HAL
open science

Stochastic observers on Lie groups: a tutorial

Axel Barrau, Silvère Bonnabel

To cite this version:

Axel Barrau, Silvère Bonnabel. Stochastic observers on Lie groups: a tutorial. IEEE Conference on Decision and Control, IEEE, Dec 2018, Miami, United States. hal-01826025v2

HAL Id: hal-01826025

https://minesparis-psl.hal.science/hal-01826025v2

Submitted on 7 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic observers on Lie groups: a tutorial

Axel Barrau, Silvère Bonnabel

***Abstract*— In this tutorial paper, we discuss the design of geometric observers on Lie groups in the presence of noise. First we review Lie groups, and the mathematical definition of noises on Lie groups, both in discrete and continuous time. In particular, we discuss the Itô-Stratonovich dilemma. Then, we review the recently introduced notion of group affine systems on Lie groups. For those systems, we discuss how using the machinery of Harris chains, (almost) globally convergent deterministic observers might be shown to possess stochastic properties in the presence of noise. We also discuss the design of (invariant) extended Kalman filters (IEKF), and we recall the main result, i.e., the Riccati equation computed by the filter to tune its gains has the remarkable property that the Jacobians (A,C) with respect to the system's dynamics and output map are independent of the followed trajectory, whereas the noise covariance matrices that appear in the Riccati equation may depend on the followed trajectory. Owing to this partial independence, some local deterministic convergence properties of the IEKF for group-affine systems on Lie groups may be proved under standard observability conditions.**

I. Introduction

Over the past decade, the design of observers on Lie groups has been a vibrant research topic, see [9], [25], [24], [23], [21], [40], to cite a few. In the presence of noise, observer design is generally referred to as a "filtering problem". Several approaches to filtering for systems possessing a geometric structure have been historically developed. For stochastic processes on Riemannian manifolds [22] some results have been derived, see e.g., [28]. The specific situation where the process evolves in a vector space but the observations belong to a manifold has also been considered, see e.g. [16], [31] and more recently [32]. For systems on Lie groups harmonic analysis has been used [37], [34], notably in the case of bilinear systems [36] and estimation of a Brownian motion [15].

Whereas this paper is concerned with the stochastic approach, an alternative stream of research consists of geometric deterministic approaches that do account for uncertainty, by solving a least squares (i.e., minimal energy) problem on the Lie group, see [18], [39]. Minimal energy filtering on groups compares favorably to state-of-the art, see e.g., [8]. Note that, they are related to optimal control theory where geometric methods have long played an important role [12].

S. Bonnabel is with MINES ParisTech, PSL Reasearch University, Centre for Robotics, 60 bd Saint-Michel, 75006 Paris, France. silvere.bonnabel@mines-paristech.fr.

A. Barrau is with SAFRAN TECH, Groupe Safran, Rue des Jeunes Bois - Chateaufort, 78772 Magny Les Hameaux CEDEX, France, axel.barrau@safrangroup.com.

II. Primer on Lie groups

For a tutorial and robotics oriented introduction to matrix Lie groups, see [1], chapter 7. For a more mathematical and differential geometry based treatment, see [20].

Matrix Lie group: In this paper, we will only consider matrix Lie groups, although the results remain valid for general Lie groups. Indeed, matrix Lie groups provide a more concrete picture to the unfamiliar reader. A matrix Lie group G is subgroup of $GL_N(\mathbb{R})$ (invertible square $N \times N$ matrices) and an embedded manifold of $\mathbb{R}^{N\times N}$. The first condition means G is a subset of $GL_N(\mathbb{R})$ containing identity ($I_N \in G$) and stable by multiplication and inversion: for χ_1, χ_2 matrices of G, we have $\chi_1\chi_2 \in G$, and $\chi_1^{-1} \in G$. The second condition means that an affine subspace $\chi + T_\chi G$ "tangent to G at χ" can be defined for any point $\chi \in G$ as on Figure 1 (the base "point" χ being a matrix and $T_\chi G$ a sub-vector space of $\mathbb{R}^{N\times N}$). The vector space $T_\chi G$ is called "tangent space at χ". The dimension of this space, independent of the chosen point χ, is the dimension d of G. One of the most basic examples is $SO(3) = \{R \in \mathbb{R}^{3\times 3} \mid \det R = 1,\ R^T R = I_3\}$ where I_3 denotes the identity matrix of $\mathbb{R}^{3\times 3}$. $SO(3)$ is a 3 dimensional Lie group, that is, $d = 3$ and $N = 3$.

Lie algebra: The identity matrix $I_N \in G$ is the neutral element for the group composition. The tangent space $T_\chi G \in \mathbb{R}^{N\times N}$ mentioned avove, taken at $\chi = I_N$, can be defined as the vector space spanned by all initial velocity vectors $\frac{d}{dt}\gamma(0)$ for curves $\gamma(t)$, that start at $\gamma(0) = I_N$ and entirely lie in G, see Figure 1, left plot. The tangent space $T_{I_N}G$ is called the Lie algebra of G, and is denoted by $\mathfrak{g}$, i.e.,

$$T_{I_N}G := \mathfrak{g} \subset \mathbb{R}^{N\times N}.$$

It is a vector subspace of $\mathbb{R}^{N\times N}$, having dimension $d < N^2$.

Identification of the Lie algebra with $\mathbb{R}^d$: $\mathfrak{g}$ being a d-dimensional vector subspace of $R^{N\times N}$, identifying it to the vector space $\mathbb{R}^d$ makes all computations easier. A map, $\mathbb{R}^d \to \mathfrak{g}$ filling this task[1] will be denoted by $\xi \mapsto \xi^\wedge$. Thus

$$\xi \in \mathbb{R}^d \text{ is a vector, and } \xi^\wedge \in \mathbb{R}^{N\times N} \text{ is a matrix.}$$

This is illustrated on Figure 1, right plot. We will denote the inverse of $^\wedge$ by $^\vee$, that is $(\xi^\wedge)^\vee = \xi$.

Exponential map: There is a natural map between $\mathfrak{g}$ and G, based on the matrix exponential $\exp_m$. As $\mathfrak{g}$ can be identified to $\mathbb{R}^d$, we also have a local parameterization of G in the neighborhood of I_N by vectors of $\mathbb{R}^d$: $\xi \to \exp_m(\xi^\wedge)$ which proves very useful in practice. In the remainder of the paper, we will define and refer to the exponential map exp :

[1] Note that, specifying the map $^\wedge$ amounts in fact to choosing a basis of the Lie algebra.

$\mathbb{R}^d \to G$ as the map defined by $\exp(\xi) := \exp_m(\xi^\wedge)$. Using a first order Taylor expansion of the matrix exponential, we obtain the local approximation $\exp(\xi) = I_N + \xi^\wedge + O(||\xi||^2)$ in $\mathbb{R}^{N\times N}$. Alternatively, $\exp(\xi)$ is the solution at time $t=1$ to $\frac{d}{dt}\chi_t = \chi_t \xi^\wedge$, for $\chi_0 = I_N$. It defines a diffeomorphism from $\mathbb{R}^d$ to G, invertible at least in an open subset of $\mathbb{R}^d$ containing 0. Its local inverse is called the Lie logarithm $\log \chi$.

Linearization on groups: Similarly, the tangent space $T_\chi G$ at arbitrary $\chi \in G$ can be identified to $\mathbb{R}^d$ as multiplication by χ transports $T_{I_N}G$ to $T_\chi G$. Indeed, $T_\chi G = \chi\mathfrak{g} = \mathfrak{g}\chi$. This means vectors of $T_\chi G$ can be written as $\chi\xi^\wedge$ with $\xi \in \mathbb{R}^d$, or $(\tilde{\xi}^\wedge)\chi$ with a different $\xi \neq \tilde{\xi}$ (see Figure 1, center plot), but we will privilege left multiplications, for exposition purposes. To linearize a function $h : G \mapsto \mathbb{R}^p$, at an arbitrary point $\chi \in G$, we can evaluate how it changes by infinitesimally following an arbitrary tangent vector $\chi\xi^\wedge$ at χ. The left linear approximation to $h : G \to \mathbb{R}^p$ at χ in the direction $\xi \in \mathbb{R}^d$ can then be defined as the matrix $H \in \mathbb{R}^{p\times d}$ such that $h(\chi\exp(\xi)) - h(\chi) = H\xi + O(||\xi||^2)$, that is, $H\xi = \frac{d}{ds}h(\chi\exp(s\xi))\,|_{s=0}$. It is usually called the left-trivialized tangent map. If h is defined for any matrix $M \in \mathbb{R}^{N\times N}$, with differential $Dh_\chi : \mathbb{R}^{N\times N} \to \mathbb{R}^p$, the first order approximation of the exponential above thus yields $H\xi = Dh_\chi(\chi\xi^\wedge)$. One must then bear in mind that infinitesimal shifts at any $\chi \in G$ are thus always represented by elements of $\mathbb{R}^d$.

Adjoint representation: The operator Ad_χ, encoded by a matrix of $\mathbb{R}^{d\times d}$, is called the adjoint representation of the Lie group G, and is defined by $(Ad_\chi\xi)^\wedge := \chi(\xi^\wedge)\chi^{-1}$. We have the useful relation $\chi\exp(\xi) = \exp(Ad_\chi\xi)\chi$.

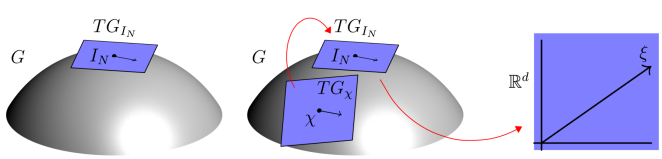

Fig. 1. Left and right multiplications offer two ways to identify the tangent space $T_\chi G$ at χ with the tangent space at Identity $T_{I_N}G$, called the Lie algebra and denoted $\mathfrak{g}$. In turn, the application $\xi \mapsto \xi^\wedge$ provides a linear bijection between the Euclidean space $\mathbb{R}^d$ and $\mathfrak{g}$.

III. Noisy dynamics on groups

Dynamical systems on a matrix Lie group G are easily defined. However, to account for dispersion under the effect of noise, the usual method that resorts to additive Gaussian white noise in the Euclidean setting does not directly carry over to Lie groups, as G is not a vector space.

A. Discrete time noisy dynamics

Consider a system evolving on G in discrete time:

$$\chi_{t+1} = f(\chi_t), \quad \text{with } f : G \to G. \tag{1}$$

Suppose we want to add a centered d-dimensional Gaussian noise $\xi \sim \mathcal{N}(0,Q)$ to the dynamics, to reflect uncertainties in the evolution model. As $f(\chi_t) \in G$, standard addition $f(\chi_t) + \xi \notin G$ does not make sense: first we must find a group counterpart to $\xi \in \mathbb{R}^d$, and then "add" it to $f(\chi_t)$.

"Gaussian" noise on G: A natural counterpart to $\xi \sim \mathcal{N}(0,Q)$ on G is easily obtained defining $W = \exp(\xi) \in G$ as a random element of G, as largely advocated in [11], [2], [1]. The law of W is often referred to as wrapped normal distribution. See also [13] for a slightly different definition.

"Adding" noise on G: Now that we have a notion of "centered" random "Gaussian" variable $\exp(\xi)$ on G, we need to apply it to $f(\chi_t)$ to model dispersion with respect to the deterministic dynamics (1). To do so, standard addition may be replaced with group multiplication. Depending on the choice of left or right multiplication, this yields two alternative noisy dynamics

$$\chi_{t+1} = f(\chi_t)\exp(\xi) \quad \text{with} \quad \xi \sim \mathcal{N}(0,Q) \tag{2}$$

$$\underline{\text{or}} \quad \chi_{t+1} = \exp(\xi)f(\chi_t) \quad \text{with} \quad \xi \sim \mathcal{N}(0,Q) \tag{3}$$

"Isotropic" noise: As $\chi\exp(\xi) = \exp(Ad_\chi\xi)\chi$, we speak of "isotropic" (or Ad-invariant or bi-invariant) noise $\xi \sim \mathcal{N}(0,Q)$ if for all $\chi \in G$ we have $Ad_\chi Q Ad_\chi = Q$. In this case both dynamics (2) and (3) define processes having identical probability distributions for noise-free $f(\chi_t)$.

B. Continuous time noisy dynamics

Consider a system evolving on G in continuous time:

$$\frac{d}{dt}\chi_t = f(\chi_t) \tag{4}$$

with $f : G \to T_{\chi_t}G$, such that $\frac{d}{dt}\chi_t$ is a tangent vector to G at χ_t, see Figure 1, central plot.

Suppose we want to add a centered d-dimensional white noise with covariance Q to the dynamics, to reflect some uncertainties in the evolution model. As the tangent space $T_{\chi_t}G$ is a vector space, and as $\chi(w^\wedge) \in T_\chi G$ for any vector $w \in \mathbb{R}^d$, a mere addition might work, that is,

$$\frac{d}{dt}\chi_t = f(\chi_t) + \chi_t(w_t^\wedge) \in T_{\chi_t}G \tag{5}$$

where we let w_t be a continuous centered white noise in $\mathbb{R}^d$ with covariance matrix Q. Note that, here again, we could have instead added a term of the form $(\tilde{w}_t^\wedge)\chi_t$, as the latter is also an element of $T_{\chi_t}G$. Owing to the definition of Ad_χ, we see that if $\tilde{w}_t = Ad_{\chi_t}w_t$ then both equations are identical. As a result, we only need consider (5), at least at this stage.

The problem is that (5) is a stochastic differential equation (SDE) on G, since it contains the stochastic term $\chi_t(w_t^\wedge)$, and it should be given a proper mathematical meaning. It turns out that, adding stochastic terms to ordinary differential equations generates some (apparently inevitable) mathematical issues whereas adding uncertainty is relatively straightforward in the discrete time case. This prompts the following paragraphs, that are meant to be a gentle, but rather mathematically loose, introduction to the subject.

C. Stochastic differential equations in $\mathbb{R}^d$

For a physical introduction to stochastic differential equations (SDE) see [35]. Consider a stochastic differential equation (SDE) in $\mathbb{R}^d$ of the form:

$$\frac{d}{dt}x_t = g(x_t) + \bar{w}_t \tag{6}$$

where $\bar{w}_t$ is a continuous centered white noise in $\mathbb{R}^d$ with covariance matrix Q. It can be re-written as:

$$\frac{d}{dt}x_t = g(x_t) + Q^{1/2}w_t \tag{7}$$

with w_t a white noise with identity covariance matrix I_d.

White noise: The noise w_t cannot be mathematically defined as a a random function of time t, only its integral (the noise cumulated over a time interval) is: for all $t_1, t_2 \geq 0$ we have $\int_{t_1}^{t_2} w_s ds = W_{t_2} - W_{t_1}$, with W_t a d-dimensional Wiener process, also known as d-dimensional Brownian motion (and the white noise w_t can be *interpreted* as its time derivative). The Wiener process or Brownian motion is a well-known stochastic process, characterized by the fact that it is continuous, centered, with statistically independent increments and such that $W_{t+\Delta t} - W_t \sim \sqrt{\Delta t}\mathcal{N}(0, I_d)$.

Discretization scheme: Eq. (7) can thus be given a proper mathematical meaning through its integral counterpart, that is, the well-defined equation (W_t is a d-Wiener process):

$$x_t = x_0 + \int_0^t g(x_s)ds + Q^{1/2}W_t. \tag{8}$$

We can then define a solution to SDE (7) as any $(x_t)_{t\geq 0}$ that satisfies the integral equation (8). Moreover we can write the increment $x_t - x_0$ as the sum of all the variations over intervals of length Δt, that is, $\sum_{i=0}^{(t/\Delta_t)}[x_{(i+1)\Delta t} - x_{i\Delta t}]$. Using in (8) that $\int_{i\Delta t}^{(i+1)\Delta t} g(x_s)ds = g(x_{i\Delta t})\Delta t + O(\Delta t^2)$, and owing to the property of the Wiener process above, we have thus

$$x_t - x_0 = \lim_{\Delta t \to 0} \sum_{i=0}^{(t/\Delta_t)} [g(x_{i\Delta t})\Delta t + \sqrt{\Delta t}Q^{1/2}\xi_i]$$

in distribution, where the $\xi_i \sim \mathcal{N}(0, I_d)$ are i.i.d standard Gaussians. More prosaically (and quite loosely speaking) it means that the solution to the differential equation (7), and thus to (6), may be defined as the limiting process as $\Delta t \to 0$ of the following (Euler-Maruyama) discretization method;

$$x_{t+\Delta t} = x_t + \Delta t g(x_t) + \sqrt{\Delta t}\xi$$

with the $\xi's$ i.i.d Gaussians with distribution $\mathcal{N}(0, Q)$.

Mathematical problems arise, though, with equation

$$\frac{d}{dt}x_t = g(x_t) + r(x_t)w_t \tag{9}$$

where $r : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ an arbitrary function of x_t. There are then two ways to cope with the multiplicative term $r(x_t)$.

Itô's approach: Roughly speaking, it consists in defining the solution to (9) as the limit as $\Delta t \to 0$ of the following (Euler-Maruyama) discretization method

$$x_{t+\Delta t} = x_t + \Delta t g(x_t) + \sqrt{\Delta t}\ r(x_t)\xi \tag{10}$$

with the $\xi's$ i.i.d Gaussians with distribution $\mathcal{N}(0, Q)$.

Stratonovich's approach: Roughly speaking, it consists, in contrast, in defining the solution to (9) as the limit as $\Delta t \to 0$ of the following discretization method

$$x_{t+\Delta t} = x_t + \Delta t g(x_t) + \sqrt{\Delta t}\ r((x_t + x_{t+\Delta t})/2)\,\xi$$

with the $\xi's$ i.i.d Gaussians with distribution $\mathcal{N}(0, Q)$.

Although Itô's approach seems more natural, as it leads to an explicit integration scheme instead of an implicit one, Stratonovich's approach has the merit to be compatible with chain's rule, i.e., if F is any smooth function, and x_t is the solution to (9) in the sense of Stratonovich, then $z_t = F(x_t)$ is the solution in the sense of Stratonovich to equation $\frac{d}{dt}z_t = DF(x_t)\big(g(x_t) + r(x_t)w_t\big)$. Unfortunately, this is not true as concerns the solution to (9) in the sense of Itô.

D. Stochastic differential equations (SDE) on Lie groups

Hence, there are two ways to interpret the stochastic term $\chi_t w_t^\wedge$ of (5). The first one is Itô's intepretation of SDE (5). Roughly speaking, it means that the solution to the differential equation (5) corresponds to the limit as $\Delta t \to 0$ of the following (Euler-Maruyama) discretization method

$$\chi_{t+\Delta t} = \chi_t + \Delta t f(\chi_t) + \sqrt{\Delta t}\chi_t(\xi_t^\wedge) \tag{11}$$

with ξ's independent Gaussians with distribution $\mathcal{N}(0, Q)$. Stratonovich's interpretation leads instead to

$$\chi_{t+\Delta t} = \chi_t + \Delta t f(\chi_t) + \sqrt{\Delta t}\frac{(\chi_t + \chi_{t+\Delta t})}{2}(\xi_t^\wedge) \tag{12}$$

The question is whether the limiting processes $(\chi_t)_{t\geq 0}$ of discretizations (11) and (12) as $\Delta t \to 0$ actually remain inside G at all times, or if we end up with matrices of the ambiant space $\mathbb{R}^{N\times N}$ that lie outside G.

Itô vs. Stratonovich: It turns out that, only the Stratonovich approach guarantees that the solution $(\chi_t)_{t\geq 0}$ remains in G. Indeed, *Stratonovich's approach is the only one that is compatible with the geometry*. From Von Neumann's closed subgroup theorem, we know that any matrix Lie group G corresponds to a level set of a submersion $F : \mathbb{R}^{N\times N} \to \mathbb{R}^{N^2}$. Thus, the the tangent space at χ is the set $T_\chi G = \{v \in \mathbb{R}^{N\times N} \mid DF(\chi)v = 0\}$. This implies that, if at all times $\frac{d}{dt}\chi_t = v_t$ is a (possibly random) element of $T_{\chi_t}G$ then, and χ_t is defined as the solution in the Stratonovich sense, we have $\frac{d}{dt}F(\chi_t) = DF(\chi_t)(\frac{d}{dt}\chi_t)$ since Stratonovich's stochastic calculus is compatible with the chain rule, and thus $\frac{d}{dt}F(\chi_t) = 0$ at all times; so that x_t remains in G. As an exemple, we may choose $F(\chi) = (\chi^T\chi, \det\chi)$, and matrices of $SO(3)$ correspond to the level set $F(\chi) = I_3, \det(\chi) = 1$. For more details, see Section 20.4 of [13].

With Itô's approach, the solution unfortunately steps out of the group almost surely, see e.g. [4].

Implementation of SDE on Lie groups: Another route is as follows. Note that, $f(\chi_t) \in T_{\chi_t}G$ and thus it is of the form $f(\chi_t) = \chi_t(\omega_t^\wedge)$ where $\omega_t^\wedge \in \mathfrak{g}$, and where $\omega_t \in \mathbb{R}^d$ is defined as $\omega_t = [\chi_t^{-1}f(\chi_t)]^\vee$. With this notation, Equation (5) then becomes

$$\frac{d}{dt}\chi_t = \chi_t(\omega_t + w_t)^\wedge. \tag{13}$$

As $\omega_t + w_t \in \mathbb{R}^d$, one could be tempted to discretize this equation using a group multiplicative counterpart of Itô's approach (10), that is, using i.i.d. Gaussians $\xi \sim \mathcal{N}(0, Q)$:

$$\chi_{t+\Delta t} = \chi_t \exp_m(\Delta t(\omega_t)^\wedge + \sqrt{\Delta t}(\xi^\wedge)) = \chi_t \exp(\Delta t\omega_t + \sqrt{\Delta t}\xi), \tag{14}$$

and the limiting process as the discretization step $\Delta t \to 0$ may serve as a direct definition for the solution of (10) on G. Indeed, we readily see the scheme (14) defines a stochastic process that does remain in G at all times, since $\Delta t(\omega_t)^\wedge + \sqrt{\Delta t}(\xi^\wedge) \in T_{I_N}G = \mathfrak{g}$, so that $\exp_m(\Delta t(\omega_t)^\wedge + \sqrt{\Delta t}(w_t^\wedge)) \in G$. Moreover, it reminds much of the discrete time case of Section III-A. It turns out this is a valid route indeed, owing to results by McKean [26] and related to his multiplicative stochastic integral, and more remotely to Brownian motion on Lie groups, [29], [22], [38].

Proposition 1 ([26], see also [13]): The discretized solution of the multiplicative scheme (14) converges in probability to the solution in the sense of Stratonovich to (13) and thus to the solution to SDE (5).

Corollary 1: Discretization (14) thus serves as an explicit convergent implementation scheme with time step Δt of the EDS (13) and thus (5), that guarantees the solution remains in the group G at all times, regardless of the chosen step Δt.

Numerical schemes: A wide variety of powerful methods exist and the reader is referred to e.g., Runge-Kutta-Munthe-Kaas (RKMK) methods [27] and more generally geometric numerical integration [17], [30].

IV. Stochastic observers on Lie groups

A. Considered stochastic observer problem

On the matrix Lie group $G \subset \mathbb{R}^{N\times N}$, we consider continuous time (not necessarily invariant) dynamics of the form

$$\frac{d}{dt} f_{u_t}(\chi_t) \tag{15}$$

here $f_{u_t} : G \to T_{\chi_t}G$. This echoes (4), except that to be more relevant to applications we allow f to also depend on a time-varying input $t \mapsto u_t$, where $u_t \in \mathscr{U}$ lives in some space $\mathscr{U}$. To model possible uncertainties in the dynamics, noise must be added in the tangent space along the lines of Section III-B.

Stochastic dynamics: This yields the two noisy dynamics

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t) + \chi_t(w_t^\wedge) \in T_{\chi_t}G \tag{16}$$

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t) + (\bar{w}_t^\wedge)\chi_t \in T_{\chi_t}G \tag{17}$$

The choice between (16) and (17) is generally dictated by the application. In many applications, χ_t represents the configuration of a vehicle in space, i.e., a transformation that maps the body (i.e., vehicle) frame to an earth fixed frame. In this case when confronted with relations of the form

$$\alpha = \chi_t\beta, \quad \alpha,\beta \in \mathbb{R}^N, \quad \chi_t \in G$$

we generally say that β is a vector of the body frame, whereas α is a vector of the fixed frame. In the main applications, u_t represents the measurements from motion sensors, such as accelerometers, gyrometers, and odometers, which are all attached to the body, and w_t then represents sensor noise, and is thus also a vector of the body frame, so that (16) are the dynamics to consider. The typical example is when $\chi_t = R_t \in SO(3)$ denotes a rotation matrix, in which case $u_t \in \mathbb{R}^3$ denotes the (bias free) gyrometers measurements, and $w_t \in \mathbb{R}^3$ the gyrometers' noise, that models the (random) discrepancy between measured angular velocity and the true angular velocity. The dynamics then write $\frac{d}{dt}R_t = R_t((u_t + w_t)^\wedge) = R_t(u_t^\wedge) + R_t(w_t^\wedge)$ indeed.

Measurement model: Suppose that there are noisy partial measurements of the state, available at discrete times $t_0 < t_1 < t_2 \cdots$, and which write

$$Y_n = h(\chi_{t_n}) + V_n \tag{18}$$

with $h : G \to \mathbb{R}^p$ an output map. The stochastic observer problem consists in providing the best estimate of χ_t given all past measurements $Y_1, \cdots, Y_n$ where $t_n \le t < t_{n+1}$.

B. Gaussian observers

The main application of the present Lie group based framework is designing EKF-like observers for systems without a proper vector space structure. In an EKF, the true state is modeled as $x_t = \hat{x}_t + e_t$, with e_t a Gaussian error variable. A Lie group counterpart may be written as $\chi_t = \hat{\chi}_t \exp(e_t)$ with $e_t \in \mathbb{R}^d$. Then, as in standard EKF, first-order expansion in e_t is performed. The most convincing application is attitude estimation, where the roll-pitch-yaw parameterization is obviously unadapted to EKF design, notably because of singularities. Interestingly, it has been noticed since, at least, the 1970s that the most efficient implementations of EKF for this problem relied on linearizing a multiplicative error parameterized by a small rotation vector (or a quaternion equivalent). In other words, the Lie group approach was implicitly adopted from the beginning, and preferred to the roll-pitch-yaw vector parameterization. See [14] for a survey.

C. Invariant observers

Invariant observers on Lie groups, or more generally nonlinear observers on Lie groups, are estimators that notably ensure the estimate $\hat{\chi}_t$ remains in the group G at all times. Referring to, e.g., [9], [25], [24], [23], [21], [40], they are continuous time and are generally of the form $\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t) + \hat{\chi}_t L(y(t) - h(\chi_t))$ ("left-invariant" observer), or alternatively of the form $\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t) + L(y(t) - h(\chi_t))\hat{\chi}_t$ ("right-invariant" observer), where $y(t)$ is a continuous time output noise-free measurement and $L : \mathbb{R}^p \mapsto \mathfrak{g}$ some nonlinear map to be tuned by the user. Note that, as $\hat{\chi}_t L(y(t) - h(\chi_t)) \in T_{\hat{\chi}_t}G$ the estimates are garanteed to remain inside G. However, noisy discrete-time measurements of the form (18) are more easily defined that continuous-time noisy measurements. Moreover, in practice, measurements always come in discrete time. The continuous-discrete counterpart of nonlinear observers on Lie groups is of the form

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \le t < t_n \tag{19}$$

$$\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n}\exp(L_n(Y_n - h(\chi_{t_n}))), \quad t = t_n \tag{20}$$

with $L_n : \mathbb{R}^p \to \mathbb{R}^d$ a nonlinear map to be tuned by the user.

Remark 1: The choice between left and right multiplication leads to two different families of observers. In (20), we have opted for left multplication by $\hat{\chi}_{t_n}$. However, right multiplications by $\hat{\chi}_{t_n}$ might prove suited to some applications. For more details, and guidelines see [6].

Step (19) is called the propagation step, and Step (20) the update step. To evaluate the accuracy of the estimates, consider the error $\eta_t = \chi_t^{-1}\hat{\chi}_t$ in the sense of group multplication, between the estimate and the true state. An absence of error leads to $\eta_t = I_N$. Of course, due to noise it is not possible to have $\eta_t \to I_N$, but one can try to minimize the dispersion of η_t under the effect of noise.

D. Group affine systems

The Lie group formalism has been utilized for the ease of parametrization of the state space it provides, see Section IV-B. It has also lead to powerful nonlinear estimation methods: the literature being too broad to be covered see e.g., [2], [23], [24]. In the next sections we will focus on specific systems for which the Lie group framework allows generalizing some properties of the linear case, as initiated in [9].

Over the past decade, many convergence results have been made possible by the fact one was dealing with systems being left or right-invariant, as in [9]. In [6], the following "group affine" property was shown to bring striking properties, and to be the proper framework generalizing linear systems:

$$\forall u \in \mathscr{U},\ a,b \in G \quad f_u(ab) = af_u(b) + f_u(a)b - af_u(Id)b. \tag{21}$$

In this equation $f_u(ab) \in T_{ab}G,\ f_u(a) \in T_aG,\ f_u(b) \in T_bG$. For example the left-invariant dynamics $\frac{d}{dt}\chi_t = \chi_t\omega_t$, the right-invariant ones $\frac{d}{dt}\chi_t = \omega_t\chi_t$ and the mixed-invariant ones $\frac{d}{dt}\chi_t = \chi_t\omega_t^{(1)} + \omega_t^{(2)}\chi_t$ all are group-affine dynamics. However they are not limited to them: the usual navigation equations[2] on flat earth with an IMU that consists of unbiased accelerometers and gyrometers are neither left nor right nor mixed-invariant, but they satisfy (21) as proved in [6].

The fundamental assumption (21) has been shown to be valid for a number of systems that revolve around navigation and simultaneous localization and mapping (SLAM) [5]. It is yet possibly limiting: if, e.g., biases of accelerometers and gyrometers are added in the navigation example just mentioned, the system is not group-affine anymore. However, the obtained system is "close" to being group-affine, ensuring the success of geometric methods in practice, see [7], [19].

Fundamental property of invariant filtering: Consider two trajectories $\chi_t, \tilde{\chi}_t$ of the noise free system (15). Let $\eta_t = \chi_t^{-1}\tilde{\chi}_t \in G$ be the error between those two solutions in the sense of group multiplication. It is the group analog of the linear error $\chi_t - \tilde{\chi}_t$, which does not make sense on the group G, since it is not a vector space. We have

Theorem 1 ([6]): f_{u_t} satisfies Equation (21) if and only if there exists a map g_{u_t} such that $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t)$. Moreover, we have necessarily that $g_{u_t}(\eta) = f_{u_t}(\eta) - f_{u_t}(I_N)\eta$.

Corollary 2 (of Thm 1, see[6]): Consider (16), with f_{u_t} group-affine, and observer (19)-(20). During the propagation step, we have $\frac{d}{dt}\eta_t = g_{u_t}(\eta_t) - (w_t^\wedge)\eta_t$, that is the evolution of the error does not depend explicitly on the trajectory $\hat{\chi}_t$.

[2]the equations then write $\frac{d}{dt}R_t = R_t\omega_t^\wedge$ $\frac{d}{dt}P_t = V_t$, $\frac{d}{dt}V_t = g + R_tu_t$ with g the earth gravity vector, ω_t, u_t the gyro and accelero outputs, R_t the orientation, P_t the position and V_t the velocity of the body that carries the IMU.

Proof: We write $\frac{d}{dt}\eta_t = (\frac{d}{dt}\chi_t^{-1})\hat{\chi}_t + \chi_t^{-1}(\frac{d}{dt}\hat{\chi}_t)$ and then we use the general matrix equality $\frac{d}{dt}\chi_t^{-1} = -\chi_t^{-1}(\frac{d}{dt}\chi_t)\chi_t^{-1}$. This yields $\frac{d}{dt}\eta_t = \chi_t^{-1}[f_{u_t}(\hat{\chi}_t) - f_{u_t}(\chi_t)\eta_t] - (w_t^\wedge)\eta_t = f_{u_t}(\eta_t) - f_{u_t}(I_N)\eta_t - (w_t^\wedge)\eta_t = g_{u_t}(\eta_t) - (w_t^\wedge)\eta_t$ where we have used (21) that implies $f(\hat{\chi}) = f(\chi\eta) = \chi f(\eta) + f(\chi)\eta - \chi f(I_N)\eta$. ■

The latter corollary ensures the error evolution during the propagation step is independent from the trajectory, and with slight abuse it is referred to as "autonomous error" evolution. Moreover, the group-affine property automatically implies g_{u_t} is linear under a suitable change of coordinates [6]. This property is referred to as "log-linearity" of the error equation.

E. Harris chains and gradient-like observers

The paper [24] discusses systems on Lie groups having a synchrony property. In the present formalism this property means there exists input trajectories u_t such that $g_{u_t}(\eta) = 0\ \forall \eta, t$. In this case Eq. (24) defines a Brownian motion on G (i.e. a "random walk" on G), and if $\bar{L}_n(\cdot) = K(\cdot)$ is minus a gradient term that tends to make the output error $Y_n - h(\hat{\chi}_{t_n})$ decrease, inspiring from gradient-like observers [24], then there are hopes that the distribution of the error η_t will asymptotically converge to a stationary distribution, resorting to the machinery of Harris chains, as proved in [3] for the $SO(3)$ case, allowing optimal offline tuning of the gains w.r.t. the stationary error distribution.

F. Invariant extended Kalman filter (IEKF)

Suppose that, moreover, $h(\chi)$ is of the form $h(\chi) = \chi\bar{d}$ with $\bar{d} \in \mathbb{R}^N$ a vector. Resorting to the methodology of invariant extended Kalman filtering, originally introduced in [10], the Left-invariant EKF (LIEKF) for continuous-discrete group affine systems with left-equivariant output writes [6]:

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \le t < t_n \tag{22}$$

$$\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n}\exp(K_n(\hat{\chi}_{t_n}^{-1}Y_n - \bar{d})), \quad t = t_n \tag{23}$$

with $K_n \in \mathbb{R}^{d\times N}$ a gain matrix to be defined in the sequel. Then, using Cor. 2 and elementary computations, we see the error $\eta_t = \chi_t^{-1}\hat{\chi}_t$ satisfies the error equation:

$$\frac{d}{dt}\eta_t = g_{u_t}(\eta_t) - (w_t^\wedge)\eta_t, \quad t_{n-1} \le t < t_n \tag{24}$$

$$\eta_{t_n}^+ = \eta_{t_n}\exp(K_n(\eta_{t_n}^{-1}\bar{d} - \bar{d} - \hat{\chi}_{t_n}^{-1}V_n)), \quad t = t_n \tag{25}$$

Using linearization on Lie groups discussed at Section II, the error system (24)-(25) may be linearized around $\eta = I_N$ writing $\eta = \exp(\xi)$. Let A_t be defined by $g_{u_t}(\exp(\xi)) = (A_t\xi)^\wedge + O(||\xi||^2)$ and C by $\exp(\xi)^{-1}\bar{d} - \bar{d} = C\xi$ the linearized error equation writes for $\xi_t = \log(\eta_t)$

$$\frac{d}{dt}\xi_t = A_t\xi_t - w_t, \quad t_{n-1} \le t < t_n \tag{26}$$

$$\xi_{t_n}^+ = \xi_{t_n} + K_n(C\xi_{t_n} - \hat{V}_n)), \quad t = t_n \tag{27}$$

where $\hat{V}_n = \hat{\chi}_{t_n}^{-1}V_n$ and as in the EKF methodology the gain may be defined through a Riccati equation

$$\frac{d}{dt}P_t = A_tP_t + P_tA_t^T + \hat{Q}_t,\ P_{t_n}^+ = (I_d - K_nC)P_{t_n}$$

with $S_n = CP_{t_n}C^T + \hat{N}_n$ and $K_n = P_{t_n}C^T S^{-1}$ where $Q_t = \mathrm{cov}(w_t)$ and $\hat{N}_n = \mathrm{cov}(\hat{V}_n)$.

Stochastic first-order optimality: as the conventional EKF, the IEKF is stochastically optimal for the linearized system around the estimate, with respect to chosen noises. This property is important in practice, but provides no guarantee regarding the nonlinear system. Yet, nonlinear guarantees exist for the IEKF in a deterministic setting.

Deterministic nonlinear convergence properties: Note that, owing to the group-affine property of f_{u_t}, the Jacobian A_t is independent from $\hat{\chi}_t$, and so is C owing to the form of output assumed. This reminds much of the linear Kalman filter theory. The fact that covariance matrix $\hat{N}_n$ may depend on the estimates, impacts only mildly the stability of the error system. As a result, it could recently be proved in [6] that, when used as non-linear observer the IEKF enjoys local asymptotic convergence properties for all (observable) group-affine systems. This is a non-trivial result as there are few local convergence results for the conventional EKF, unless some impractical assumptions on the stability of the covariance matrix are made. see e.g., [33].

For comparisons of invariant EKF with other methods, the reader is referred to e.g., [7], [19], [5].

Acknowledgments

The authors thank Salem Said, for proofreading and advice, and Tarek Hamel and Robert Mahony for inviting them to participate in this tutorial session at CDC.

References

[1] T. D Barfoot. *State Estimation for Robotics*. Cambridge University Press, 2017.

[2] T. D Barfoot and P. T Furgale. Associating uncertainty with three-dimensional poses for use in estimation problems. *IEEE Transactions on Robotics*, 30(3):679–693, 2014.

[3] A. Barrau and S. Bonnabel. Intrinsic filtering on Lie groups with applications to attitude estimation. *IEEE Transactions on Automatic Control*, 60(2):436 – 449, 2015.

[4] A. Barrau and S. Bonnabel. Intrinsic filtering on SO (3) with discrete-time observations. *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on* , pp 3255–3260, 2013.

[5] A. Barrau and S. Bonnabel. An EKF-SLAM algorithm with consistency properties *arXiv preprint arXiv:1510.06263*, 2015.

[6] A. Barrau and S. Bonnabel. The invariant extended Kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4):1797–1812, 2017.

[7] A. Barrau and S. Bonnabel. Invariant Kalman Filtering. *Annual Review of Control, Robotics, and Autonomous Systems*, Vol 1, pp:237-257, 2018.

[8] J. Berger, A. Neufeld, F. Becker, F. Lenzen, C. Schnorr. Second Order Minimum Energy Filtering on SE(3) with Nonlinear Measurement Equations, *International Conference on Scale Space and Variational Methods in Computer Vision, Springer*, pp: 397-409. 2015.

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

[10] S. Bonnabel. Left-invariant extended Kalman filter and attitude estimation. In *IEEE conf. on decision and control*: 1027–1032, 2007.

[11] G. Bourmaud, R. Mégret, A. Giremus, and Y. Berthoumieu. Discrete extended Kalman filter on Lie groups. In *EUSIPCO, 2013 Proceedings of the 21st European*, pages 1–5, 2013.

[12] R. W. Brockett. Lie algebras and Lie groups in control theory. In *Geometric methods in system theory*, pages 43–82. Springer, 1973.

[13] G. S Chirikjian. *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, volume 2. Springer Science & Business Media, 2011.

[14] J. L. Crassidis, F. L. Markley, Y. Cheng. Survey of nonlinear attitude estimation methods. *Journal of guidance, control, and dynamics*, 30(1), 12-28. 2007

[15] T.E. Duncan. An estimation problem in compact Lie groups. *Systems & Control Letters*, 10(4):257–263, 1988.

[16] T. E Duncan. Some filtering results in Riemann manifolds. *Information and Control*, 35(3):182–195, 1977.

[17] E. Hairer, C. Lubich, C., and G. Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations* Springer Science & Business Media, 2006.

[18] Y. Han and F.C. Park. Least squares tracking on the Euclidean group. *Automatic Control, IEEE Transactions on*, 46(7):1127–1132, 2001.

[19] R. Hartley, M. G. Jadidi,J. W. Grizzle and R. M. Eustice. Contact-Aided Invariant Extended Kalman Filtering for Legged Robot State Estimation. *arXiv preprint arXiv:1805.10410.* 2018.

[20] S. Helgason *Differential Geometry, Lie Groups and Symmetric Spaces* Academic, New York, 1978.

[21] M.-D. Hua, P. Martin, and T.Hamel. Stability analysis of velocity-aided attitude observers for accelerated vehicles. *Automatica*, 63:11–15, 2016.

[22] K. Itô. Stochastic differential equations in a differentiable manifold. *Nagoya Mathematical Journal*, 1:35–47, 1950.

[23] M. Izadi, A. K Sanyal. Rigid body attitude estimation based on the Lagrange-dAlembert principle. *Automatica*, 50(10):2570–2577, 2014.

[24] C. Lagemann, J. Trumpf, and R. Mahony. Gradient-Like Observers for Invariant Dynamics on a Lie group. *IEEE Trans. on Automatic Control*, 55:2:367 – 377, 2010.

[25] R Mahony, T. Hamel, J. Trumpf, and C. Lageman. Nonlinear attitude observers on so (3) for complementary and compatible measurements: A theoretical study. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 6407–6412. IEEE, 2009.

[26] H. P. McKean. *Stochastic integrals*, volume 353. American Mathematical Soc., 1969.

[27] H. Munthe-Kaas. High order Runge-Kutta methods on manifolds, *Applied Numerical Mathematics*, 29(1):115–127, 1999.

[28] S. K Ng and P. E Caines. Nonlinear filtering in Riemannian manifolds. *IMA journal of math. control and information*, 2(1):25–36, 1985.

[29] F. Perrin. *Étude mathématique du mouvement brownien de rotation*. PhD thesis, Faculté des sciences de Paris, February 1928.

[30] S.M. Persson, and I. Sharf Invariant trapezoidal Kalman filter for application to attitude estimation, In *Journal of Guidance, Control, and Dynamics*, 36(3), pp 721–733, 2013.

[31] M. Pontier and J. Szpirglas. Filtering on manifolds. In *Stochastic Modelling and Filtering*, pages 147–160. Springer, 1987.

[32] S. Said and J. H. Manton. On filtering with observation in a manifold: Reduction to a classical filtering problem. *SIAM Journal on Control and Optimization*, 51(1):767–783, 2013.

[33] Y. Song and J. W. Grizzle. The extended Kalman filter as a local asymptotic observer for nonlinear discrete-time systems. In *American Control Conference*, pp. 3365-3369, 1992.

[34] A. S. Willsky. Some estimation problems on Lie groups. In D.Q. Mayne and R.W. Brockett, editors, *Geometric Methods in System Theory*, volume 3 of *NATO Advanced Study Institutes Series*, pages 305–314. Springer Netherlands, 1973.

[35] Van Kampen, N. G. *Stochastic processes in physics and chemistry*. Elsevier, 1992.

[36] A. S. Willsky and S. I. Marcus. Estimation for bilinear stochastic systems. In *Variable Structure Systems with Application to Economics and Biology*, pages 116–137. Springer, 1975.

[37] A.S. Willsky. *Dynamical Systems Defined on Groups: Structural Properties and Estimation.* PhD thesis, MIT Dept. of Aeronautics and Astronautics, May 1973.

[38] K. Yosida. On brownian motion in a homogeneous riemannian space. *Pacific Journal of Mathematics*, 2(2):263–270, 1952.

[39] M. Zamani, J. Trumpf, and R. Mahony. Minimum-energy filtering for attitude estimation. *IEEE Transactions on Automatic Control*, 58(11):2917–2921, 2013.

[40] D. E. Zlotnik, J. R. Forbes. Gradient-based observer for simultaneous localization and mapping. *IEEE Trans. on Automatic Control*, 2018.