{ad}(\mathbf{v}(y))\mathrm{ad}(\mathbf{v}(y))\mathbf{v}(x)\big) + \frac{1}{24}\mathrm{ad}(\mathbf{v}(x))\mathrm{ad}(\mathbf{v}(y))\mathrm{ad}(\mathbf{v}(y))\mathbf{v}(x) + \ldots \quad (31)\end{aligned}$$

We note that due to the linearity of the Lie bracket and the $\mathrm{ad}(\cdot)$ operator, one is able to write

$$\mathrm{ad}(\mathbf{v}(x)) = \sum_{i=1}^{n}[\mathbf{v}(x)]_i \,\mathrm{ad}\left((E_i)^\vee\right),$$

where E_i is the ith basis element of the associated Lie algebra and $[\mathbf{v}(x)]_i$ is the ith entry in the vector $\mathbf{v}(x)$. It is often convenient to distinguish the basis elements of different Lie algebras from one another. Therefore, let $\{\bar{E}_i\}$, $\{\hat{E}_i\}$, and $\{\acute{E}_i\}$ represent the basis elements of $so(3)$, $se(3)$, and $se(2)$, respectively.

A.1. The Adjoint Matrix for $so(3)$

The Lie algebra, $so(3)$, consists of skew-symmetric matrices of the form:

$$\Omega = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} = \sum_{i=1}^{3}\omega_i \bar{E}_i$$

where $\omega_i = [\Omega^\vee]_i$.

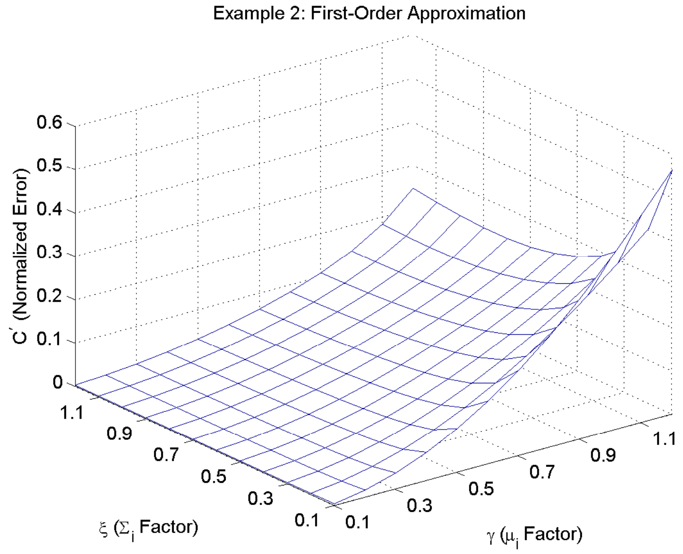

Figure 5: Normalized error, C', plotted versus γ and ξ for the second example using a first-order approximation.

If $X, Y \in so(3)$, then

$$[X,Y]^\vee = X^\vee \times Y^\vee$$

where $\times : \mathbb{R}^3 \to \mathbb{R}^3$ is the traditional cross product. This leads to the fact that

$$\mathrm{ad}(X^\vee) = X.$$

A.2. The Adjoint Matrix for $se(3)$

The Lie algebra, $se(3)$, consists of "screw" matrices of the form

$$X = \begin{bmatrix} 0 & -x_3 & x_2 & x_4 \\ x_3 & 0 & -x_1 & x_5 \\ -x_2 & x_1 & 0 & x_6 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \sum_{i=1}^{6} x_i \hat{E}_i$$

where $X^\vee = [x_1, x_2, x_3, x_4, x_5, x_6]^T$.

The adjoint matrix for $se(3)$ is given by

$$\mathrm{ad}(X^\vee) = \begin{bmatrix} 0 & -x_3 & x_2 & 0 & 0 & 0 \\ x_3 & 0 & -x_1 & 0 & 0 & 0 \\ -x_2 & x_1 & 0 & 0 & 0 & 0 \\ 0 & -x_6 & x_5 & 0 & -x_3 & x_2 \\ x_6 & 0 & -x_4 & x_3 & 0 & -x_1 \\ -x_5 & x_4 & 0 & -x_2 & x_1 & 0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{3}(x_i \bar{E}_i) & \mathbf{0} \\ \sum_{i=4}^{6}(x_i \bar{E}_i) & \sum_{i=1}^{3}(x_i \bar{E}_i) \end{bmatrix}.$$

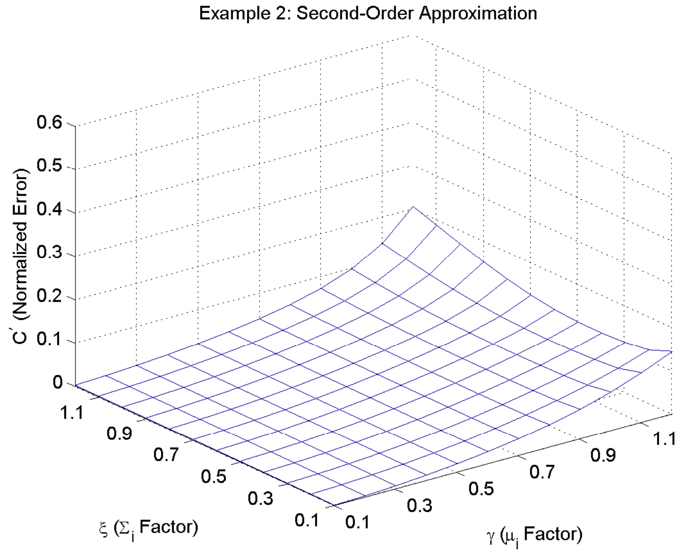

Figure 6: Normalized error, C', plotted versus γ and ξ for the second example using a second-order approximation.

A.3. The Adjoint Matrix for $se(2)$

Matrices of the form

$$X = \begin{bmatrix} 0 & -x_1 & x_2 \\ x_1 & 0 & x_3 \\ 0 & 0 & 0 \end{bmatrix} = \sum_{i=1}^{3} x_i \acute{E}_i$$

comprise $se(2)$ where $X^\vee = [x_1, x_2, x_3]^T$. From this, it is easily verified that the adjoint matrix for $se(2)$ is given by

$$\mathrm{ad}(X^\vee) = \begin{bmatrix} 0 & 0 & 0 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}.$$

B. Appendix: Derivation of the First-Order and Second-Order Constraints

Consider the definition of a Gaussian distribution given in (14). This combined with the version of the Baker-Campbell-Hausdorff given in (31) can be used to generate the first-order constraints (19) and (20) and the second-order constraints (23) and (24). This is done by allowing $X = -\log \epsilon_i$ and $Y = \log(\hat{\mu}^{-1}_{1,2,\ldots,n} \circ g)$. To simplify some of the derivation below, we will substitute h for $(\hat{\mu}^{-1}_{1,2,\ldots,n} \circ g)$; therefore $Y = \log(h)$.

B.1. First-Order Constraints

Since we are only concerned with terms that are at most linear in ϵ_i and at most quadratic in h we can remove a number of higher order terms in the BCH formula. This leaves

$$\Big(\log\big(\exp[-\log(\epsilon_i)]\exp[\log(h)]\big)\Big)^\vee \approx -\mathbf{v}(\epsilon_i) + \mathbf{v}(h) - \frac{1}{2}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h). \tag{32}$$

The terms of (18) can then be approximated as:

$$\begin{aligned}
[\mathbf{v}(\epsilon_i^{-1}\circ h)]^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i^{-1}\circ h) \approx & \left(-\mathbf{v}(\epsilon_i)^T + \mathbf{v}(h)^T - \frac{1}{2}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\right)\Sigma_i^{-1} \\
& \left(-\mathbf{v}(\epsilon_i) + \mathbf{v}(h) - \frac{1}{2}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\right) \\
=& \mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i) - 2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(h) + \mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) \\
& + \mathbf{v}(h)^T\Sigma_i^{-1}\mathbf{v}(h) - \mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) \\
& + \frac{1}{4}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)
\end{aligned} \tag{33}$$

Removing the higher order terms of (33) leaves

$$\begin{aligned}
[\mathbf{v}(\epsilon_i^{-1}\circ h)]^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i^{-1}\circ h) \approx & -2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(h) + \mathbf{v}(h)^T\Sigma_i^{-1}\mathbf{v}(h) \\
& - \mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)
\end{aligned}$$

If one now considers (15) it should be clear that

$$\begin{aligned}
& \mathbf{v}(h)^T\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1}\mathbf{v}(h) - \mathbf{v}(h)^T\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1}\mathrm{ad}\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)})\right)\mathbf{v}(h) \\
& \quad - 2\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(1)})\right)^T\left(\Sigma_{1,2,\ldots,n}^{(1)}\right)^{-1}\mathbf{v}(h) = \\
& \qquad \sum_{i=1}^{n}\Big(-2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(h) + \mathbf{v}(h)^T\Sigma_i^{-1}\mathbf{v}(h) - \mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\Big)
\end{aligned} \tag{34}$$

In (34), equating the terms linear in h gives rise to (19). Similarly, equating terms quadratic in h yields (20).

B.2. Second-Order Constraints

Analogous to the first-order derivation, terms of cubic order and higher in either ϵ_i or h can be disregarded in the BCH formula when establishing constraints for the second-order theory. This allows one to write

$$\begin{aligned}
\Big(\log\big(\exp[-\log(\epsilon_i)]\exp[\log(h)]\big)\Big)^\vee \approx & -\mathbf{v}(\epsilon_i) + \mathbf{v}(h) - \frac{1}{2}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) \\
& + \frac{1}{12}\big(\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) - \mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i)\big).
\end{aligned} \tag{35}$$

The expansion of (18) is then taken as

$$[\mathbf{v}(\epsilon_i^{-1}\circ h)]^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i^{-1}\circ h) \approx \mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i) - 2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(h) + \mathbf{v}(\epsilon_i)\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)$$

$$
\begin{aligned}
&- \frac{1}{6}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) + \frac{1}{6}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i)\\
&+ \mathbf{v}(h)^T\Sigma_i^{-1}\mathbf{v}(h) - \mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&+ \frac{1}{6}\mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) - \frac{1}{6}\mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i)\\
&+ \frac{1}{4}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&- \frac{1}{12}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&+ \frac{1}{12}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i)\\
&+ \frac{1}{144}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&- \frac{1}{144}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i)\\
&+ \frac{1}{144}\mathbf{v}(\epsilon_i)^T\mathrm{ad}(\mathbf{v}(h))^T\mathrm{ad}(\mathbf{v}(h))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i). \qquad (36)
\end{aligned}
$$

If the higher order terms are removed from (36), one is left with

$$
\begin{aligned}
[\mathbf{v}(\epsilon_i^{-1}\circ h)]^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i^{-1}\circ h) \approx &\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i) - 2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(h) + \mathbf{v}(\epsilon_i)\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&+ \frac{1}{6}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(h))\mathrm{ad}(\mathbf{v}(h))\mathbf{v}(\epsilon_i) + \mathbf{v}(h)^T\Sigma_i^{-1}\mathbf{v}(h)\\
&- \mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h) + \frac{1}{6}\mathbf{v}(h)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h)\\
&+ \frac{1}{4}\mathbf{v}(h)^T\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathbf{v}(h). \qquad (37)
\end{aligned}
$$

For fixed values of ϵ_i and Σ_i, $\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathbf{v}(\epsilon_i)$ is a constant and does not need to be included in the constraints as the resulting Gaussian can be normalized after $\epsilon_{1,2,\ldots,n}^{(2)}$ and $\Sigma_{1,2,\ldots,n}^{(2)}$ are determined. Now equating linear terms in h for (37) yields

$$
\begin{aligned}
&\left(-2\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)})\right)^T\left(\Sigma_{1,2,\ldots,n}^{(2)}\right)^{-1} + \mathbf{v}(\epsilon_{1,2,\ldots,n})\left(\Sigma_{1,2,\ldots,n}^{(2)}\right)^{-1}\mathrm{ad}\left(\mathbf{v}(\epsilon_{1,2,\ldots,n}^{(2)})\right)\right)\mathbf{v}(h) =\\
&\qquad \sum_{i=1}^{n}\left(-2\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1} + \mathbf{v}(\epsilon_i)\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\right)\mathbf{v}(h). \qquad (38)
\end{aligned}
$$

The constraint in (23) is then obtained by using (38) with the understanding that (38) must hold for all $h \in G$.

Now consider the terms of (37) that are quadratic in h. Using the linearity of the adjoint operator and letting $h_j = [\mathbf{v}(h)]_j$, it is easily verified that these terms can be expressed as

$$
\frac{1}{6}\sum_{j=1}^{d}\sum_{k=1}^{d}h_jh_k\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_i)+
$$

$$\mathbf{v}(h)^T \left(\Sigma_i^{-1} - \Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{6}\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{4}\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i)) \right) \mathbf{v}(h)$$

where d is the dimension of the Lie algebra. Now let

$$\mathcal{M}_i = \Sigma_i^{-1} - \Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{6}\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))\mathrm{ad}(\mathbf{v}(\epsilon_i)) + \frac{1}{4}\mathrm{ad}(\mathbf{v}(\epsilon_i))^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{v}(\epsilon_i))$$

and

$$\begin{aligned}\mathcal{M}^{(2)}_{1,2,\ldots,n} = \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} - \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right) \\ + \frac{1}{6}\left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right) \mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right) \\ + \frac{1}{4}\left[\mathrm{ad}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)\right]^T \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}(\mathbf{v}(\epsilon_i)).\end{aligned}$$

These terms quadratic in h can then be expressed as

$$\sum_{j=1}^{d}\sum_{k=1}^{d} h_j h_k \Big(\frac{1}{6}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_i) + [\mathcal{M}_i]_{jk} \Big)$$

Then considering (15) with the approximation given by (37) it should be apparent that we can equate the symmetric portion of the quadratic terms in h such that:

$$\begin{aligned}\frac{1}{6}\left(\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n})\right)^T \left(\Sigma^{(2)}_{1,2,\ldots,n}\right)^{-1} \mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon^{(2)}_{1,2,\ldots,n}) + \frac{1}{2}\left[\mathcal{M}^{(2)}_{1,2,\ldots,n} + \left(\mathcal{M}^{(2)}_{1,2,\ldots,n}\right)^T\right]_{jk} = \\ \sum_{i=1}^{n}\left(\frac{1}{6}\mathbf{v}(\epsilon_i)^T\Sigma_i^{-1}\mathrm{ad}(\mathbf{e}_j)\mathrm{ad}(\mathbf{e}_k)\mathbf{v}(\epsilon_i) + \frac{1}{2}\left[\mathcal{M}_i + \mathcal{M}_i^T\right]_{jk}\right)\end{aligned}$$

for $1 \leq j, k \leq d$. This is equivalent to (24).

C. Appendix: Nomenclature

$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbf{x}$	a vector in $\mathbb{R}^d$
$\mathbf{e}_i$	the ith natural unit basis vector for $\mathbb{R}^d$
$\lvert\cdot\rvert$	the determinant if the argument is a matrix or the magnitude if the argument is a scalar
$\lVert\cdot\rVert$	the Euclidean norm if the argument is a vector or the Frobenius norm if the argument is a matrix
G	a connected unimodular Lie group
$g \in G$	a generic element of G
$\mathcal{G}$	the Lie algebra corresponding to G

$X \in \mathcal{G}$	a generic element of $\mathcal{G}$
d	the dimension of G and $\mathcal{G}$
E_i	the ith basis element of the Lie algebra $\mathcal{G}$
$(\cdot)^\vee$	a linear function $\vee : \mathcal{G} \to \mathbb{R}^d$ such that $(E_i)^\vee = \mathbf{e}_i$
$(\cdot)^\wedge$	a linear function $(\cdot)^\wedge : \mathbb{R}^d \to \mathcal{G}$ such that $(\mathbf{e}_i)^\wedge = E_i$
$f(\cdot)$	a probability density function (on $\mathbb{R}^d$ or G)
$J(\cdot)$	the Jacobian relating exponential coordinates to the associated group element, $J(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$
$\boldsymbol{\mu} \in \mathbb{R}^d$	the mean of $f(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$
$\tilde{\mu} \in G$	the mean of $f(\cdot)$ given by (12)
$\tilde{\Sigma} \in \mathbb{R}^{d\times d}$	the covariance of $f(\cdot)$ given by (13)
$\mu_i \in G$	the first set of parameters that define $f_i(g) = f(g; \mu_i, \Sigma_i)$
$\Sigma_i \in \mathbb{R}^{d\times d}$	the second set of parameters that define $f_i(g) = f(g; \mu_i, \Sigma_i)$
$\mu_{1,2,\ldots,n} \in G$	the first set of parameters in $f(\cdot; \mu_{1,2,\ldots,n}, \Sigma_{1,2,\ldots,n})$ to best match with $\prod_{i=1}^n f(\cdot; \mu_i, \Sigma_i)$ in the sense of (11)
$\Sigma_{1,2,\ldots,n} \in \mathbb{R}^{d\times d}$	the second set of parameters in $f(\cdot; \mu_{1,2,\ldots,n}, \Sigma_{1,2,\ldots,n})$ to best match with $\prod_{i=1}^n f(\cdot; \mu_i, \Sigma_i)$ in the sense of (11)
$\mu^{(k)}_{1,2,\ldots,n} \in G$	the kth-order approximation of $\mu_{1,2,\ldots,n}$
$\Sigma^{(k)}_{1,2,\ldots,n} \in \mathbb{R}^{d\times d}$	the kth-order approximation of $\Sigma_{1,2,\ldots,n}$
$\hat{\mu}_{1,2,\ldots,n} \in G$	the initial estimate of $\mu_{1,2,\ldots,n}$ given by (16)
$\hat{\Sigma}_{1,2,\ldots,n} \in \mathbb{R}^{d\times d}$	the initial estimate of $\Sigma_{1,2,\ldots,n}$ given by (28)
$\epsilon_i \in G$	defined such that $\mu_i = \hat{\mu}_{1,2,\ldots,n} \circ \epsilon_i$
$\epsilon_{1,2,\ldots,n} \in G$	defined such that $\mu_{1,2,\ldots,n} = \hat{\mu}_{1,2,\ldots,n} \circ \epsilon_{1,2,\ldots,n}$
$\epsilon^{(k)}_{1,2,\ldots,n} \in G$	defined such that $\mu^{(k)}_{1,2,\ldots,n} = \hat{\mu}_{1,2,\ldots,n} \circ \epsilon_{1,2,\ldots,n}$

References

[1] G. S. Chirikjian. Information-theoretic inequalities on unimodular lie groups. *Journal of Geometric Mechanics*, 2(2):119–158, June 2010.

[2] G. S. Chirikjian. Modeling loop entropy. *Methods in Enzymology*, 487:99–132, 2011.

[3] G. S. Chirikjian and A. B. Kyatkin. *Engineering Applications of Noncommutative Harmonic Analysis*. CRC Press, Boca Raton, FL, 2000.

[4] P. Diaconis. *Group Representations in Probability and Statistics*. Lecture Notes-Mongraph Series. Institute of Mathamatical Statistics, Hayward, CA, 1988.

[5] H. D. Fegan. The heat equation on a compact lie group. *Transactions of the American Mathematical Society*, 246:339–357, 1978.

[6] H. D. Fegan. The fundamental solution of the heat equation on a compact lie group. *Journal Differential Geometry*, 18:659–668, 1983.

[7] S. Fiori and T. Tanaka. An algorithm to compute averages on matrix lie groups. *IEEE Transactions on Signal Processing*, 57(12):4734–4743, 2010.

[8] C. D. Gorman. Brownian motion of rotation. *Transactions of the American Mathematical Society*, 94:103–117, 1960.

[9] U. Grenander. *Probabilities on Algebraic Structures*. 1963 (reprinted by Dover, 2008).

[10] H. Heyer. *Probability Measures on Locally Compact Groups*. Springer-Verlag, New York, 1977.

[11] J. Kwon, M. Choi, F. C. Park, and C. Chun. Particle filtering on the euclidean group: Framework and applications. *Robotica*, 25:725–737, 2007.

[12] J. T.-H. Lo and L. R. Eshleman. Exponential fourier densities on s^2 and optimal estimation and detection for directional processes. *IEEE Transactions on Information Theory*, IT-23:321–336, May 1977.

[13] J. T.-H. Lo and L.R. Eshleman. Exponential fourier densities and optimal estimation for axial processes. *IEEE Transactions on Information Theory*, IT-25(4):463–470, May 1979.

[14] J. T.-H. Lo and L.R. Eshleman. Exponential fourier densities on $so(3)$ and optimal estimation and detection for rotational processes. *SIAM Journal of Applied Mathematics*, 36(1):73–82, February 1979.

[15] V. M. Maksimov. Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicompact groups. *Theory Of Probability And Its Applications*, 12(2):267–280, 1967.

[16] J. McConnell. *Rotational Brownian Motion and Dielectric Theory*. Academic Press, New York, 1980.

[17] H. P. McKean, Jr. Brownian motions on the 3-dimensional rotation group. *Memoirs of the College of Science, University of Kyoto, Series A*, 33(1):2538, 1960.

[18] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. S. Chirikjian. Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map. *Robotica*, 26(4):419–434, 2008.

[19] W. Park, Y. Wang, and G. S. Chirikjian. The path-of-probability algorithm for steering and feedback control of flexible needles. *International Journal of Robotics Research*, 29(7):813–830, 2010.

[20] P. F. Perrin. Étude mathématique du mouvement brownien de rotation. *Annales Scientifiques de L' École Normale Supérieure*, 45:1–51, 1928.

[21] K. K. Roy. Exponential families of densities on an analytic group and sufficient statistics. *Sankhya: The Indian Journal of Statistics*, 37(1):82–92, January 1975.

[22] A. Skliros, W. Park, and G. S. Chirikjian. Position and orientation distributions for non-reversal random walks using space-group fourier transforms. *Journal of Algebraic Statistics*, 1(1):27–46, 2010.

[23] R. C. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research*, 5(4):56–68, 1986.

[24] S. Su and C. S.G. Lee. Manipulation and propagation of uncertainty and verification of applicability of actions in assembly tasks. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1376–1389, 1992.

[25] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, Cambridge, MA, 2005.

[26] Y. Wang and G. S. Chirikjian. Error propagation on the euclidean group with applications to manipulator kinematics. *IEEE Transactions on Robotics*, 22:591–602, August 2006.

[27] Y. Wang and G. S. Chirikjian. Nonparametric second-order theory of error propagation on the euclidean group. *International Journal of Robotics Research*, 27(11–12):1258–1273, 2008.

[28] G. Wyllie. Random motion and brownian rotation. *Physics Reports*, 61(6):327–376, 1980.