Stochastic observers on Lie groups: a tutorial

Axel Barrau, Silvere Bonnabel

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 Ito-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 (Q,R) 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 [7], [21], [20], [19], [17], [32], to cite a few. Yet, most of these approaches do not explicitly account for uncertainties in the model and in the measurements, that is, they are deterministic.

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. 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 stochastic properties of geometric observers might be obtained in the presence of noise. We also discuss the design of (invariant) extended Kalman filters (IEKF) [9], [8], and we recall that 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 (Q,R) that appear in the Riccati equation may depend on the followed trajectory. Owing to this partial independence, some local convergence properties of the IEKF for group-affine systems on Lie groups may be proved under standard observability conditions in [5].

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.

A. Literature review and historical perspectives

In the presence of noise, the synthesis of observers 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 [18] some results have been derived, see e.g., [23]. 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. [15], [25] and more recently [26]. For systems on Lie groups powerful tools to study the filtering equations - such as harmonic analysis [29], [27] - have been used, notably in the case of bilinear systems [28] and estimation of the initial condition of a Brownian motion in [14]. A somewhat different but related approach to filtering consists of finding the path that best fits the data in a deterministic setting. It is thus related to optimal control theory where geometric methods have long played an important role [12]. A certain class of least squares problems on the Euclidean group has been tackled in [16], see also [31].

II. Primer on Lie groups

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 for 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 $\chi + TG_\chi$ "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 TG_χ a sub-vector space of $\mathbb{R}^{N\times N}$). The vector space TG_χ is called "tangent space at χ". The dimension d 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 $I_N \in G$ plays a central role, as it 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 $\mathbb{R}^{N\times N}$, with dimension $d < N^2$.

Identification of the Lie algebra to $\mathbb{R}^d$: $\mathfrak{g}$ 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 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 $(\xi^\wedge)^\vee = \xi$. Note that, specifying the map $^\wedge$ is equivalent to choosing a basis of the Lie algebra. There is often a basis that is more suitable than others. Note that, the covariance matrices below reflect a dispersion with respect to the choice of basis induced by $^\wedge$.

Exponential map: There is a natural map between $\mathfrak{g}$ and G, given by the matrix exponential $\exp_m$. As $\mathfrak{g}$ can be identified to $\mathbb{R}^N$, 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 $\exp : \mathbb{R}^d \to G$ 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 $\exp(\xi) = I_N + \xi^\wedge + O(||\xi||^2)$. The exponential 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 $\mathbb{R}^d$ for we have $TG_\chi = \chi\mathfrak{g} = \mathfrak{g}\chi$. This means vectors of TG_χ can be written as $\chi\xi^\wedge$ with $\xi \in \mathbb{R}^d$, or $(\xi^\wedge)\chi$ with a different $\xi \in \mathbb{R}^d$ (see Figure 1, center plot). Throughout this paper, 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, $H\xi = \frac{d}{ds}h(\chi\exp(s\xi))\mid_{s=0}$. 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 gives $H\xi = Dh_\chi(\chi\xi^\wedge)$. 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, and is defined by $(Ad_\chi\xi)^\wedge := \chi(\xi^\wedge)\chi^{-1}$. We have the useful relation $\chi\exp(\xi) = \exp(Ad_\chi\xi)\chi$.

III. Noisy dynamics on groups

Dynamical systems on a matrix Lie group G are easily defined. However, if one wants to put some random variability in the system, 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. In this section, we discuss the usual methods for "adding" white noise on G.

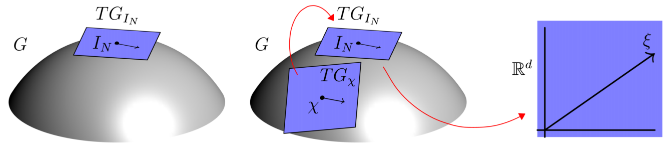

Fig. 1. Left and right multiplications offer two ways to identify the tangent space $T_\chi G$ at χ with the tangent space at Identity $T_{I_N}G$, called the Lie algebra $\mathfrak{g}$. In turn, the application $\xi \mapsto \xi^\wedge$ provides a linear bijection between the Euclidean space $\mathbb{R}^d$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

A. Discrete time noisy dynamics

Consider a system evolving on G in discrete time:

$$\chi_{t+1} = f(\chi_t) \tag{1}$$

with $f : G \to G$. Note that, the system is general at this stage and does not need to be invariant.

Suppose we want to add a centered d-dimensional Gaussian noise $\xi \sim \mathcal{N}(0,Q)$ to the dynamics, to reflect some uncertainties in the evolution model. As $f(\chi_t)$ is a group element, a mere addition $f(\chi_t) + \xi$ does not make sense. First we must find some group counterpart to $\xi \in \mathbb{R}^d$, and then we must "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 using the exponential map, and defining $W = \exp(\xi) \in G$ as a random element of G, as largely advocated in [11], [2], [1]. Note that, an alternative route may consist in defining some measure on G and then to use it to characterize "Gaussian" laws, see [13].

"Adding" noise on G: Now that we have a "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

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

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

Note that these two are related as $\chi\exp(\xi) = \exp(Ad_\chi\xi)\chi$, with Ad_χ the adjoint. As a result (4) and (5) define stochastic processes having identical distribution:

$$\chi_{t+1} = f(\chi_t)\exp(\xi) \quad \text{with} \quad \xi \sim \mathcal{N}(0,Q) \tag{4}$$

$$\chi_{t+1} = \exp(\eta)f(\chi_t) \quad \text{with} \quad \eta \sim \mathcal{N}(0,Ad_{\chi_t}QAd_{\chi_t}^T) \tag{5}$$

with T denoting the transposed matrix. Note that, the transpose of Ad_{χ_t} is well defined here since we have chosen a basis of the Lie algebra when we have introduced the map $^\wedge$. And the matrix Q is naturally also defined using this basis.

"Isotropic" noise: As a result, we will say a noise $\xi \sim \mathcal{N}(0,Q)$ is "isotropic" if for all $\chi \in G$ we have $Ad_{\chi_t}QAd_{\chi_t} = Q$. In this case both dynamics (2) and (3) define processes having identical probability distributions. It is also called Ad-invariant noise or sometimes bi-invariant noise.

B. Continuous time noisy dynamics

Consider a system evolving on G in continuous time:

$$\frac{d}{dt}\chi_t = f(\chi_t) \tag{6}$$

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. Note that, the system is not necessarily invariant. In (6), $f(\chi)$ is just a vector field on G.

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,

$$\frac{d}{dt}\chi_t = f(\chi_t) + \chi_t(w_t^\wedge) \in T_{\chi_t}G \tag{7}$$

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 (7), at least at this stage.

The problem is, that (7) 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, which may be somehow surprising, since adding uncertainty is so easy in the discrete time case. This prompts the following subsections, that are meant to be a gentle, but rather mathematically loose, introduction to the subject.

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

Consider a stochastic differential equation (SDE) in $\mathbb{R}^d$ of the form:

$$\frac{d}{dt}x_t = g(x_t) + \bar{w}_t \tag{8}$$

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

$$\frac{d}{dt}x_t = g(x_t) + Q^{1/2}w_t \tag{9}$$

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: Equation (9) can thus be given a proper mathematical meaning through its integral counterpart, that is, the well-defined equation:

$$x_t = x_0 + \int_0^t g(x_s)ds + Q^{1/2}W_t. \tag{10}$$

We can then define a solution to SDE (9) as any $(x_t)_{t\geq 0}$ that satisfies the integral equation (10). 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 (10) 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 (9), and thus to (8), 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

$$\frac{d}{dt}x_t = g(x_t) + r(x_t)w_t \tag{11}$$

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 $r(x_t)$.

Ito's approach: Roughly speaking, it consists in defining the solution to (11) as the limit as $\Delta t \to 0$ of the following (Euler-Maruyama) discretization method

$$x_{t+\Delta t} = x_t + \Delta t g(x_t) + \sqrt{\Delta t}\ r(x_t)\xi \tag{12}$$

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 (11) as the limit as $\Delta t \to 0$ of the following discretization method

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

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

Although Ito'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 (11) 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 (11) in the sense of Ito.

D. Stochastic differential equations on Lie groups

Hence, there are two ways to interpret the stochastic term $\chi_t w_t^\wedge$ of (7). The first one is Ito's intepretation of SDE (7). Roughly speaking, it means that the solution to the differential equation (7) corresponds to the limit as $\Delta t \to 0$ of the following (Euler-Maruyama) discretization method

$$\chi_{t+\Delta t} = \chi_t + \Delta t f(\chi_t) + \sqrt{\Delta t}\chi_t(\xi_t^\wedge) \tag{13}$$

with ξ's independent Gaussians with distribution $\mathcal{N}(0, Q)$. Stratonovich's interpretation leads to (13) *but* where χ_t is replaced with $(\chi_t + \chi_{t+\Delta t})/2$. The question is whether the limiting process $(\chi_t)_{t\geq 0}$ of discretization (13) as $\Delta t \to 0$

actually remains inside G at all times, or if it results in matrices of the ambiant space $\mathbb{R}^{N\times N}$, that lie outside G.

Ito versus 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. To fix ideas, in the case of rotation matrices, we choose $F(\chi) = \chi^T\chi$, and $SO(3)$ corresponds to the level set $F(\chi) = I_3, \det(\chi) = 1$. For more details, see Section 20.4 of [13]. With Ito's approach, the solution steps out of the group almost surely, see e.g. [4].

Implementation of EDS on Lie groups: Another idea to ensure that χ_t remains in G may be to define a group counterpart to Ito's discretization (12), by replacing addition with group multiplication. This can be done as follows. $f(\chi_t) \in T_{\chi_t}G$ 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 (7) then becomes

$$\frac{d}{dt}\chi_t = \chi_t(\omega_t + w_t)^\wedge. \tag{14}$$

As $\omega_t + w_t \in \mathbb{R}^d$, one could be tempted to discretize this equation using a group multiplicative counterpart of Ito's approach (12), that is, using i.i.d. Gaussians $\xi \sim \mathcal{N}(0,Q)$:

$$\chi_{t+\Delta t} = \chi_t \exp_m(\Delta t(\omega_t)^\wedge + \sqrt{\Delta t}(\xi^\wedge)) = \chi_t \exp(\Delta t\,\omega_t + \sqrt{\Delta t}\xi), \tag{15}$$

and the limiting process as the discretization step $\Delta t \to 0$ may serve as a direct definition for the solution of (12) on G. Indeed, we readily see the scheme (15) defines a stochastic process that remains 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 is reminds much of the discrete time case of Section III-A. It turns out this is a viable route indeed, owing to results by McKean [22] and related to his multiplicative stochastic integral, and more remotely to Brownian motion on Lie groups, [24], [18], [30].

Proposition 1 ([22], see also [13]): The discretized solution of the multiplicative scheme (15) converges in probability to the solution in the sense of Stratonovich to (14) and thus to the solution to SDE (7).

Corollary 1: Discretization (15) thus serves as an explicit convergent implementation scheme with time step Δt of the EDS (14) and thus (7), that guarantees the solution remains in the group G at all times, regardless of the chosen step Δt.

IV. STOCHASTIC OBSERVERS ON LIE GROUPS

Due to space limitation, we focus on the continuous time case with discrete time measurements. As concerns the discrete time case the reader is referred to [6] which provides a tutorial introduction to the discrete time case. In this section we essentially review the results of our main paper on the subject [5], yet adopting a perspective more oriented towards stochastic aspects.

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

$$\frac{d}{dt}f_{u_t}(\chi_t) \tag{16}$$

here $f_{u_t} : G \to T_{\chi_t}G$. This echoes (6), except that to be more relevant to applications we allow f to also depend on a time-varying input $t \mapsto u_t$, where $u_t \in \mathcal{U}$ lives in some space $\mathcal{U}$. 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

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

$$\frac{d}{dt}\chi_t = f_{u_t}(\chi_t) + (\bar{w}_t^\wedge)\chi_t \in T_{\chi_t}G \tag{18}$$

The choice between (17) and (18) 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 (17) 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

$$Y_n = h(\chi_{t_n}) + V_n \tag{19}$$

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 $t_n \leq t < t_{n+1}$.

B. Group affine systems

In the sequel, we will systematically suppose that f_{u_t} possesses the "group affine" property [5]:

$$\forall u \in \mathcal{U},\ a,b \in G \quad f_u(ab) = af_u(b) + f_u(a)b - af_u(Id)b. \tag{20}$$

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.

Fundamental property of invariant filtering: Consider two trajectories $\chi_t, \tilde{\chi}_t$ of the noise free system (16). 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 ([5]): f_{u_t} satisfies Equation (20) 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 $g_{u_t}(\eta) = f_{u_t}(\eta) - f_{u_t}(I_N)\eta$.

C. Invariant observers and autonomous error equations

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., [7], [21], [20], [19], [17], [32], 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 (19) 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

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \le t < t_n \tag{21}$$

$$\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n} \exp(L_n(Y_n - h(\chi_{t_n}))), \quad t = t_n \tag{22}$$

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

Remark 1: As always, a choice between left and right multiplication must be made, leading to two different families of observers. In (22), we have opted for left multplication by $\hat{\chi}_{t_n}$. However, right multiplications by $\hat{\chi}_{t_n}$ might prove more suited in various applications. For tutorial purposes, we will stick with the left version (22), but the reader should bear in mind the other family of observers. For more details, and guidelines see [5].

Step (21) is called the propagation step, and Step (22) 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.

Corollary 2 (of Thm 1, see[5]): Consider (17), with f_{u_t} group-affine, and observer (21)-(22). 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 $\hat{\chi}_t$.

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 (20) that implies $f(\hat{\chi}) = f(\chi\eta) = \chi f(\eta) + f(\chi)\eta - \chi f(I_N)\eta$. ∎

Proposition 2 ([5]): Suppose that, moreover, $h(\chi)$ is of the form $h(\chi) = \chi\bar{d}$ with $\bar{d} \in \mathbb{R}^N$ a vector, or a collection of such measurements, and the measurement noise is isotropic. Let $L_n(Y_n - h(\chi_{t_n}))) = \bar{L}_n(\hat{\chi}_{t_n}^{-1}[Y_n - h(\chi_{t_n})])$ in (22) for some function $\bar{L}_n$ that does not depend on $\hat{\chi}_{t_n}$. Then, the error after update writes (24), and thus neither does it explicitly depend on the current estimate.

Proof: We have $\hat{\chi}_{t_n}^{-1}[Y_n - h(\chi_{t_n})] = \eta_{t_n}^{-1}\bar{d} - \bar{d} - \hat{\chi}_{t_n}^{-1}V_n$ which is equal in distribution to $\eta_{t_n}^{-1}\bar{d} - \bar{d} - V_n$ since the noise was supposed to be isotropic. And we see the updated error $\eta_{t_n}^+$ is a function of η_{t_n} only. ∎

Autonomous error equation: Under the assumptions that the dynamics are group affine, and the measurements as in Proposition 2, we have shown using Cor. 2 and Prop. 2, that the error $\eta_t = \chi_t^{-1}\hat{\chi}_t$ satisfies the error equation:

$$\frac{d}{dt}\eta_t = g_{u_t}(\eta_t) - (w_t^\wedge)\eta_t, \quad t_{n-1} \le t < t_n \tag{23}$$

$$\eta_{t_n}^+ = \eta_{t_n} \exp(\bar{L}_n(\eta_{t_n}^{-1}\bar{d} - \bar{d} - V_n)), \quad t = t_n \tag{24}$$

that can be called "autonomous" in the sense that its evolution does not depend on the trajectory followed by the estimate $\hat{\chi}_t$. This opens up for possible analysis of the stochastic behavior of the error, without having to care about the particular behavior of $\hat{\chi}_t$, or of the true trajectory χ_t.

D. Benefits of the autonomous error property

Harris chains and gradient-like observers: The paper [20] 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. (23) 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 $\eta_{t_n}^{-1}\bar{d} - \bar{d}$ decrease, inspiring from gradient-like observers [20], 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. Once this is established, the gain function $K(\cdot)$ can advantageously be tuned offline through Monte-Carlo experiments so as to minimize the asymptotic dispersion of η_t. This kind of optimal offline tuning is made possible only by the fact the error equation (23)-(24) does not depend on the actual trajectory, a well known feature of linear systems, that proves to carry over to group-affine systems.

Invariant EKF (IEKF): Unfortunately, in many applications of interest, g_{u_t} depends on the time t, through the input u_t, and the noises are not isotropic. This makes the analysis of the error system very difficult, albeit autonomous. To cope with dependency with respect to time, we can resort to extended Kalman filtering (EKF), which is the most widespread approach to observer design for non-linear (time varying) systems. However, the EKF is designed for systems on vector spaces, and it needs to be transposed in the Lie group setting.

This yields the invariant EKF (IEKF), introduced in [9], [8]. The Left-invariant EKF (LIEKF) for continuous-discrete group affine systems with left-equivariant output writes [5]:

$$\frac{d}{dt}\hat{\chi}_t = f_{u_t}(\hat{\chi}_t), \quad t_{n-1} \leq t < t_n \tag{25}$$

$$\hat{\chi}_{t_n}^+ = \hat{\chi}_{t_n} \exp(K_n(\hat{\chi}_{t_n}^{-1} Y_n - \bar{d})), \quad t = t_n \tag{26}$$

with $K_n \in \mathbb{R}^{d \times N}$ a matrix tuned through a Riccati equation. It turns out that, the Jacobians (A_t, C_t) in the sense of group linearization (see Section I) with respect to the system's dynamics and output map are independent of the followed trajectory, a remarkable property owing to the autonomy of the error equation, due in turn to the group-affine property, whereas the noise covariance matrices (Q_t, R_t) that appear in the Riccati equation may depend on the followed trajectory when the noises are not isotropic.

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: The fact that the Jacobians (A_t, C_t) do not depend on the estimates $\hat{\chi}_t$ reminds much of the linear (time-varying) Kalman filter theory. This implies great stability of the linearized error system. The fact that covariance matrices (Q_t, R_t) may depend on the estimates, impacts only mildly the stability of the error system. As a result, it could recently be proved in [5] that, when used as non-linear observer the IEKF enjoys local asymptotic convergence properties for all (observable) group-affine systems with equivariant outputs. 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., [10]. Moreover, group affine systems are a large class of nonlinear systems on Lie groups with relevant applications, see [6].

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. The invariant extended Kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4):1797–1812, 2017.

[6] A. Barrau and S. Bonnabel. Invariant Kalman Filtering. *Annual Review of Control, Robotics, and Autonomous Systems*, 1(1), 2018.

[7] 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.

[8] S. Bonnabel, Ph. Martin, and E. Salaun. Invariant extended Kalman filter: theory and application to a velocity-aided attitude estimation problem. 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 1297–1304. IEEE, 2009.

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

[10] S. Bonnabel and J. -J. Slotine. A contraction theory-based analysis of the stability of the deterministic extended Kalman filter. *IEEE Transactions on Automatic Control*, 60(2):565–569, 2015.

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

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

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

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

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

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

[20] 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.

[21] 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.

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

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

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

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

[26] 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.

[27] 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.

[28] 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.

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

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

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

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