53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Gaussian Approximation of Non-linear Measurement Models on Lie Groups.

Gregory Chirikjian[1] and Marin Kobilarov[1]

Abstract— **Extended Kalman filters on Lie groups arise naturally in the context of pose estimation and more generally in robot localization and mapping. Typically in such settings one deals with nonlinear measurement models that are handled through linearization and linearized uncertainty transformation. To circumvent the loss of accuracy resulting from the typical coordinate-based linearization, this paper develops a method for accurately describing the probability density associated with nonlinear measurement models by a second-order approximation of a distribution defined directly on the Lie group configuration space. We show that, like the case of linearized measurement models, this density can be described well as a Gaussian distribution in exponential coordinates (though with different mean and covariance than those that result from linearized measurement models). And therefore previously developed methods for propagation of uncertainty and fusion of measurements can be applied to this generalized formulation without the a priori assumption of linearized measurement. A case study using a range-bearing model in planar robot localization is presented to demonstrate the method.**

I. Introduction

Estimation/filtering algorithms have three fundamental parts: (1) uncertainty propagation; (2) formulation of measurement models; and (3) updating (or fusion) of prior state estimates (or their corresponding distributions) with current measurements. The Kalman filter implicitly makes extensive use of the properties of Gaussian distributions such as closure under convolution (for propagation) and closure under conditioning and marginalization (for fusion). The extended Kalman filter (EKF) is often used for nonlinear systems and measurement models through local linearization. A number of recent works have developed extended Kalman filters for Lie groups in which the intricacies of (2) are overlooked. These works typically use the properties of exponential coordinates, as do we. The difference is that usually in existing formulations a linear (or linearized) measurement model is used from the outset. While this assumption can be justified for particular measurement modalities (such as magnetometer data in spacecraft attitude estimation), in many practical applications, measurement models can be far from linear. Therefore, we develop a measurement update step for Lie groups that captures higher order terms in the measurement model.

Exponential coordinates for SE(2) and SE(3) are used extensively in the context of robotics problems, as described in [4], [5], [9], [11], [10]. The idea of modeling measurement distributions on Lie groups is by no means new (see e.g., [1], [2]), and a variety of different models for the resulting probability densities continue to be proposed [3]. The recent literature is particularly rich with methods for filtering on Lie groups that use exponential coordinates. In attitude estimation the group of rotations, SO(3), is of interest [12], [13], [29] and works on filtering in this context include [14], [16], [23], [24], [25], [26], [28]. And even outside of the context of attitude estimation, the group SO(3) arises in other applications [15]. In mobile robot localization [30], the groups of rigid-body motions of the plane and of 3D space, SE(2) and SE(3), are of interest [8], [7], [18], [21]. Other recent works on Lie-group filtering in a more abstract settings and for other Lie groups include [27], [17]. Often in the process of fusion, an optimization step is required for which the methods developed in [22] are useful. We note that generic nonlinear filtering methods such as those reviewed in [19], [20] can also be applied to the Lie group setting, but the emphasis in this paper is to determine how far the Lie group formalism can be taken as an alternative.

This paper focuses on deriving accurate measurement models from first principles (rather than assuming a priori a linearized model) and showing that even in the nonlinear setting, Gaussian distributions in exponential coordinates constitute a good model. The remainder of this paper is organized as follows. Section II introduces measurements models and their approximations as Lie-group Gaussians. Section III defines Lie-group Gaussians and formulates their optimal estimation. Section IV then provides a motivating example as to why linear measurement models cannot be assumed from the outset when doing filtering. Section V then generalizes this to the context of nonlinear measurement models on Lie groups, and develops a methodology for matching the best-fit Lie-group Gaussian distribution to capture these nonlinear measurement models.

II. Measurement Models

Consider an autonomous vehicle such as a mobile robot, satellite, underwater or aerial vehicle operating in a workspace $\mathcal{W}$, where $\mathcal{W}\subset\mathbb{R}^2$ when the vehicle is moving a plane, or $\mathcal{W}\subset\mathbb{R}^3$ in the 3-D case. The vehicle is described by its pose $g\in G$, where $G=SE(2)$ in the planar case or $G=SE(3)$ in the 3-D case. The pose has the general form

$$g=\left[\begin{array}{cc} R & x \\ 0 & 1 \end{array}\right],$$

where R is the rotation matrix and $x\in\mathcal{W}$ is the position. The vehicle obtains noisy sensor measurements by observing

[1]Gregory Chirikjian and [1]Marin Kobilarov are with Faculty of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA gregc|marin@jhu.edu

point features relative to its body-fixed frame of reference. Such measurements, denoted by the random variable z and taking values in $\mathbb{R}^m$, are typically defined using a nonlinear function $\bar{h}:\mathcal{W}\to\mathbb{R}^m$ according to

$$z=\bar{h}(g^{-1}\cdot\ell)+H(g)n,$$

where n is a noise vector and $H(g)$ is a coupling matrix. The notation $g^{-1}\cdot\ell$ should be understood as a left action of G on $\mathcal{W}$, or simply $g^{-1}\cdot\ell=R^T(\ell-x)$. We will assume that $H(g)$ is square and invertible, and hence n also takes values in $\mathbb{R}^m$. The sensor model can then be expressed in a standard form

$$z=h(g)+H(g)n.$$

through the function $h(g)=\bar{h}(g^{-1}\cdot\ell)$. If the probability density function of the noise n is $q_n:\mathbb{R}^m\to\mathbb{R}_{\geqslant 0}$, then the measurement probability density for given $g\in G$ is defined by

$$\rho(z\,|\,g)\;=\;\frac{1}{|H(g)|}\,q_n\left([H(g)]^{-1}[z-h(g)]\right). \qquad (1)$$

This equation is completely general and makes no assumptions about the nature of the noise.

In many practical problems, the noise is Gaussian and zero mean, and $H(g)$ does not depend on g. In this case, H can be absorbed into the definition of q_n by taking the covariance as $N=HH^T$, and setting $H=I$ in (1). Hence, for our purposes,

$$\rho(z|g)=\frac{1}{(2\pi)^{\frac{m}{2}}|N|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}[z-h(g)]^TN^{-1}[z-h(g)]\right). \quad (2)$$

This is a probability density function (pdf) in z, but is not a pdf in g. However, we can generate a pdf by normalizing:

$$\rho_n(g)\equiv\rho(g|z)\;\doteq\;\frac{\rho(z\,|\,g)}{\int_G\rho(z\,|\,g)dg},$$

where the dependence on z is suppressed in the notation $\rho_n(g)$. Note, that $\rho_n(g)$ is well defined under the assumption that $h(g)$ is non-degenrate. In addition, if g has a known nominal distribution $\rho_0(g)$ (i.e. a prior on G) then we have

$$\rho(g|z)\;\doteq\;\frac{\rho(z\,|\,g)\rho_0(g)}{\int_G\rho(z\,|\,g)\rho_0(g)dg}. \qquad (3)$$

Our main focus is to obtain the optimal parametric (i.e. Gaussian) approximation of $\rho(g|z)$ and to determine whether a Lie group representation has advantages over the traditional coordinate-based parametric forms. The expression in (2) is Gaussian in z, but it is clearly not Gaussian in g. Indeed, the concept of a Gaussian on a Lie group has not even been defined at this point. We next provide this definition to approximate (2) by a Lie-group Gaussian.

Example Models.: We are concerned with measurement models providing information about the vehicle pose. A common model in the planar case is the range-bearing (RB) model defined by

$$\bar{h}_{RB}(y)=\begin{pmatrix}\|y\|\\ \arctan2(y_2,y_1)\end{pmatrix},$$

for a given $y\in\mathbb{R}^2$, which is often employed for range sensors such as Lidars. The 3-D version (for 3-D Lidars) can be similarly constructed using range and two angular measurements (e.g. elevation and azimuth). Another model of interest in the 3-D case is the monocular camera (MC):

$$\bar{h}_{MC}(y)=\frac{y}{\|y\|},$$

for a given $y\in\mathbb{R}^3$, using a spherical projection model. Note that this model has one-to-one correspondence to the more standard perspective projection model. Here, $\bar{h}_{MC}(y)$ simply returns a body-fixed unit vector pointing from the camera center of projection to the feature ℓ, which is computed by shifting and normalizing the raw image pixel coordinates.

III. Gaussians on Lie Groups and their Optimal Estimation

A concentrated Gaussian on an d-dimensional Lie group can be defined as

$$f(g;\mu,\Sigma)\doteq\frac{1}{(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left\|\log(\mu^{-1}g)^\vee\right\|^2_{\Sigma^{-1}}\right), \; (4)$$

where $\log:G\to\mathfrak{g}$ is the Lie group logarithm, and $\mathfrak{g}$ denotes the Lie algebra. The map $(\cdot)^\vee:\mathfrak{g}\to\mathbb{R}^d$ converts Lie algebra elements into their corresponding vector of elements in a chosen algebra basis. By "concentrated" we mean that the eigenvalues of Σ are all small enough so that the tails of the distribution decay to zero along every geodesic path leading away from the identity. In other words, the pdf diminishes to a value close to zero on a small sphere centered at the mean. And therefore the global topological properties of G are not relevant.

From previous work, [8], [9], it has been shown that this sort of Gaussian distribution captures very well the propagated pdf for noisy kinematic systems. However, the measurement model in (2) is not expressed in this form. Even if a linear measurement model of the form $h(g)=g^{-1}\cdot\ell$ were used, the resulting distribution in g would not be a Gaussian in the sense of (4).

But we *can* approximate the pdfs for both linear and nonlinear measurement models by Gaussians of the form in (4). And the procedure for doing this does *not* require linearization of the measurement model from the beginning.

Suppose that $\mu_0=\mathbb{E}_{\rho_0}[g]$ is the mean of the prior distribution on G. Our main goal is to show that (3) can be captured well for values of g in neighborhood of μ_0 by a Gaussian of the form (4). And therefore, previously developed methods for fusion of pdfs in [7], [18] can be used for nonlinear measurement models.

In principle, the optimal parametric density $f(g;\mu,\Sigma)$ that is closest to $\rho(g|z)$ can be defined in the Kullback-Liebler (KL) sense, i.e. as the solution to the optimization problem

$$\min_{\mu,\Sigma}\mathrm{KL}\left(\rho(g|z)\;\|\;f(g;\mu,\Sigma)\right), \qquad (5)$$

where the KL distance between two given densities $p(g)$ and

$q(g)$ is defined by

$$\mathrm{KL}(p \parallel q) = \int_G p(g) \log \frac{p(g)}{q(g)} dg.$$

The optimization problem (5) is then equivalent to

$$\min_{\mu,\Sigma} \int_G -\rho(g|z) \log f(g;\mu,\Sigma) dg$$

and can be solved by locally parameterizing the mean according to $\mu = \mu_0 \exp(\epsilon)$ for some $\epsilon \in \mathfrak{g}$. In addition, the covariance Σ can be parametrized uniquely using an upper triangular matrix A such that $\Sigma^{-1} = A^T A$, i.e. A is the Cholesky factor of Σ^{-1}. Employing this parametrization the problem becomes

$$\min_{\epsilon,A} \Big\{ -\sum_{i=1}^{d} \log A_{ii} + \frac{1}{2} \int_G \rho(g|z) \left\| \log(\exp(-\epsilon)\mu_0^{-1} g)^\vee \right\|^2_{A^T A} dg \Big\}. \tag{6}$$

The most general (but often not the most efficient) way to optimally estimate $f(g)$ is through sampling. In particular, the problem (6) can be approximated according to

$$\min_{\epsilon,A} \Big\{ -\sum_{i=1}^{d} \log A_{ii} + \frac{1}{2} \sum_{i=1}^{N_s} \frac{\rho(z|g_i)}{\sum_{i=1}^{N_s} \rho(z|g_i)} \left\| \log(\exp(-\epsilon)\mu_0^{-1} g_i)^\vee \right\|^2_{A^T A} \Big\}, \tag{7}$$

where $g_i \in G$ are N_s i.i.d. samples from $\rho_0(g)$. The sampling-based form (7) is solved directly using a nonlinear method such as Newton's method for the unknowns (ϵ, A). The only requirement is that enough samples are chosen, and as $N_s \to \infty$ one recovers the optimal solutions.

Such sampling-based estimation can be used to compute approximately optimal Lie group Gaussians. While the technique can be employed for filtering (e.g. similarly to particle filtering in the Euclidean case), its main purpose in this work is for evaluation purposes. More specifically, we will compare such optimal parametric Lie group density to the corresponding optimal parametric Euclidean density.

A key result in this paper is then to derive an accurate parametric density not through sampling but through a local expansion and (first and second) moment matching, i.e. a local second-order perturbation approach developed in §V.

IV. Case study: capturing nonlinearities in a range-bearing model

We next consider the advantages of employing a Lie-group Gaussian measurement models as opposed to standard Gaussians in coordinates. The RB model will be used to illustrate the differences since the associated densities can be easily visualized in the three coordinates $q = (x_1, x_2, \theta) \in \mathbb{R}^3$, where the orientation angle θ is related to the rotation matrix $R \in \mathrm{SO}(2)$ by

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

Assume that the vehicle is at a true position $q = (0.1, 0, 0)$ and has a high-variance prior $\rho_0(g)$ shown in Figure 1a. The vehicle observes a landmark $\ell = (0.5, 0.5)$ with bearing-range measurement covariance given by $N = \mathrm{diag}(0.01, .2)$. The measurement pdf $\rho(z|g)$ defined in (2) is shown in Figure 1b. The pdf $\rho(g|z)$, defined in (3), combines the prior and measurement pdf's and is computed numerically using a fine grid discretization (Figure 1c). Note that all distributions shown in Figure 1 are depicted in coordinates for clarity.

Next, $N_s = 5000$ random samples are drawn from $\rho_0(g)$ to compute the approximately optimal Gaussian densities, both in standard Euclidean coordinates $f_c(q; \mu_c, \Sigma_c)$ and as a Lie group Gaussian $f(g; \mu, \Sigma)$. The parameters (μ, Σ) are computed using (7) while their coordinate counterpart (μ_c, Σ_c) using the analogous Monte Carlo Euclidean density estimation. The resulting densities are shown in Figure 2. Due to the nonlinearity in the models, the original density $\rho(g|z)$ "twists" along the vertical θ-axis. Such behavior clearly cannot be captured by ellipsoidal geometry in $\mathbb{R}^3$ associated to $f_c(q)$. In contrast, the Lie group Gaussian $f(g)$ is able to match the distribution $\rho(g|z)$ with surprisingly good accuracy. This could be explained through the built-in curvature information in the exponential map which captures the observed nonlinear behavior.

More specifically, the resulting distributions were compared numerically in terms of KL-divergence as well as difference between computed mean q and true mean q_0:

Lie group Gaussian	Euclidean Gaussian
$\mathrm{KL}(\rho(g\|z) \parallel f(g)) \approx 0.67$	$\mathrm{KL}(\rho(g\|z) \parallel f_c(q)) \approx 1.42$
$\\|q_0 - \hat{q}\\| \approx .013$	$\\|q_0 - \hat{q}\\| \approx .025$

The marked improvement in employing a Lie group Gaussian serves as a motivation to develop perturbation methods for approximating $\rho(g|z)$ which do not rely on samples but employ local expansion and moment matching, as described next.

V. Measurement update using second-order perturbation

A second-order Taylor series for $h(g)$ can be written as

$$h(g) \approx h(\mu) + \sum_{i=1}^{d} \eta_i (\partial_i h)(\mu) + \frac{1}{2} \sum_{i,j=1}^{d} \eta_i \eta_j (\partial_i \partial_j h)(\mu) \tag{8}$$

where here ∂_i is shorthand defined as

$$(\partial_i h)(\mu) = \frac{d}{ds} h(\mu\, e^{sE_i}) \Big|_{s=0}.$$

Note that (8) is not a linear measurement model, but rather a quadratic approximation of a nonlinear measurement model.

If we substitute (8) back into (2), and keeping up to quadratic terms, the exponent can be written in the form

$$c(\eta) \doteq -\frac{1}{2}(a + 2b^T \eta + \eta^T K \eta) \tag{9}$$

where

$$a = [h(\mu) - z]^T N^{-1} [h(\mu) - z]$$

$$b_i = [h(\mu) - z]^T N^{-1} (\partial_i h)(\mu)$$

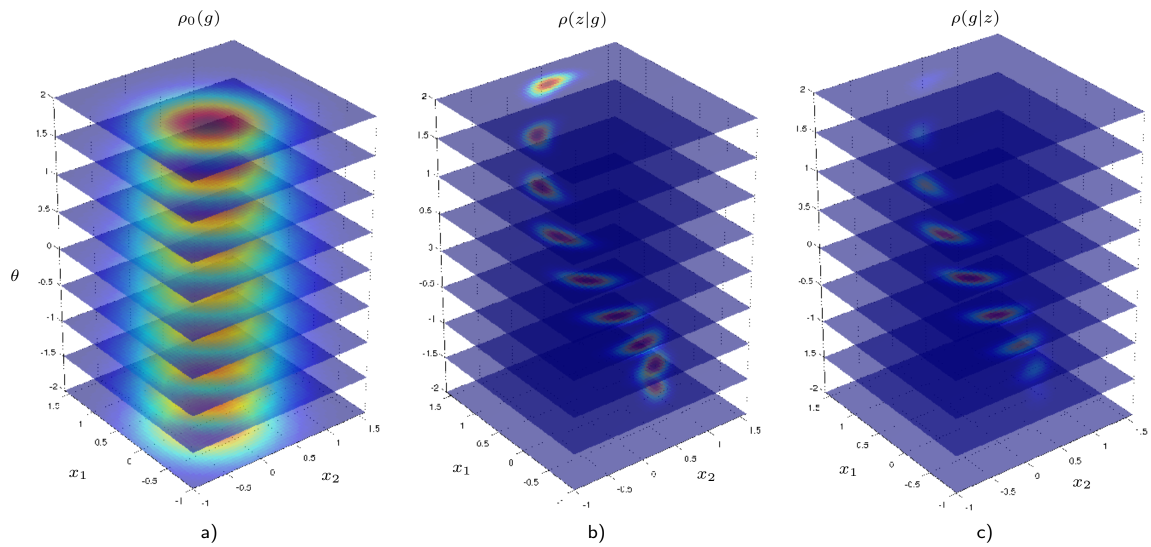

Fig. 1. Densities displayed in $q=(x_1,x_2,\theta)$ coordinates corresponding to prior $\rho_0(g)$ (shown in a), measurement model $\rho(z|g)$ (shown in b), and combined density $\rho(g|z)$ (shown in c). The density $\rho(g|z)$ is the full (non-parametric) nonlinear (and non-Gaussian in pose space) density that we aim to approximate.

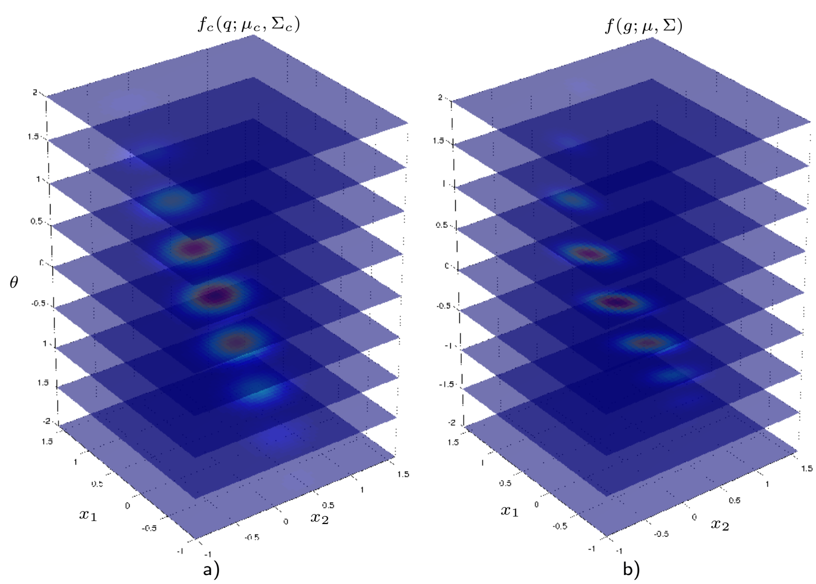

Fig. 2. Optimal parametric approximations of $\rho(g|z)$ from Figure 1: a) a Gaussian $f_c(q;\mu_c,\Sigma_c)$ in coordinates $q=(x_1,x_2,\theta)$; b) a Lie-group Gaussian $f(g;\mu,\Sigma)$ for $g\in SE(2)$.

$$K_{ij}=[(\partial_i h)(\mu)]^T N^{-1}(\partial_j h)(\mu)+[h(\mu)-z]^T N^{-1}(\partial_i\partial_j h)(\mu)$$

Keeping all terms up to quadratic (and no terms higher than this) is justified because of the assumption that the eigenvalues of N are all small. Note that if a linear measurement model had been assumed a priori, then the second term in K_{ij} would not be present. Higher-order Taylor series approximations are also possible, but then the measurement pdf would be defined by higher moments and would no longer be Gaussian.

A. Measurement PDFs Described as G-Gaussians

We now ask the following question: What G-Gaussian best approximates

$$\rho_n(\mu\circ e^{\eta})\approx\frac{|K|^{\frac{1}{2}}}{(2\pi)^{d/2}}e^{-\frac{c(\eta)}{2}}\quad ?$$

In other words, we seek

$$f(\mu\circ e^{\eta};\mu_n,\Sigma_n)=f(e^{\eta};\mu^{-1}\circ\mu_n,\Sigma_n)$$

to match to $\rho_n(\mu\circ e^{\eta})$ when $\mu^{-1}\circ\mu_n$ is close to the identity and the eigenvalues of Σ_n are small.

Under these conditions, the exponent in (4) can be expanded using the Baker-Campbell-Hausdorf (BCH) formula, and put in a form like (9). This provides equations that can be solved to obtain the equivalent $(\mu^{-1}\circ\mu_n,\Sigma_n)$ (and hence (μ_n,Σ_n)) that define a G-Gaussian approximation to the measurement model.

Let

$$e^{\epsilon}=\mu^{-1}\circ\mu_n\quad\text{and}\quad e^{\eta}=\mu^{-1}\circ g.$$

Then

$$\mu_n^{-1}\circ g=e^{-\epsilon}\circ e^{\eta}$$

where ϵ and η are both small. The BCH gives

$$\log^\vee(e^{-\epsilon}\circ e^{\eta})\approx -\epsilon^\vee+\eta^\vee-\frac{1}{2}ad(\epsilon)\eta^\vee + \frac{1}{12}[ad(\epsilon)ad(\epsilon)\eta^\vee - ad(\eta)ad(\eta)\epsilon^\vee]. \tag{10}$$

This results from truncating the BCH series at two nested Lie brackets and is called "second order" in [8]. It is quadratic in each of η and ϵ independently. Moreover, it is cubic in these variables jointly, and hence it could also be called a "third order" approximation from this perspective. The next term in the BCH involves three nested brackets and is quadratic in each η and ϵ independently, and is quartic in them jointly. There is no reason to use this "fourth order" approximation here because the information it contains pertains to moments higher than covariances. If the Gaussian model is good, then higher-order moments can be obtained from lower moments. Or, put another way, if the quartic approximation is used, it will contribute cubic terms (and higher) in η in the exponent of the G-Gaussian that will have nothing to match to in the exponent of ρ_n. If one desired to keep higher order terms, it may be possible, but this would only make sense in the context of a higher order Taylor series expansion than that in (8), and a non-Gaussian approximation.

Using the approximation in (10), the exponent in the G-Gaussian $f(e^{\eta}; e^{\epsilon}, \Sigma_n)$ can be expanded as

$$-\frac{1}{2}[\log^\vee(e^{-\epsilon}e^{\eta})]^T\Sigma_n^{-1}\log^\vee(e^{-\epsilon}e^{\eta})] \approx -\frac{1}{2}(a_n+2b_n^T\eta+\eta^T K_n\eta). \tag{11}$$

The right-hand side of (11) is a simple form to which we can match the corresponding approximation in the measurement model expressed in (9). In other words, *our main goal* becomes how to perform the matching

$$a=a_n,\ b=b_n,\ K=K_n,$$

which of course first requires writing analytical expressions for a_n, b_n, and K_n.

In finding a_n, b_n, and K_n in terms of ϵ and Σ_n, it will be convenient to use the linearity of $ad(\cdot)$. That is, if $\eta=\sum_{i=1}^d \eta_i E_i$ where E_i is an orthonormal basis for the Lie algebra } with respect to the inner product $(\cdot,\cdot)$, then

$$ad(\eta)=\sum_{i=1}^d \eta_i\, ad(E_i).$$

Using this, (10) becomes

$$\log^\vee(e^{-\epsilon}\circ e^{\eta})\approx -\epsilon^\vee+\left[\mathbb{I}_d-\frac{1}{2}ad(\epsilon)+\frac{1}{12}ad(\epsilon)ad(\epsilon)\right]\eta^\vee - \frac{1}{12}\sum_{i,j=1}^d \eta_i\,\eta_j\,ad(E_i)\,ad(E_j)\,\epsilon^\vee. \tag{12}$$

and we find that

$$(\epsilon^\vee)^T\Sigma_n^{-1}\epsilon^\vee=a \tag{13}$$

$$-(\epsilon^\vee)^T\Sigma_n^{-1}\left[\mathbb{I}_d-\frac{1}{2}ad(\epsilon)+\frac{1}{12}ad(\epsilon)ad(\epsilon)\right]=b^T \tag{14}$$

and

$$\left[\mathbb{I}_d-\frac{1}{2}ad(\epsilon)+\frac{1}{12}ad(\epsilon)ad(\epsilon)\right]^T\Sigma_n^{-1} \cdot\left[\mathbb{I}_d-\frac{1}{2}ad(\epsilon)+\frac{1}{12}ad(\epsilon)ad(\epsilon)\right]+M=K \tag{15}$$

where

$$M_{ij}=\frac{1}{12}(\epsilon^\vee)^T\left[ad_{E_j}^T ad_{E_i}^T\Sigma_n^{-1}+\Sigma_n^{-1}ad_{E_i}ad_{E_j}\right]\epsilon^\vee. \tag{16}$$

In keeping quadratic terms in η in the BCH, cross terms are present (such as those in M) that would not be if we had stopped at the linear approximation. We also get higher order terms in ϵ.

Referring back to (9) and matching $a=a_n$, $b_i=(b_n)_i$, and $K_{ij}=(K_n)_{ij}$ provides conditions to specify the values of ϵ and Σ_n in terms of properties obtained from the original measurement pdf. The nonlinear equations $b_n=b$ may not have exact closed-form solutions (and exact solutions may not even be possible). But a good initial guess that can be used to minimize $\|b_n=b\|^2$ with respect to ϵ is $\epsilon^\vee(0)=\Sigma_n b_n$, though Σ_n is unknown. A good initial guess for Σ_n such that $K_n=K$ is $\Sigma_n(0)=K^{-1}$.

Substituting both of these back in, the next better approximation can be obtained by letting $\epsilon^\vee(1)=\epsilon^\vee(0)+\nu$, where $\nu\in\mathbb{R}^d$ is small, and solving the associated quadratic minimization in closed form. Then substituting this known value into the expression for K_n reduces the search for the next Σ_n to one which involves inverting a linear equation in the entries of Σ_n^{-1}.

With both the propagation and measurement pdfs put in this common description, both fusion and future propagations can be handled in a single unified framework.

B. Solving the Matching Equations

Given the constraint equations (13)-(15) we seek Σ_n and ϵ. Our solution is based on approximations of increasing order of accuracy.

1) Zeroth Order Approximation: The simplest zeroth-order approximation from (14) and (15) results in $\Sigma_n\approx K^{-1}$ and $b_n\approx -K^{-1}b$. Such approximation is valid under the assumption that both $\|\Sigma_n\|$ and $\|\epsilon\|$ are small relative to 1.

2) Case 1: $\nu=O(\|\Sigma_n\|)=O(\|\epsilon\|)$.: Following a standard perturbation approach one can show the a first-order approximation requires the solution of the equations

$$\Sigma_n^{-1}=(\mathbb{I}-A_1^T+B^T)K(\mathbb{I}-A_1+B), \tag{17}$$

$$\epsilon^\vee=-(\mathbb{I}+A_1+C)K^{-1}b \tag{18}$$

where

$$A_1=\frac{1}{2}ad\left(\widehat{K^{-1}b}\right),\quad A_2=\frac{1}{12}ad\left(\widehat{K^{-1}b}\right)ad\left(\widehat{K^{-1}b}\right)$$

The matrix B is computed from the linear relationship

$$B^TK+KB=[-A_2+A_1^2]^TK[-A_2+A_1^2]-M. \tag{19}$$

after which the matrix C is computed to satisfy the equation

$$B^T - K(A_1^2 - B - C)K^{-1} = [-A_2 + A_1^2]^T. \qquad (20)$$

The procedure can be performed once or iterated multiple times until the variables Σ_n, ϵ converge. These terms are initialized using the zeroth order solution.

3) Case 2: $\|\Sigma_n\| = O(\nu)$ and $\|\epsilon\| = O(\nu^2)$.: We again start with (17) and (18) and the same lowest order approximations $\Sigma_n \approx K^{-1}$ and $b_n \approx -K^{-1}b$. But in this scenario, we take $A_2 = \mathbb{O}$ since ϵ is already $O(\nu)$-times smaller than Σ_n. Therefore, in the first order matching does not appear and we solve the following linear equation (which is a modified version of (20) given the above constraints)

$$B^T - KBK^{-1} = -A_1^T$$

for B, which is the only second-order correction, along with

$$\Sigma_n = K^{-1} - BK^{-1} - K^{-1}B^T, \quad \epsilon^\vee = -K^{-1}b.$$

4) Case 3: $\|\Sigma_n\| = O(\nu^2)$ and $\|\epsilon\| = O(\nu)$.: In this scenario $B = \mathbb{O}$ because corrections at this level are not required for Σ_n. We then have

$$\Sigma_n = K^{-1} + A_1K^{-1} + K^{-1}A_1^T, \quad \epsilon^\vee = -(\mathbb{I} + A_1 + C)K^{-1}b.$$

VI. Conclusion

A general methodology was presented to capture measurement probabilities associated with nonlinear measurement models on Lie groups as Gaussian distributions in exponential coordinates. Unlike previous Lie-group methods that assume a priori linearized measurement models, we build fully nonlinear measurement models from first principles. Nevertheless, and somewhat surprisingly, the probability densities associated with these measurement models can be captured quite well as Gaussian distributions in exponential coordinates. Future work will develop general recursive filters based on the proposed formulation and moment matching solution, and study their performance in comparison to standard extended or unscented Kalman filters.

References

[1] Lo, J. T.-H., Eshleman, L.R., "Exponential Fourier Densities on $SO(3)$ and Optimal Estimation and Detection for Rotational Processes," *SIAM J. Appl. Math.*, Vol. 36, No. 1, pp. 73-82, Feb. 1979.

[2] Lo, J.T.-H., "Optimal Estimation for the Satellite Attitude Using Star Tracker Measurements," *Automatica*, Vol. 22, No. 4, pp. 477-482, 1986.

[3] Glover, J., Rusu, R., Bradski, G., "Monte Carlo Pose Estimation with Quaternion Kernels and the Bingham Distribution," *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

[4] Murray, R.M., Li, Z., Sastry, S.S., *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, 1994.

[5] Park, F.C., *The Optimal Kinematic Design of Mechanisms*, Ph.D. Thesis, Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 1991.

[6] Kobilarov, M. B., Marsden, J. E. (2011). "Discrete geometric optimal control on Lie groups," *IEEE Transactions on Robotics*, 27(4), 641-655.

[7] Wolfe, K.C., Mashner, M., Chirikjian, G.S., "Bayesian Fusion on Lie Groups," *Journal of Algebraic Statistics* Vol. 2, No. 1, 2011, 75-97

[8] Wang, Y., Chirikjian, G.S., "Nonparametric Second-Order Theory of Error Propagation on the Euclidean Group," *International Journal of Robotics Research*, Vol. 27, No. 1112, November/December 2008, pp. 12581273

[9] Park, W., Liu, Y., Zhou, Y., Moses, M., Chirikjian, G.S., "Kinematic State Estimation and Motion Planning for Stochastic Nonholonomic Systems Using the Exponential Map," *Robotica*, 26(4), 419-434. July-August 2008

[10] Chirikjian, G.S., *Stochastic Models, Information Theory, and Lie Groups: Volume 2 - Analytic Methods and Modern Applications*, Birkhäuser, (450 pages), December 2011.

[11] Chirikjian, G.S., Kyatkin, A.B., *Engineering Applications of Noncommutative Harmonic Analysis*, CRC Press, October 2000.

[12] Marandi, S.R., Modi, V.J., "a Preferred Coordinate System and the Associated Orientation Representation in Attitude Dynamics," *Acta Astronautica*, Vol. 15, pp. 833-843, 1987.

[13] Schaub, H., Tsiotras, P., Junkins, J.L., "Principal Rotation Representations of Proper $N \times N$ Orthogonal Matrices," *International Journal of Engineering Science*, Vol. 33, No. 15, pp. 2277-2295, Dec. 1995.

[14] Lee, T., Leok, M. ; McClamroch, N.H. "Global symplectic uncertainty propagation on SO(3)," *47th IEEE Conference on Decision and Control*, Dec. 9-11 2008, pp. 61-66.

[15] Swensen, J.P., Cowan, N.J. "An almost global estimator on $SO(3)$ with measurement on S^2," *American Control Conference (ACC)*, 27-29 June 2012, pp. 1780 - 1786

[16] Markley, F.L., "Attitude Filtering on SO(3)," *The Journal of the Astronautical Sciences* 54(3-4), July-Dec. 2006, pp. 391-413.

[17] Malis, E., Hamel, T., Mahony, R., Morin, P. "Dynamic estimation of homography transformations on the special linear group for visual servo control," *IEEE International Conference on Robotics and Automation, 2009. ICRA'09.* (pp. 1498-1503).

[18] Long, A., Wolfe, K. C., Mashner, M., Chirikjian, G. S. (2012). "The Banana Distribution is Gaussian: A Localization Study with Exponential Coordinates." In *Robotics: science and systems.*

[19] Julier, S. J., Uhlmann, J. K. (2004). "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, 92(3), 401-422.

[20] Daum, F. (2005). "Nonlinear filters: beyond the Kalman filter," *IEEE Aerospace and Electronic Systems Magazine*, 20(8), 57-69.

[21] Barfoot T D and Furgale P T. "Associating Uncertainty with Three-Dimensional Poses for use in Estimation Problems," *IEEE Transactions on Robotics*, 2014 (to appear)

[22] Absil, P. A., Mahony, R., Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds.* Princeton University Press.

[23] Mahony, R., Hamel, T., Pflimlin, J. M. (2008). "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on Automatic Control*, 53(5), 1203-1218.

[24] Bonnabel, S., Martin, P., Rouchon, P. (2009). "Non-linear symmetry-preserving observers on Lie groups," *IEEE Transactions on Automatic Control* 54(7), 1709-1713.

[25] Bonnabel, S., Martin, P., Salaun, E. (2009, December). "Invariant Extended Kalman Filter: theory and application to a velocity-aided attitude estimation problem," *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC* 2009. (pp. 1297-1304).

[26] Bonnabel, S. (2007, December). "Left-invariant extended Kalman filter and attitude estimation," *IEEE conference on decision and control* (pp. 1027-1032).

[27] Said, S., Manton, J. H. (2013). "On Filtering with Observation in a Manifold: Reduction to a Classical Filtering Problem," *SIAM Journal on Control and Optimization*, 51(1), 767-783.

[28] Bourmaud, G., Mgret, R., Giremus, A., Berthoumieu, Y. (2013, September). "Discrete extended Kalman filter on Lie groups," *21st European Signal Processing Conference* 2013 (EUSIPCO 2013).

[29] Lefferts, E. J., Markley, F. L., Shuster, M. D. (1982). "Kalman filtering for spacecraft attitude estimation," *Journal of Guidance, Control, and Dynamics*, 5(5), 417-429.

[30] Thrun, S., Burgard, W., Fox, D. (2005). *Probabilistic robotics*. MIT press.